Universal biases in the perception of Mandarin tones:

From infancy to adulthood

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006

e-mail: lot@uu.nl
http://www.lotschool.nl

ISBN: 978-94-6093-099-7
NUR 616

Universal biases in the perception of Mandarin tones: From infancy to adulthood

Universele bias in de perceptie van Mandarijn-Chinese tonen: Van babies naar volwassenen
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr.G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op
donderdag 10 januari 2013 om 12.45 uur

door

Ao Chen
geboren op 1 juli 1985 te Beijing, China

Promotor:
Prof.dr. R.W.J. Kager.

Dit proefschrift werd mede mogelijk gemaakt door het onderzoeksprogramma “The Early Acquisition of Speech Prosody, A comparative study of Dutch and Chinese”, dat gefinancierd is door de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) en Chinese Academy of Social Sciences (CASS)

Acknowledgement

Without the help of many people, this dissertation could not have been finished.

First and most of all, I need to thank my supervisor René Kager. Ever since the first day that I met him, for all these years that I have worked with him, he has given me so much support and valuable suggestions. I really appreciate that he, and Uil OTS have given me the freedom to work on the topics that I am really interested in, without limiting the scope. It is in the last year of my PhD that I started working on infants' perception of musical pitch and lexical tones, and even the time is really getting tight, René gives me the chance to focus on this issue that I like. Another thing that I appreciate very much is that although he is extremely busy, he is always available, and willing to listen to the students (who sometimes just ask and propose stupid things). Every time that I had new thoughts or new results, I could easily find him and talk to him, and he always helped me to clear up my minds. Throughout my PhD, he has always been keen in pinpointing core questions. He has taught me how to sort out the core issues from peripheral questions, and how to make the arguments coherent. Needless to mention, his thorough knowledge of the field has always been the encyclopaedia that I could rely on. During the last a few weeks before I finished the dissertation, I wrote till 10pm, and he read it till 2am. Even his Saturdays and Sundays were spent on reading the manuscript. I cannot overstate my thanks to him. He is the example that I would like to follow in my career.

I owe many thanks to the Babylab of Utrecht University. Without the help from them, my dissertation could not have been written so smoothly. I need to thank, Frank Wijnen, Maartje de Klerk, Annemarie Kerkhoff, Desiree Capel, Elise de Bree, Carolien van den Hazekamp, Ileana Grama, and Nadya Goldberg, who gave me inspiring thoughts and valuable suggestions about the experiments, from design to results. It is Maartje de Klerk, our babylab manager, who co-ordinates making of appointments with parents, sending out brochures of the lab, buying the gifts for the babies, and rearranging the experiments if the baby was sick, if the parents forgot the appointment, and all other unexpected ifs. Without her excellent coordinating, it would just have been impossible for me to test so many babies. To make sure that I could finish the dissertation on time, I was given the priority to test the young infants in my last year, and I'm really grateful for people's understanding. I also need to thank our student assistants who helped calling the parents, Jorinde Timmer, Lorijn Zaadnoordijk, Natalia Davidson, and Sule Kurtcebe, and also Eline van Baal, who replaced Maartje during her maternal leave. Especially Sule, who helped me to test the babies during the last year, and without her, I would not have had time to focus on writing.

Liquan Liu deserves a paragraph of my thanks. He shared the office with me, and is the person that I spent most time with in my PhD. There have been countless days and nights that we tested, debated, wrote, ate, and relaxed together in Janskerkhof. I can't imagine how life could have been without him being there. We shared thoughts for work, and we shared mood for life. I will always remember the Dutch rainy winter days that we each took a takeaway wok or AH meal for dinner, ate in the office, kept working a little bit, and left together at 8pm. Then, the next day started again with meeting him downstairs in the lab. For many days, he was the first and last person that I talked to in a day.

The experimental phonology people in Utrecht are wonderful colleagues. Natalie Boll-Avetisyan, Frans Adriaans, Tom Lentz, Diana Apoussidou, Brigitta Keij, Sandrien van Ommen, Aoju Chen, Anna Sara Romoren, Anne-France Pinget, Xiaoli Dong, Jingwei Zhang and Liquan Liu, The discussion with them has always been inspiring. They helped me to realize the points that I have missed. Also, when I encountered the bottleneck, their opinions and suggestions helped me to break through. Almost all my experiments are discussed and criticized with these colleagues.

The presence of the Chinese people in Utrecht makes research on Mandarin easy. Especially the Chinese people in Uil OTS, Min Que, Xiaoli Dong, Fang Li, Jingwei Zhang, Yueru Ni, Yuning Sun, Yipu Wei, Mengru Han, Haifan Lan. They are linguists who are able produce and listen to the stimuli, and they are the intelligent little mice who are not only able to participate in experiments but also give advices about how to improve the experiments.

I also need to thank Theo Veenker and Iris Mulders. They helped me with recruiting the participants, and teaching me how to use the lab facilities. Iris is always fast in arranging everything, and the recruitment of Dutch participants was always finished very efficiently. Theo has spent so much time in scripting the experiments and maintained all the lab equipment in good state. He also provided very valuable instructions and diagrams when I set up the babylab in Beijing. It took him hundreds of emails to teach me how to connect the lights of the HPP paradigm, and how to install everything under a linux environment. Without this experience, I would have never been able to imagine how much work there is for a lab to function properly.

Many thanks to KNAW for sponsoring the China Exchange Program--*The Early Acquisition of Speech Prosody: A comparative study of Dutch and Chinese*, and many thanks to CASS for their support of this project. It is within this project that René and I were able to make well targeted the research questions and plan things in an executable and trackable way. Without this framework and financial support, the testing in Beijing would not have been possible. I also need to thank Uil OTS for financially supporting my PhD program and giving me the opportunities to take courses outside Utrecht and go to international conferences.

My colleagues in Beijing have also given tremendous help for my research. I started cooperating with Aijun Li of China Academy of Social Science (CASS) in my master. She shared with me the corpus analysis, which produced important information about the distributional pattern of Mandarin. Also, without her effort, the babylab in Beijing could not have come true. She spends so much time and energy to make sure that parents are willing to come to the lab. In a city where infant research just started, her job is far from easy. Every time that I visited her lab, she did her best to provide me with things that I needed. Jianhua Hu has also given much support to the cooperation between

Utrecht Institute of Linguistics and CASS. Many thanks to Qiang Fang and Zhigang Yin, who helped with technical issues and makes running experiments practically possible. Thanks to Yuan Jia, she shared with me not only her knowledge of phonology, but also practical information about where to enjoy life. I need to thank Jun Gao and Zhao Zhang for their active involvement in cooperation between Utrecht and CASS. I also owe many thanks to Fengmei Zhang, Yao Lu, and Yang Xing for their assistance in recruiting and testing the babies in Beijing. For practical reasons, frequently had they to work on weekends, but they never complained a word. Without the time that they spent in the lab, the data of the Chinese babies could not have been accumulated. I also need to thank Guohong Tian, who took care of the administrative issues for my and René‘s stay in Beijing, and always gave us the cheerful smile.

The people from Nijmegen have also contributed to this dissertation. I thank Paula Fikkert, Christina Bergmann, Helen Buckler, Nienke Dijkstra, Stefanie Ramachers, Sho Tsuji, Nicole Alvater-Machensen for all their comments and suggestions on my research. The Utrecht-Nijmegen meetings gave me the chance to look at the same thing from different angles.

I also need to thank Yiya Chen for her comments and continuous support of my research. Thanks to Rushen Shi for sharing her opinion of the infant research with me. Thanks to James Morgan, every time we met, he gave me new inspirations. I thank Alexis Dimitriadis for helping me with the programming, and I thank Makiko Sadakata for discussing the experiments with me. Many thanks to Marlisa Hommel for her very careful proofreading in a short time and translating the summary. The dissertation looked much better after her work. Thanks to Rongjia Cui, Qian Li, Wei He, Yan Gu, Xuliang He, Wencui Zhou, who helped me in designing and running experiments. Thanks to the people who arrange the EMLAR, Elitu, Uitjes Dag, etc, which make our academic life in Utrecht so vibrant.

I sincerely appreciate the careful work of the reading committee. Sergey Avrutin, Denis Burnham, Yiya Chen, James McQueen, and Wim Zonneveld spent a lot of time reading and evaluating my dissertation, and they have given extremely helpful and valuable comments and suggestions. From their comments I can see how much more I still need to learn.

Many thanks to the PhD information center, Maaike Schoorlemmer, who's always ready to help and who really helps. Also thanks to our secretaries, they are always nice and efficient in their work.

I really appreciate the efforts that the participants made to join my experiments, especially the parents. They finish the survey very carefully, and take their young babies to the lab for the experiments, regardless of how infant-unfriendly the weather was. Many times, they arrived with water drizzling down from their hair, but still happily did the experiment. Without their understanding and contribution, infant research could not have been possible. Participants are punctual, which makes it easy to estimate and schedule the time that we need for the experiments.

PhD is more than doing research. I have to say that I am so lucky to have many friends surrounding me, who comfort me during my low times and share the joy with me. Fang and Abu are the two that put me back together when I fell apart. We witness each other's life and change for all these years in the Netherlands. Thanks to Haifan, Jingwei and Qiulan, I really enjoy all the chatting, cooking, shopping and laughing we had. Thanks to Yigen Huang, Zheng Guan, Yiwei Gong, Chenguang Wang, Dong Guo, Yuning Huang for the wonderful time that we had together. It is actually during our relaxing meals that Zheng Guan saved me twice before I went to a wrong direction for running the statistics, and his mean jokes work better than espresso.

I owe a lot to Hans van de Velde, who's not only a linguist, but also my supervisor for moving. Without him, I could have never succeeded in measuring, cleaning, and painting my apartment, not mentioning to make the apartment home.

I need to thank Shaoyu Yin for drawing the cover for me. I not only owe him thanks, but also sorrys for my bad temper when I was stressed out writing the dissertation.

My parents have been and will forever be my strongest support. I have never had any pressure ever from them, and they just support my decisions and let me live my own life.

To my parents

至我的父母

TABLE OF CONTENTS

Chapter 1 Literature Review

1.1 THE TONE SANDHI PHENOMENON

Most languages in the world are tone languages (Yip 2002). In tone languages, lexical tones are mainly realized by pitch variations, and they function in a phonemic way to distinguish meaning lexically. For example, the widely cited example of Mandarin is that the syllable /ma/, when carrying a high level tone, mid rising tone, low dipping tone, and high falling tone, means ‘mom, ‘hemp’, ‘horse’ and ‘scold’, respectively; in Cantonese, the single syllable /si/ can be realized with six different tones, namely high level, mid rising, mid level, low falling, low rising, and low level, meaning ‘poem’, ‘history’, ‘try’, ‘hour’, ‘market’, and ‘yes’, respectively. In the lexical tone languages spoken in China and surrounding areas, tones are involved in productive alternations when specific tones co-occur in connected speech, which is known as “tone sandhi” (Chen 2000). The functional factors that shape tone sandhi processes are not well-understood; in particular, little is known about its perceptual factors. This dissertation is intended as a contribution to this area.

Among all the sandhi phenomena, T3 sandhi in Mandarin is the most studied case. In Mandarin, there are four lexical tones, namely high level (T1), mid rising (T2), low dipping (T3), and high falling (T4). When two T3's occur next to one another within a prosodic word, the T3 sandhi rule requires the first T3 to change into a T2 obligatorily. Thus a sequence of T3T3 changes into a T2T3. For example, /haomi/ ‘good rice’, in which both syllables carry a lexical tone T3, is sandhied into /hao/T2 /mi/T3. In Mandarin, the sequence /haomi/ in which the first syllable carries a T2 and the second syllable carries a T3 means ‘millimeter’. Perceptually, ‘good rice’ and ’millimeter’ sound the same. The T3 sandhi is a well documented contextual phonological rule, and numerous studies have tried to give phonological accounts of this rule. However, these accounts are not yet satisfactory (detailed discussion will be given in section 1.3.3). Considering the fact that tone sandhi occurs in connected speech, one may easily link tone sandhi with coarticulation. However, regarding T3 sandhi in Mandarin, no convincing articulatory motivation has been found either (as will be discussed in section 1.3.3).

Considering the failure of previous articulatory accounts, this dissertation will search elsewhere, namely in the area of perception. Earlier perceptual studies have focused on the result of T3 sandhi, namely how the derived T2 is perceived (I will discuss these studies in detail in section 1.4). Surprisingly, no previous studies addressed the perceptual factors that may drive the T3 sandhi process; a perceptual account of T3 sandhi is completely missing. With regard to the perception of T3 sandhi, several important questions remain unanswered: to what extent is T3 sandhi grounded in possibly universal perceptual factors? What is the nature of native listeners' perceptual knowledge of T3 sandhi? How does this knowledge develop as a result of language experience? Nevertheless, it is not a new idea that perception may play a role in shaping phonology (Ohala 1981, 1986, 1993, Steriade 1995, 1997, among others, which will be discussed in section 1.2). Hence, in the current dissertation, I will study T3 sandhi from the perspective of perception, and try to find out whether there are universal perceptual biases that may favor it to occur. Importantly, such perceptual biases will be studied from different angles with different populations, namely native versus non-native listeners and infant versus adult listeners, as well as from a domain-specific (language) versus domain-general (speech and music) perspective.

Before we move on to studies that discuss the relation between perception and phonology, I would like to give a short overview of the factors that may be involved in the perceptual biases (if any) favoring T3 sandhi, and discuss each of them in detail in due course. First, it has to be emphasized that T3 sandhi occurs in an asymmetrical and unidirectional way. Only the first syllable of a T3T3 sequence alternates, while the second stays intact. In order to understand how perceptual factors may have shaped the T3 sandhi pattern, asymmetries in the perception of pitch sequences need to be examined. Regarding the perceptual biases, if these are claimed to have a universal basis, then a cross linguistic perspective has to be taken. By only studying the perception of native Mandarin listeners, it is difficult to assess whether their performance is to be attributed to innate biases or to language experience. Hence, listeners naïve to Mandarin need to be studied in order to identify the natural responses to Mandarin tones. Also, the fact that non-tone languages lack the representation of lexical tones, while native listeners of such languages must still be sensitive to pitch variations due to its role in intonation, makes the perception of lexical tones a good topic to address universal biases in speech prosody. In sum, to test non-tone language listeners who have no representation of lexical tones whatsoever on their perception of tonal patterns will shed light on the nature of universality in lexical tone perception. The comparison between native tone language listeners and non-tone language listeners will give more information about the innateness versus experience in terms of speech processing. Third, the question how lexical tones are perceived and acquired by infants cannot be ignored if we are to understand the innate restrictions on tone perception at the very initial state. Infants are listeners who have not had substantial knowledge of any language yet, and their performance will best reflect the natural perceptual pattern when first encountering tones. A comparison between Mandarin learning infants and non-tone language learning infants may thus offer valuable information about the initial state of tone perception and how a native language starts to shape speech perception. Last, lexical tones are mainly realized by pitch variations, and pitch functions across different domains in that both music and speech use pitch to convey meaning. Hence, when we consider innate biases in the perception of lexical tones, it may give us a more complete picture to put it into a broader scope, namely how the general sensitivity to pitch restricts lexical tone perception.

In this chapter, I will first review the literature that deals with the relation between perception and phonology in section 1.2. Next, in section 1.3 I will give a brief introduction of the phonology and phonetics of Mandarin tones, which includes their coarticulation as well as T3 sandhi. In section 1.4, I will review the literature on native adults' perception of lexical tones, including studies on the perception of T3 sandhi. In section 1.5, I will discuss studies that target early perception of lexical tones in infancy. Considering the main focus of this dissertation is to identify innate perceptual biases favoring T3 sandhi, in section 1.6, I will review the literatures dealing with early speech perception in infancy, which may reflect innate biases. Finally in section 1.7 I will review the studies dealing with pitch perception in music and from a domain-general perspective. After reviewing the literatures, in section 1.8, I will list the research questions and the outline of the current dissertation. The current chapter will end in a statement of the research methods of the current dissertation (section 1.9).

1.2 PERCEPTUAL BIASES AND PHONOLOGY

The phonologies of individual languages differ substantially; accordingly, the perception of speech sounds is, to a large extent, language-specific. For example, native English adults are not able to discriminate between dental voiceless aspirated /t^h/ and voiced aspirated /d^h/ while native Hindi adults are (Werker, Gilbert, Humphrey, & Tees 1981). It has also been observed that listeners speaking different languages map a same vowel into different vowel categories as a result of different distribution of prototypical vowels in the languages (Kuhl 1991). Despite the language specificity observed in phonology, however, a great deal of universal preference towards certain structures has been observed in perception.

Regarding how to interpret these perceptual biases, it seems unlikely that the debate between Universal Grammar and the emergentist viewpoints will be settled in the near future. Yet proponents of both views do accept that there are biases, and that such biases may shape the formation of phonology. In other words, these biases are reflected in phonological patterns. Important contributions have been made from both sides.

Universal grammar (UG) proponents assume that a language faculty is hard-wired in the human mind innately. The language faculty is not a result of experience, and language users are pre-specified with language devices favoring certain regularities regardless of which specific native language he or she will acquire (Chomsky & Halle 1968, Chomsky 1981). The notion of "markedness" well embodies the idea that some patterns tend to be favored. In UG, it has been a long tradition that, when describing languages, phonologists distinguish between "marked" and "unmarked" structures (Trubetzkoy 1931, 1939; Jakobson 1941; Chomsky & Halle 1968; Stampe 1972; Prince & Smolensky 1993), where "'markedness' refers to the tendency of languages to show preferences for particular structures or sounds. This bias towards "unmarked" elements is consistent within and across languages, and tells us a great deal about what languages can and cannot do." (De Lacy 2006:1).

After the advent of optimality theory (OT) (Prince & Smolensky 1993), sound patterns that occur in different languages have been considered as individualized compromises between *markedness* constraints and *faithfulness* constraints. While “markedness” constraints give priority to patterns that are widely present both within a language and cross-linguistically and are accordingly easier to process and acquire, “faithfulness” constraints assure contrastiveness in a language, favoring preservation of lexical contrasts. Thus markedness constraints tend to favor unmarked structures in the grammar's output while faithfulness constraints tend to preserve contrasts in the grammar's output that occur in the input, irrespective of markedness.

Taking a UG point of view, Berent, Steriade, Lennertz & Vaknin (2007) tested English native listeners on their perception of onset clusters differing in their sonority shape. Typological work of Greenberg (1978) found that cross-linguistically, within a consonant cluster, a sonority fall such as /lb/ is less preferred than a sonority plateau such as /bd/ or a sonority rise such as /bl/. Berent et al. (2007) found that, although neither /lb/ nor /bd/ actually occur in English, native English listeners were more likely to misperceive monosyllabic non-words as bisyllabic if the non-word began with a universally dispreferred cluster with falling sonority, such as /lbif/, as compared to a universally more preferred cluster with a sonority plateau, such as /bdif/. Listeners also misjudged non-word monosyllables with sonority falls more frequently as being identical to their epenthetic counterpart /lebif/. The authors argued that the misperception was a result of a universal dispreference against ill-formed structure regardless of listeners' language background. In a follow-up study, Berent, Lennertz, Smolensky, & Vaknin-Nusbaum (2009) found that for illegal onset clusters that begin with a nasal, English listeners perceived an onset cluster with a rising sonority more accurately than those with a falling sonority. As the performance of English listeners cannot be explained by structures in native input, Berent et al. (2007, 2009) argued that the ill-formedness of a sonority fall in onset clusters is universal in grammar regardless of language background.

On the other hand, emergentists tend to believe that a functionalist perspective supports the idea that certain regularities occur in order to suit functional needs such as perception, production, the motoric function system, and the social needs of human beings, which are not specific to language (MacWhinney 2001, Ellis 1998). For example, regarding the categorical perception of speech sounds (Aslin & Pisoni 1980), which is essential for processing the speech signal efficiently, it has been found that when presented with a continuum of sound change, categorical perception was not only observed in human adults (Liberman, Harris, Eimas, Lisker, & Bastian 1967), but also in human infants (Eimas, Siqueland, Jusczyk, Vigorito 1971), macaques (Kuhl & Padden 1982), chinchillas (Kuhl & Miller 1975) and parakeets (Dent, Brittan-Powell, Dooling, & Pierce 1997). Hence, Blevins (2006) argued that sounds in human speech possibly have developed into discrete contrastive categories as a result of physical constraints of the perceptual system.

From the emergentist point of view, abundant evidence has been found supporting the perceptual motivations for phonological regularities. Ohala (1981, 1986, 1993) explicitly linked the phonological grammar to perception by arguing that the listener is a source of sound change. He considered diachronic phonological variation as the result of the listeners' failure to recover the co-articulation induced variation from the speech signal, and argued that systematic misapprehension triggers the phonologization of the variants as perceived. Later, the work of Steriade (1995, 1999) argued that the neutralization of laryngeal categories occurred in a context where perceptual cues to this contrast would be diminished, or where additional articulatory maneuvers were required to obtain the perceptibility of the contrast (Hume 1998, 2001).

Besides the pioneering work of Steriade, quite some other studies have argued that to some extent, the phonological grammar tends to suit the listeners. In other words, the perceptual biases of the listeners are phonologized into grammars. For example, it has been observed that cross-linguistically, vowel systems tend to occupy the vowel space in a way that maximum contrastiveness is maintained (Liljencrants & Linblom 1972, Lindblom & Engstrand 1989, Lindblom 1990). Meanwhile, the contrasts with low perceptibility tend to be either repaired for stronger contrastiveness, or to be merged. In the former case, multiple strategies such as epenthesis, dissimilation or metathesis could be applied (Hume & Johnson 2001). For instance, Abramson (1987) argued that the perceptibility of word-initial stop geminates was weak, and in Maltese, the epenthesis of [i] before word-initial geminate consonants serves to enhance the perceptibility of consonant length (Aquilina 1959, Hume 1996). In Greek, consecutive stops or consecutive fricatives in a consonant cluster are optionally replaced by a combination of one stop and one fricative (e.g. [pt] to [ft] in 'epta' to 'efta', which means 'seven'), which is an example of increasing contrastiveness by dissimilation. Metathesis is applied to assure perceptibility in the Faroese sequence /sk/ when it occurs before a stop consonant (e.g. /baisk +t/ [baikst] *[baiskt]). The position of the stop /k/ is changed to a postvocalic position for better releasing the stop place cues. As to the merging case, for example, Korean /n+l/ and /l+n/ are realized to /ll/ for the low salience of the contrast between nasal and lateral in this case (Seo 2001). Specifically regarding the merging situation, however, the phonological system does not like alternations that are noticeable. In other words, the listeners make an effort to stay as close as possible to the intended target produced by the speaker. Taking this perspective, regarding why T3 sandhi occurs between T2 and T3, Huang (2001) argued that T2 and T3 are the most similar tones perceptually, and therefore the alternation is less noticeable.[1] Nevertheless, Huang (2001) did not answer the question why T3 sandhi occurs in an asymmetrical way. In other words, if T2 and T3 are perceptually similar, why does not a T3T3 change to a T3T2 or T2T2? Why does a T3T3 sandhi to a T2T3 specifically?

Regarding lexical tones, it has also been argued that perception may have played a role in tonogenesis (Hombert, Ohala, Ewan 1979). In this paper, the authors reviewed abundant typological work regarding tonogenesis, and consistently found that tonal contrasts tended to originate from the voiced/voiceless distinction of the prevocalic stop consonants or from postvocalic glottal consonants. On the contrary, postvocalic non-glottal consonants and vowel height barely played a role in tonogenesis. Besides the influence from articulation, the authors attributed the success of tonogenesis to the perceptibility of the listeners, meaning that because the pitch perturbation introduced by these consonantal differences was perceptible for the listeners, tone derived from a voicing contrast of prevocalic consonants and postvocalic glottal consonants In the other two cases, however, the pitch introduced by the consonant was too subtle for the listeners to track.

---

[1] Examples cited from Hume & Johnson (2001).

If we zoom in on the case of T3 sandhi, in which two consecutive T3s are required to change to a T2T3 sequence, in terms of interaction between the phonological system and perception, two things are worth noticing. First, T3 sandhi is a dissimilation process, in which a combination of two identical tones is changed to a combination of two different tones. Second, T3 sandhi is asymmetrical. It occurs in a way that only the first tone gets involved in alternation, while the second tone stays intact. Both dissimilation and asymmetrical alternation have been interpreted from the perspective of perception. Ohala (1990, 1993, 2012) has considered the dissimilation as result of listeners' "hyper-correction". In other words, when two similar sounds occur in a row, the listeners may be tempted to consider the perceptual similarity as a result of coarticulation, and tend to compensate the illusionary coarticulation by dissimilating the two sounds. Referring to the asymmetry, it has been found that [ki] is often confused with [ti], but not the other way around. Chang, Plauché, Ohala (2001) demonstrated that such confusion asymmetry is specific to a high, front vocalic environment. They suggested that it was the relatively tight constrictions of these vowels that resulted in a slightly longer VOT or slight frication, which led the listeners to perceive a change in manner and place of articulation. Hence, it is the acoustical property of the sounds that biased the perception of the listeners, which triggered the discrimination asymmetry observed. Coming back to the issue of T3 sandhi, though for now it is hard to tell whether the explanations from Ohala (2012) and Chang et al. (2001) could be applied to T3 sandhi, at least, it is justifiable to speculate that the perception of the listeners may have played a role in shaping the T3 sandhi rule. It is plausible that T3 sandhi merges T2T3 and T3T3 due to the lack of perceptibility, as the difference between T2 and T3 are not perceptible when these tones are followed by another T3. To go one step further, there might be some innate biases that favor the discrimination between T2 and T3 in one direction but not the other. It should be mentioned that very little is known about the historical origins of T3 sandhi, and no convincing evidence can be found about its emergence from tonal variation in earlier stages of the language. Yet, starting from the synchronical situation of T3 sandhi, we can address the issue of how innate biases may play a role in shaping the phonology of tone.

To summarize, it is widely agreed that listeners, or the perceptual biases, can play a role in shaping phonological regularities (UG proponents argue that these perceptual biases reflect the proposed markedness factors). To support this claim, evidence has been extensively found from, but not limited to, segmental operations, tone genesis, and language evolution. However, if we look at the studies dealing with T3 sandhi, except the one of Huang (2001) which hinted that perceptibility may be the reason that T2 and T3 are involved in the sandhi process, barely any other studies have tried to explain T3 sandhi from a perception perspective. To fill this void, the current dissertation will explore Mandarin T3 sandhi from a perceptual angle, addressing the issue of how innate perceptual biases may have shaped its formation. To study the tonal grammar from the perspective of perception will enrich our understanding of why the Mandarin language is structured in the way as it is now. To observe how listeners perceive the tone sandhi pattern will also help to reveal the nature of perceptual biase that may underly tonal processes. Moreover, examining the role of listeners in the tonal grammar will give crucial information for understanding the suprasegmental structures of the language.

1.3. INTRODUCTION OF MANDARIN LEXICAL TONES

1.3.1 Phonology of Mandarin lexical tones

As is well-known, Mandarin Chinese is a tone language. A tone language is a language in which "lexical tones", or pitch variations, function in a phonemic manner so as to distinguish lexical meanings, regardless of their syntactic or morphological status. Depending on the pitch contours of the tones, the four citation tones in Mandarin could be categorized as either high level (T1), low rising (T2), low dipping (T3), or high falling (T4). An often cited example of this tonal manifestation is that in Mandarin, the syllable /ma/ may mean either 'mom', 'hemp', 'horse', or 'scold', depending on the lexical tone. Numerous other examples could be given in which the same syllable means different things when carrying different tones. Just to list a few, /ta/ can mean 'build', 'arrive', 'hit', 'big' when carrying T1, T2, T3, T4, respectively, and /t^haŋ / can mean 'soup', 'sugar', 'lie', 'hot' when carrying each of the four tones. Therefore, in Mandarin, lexical tone is crucial in determining the meaning of a syllable.

It is the pitch contour that distinguishes the different tones, and it is the F0 that defines the pitch contour. Recently, the F0 contours of the four Mandarin lexical tones have been depicted by averaging the production of multiple tokens by multiple speakers. Figure 1.1 gives the F0 contours of the four lexical tones, extracted from the six tokens of each tone, produced by eight native speakers (cited from Xu 1997). Other renderings of the four tones can be found in Wu (1982), Moore & Jongman (1997), Wang, Jongman & Sereno (2001), Gandour (1982) among others. As illustrated in Figure 1.1, Mandarin lexical tones may differ in both onset and offset. For example, T1 and T4 both start with a high onset while the T2 and T3 both start with a low onset. Concerning the offset, T4 and T3 both end with a low offset, whereas T1 and T2 end with a high offset. Mandarin lexical tones also differ in terms of contour. For example, T1 stays high throughout the syllable while T3 exhibits a dipping contour.

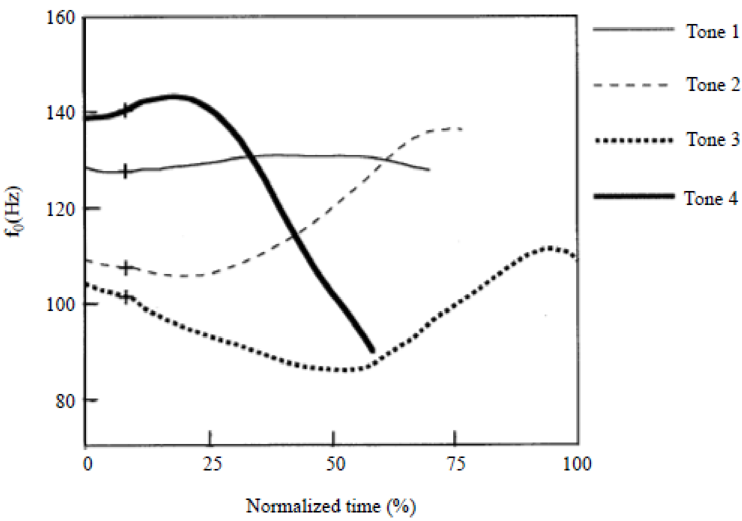


Figure 1.1. The mean F0 contour of four Mandarin tones in the /ma/ monosyllable produced in isolation. The time is normalized, with all tones plotted with their average duration proportional to the average duration of T3 (Cited from Xu 1997: 67).

Based on the pitch of the lexical tones, Zhao (1930) proposed a tone letter system to categorize Mandarin lexical tones. The pitch range of all tones were divided into five scales, namely from 1 to 5, in which 1 was defined as the lowest pitch, while 5 was defined as the highest pitch. Zhao described the four distinctive Mandarin tones as 55 (high tone), 35 (rising tone), 214 (low-dipping tone), and 51 (falling tone). Later Wang (1967) introduced features into the description of lexical tones. The features, such as [high]/[low], [rising]/[falling], and [level]/[contour], depict the pitch height as well as the pitch contour of the lexical tones. These features proposed by Wang (1967) have captured the acoustical quality of the different tones to a large extent, and are still widely used today. Table 1.1 summarizes the different descriptions of Mandarin lexical tones using tone scale, tone letter, and tone features.

Tones	Tone scale	Tone letter	Tone feature
T1	55	˥	high level
T2	35	˧˥	low rising
T3	214	˨˩˦	low dipping
T4	51	˥˩	high falling

Table 1.1 Descriptions of Mandarin lexical tones in tone scales, tone letters, and tone features.

When comparing the acoustic contours of the four tones in Figure 1.1 with the traditional (Chao's) phonological classification of the tones in Table 1.1, the following observations can be made. T1 starts with a high onset (5 in Chao’s tone scale), and stays high throughout the tone (55 in Chao’s tone letter), hence the classification ‘high level’. T2 has a low onset (a mid onset in Chao’s tone scale, which he denotes as 3) and rises towards the offset (5 in Chao’s tone scale), hence it is classified as ‘low rising’. T3 starts with a low onset (2 in Chao’s tone scale), goes down towards the middle (1 in Chao’s tone letter), and then rises towards the offset (4 in Chao’s tone letter). For this reason, T3 is classified as a ‘low dipping’ tone. T4 starts with a high onset (5 in Chao’s tone scale), and drops towards the tone’s offset (1 in Chao’s tone letter), and it is classified as ‘high falling’. It should be noted that Chao’s tone scale is intuitive, and it does not capture the acoustical contour of lexical tones very accurately. It can be seen from Figure 1.1 that the offset of T3 does not rise to a pitch level matching Chao's level 4, while T2 starts lower than a pitch level matching Chao's level 3. Nevertheless, the tone scale gives a direct visualization of the tones, and hence the scale is still widely used nowadays to describe the tone.

Specifically regarding T3, its rising offset is only observed when it occurs before a pause, and otherwise the high feature is deleted. Therefore, it is alternatively characterized as a [low] tone by some analysts (Wu 1982, Yip 2000, Shih 2007). Phonetic evidence of T3 being low can be seen from Figure 1.1: the offset of T3 is also much lower than the high offset of T1 and T2.

Although the crucial role of F0 in determining the lexical tone is without question, Mandarin lexical tones do differ in other dimensions. Among the four tones, T2 and T3 are claimed to be longer than the other two tones, while T4 tends to be the shortest (Dreher & Lee 1966, Kratochvil 1971, Chuang & Hiki 1972, Howie 1976, Nordenhake & Svantesson 1983). In terms of amplitude, T4 reveals the highest overall amplitude, while T3 presents the lowest amplitude (Chuang & Hiki 1972). In addition, Lin (1988) found that the double peaks pattern of the amplitude contour occurs consistently among T3.

Regarding how the intonation contour may affect the realization of lexical tones, it has been noted that lexical tones tend to preserve their pitch contour in running speech. Following Zhao, Chinese phoneticians have widely agreed that lexical tones in Mandarin Chinese are realized on monosyllables on a lexical level, whereas intonation is exhibited on a sentence level. Importantly, the realization of a lexical tone on the word level is fairly stable regardless of the sentence intonation (Shen 1992, Cao 2002, Lin 2006, Liu 2007). Wu (1995) proposed that when realized in words and phrases, the tone of each monosyllable tends to assimilate to its citation form.

Besides F0 variations, four other aspects regarding Mandarin lexical tones are worth mentioning. The first is the tone-bearing unit. Lexical tones are mainly realized by pitch variations, which reflect changes of fundamental frequency. Theoretically, a tone is expected to start from the onset of the voicing of a syllable (Howie 1974). Yet it should be noted that Mandarin has aspirated consonant onsets, which interrupt the realization of a pitch contour, in which case the tone should start at the voiced portion of the syllable, excluding the aspirated consonant. However, as Mandarin only allows nasal consonants /n/ and /ŋ/ for codas, and there are no onset or coda consonant clusters, syllabification is quite clear. In addition, the morpheme of Chinese (characters orthographically), which constitutes the minimal unit used for forming words, is also a syllable phonetically. In addition, previous experimental work has also argued for the alignment between the lexical tone and the syllable (e.g. Xu 1997, 1998). Therefore, I will use the syllable as the tone-bearing unit in the current dissertation. The second aspect is the “neutral tone”. A small amount of Chinese syllables do not have an underlying tonal form. Most often these syllables are function words or suffixes that do not occur at the beginning in bi-syllabic or multi-syllabic words (Yip 2001, Duanmu 2002). Compared to their tonal counterparts, neutral tones are always shorter in terms of duration and weaker in terms of amplitude, with their realization being highly influenced by the tonal context (Zhao 1968, Yip 2001, Duanmu 2001). Quite some work has been done in the research on neutral tones (e.g. Chen & Xu 2006). However, as the theme of this dissertation is to study the mechanism by which the contrastive role of lexical tones is processed and acquired, I will not go deep into neutral tones either. The third aspect is the tone register. “Register refers to the effects of initial consonants on the tones of the syllables in which they occur; typically, voiced initials condition a lower pitch and voiceless initials a higher pitch” (Norman 1988:53). In some Chinese dialects, the tones could be different when realized in different registers, but as I am focusing on Mandarin Chinese in this dissertation, which lacks the contrast on a register level (Wu 1969, Duanmu 2002, Wang 1999), I will not discuss tonal register in this dissertation. Fourth and last, Mandarin lexical tones may differ in phonation type. Duanmu (2006) claimed that T3 is murmured while other tones are not. Also, creaky voice may be present in tones T3 and T4 (e.g. Belotel-Grenié & Grenié 2004, Keating & Esposito 2006). However, Duanmu (2006) did not offer phonetic evidence of the murmuring, and neither Belotel-Grenié & Grenié (2004) nor Keating & Esposito (2006) argued that creaky voice is intrinsic to T3 and T4. Therefore, in the current dissertation, different phonation types will not be discussed.

Although the lexical tones display consistent pitch contours in Mandarin, in running speech, lexical tones do undergo contextual variations. Two types of tonal contextual variations should be considered separately: those introduced by articulatory factors, such as co-articulation, and those that cannot be fully accounted for by articulation mechanisms due to their phonologization, such as T3 sandhi. In the former case, identifying the co-articulated tones may depend on the experience with the input language and the knowledge of the frequency of tone co-occurrence in the input (Saffran, Aslin, & Newport 1996a, Saffran et al. 1996b). The latter case is of greater interest for understanding the possible biases in tone perception, as it involves a solid tonal grammar but yet its motivation has not been fully understood. Next, in section 1.3.2, I will briefly discuss the co-articulation of lexical tones in Mandarin, and in section 1.3.3, I will examine the different, yet unsatisfactory phonological interpretations of T3 sandhi before proceeding to review the perception studies on lexical tones in section 1.4.

Regarding T3 sandhi, this dissertation hypothesizes that universal perceptual biases underlie the occurrence of T3 sandhi, and aims to identify these biases. Native listeners' perception of T3 sandhi is expected to be simultaneously affected by universal biases and native grammatical knowledge (e.g., the frequency of tonal sequences in the native spoken input). Furthermore, according to the hypothesis that a perceptual bias favors the occurrence of T3 sandhi, we would expect that non-native listeners without any Mandarin input would also display a consistent preference for certain tonal structures.

1.3.2 Co-articulation of lexical tones

In Wu's (1982) qualitative analysis of the co-articulation of lexical tones, he claimed that the realization of lexical tones is fairly stable across different tonal contexts, which assures their phonemic function. He considered monosyllables and co-articulated bi-syllables to function as the minimal prosodic domain in Mandarin. Wu also pointed out that tonal co-articulation occurs primarily in syntactically immediate neighbors, or within a prosodic word. Shen (1992) contended that the co-articulation of lexical tones was a process of bi-directional assimilation, meaning that the preceding tone not only exerts influence on the following tone, but the realization of the preceding tones also varies due to the anticipation of the following tone. She also pointed out that the onset and the offset of the tone, as well as the overall height, are affected by the tonal context surrounding it. Lin (1992), on the other hand, supported a narrower domain of tone coarticulation, arguing that the carry-over effect from the preceding tone only influences the onset of the following tone, while the anticipatory effect of the following tone is limited to the offset of the preceding tone. His experiments also demonstrated that the change of pitch height could be seen as a result of contextual variations, and the pitch contour of the target tone was not influenced dramatically.

More recently, experimental studies have been conducted in order to test the phonetic implementation of tone co-articulation. In Xu (1994), native Mandarin speakers produced either T2 or T4 as the middle syllable in tri-syllabic sequences. The data showed that as the middle tone is less salient than the first or the third tone, its realization is largely influenced by the tonal context. The author divided the tonal context surrounding the middle tone as either compatible or conflicting. Here, "compatible" was defined as the F0 offset of the preceding syllable resided on the same or similar height as the F0 onset of the middle syllable, while the F0 onset of the following syllable resided on the same or similar height as the F0 offset of the middle syllable (a schematic illustration would look like ¯ \ _). "Conflicting" refers to the combinations in which the onset and offset of the middle tone maintain heights that are opposite to its pre and post context (a schematic illustration would look like ¯/_). It could be observed from the production that the middle tones produced in a "compatible" context generally preserved their underlying forms, whereas wide variation and even an opposing direction of the F0 contour was observed for the middle tones in the "conflicting" context.

In 1997, Xu examined the production data of the bi-syllabic sequence /mama/ for all 16 possible tonal combinations. In his experiment, eight male speakers produced /mama/ sequences embedded in four carrier sentences which differed in pre-target and post-target pitch height. To put it differently, taking the /mama/ sequences as the target, the F0 offset of the syllable preceding the target could either be high or low, and the F0 onset of the syllable following the target could either be high or low too. The carrier sentences were constructed in such a way that /mama/ functioned as a prosodic word in object position. The target surrounding tones only introduced a slight effect: a high offset of the preceding context resulted in a higher F0 contour of the first /ma/ in the /mama/ sequence, while a high onset of the following context lowered the F0 of the second /ma/ in the /mama/ sequence. On the other hand, when inspecting the contextual variation locally within the /mama/ sequence, substantial deviation from the underlying tonal contour occurred. The F0 contour of the second tone showed enormous variation due to the carry-over effect from the first syllable, whereas the realization of the first tone was fairly congruent to its underlying form. As the second tone started from the offset of the previous syllable as a continuation, the largest difference in pitch contour due to the carry-over effect occurred at the onset of the second syllable. The deviation decreased over time, and at the offset of the second syllable the pitch contours of the same tone tended to cluster. For the anticipatory effect, slight deviations from the underlying form were observed, such that a high onset of the second syllable lowered the F0 contour of the first syllable while a low onset of the second syllable raised the F0 contour of the first syllable. Xu summarized the local carry-over effect as assimilation, in which the onset of the second syllable is assimilated to the offset of the preceding syllable. On the contrary, the anticipatory effect is dissimilatory in that the first syllable varies in the opposite direction from the onset F0 height of the following syllable. To summarize, the co-articulation of Mandarin lexical tones is mainly observed within syntactical domains and within prosodic words, the carry-over effect is substantial and mainly assimilatory, whereas the anticipatory effect is only small and mainly dissimilatory.

Based on the results obtained from these studies, it is clear that lexical tones in Mandarin are subject to substantial contextual variations. These variations mainly occur within the prosodic words consisting of two or three syllables. The co-articulation of lexical tones in natural speech suggests that, to successfully establish representations of lexical tones, listeners have to normalize variations of the identical tone encountered in the input to the same category. Specifically regarding the lexical tone normalization, how listeners perceive the variations of lexical tones, and how infants learn to categorize Mandarin tones may shed light on the natural perceptual space of specific tones. The categorization of individual tones may serve as a basis for understanding the perceptual biases in perceiving Mandarin T3 sandhi. Therefore, one main focus of the current dissertation is the categorization of Mandarin lexical tones.

1.3.3 T3 Sandhi in Mandarin

In Mandarin, T3, the low dipping tone, goes through the most contextual variations among all four tones, and it shows quite consistent variation patterns.

As a result, the contextual variation of T3 has been the most studied aspect of tone sandhi. Two different sandhi processes apply to T3: first, if a T3 is followed by any tone other than a T3 or a neutral tone, it changes into a low-falling tone without a rising offset. In other words, T3 changes from fall-rising to falling, where the rising part of the contour is deleted. Second, if two T3s occur together, it is required that the first T3 in the T3T3 sequence changes into a rising tone. Acoustically, the derived rising tone is highly similar to T2 in terms of its pitch contour. The latter process is referred to as 'T3 sandhi'. Consider an example mentioned earlier: in the bisyllabic sequence /haomi/ ‘good rice’, both syllables have an underlying T3. Due to T3 sandhi, this sequence is realized as /hao/ T2 followed by /mi/ T3. In Mandarin, /haomi/ with underlying T2T3 means ‘millimeter’. For Mandarin listeners, ‘good rice’ and ‘millimeter’ sound the same. Concerning the domain of T3 sandhi, Lu (1986) considered it to be restricted by speech rhythm. He vaguely pointed out that “Tone sandhi occurs when monosyllabic morphemes or words are combined into compound words or idiomatic phrases”, or “phonological words”. Yet, he did not clarify whether the phrases referred to syntactic phrases or phonological phrases. Later, Shen (1992) and Shih (1997) confirmed Lu's idea, and argued that bi-syllables within a single phrase are the domain of T3 sandhi, while a prosodic boundary, such as prosodic word boundary and phrase boundary, blocks the T3 sandhi process. Kuang & Wang (2006) again found that T3 sandhi is a dissimilation process that does not cross phrase boundaries, and they argued that large pauses indicate phrase boundaries.

So far, no consensus has been reached on the phonological interpretation of T3 sandhi. Duanmu (2007) reviewed two approaches, which he called ‘the tree-only approach’ and ‘the stressless-foot approach’. The first approach assumes that T3 sandhi is sensitive to both syntactic tree and tempo of speech (Cheng 1973, Shen 1994); the second approach suggests that T3 sandhi applies in bisyllabic units in a left-to-right direction (Shih 1986, 1997, Chen 2000). Yet, neither approach is able to account for why T3 sandhi will generate different outputs for the same syntactic structure. Duanmu (2007) proposed that T3 sandhi is cyclic starting from each foot and the process need not apply between two cyclic branches. Furthermore, he argued against the idea that T3 sandhi is constrained by the tempo of speech. However, he did not provide phonetic evidence to support this assumption. Most importantly from the viewpoint of the current dissertation, he failed to answer the crucial question why T3 sandhi occurs.

Some researchers consider T3 sandhi as a dissimilation process which can be incorporated into the Obligatory Contour Principle (OCP; Leben 1973). OCP prohibits the co-occurrence of two identical elements. Avoidance of consecutive consonants of the same place of articulation has been widely observed in Semitic languages (e.g. McCarthy 1986). OCP has been extended to the domain of lexical tones to explain the dissimilation in T3 sandhi (Cheng 1973, Yip 1908, 2002, Zhang 2007). However, OCP fails to explain why only T3T3 is subject to dissimilation while other tone sequences survive the co-occurrence of identical values, such as T1T1, T2T2 and T4T4. Other researchers hold the opinion that T3 sandhi involves no categorical change of lexical tones, but functions as a simplification by deleting part of the contour of the tone (Chen 2000, Yip 2000, Lee-Schoenfeld 2002, Lee-Schoenfeld & Kandybowicz 2008). From this point of view, the underlying form of T3 is considered to be specified as a sequence of [mid] [low] and [high] features, and the derived T2 arises by deleting the [low] feature from the underlying T3. However, the feature deletion account only succeeds in explaining T3 when only two syllables are involved in the T3 sandhi domain.

Efforts have been made from a production perspective to understand T3 sandhi. Fang (2010) considered T3 sandhi as a reorganization of articulatory gestures. In her experiment, speakers produced two T3s with a pause in between, and as speed of production increased, the pre-pause T3 finally changed into a rising tone. Therefore, she argued that under time pressure, a rising tone functions so as to synchronize a high feature and a low feature. However, this interpretation is unsatisfactory, because T3 lacks a high feature and T3 is actually distinguished from other tones by having the low feature (Yip 2000, Wu 1982). Moreover, even with a rising offset, a T3 does not go as high as a T2. If production efficiency is concerned, while the output should stay as close as possible to its underlying form, it is rather expected that the first T3 is synchronized into a low tone rather than a rising tone. Yeh (2010) found that compared to the production of T3T2 sequences, there were more pitch variations in the production of a T2T3 sequence by native Taiwan Mandarin speakers, which suggests that in terms of articulation, a T3T2 is easier than a T2T3. Her finding renders even the articulatory motivation of the asymmetrical pattern of T3 sandhi doubtful: if a T3T2 sequence is easier to produce than a T2T3 sequence, then why is a T3T3 not sandhied into a T3T2?

Even though a satisfactory account of T3 sandhi is lacking, T3 sandhi does occur systematically. T3 not only applies between Mandarin T3 syllables, but it even occurs when a Mandarin T3 syllable is followed by an English word which starts with a weak syllable (Cheng 1968). Cheng gave the example that, in the context of '/hao/ professor' ('a good professor'), the Mandarin syllable /hao/, whose underlying tone is T3, changed into a rising tone in production, whereas in the context of '/hao/ *student'* ('a good student'), /hao/ remains low. As the weak syllable /pro/ can be described as low in terms of pitch while the strong syllable /stu/ is high, and taking into consideration that the most important feature of T3 is [low] (Yip 1980, Milliken 1989, Wu 1988), it seems that once there are two consecutive [low]s, T3 sandhi is triggered. The fact that T3 sandhi occurs even in a code switch situation involving a non-Mandarin element, serves as strong evidence that at least in production, T3 sandhi is fully productive. To put it differently, once a context meets the requirements for T3 sandhi, i.e. having two successive low tones, it is compulsory for T3 sandhi to apply. However, by looking at only the native speakers, it is hard to infer whether Cheng (1968)'s production data was the result of the knowledge of grammar, or it reflected innate biases.

As we have seen so far, although there is no doubt that T3 sandhi is psychologically productive as it even applies in code switch situations, a satisfactory phonological account of T3 sandhi is still missing. To provide a clearer picture of what exactly the T3 sandhi process entails, and why consecutive T3s are sandhied into the specific pattern of [rising] + [low], we need to understand what constitutes knowledge of T3 sandhi for language users, and what drives the process in the first place. In earlier attempts to answer these questions, a great deal of efforts have been made by analyzing production data, by referring to articulatory mechanisms, and by testing the perception of a T2 derived from T3 in a T3 sandhi context in isolation (to be discussed in section

1.4.2), but an important aspect of phonological grammar formation, i.e. the influence from perception, has been largely neglected. To include this missing aspect, my own research will focus on the perceptual biases that may favor T3 Sandhi to occur.

As mentioned earlier, besides involving a change of a T3 into a T2, another crucial aspect of T3 sandhi is that T3 sandhi is positional and asymmetrical; T3 sandhi only influences the first syllable in an adjacent T3T3 sequence, but the second T3 in the sequence remains fairly intact. Regarding the directionality of tone sandhi, across various Chinese dialects (not restricted to Mandarin T3 sandhi), two patterns have been proposed: either left dominant, or right dominant. In the former situation, the left-hand syllable preserves its underlying form while the right-hand syllable goes through the sandhi; conversely, for right dominant sandhi, the right-hand syllable preserves its base tone, whereas the left-hand tone is neutralized with another tone (Yue-Hashimoto 1987, Chan & Ren 1989, Chan 1995, Chen 2000, reference from Zhang 2007). Both directions of tone sandhi are observed in dialects in China (see Zhang 2007 for a typological summary). After examining the tone sandhis in various Chinese dialects, Zhang (2007) argued that, for the left dominant tone sandhi process, it is mainly progressive coarticulation, i.e. the tone of the first syllable is carried over to the following tone. However, right dominant sandhi tone processes, such as T3 sandhi in Mandarin, frequently involved insertion of paradigmatic neutralization, meaning that the first tone is neutralized phonologically with another tone, but not assimilated to the final tone (Zhang 2007). Zhang (2007) described the directionality in tone sandhi as follows (cited from Zhang 2007, pp 275):

If a language L has both left-dominant and right-dominant tone sandhis,

a. if its left-dominant sandhi involves default insertion/paradigmatic neutralization, then its right-dominant sandhi also involves default insertion/paradigmatic neutralization

b. if its right-dominant sandhi involves tone extension, then its left-dominant sandhi also involves tone extension.

According to Zhang, default insertion/neutralization is the unmarked sandhi pattern for right dominant sandhis. He also argued that contour tones tend to occupy the final position, as the final lengthening allows the tone to fully realize the contour. Looking at T3 sandhi in Mandarin within Zhang's model, first, the contour tone T3 is favored to be at a final position; second, the neutralization of the first T3 to a T2 is the default format in a right dominant sandhi process. Hence, it seems that T3 sandhi is unmarked. Bearing in mind that unmarked structures are assumed to be universally preferable (Kager 1999), and humans are biased for such universal preference (see section 1.2), it is reasonable to suspect that the grammar of T3 sandhi may also originate from universal biases; that is, perceptibility may play a role in shaping the directionality of T3 sandhi.

1.4 LEXICAL TONE PERCEPTION BY ADULTS

1.4.1 Perceptual cues of lexical tones

As mentioned earlier, the acoustic correlates of lexical tones are fundamental frequency (F0), amplitude, and intensity, among which the F0 serves as the most important correlate that defines a tone. Therefore, in this section I will go through the studies that discuss how pitch variations realized in lexical tones are perceived by listeners. In the current section 1.4.1, I will summarize pitch information that is relevant for tone perception, and in section 1.4.2, I will review the studies dealing with tone perception by native listeners. Studies that compare the perception of lexical tones by native versus non-native listeners will be reviewed in section 1.4.3.

It is widely accepted that, consistent with the feature description of lexical tones, the lexical tone contour such as the F0 shape, F0 height, and F0 direction, serves as a perceptual cue for the identification of tones. For dynamic tones, the onset and the offset of the pitch contour determine the direction of F0, and together with the turning point of the contour (if any), they work as informative cues for the recognition of the tones (Gandour 1978, 1981, Gandour & Harshman 1978). Gandour & Harshman (1978) tested the perception of both native tone language listeners and native non-tone language listeners in 13 different pitch patterns imposed on synthesized speech-like sounds, and they found that average pitch height, pitch direction, extreme endpoint, and pitch slope function as cues for the perception of lexical tones. Comparing the discrimination of lexical tones between native listeners and non-native listeners, Gandour (1978, 1981) noted that native tone language listeners relied more on pitch slope and pitch direction for tone discrimination, while average pitch height played a more important role for non-tone language listeners. Fok (1974) tested native Cantonese listeners and found that the height of F0 together with the F0 contour and F0 direction worked as cues for tone identification. Taking into consideration that Cantonese has tonal contrast between level tones, while Mandarin lacks such contrast, it is plausible that Mandarin listeners may rely more upon the F0 direction rather than the overall F0 height for tone identification.

More specifically, regarding the perception of lexical tones in Mandarin, both F0 height and F0 contour are claimed to influence tonal recognition. Massaro, Cohen & Tseng (1985) studied the perception of Mandarin tones by native listeners and demonstrated that both F0 height and F0 contour are effective cues for the recognition of lexical tones. Gandour (1984) tested Mandarin listeners for their discrimination of lexical tones, finding that Mandarin listeners rely slightly more on F0 direction than on F0 height. As mentioned earlier, the performance of Mandarin listeners corresponded to the tonal distribution in Mandarin, which lacks the contrast between tones with same contour but different height. On the contrary, Zue (1976) argued that the average height of the F0 constitutes an essential cue for tonal identification, which suggests that native listeners possess knowledge of the absolute height of each tone. In sum, it seems that native listeners have knowledge of the shapes of tone contours as well as of the absolute height of tones, and are able to perceive the detailed acoustics of lexical tones quite accurately.

As can be seen in Figure 1.1, specifically the T2 and T3 tones share a similar F0 shape, which first falls at the beginning of the syllable and then rises towards the end of the syllable. Early studies on the perceptual cues of T2 and T3 have suggested that the duration of the pre-rise portion of the F0 contour is crucial for the discrimination between T2 and T3: if the pre-rise part is short and undetectable, listeners tend to identify that token as a T2, while on the other hand, if the turning point is far from the F0 onset while the initial fall is perceivable, native listeners would identify that token as a T3 (Zue 1976, Blicher, Diehl, & Cohen 1990). Later, Shen & Lin (1991) created a continuum ranging from T2 to T3 in which the location of the turning point, as well as the amplitude of the pre-turning point at which the F0 falls was manipulated. In a forced choice paradigm, they asked native Mandarin listeners to label the stimuli as either T2 or T3. They found that the turning point, together with the initial downward shift, forms the perceptual cue for discrimination between T2 and T3. When the correlation between the location of the turning point and the degree of the initial fall was interrupted, listeners showed difficulties in discriminating the contrast.

Based on these studies, it can be seen that multiple cues such as F0 shape, height, and direction contribute to the perception of lexical tones. Second, at least in Mandarin, the direction of the F0 contour serves as a crucial cue for tone perception, and native listeners are sensitive to even finer-grained cues, and they have knowledge about the relation between the timing of pitch direction change and the amplitude of the pitch change. Hence, in order to establish tonal categories, Mandarin learners need to be knowledgeable of the phonetic information that is relevant for distinguishing tonal categories.

1.4.2 Lexical tone perception by native adults

1.4.2.1 Perception of citation tones

Native Mandarin listeners' ability to accurately perceive Mandarin lexical tones in their citation form is widely-accepted as a fact. Howie (1976) carried out a series of experiments testing the perception of monosyllables realized with the four original tones, synthesized tones simulating the original tones, monotones, and neutralized tones in which the F0 was removed. Three out of nine participants had an accuracy rate of 100% for the identification of the original tones, and across the four different tone bearing syllables even the syllable which was identified most poorly still had an accuracy rate of 91.7%. Moreover, the listeners reached a high accuracy in identifying the synthesized tones, which were generated based on averaging the citation tones produced by native speakers. Nevertheless, listeners had difficulty in recognizing the monotones and neutralized tones. The results of Howie (1976) offered experimental evidence that native listeners are accurate and consistent in recognizing lexical tones in citation form. Lin (1988) also demonstrated that, based solely on the F0 contour, the accuracy rate of Mandarin tone recognition could be as high as 94%. Later, Lin & Repp (1989) found that native listeners identifying Taiwanese tones succeeded by processing only the F0 information. Therefore, it is safe to draw the conclusion that native tone language listeners have established stable representations of lexical tones, and that their identification of lexical tones is consistent.

1.4.2.2 Perception of lexical tones in context

In reality, however, lexical tones are hardly produced in isolation, but as is the case with segments, they are produced in running continuous speech. As covered in section 1.3.2, the F0 contour of the lexical tone could be distorted due to co-articulation; hence, the same tone presented in context could well be perceived differently compared to the same tone realized in isolation. Consequently, the recognition of co-articulated lexical tones is more vulnerable than the perception of citation tones. Lin & Wang (1985) ran identification experiments with native Mandarin listeners and found that a high level tone with a constant height tended to be perceived as a rising tone when followed by a tone whose onset was higher than the level tone. Xu (1994) tested native listeners' perception of the middle tone in tri-syllabic sequences, and found that when the original context was present, the identification of the middle tone was quite successful—with an accuracy rate above 80%— regardless of whether the context was compatible or conflicting (see also section 1.3.2). Therefore, native listeners are fairly successful in compensating lexical tone variations by relying on the context. However, when the original first and the third syllable were replaced with white noise, the identification of the middle tones in isolation remained accurate only if they were cut from the compatible context, while for middle tones cut from a conflicting context, the accuracy of identification dropped to below a chance rate. More importantly, when the first syllable and the third syllable swapped positions, the perception of the middle tone was restricted by the new context, and listeners tended to identify the middle tone as a tone that was compatible with this new context. The results of Xu (1994) demonstrate that native listeners are aware of the contextual variations of lexical tones, and that tonal co-articulation is taken into consideration when they identify a tone in continuous speech. For heavily co-articulated tones, the proper tonal context is necessary to assure a high accuracy of recognition. Hence, learners of Mandarin, who encounter not only isolated tones but also co-articulated tones, must learn to normalize the variations introduced by context in order to establish proper tonal categories.

Besides the contextual variations caused by co-articulation, another important external factor that may influence tone perception is the speaker's pitch range. It could very well be that the low tone of a speaker with a high pitch range is physically identical or overlaps with the high tone produced by a speaker with a low pitch range. As a result, in order to process the tonal information efficiently, listeners need to normalize different speakers' voices and neglect absolute tonal height variations introduced by the intrinsic voice quality of the speaker. Leather (1983) tested native listeners' identification of tones changing along two T1-T2 continua produced by two different speakers respectively. He found that when the ambiguous tones taken out of each continuum were presented in isolation, in spite of having identical pitch height and contour, listeners identified these as different categories due to the different pitch height of the speakers. Later, Moore & Jongman (1997) tested native listeners' identification of Mandarin T2 and T3 that were preceded by a precursor sentence produced by a speaker whose voice either had a high pitch or a low pitch. They created continua changing from T2 to T3, and found that the relation between the pitch of the precursor and the identification of the steps along the continua were inverted; the same step tended to be perceived as a high tone (T2) when following a low precursor, while it tended to be perceived as a low tone (T3) when following a high precursor. Moreover, they also found that for the continuum created along a temporal scale, in which only timing of the turning point was manipulated while the F0 information (onset, lowest point, and offset) remained constant, the precursor sentence failed to introduce a significant difference in tone identification. The study of Moore & Jongman (1997) clarified an important scenario in Mandarin tone perception: speaker normalization only occurs when listeners encounter speech input that varies along the same acoustical dimension as the voice of the speaker, namely in terms of pitch.

To summarize, lexical tones realized in running speech are highly variable, and tonal context as well as external information, such as the pitch range of speakers, play a crucial role in the identification of lexical tones. Native listeners need to normalize the tonal variations they encounter in the speech input in order to establish representations of tonal categories. Yet, due to the phonemic function of lexical tones, native listeners must remain sensitive to variations that may shift a token from one tonal category to another. On the other hand, for non-native listeners, perception of such tonal variations may reflect the innate perceptual pattern that has not yet been tuned by the tonal distribution in speech. Hence, studying non-native listeners' responses to tonal variation may reveal the innate perceptual biases in lexical tone processing. Indeed, native listeners and non-native listeners do perceive the tonal variations differently (a detailed discussion will come in section 1.4.3). Furthermore, to obtain a more thorough view of the role of innateness in tonal perception, it is necessary to investigate how young infant learners, who have not yet received substantial speech input of any language, learn to categorize lexical tones under contextual variations. Infant studies may provide us with valuable clues about the natural perceptibility of different tonal contrasts.

1.4.2.3 Perception of T3 sandhi

As mentioned in section 1.3.3, T3 sandhi is the phonological rule that restricts the co-occurrence of T3s, and requires the first T3 in a T3T3 sequence to change into a rising contour. T3 sandhi is the most important phonological sandhi pattern observed in Mandarin, and accordingly it has received much attention in the research on Mandarin lexical tone processing. The two tones that are involved in T3 sandhi, the T2 and T3, are claimed to be the most difficult contrast to discriminate, for both native and non-native listeners (Kiriloff 1969, Chuang & Hiki 1972, Gandour 1978, Li & Thompson 1978, Hume & Johnson 2003). Among the numerous studies addressing T2 and T3, Huang (2001) tested both native Mandarin listeners and native English listeners on their discrimination of all possible tone pairs in Mandarin. The results demonstrated that although both language groups reached a high accuracy rate when discriminating T3-T2 and T2-T3 pairs (native English listeners above 84%, native Mandarin listeners above 89%), both groups made more errors in discriminating between T2 and T3 compared to other tone pairs. Meanwhile, it took both groups the most time to give a correct response for the discrimination between T2 and T3. Hume & Johnson (2003) also tested the discrimination between T2 and T3 by native Mandarin listeners and native English listeners. They used the ratio of 1/reaction-time as a measurement of the perceptual distance, and showed that for both groups of listeners, the smallest perceptual distance was observed between T2 and T3.

Most studies on the perception of T3 sandhi have focused on the perception of the sandhied T3, which is realized with a rising pitch contour. Much attention has been given to the issue of whether or not the rising pitch contour of the sandhied T3 is perceptually identical to an underlying T2. Discrepancy still remains regarding this issue. On the one hand, it has been argued that the sandhied T3 is identical to an underlying T2. Wang, Li, & Brotzman (1963) carried out identification tasks with native Mandarin listeners, and showed that underlying T2T3 and T3T3 sequences, when produced naturally by a native speaker, are indistinguishable for native listeners, which supports the conventional idea that the sandhied T3 is identified as a T2. Speer, Shih, & Slowiaczek (1989) also reported evidence that when native listeners were presented with naturally produced non-words which were composed of two existing T3 syllables, they identified these T3T3 sequences as T2T3 half of the time. Peng (1996) found that though the sandhied T3 and an underlying T2 are acoustically different, native listeners were unable to keep track of this subtle difference, which resulted in the indistinguishability between a sandhied T3 and a T2 in perception. However, when the sandhied T3 was presented in a T3T3 context, which was necessary for triggering the sandhi process, listeners were able to categorize the derived rising contour as belonging to the T3 category. On the other hand, Zee (1980) examined the phonetic realization of the sandhied T3, and argued that the sandhied T3 had an overall lower F0 contour than an underlying T2. Chen & Yuan (2007) examined a large telephone speech corpus and found that the sandhied T3, though having a rising contour, was different from an underlying T2 in terms of the magnitude of the rising part in F0 and the time span of F0 rising. Later Chen, shen, & Schiller (2011) found evidence from the encoding of speech production that both the T3 with a low pitch contour and the sandhied T3 are stored in the mental lexicon, which supports the idea that native listeners do have separate representations of the sandhied T3 and the T2.

The above-mentioned studies are highly informative in revealing the perceptual knowledge of T3 sandhi, and researchers seem to agree on the following points. First, the sandhi process involves neutralization between the perceptually closest tones. Second, the sandhied T3 carries a rising F0 contour that is at least similar, if not identical, to the F0 contour of an underlying T2. Though listeners are able to perceive acoustical differences between a sandhied T3 and an underlying T2 in certain tasks, this does not necessarily mean that they use this difference to distinguish a sandhied T3 from an underlying T2 in online speech processing. However, bearing in mind that T3 sandhi is the process that restricts the co-occurrence of two T3s, the domain of T3 sandhi has to be at least a bi-syllabic sequence. The previous studies, instead, have focused on the perception of T2s and T3s in isolation, and have tried to establish whether or not the sandhied T3 and an underlying T2 are perceptually identical, while little attention has been paid to the mental process occurring when listeners encounter two consecutive T3s. Hence, their studies are limited to the perception of the components of the T3 sandhi process, but the perceptual knowledge of the T3 sandhi process is still unknown.

Neither is there a clear view regarding the motivation and psychological representation of T3 sandhi. Only a few studies have addressed the issue of why T3 sandhi occurs in such a form. Xu (2000, 2004) discussed the Target Approximation Model in which he assumed that the realization of lexical tones in speech goes by means of a process in which a pitch movement tries to approach a pitch target. Minimal time is needed for the human articulators to produce a sound, which constrains the speed of pitch change during production. Within this framework, Xu hypothesized that contextual tonal variations can be divided into two kinds. In the first, the tonal target implemented in context remains close to its underlying form, so that the tonal variation is due to the articulatory constraint on approaching the target. The second kind deals with the case in which the tonal target itself deviates from its underlying form and the articulatory implementation was to accomplish the new target. The Target Approximation Model is not able to generate a rising contour as the realization of a low tone. Taking into consideration that the sandhied T3 is much shorter than an underlying T3, Xu assumed that the sandhied T3 is compromised to have a rising contour similar to a T2 in order to combine the falling and rising information encompassed in an underlying T3. In this case, he assumed that the tonal target for production of the sandhied T3 was not the underlying T3, but a complex rising contour similar to T2. However, he also pointed out that if the to-be-realized target was supposed to be similar to T3 in terms of the F0 shape, then the higher pitch height and larger pitch range realized in the rising contour of the sandhied T3 could not be explained by his hypothesis. Moreover, enlightening though his idea is, Xu has not yet carried out experimental work to test whether humans behave in the manner predicted by his hypothesis.

As is evident from these studies, our understanding of T3 sandhi is still limited, and the perceptual knowledge of T3 sandhi is also unclear. By now it seems that no solid factor favoring a T3T3 to be neutralized into a T2T3 has been identified. Yet, it is reasonable to speculate that this asymmetrical pattern should be reflected in perception, and may even be driven by perception. As mentioned in Section 1.2, Ohala (1990, 1993, 2012) has attributed the dissimilation to the hypercorrection of the listeners. Whether the theory of Ohala, however, applies to the T3 sandhi case needs further testing, and it is at least reasonable to speculate that listeners may play a role in the phonologization of T3 sandhi.

To take a different perspective, the results of some perceptual studies may give us some hints about why T3 sandhi occurs in its present form. When Chuang et al. (1972) tested native listeners in order to find the perceptual cues for the four lexical tones in Mandarin, they observed that the mis-identification of a T3 as a T2 occurred twice as much as the mis-identification of a T2 as a T3. When studying how the timing of the turning point and degree of initial falling may affect the discrimination between T2 and T3, Moore & Jongman (1997) found that, as long as the initial falling is trivial, listeners could identify a token with a late turning point as a T2, while a late turning point was actually a cue for a T3. They implied that the identification of a T2 tolerates more variation than that of a T3; T3 requires a large initial falling of the F0 contour as well as a late turning point, while a small initial falling would already be enough for identifying a T2. The findings in Chuang et al. (1972) and Moore & Jongman (1997) that a T2 may have a wider perceptual category compared to a T3 came as a "by product" when answering their other research questions, and no strict experimental work has been carried out to test the distribution of the perceptual space of T2 and T3. Nevertheless, both studies suggested that for native listeners, the perceptual space of T3 is rather small when compared to that of T2, and if this is the case, we may have a new angle to look at T3 sandhi. We would expect that a T3 could be perceived as a peripheral instantiation within the T2 category, but T2 by no means could be perceived as a non-prototypical exemplar of T3. As a result, in a T2-T3 discrimination task in which the first item functions as a referent, presumably, if the referent is an exemplar representing a wider category, namely T2, listeners may accept the following T3 as a peripheral exemplar of the wider category. However, if the referent is an exemplar of a stricter category, namely T3, then the following T2 would be perceived as lying out of the category. If so, we would expect the discrimination of T2 and T3 to be easier when the discriminated pair occurs in a T3T2 order rather than in a T2T3 order. Taking this a step further, would it not be possible that T2 and T3 sound more identical when presented in a T2T3 order rather than in a T3T2 order? If this hypothesis is correct, and taking into consideration that T2 has the smallest perception distance to a T3, there would be a perceptual basis for T3 sandhi in Mandarin; T2T3 is the perceptually most similar sequence to T3T3, in that T2 is similar to T3, and perceptually a T2T3 sequence largely shares the same structure as a T3T3, in that both sequences are perceived as being composed of identical tones.

However, so far this is only an assumption, and several steps need to be taken to test whether it actually occurs during online processing as assumed. First, it must be tested whether a T2 has a less distinct boundary than a T3, and hence allows for more variation. Second, regarding the discrimination between Mandarin T2 and T3, it is necessary to test whether presenting the two tones in a T3T2 order is easier for successful discrimination than presenting them in a T2T3 order. Third, it needs to be tested whether a T2T3 sequence sounds more similar than a T3T2 compared to a T3T3. As the focus of the current dissertation is to find possible universal biases that may encourage the occurrence of T3 sandhi, and as such biases are not supposed to come from speech exposure, we need to extend the scope beyond native Mandarin listeners. As native listeners have knowledge of lexical tones, T3 sandhi, as well as distributional pattern of Mandarin tones, it would not be surprising if they tend to misperceive a T2-T3 pair as having identical tones. Hence, by only looking at the performance of native listeners, it would be difficult to tell whether this misperception is driven by an innate perceptual bias or, alternatively, by the linguistic knowledge that native listeners have acquired from the tonal distribution in ambient input. Therefore, if we were to explore the innate bias affecting the perception of Mandarin T2 and T3, we have to test naïve listeners who lack the tonal representation. In this case, non-native adult listeners serve as a suitable population. If non-native listeners, who have no representation of lexical tones and no knowledge of Mandarin T3 sandhi, demonstrate consistent patterns in perceiving Mandarin T2 and T3, then these perceptual patterns would most likely reflect innate biases rather than acquired knowledge. Even more revealing, investigating how infants without substantial speech input and knowledge of any language would perceive this tone contrast, and whether order of presentation interferes with infants' discrimination of Mandarin T2 and T3, would give more direct evidence regarding the natural biases in Mandarin tone perception.

Besides testing the adults, in this dissertation, I will also take a developmental perspective, running Mandarin lexical tone perception experiments with native and non-native infants as well as native and non-native adults. The infants serve as a good population to learn about the initial state of speech perception and possible innate biases. Comparing the performances of Mandarin and Dutch infants at an early stage, two populations whose native language will differ later in life, will assist us in finding traces of how language shapes initial perception. The adult groups have achieved the end-state perception of language, and by comparing the performance of Mandarin and Dutch listeners, we could find out how lexical tones are perceived linguistically and non-linguistically, and whether consistent perceptual patterns occur regardless of different language backgrounds. Referring to the infant data, we will have the chance to learn how the initial biases remain or are shaped by language input, and whether the phonological grammar in return shapes the initial perception of lexical tones. In the next section, 1.4.3, I will review the literature regarding native versus non-native perception of lexical tones. In section 1.5, I will introduce the studies discussing speech perception in early infancy.

1.4.3 Comparing tone perception between native and non-native listeners

As mentioned earlier, lexical tones are primarily realized by pitch variations, and the acoustic correlate of such variation is the F0. However, pitch variations are not only present in tone languages, but they are also present in non-tone languages for intonational or pragmatic purposes. Tone languages and non-tone languages differ in terms of the function of F0 variations. In tone languages, lexical tones serve the purpose of differentiating meaning on a lexical level, but in non-tone languages, lexical meaning does not change as a function of pitch variation. In psycholinguistic studies, it has been observed that the acquisition of lexical tones by non-tone language speakers is extremely difficult, both for perception and for production (e.g. Kiriloff 1969; Bluhme & Burr 1971, Shen 1989).

Processing lexical tones categorically involves three aspects: first, listeners need to have underlying representations of each tone; second, they need to ignore the variations within the same tonal category for efficient decoding of the speech signal; and third, to assure the phonemic function of lexical tones, they need to remain sensitive to the variations, subtle though these can be, that change a tone from one category to another (see Francis, Ciocca & Brenda 2003, Hallé, Chang & Best 2004). The tone perception between native and non-native listeners differs substantially in all three aspects.

First, the representation of lexical tones is an intrinsic feature of morphemes (and hence, syllables) in Mandarin, and each syllable has to have a tone. There are abundant 4-way minimal tone pairs. The high accuracy in identifying citation tones serve as evidence that natives have stable representations of lexical tones (e.g. Howie 1976). For non-tone language speakers, tone does not function distinctively in their language, and presumably, they do not have a representation of lexical tones, and are not able to categorize a syllable according to its tone.

Regarding the within- versus cross-category discrimination of sound categories, a classical paradigm to investigate native and non-native sound categories is the Categorical Perception experiment (CP). CP assumes that native listeners have formed phonologically contrastive categories of speech sounds. In order to process speech efficiently, listeners tend to ignore the phonetic variation within the same phonetic category, whereas they are highly sensitive to cross-category boundary variations. Classical CP experiments create an acoustical continuum between two endpoint tokens representing two well-formed contrastive sound categories, and along the continuum, the in-between tokens gradually change from one category to the other. Listeners are asked to identify the in-between tokens as belonging to either one or the other endpoint category and to discriminate the in-between tokens. If they perceive the two endpoint representations categorically, then their identification of the in-between tokens will abruptly change from one category to the other at a certain point along the continuum. Categorical perception is also indexed by a high discrimination accuracy between adjacent tokens that straddle the category boundary, while the discrimination accuracy of tokens within the same category is poor. CP has been shown for both consonants and vowels (Liberman, Harris, Eimas, Lisker & Bastian 1961; Liberman, Harris, Hoffman & Griffith 1957; Liberman, Harris, Kinney & Lane 1961; Pisoni 1971; Gerrits & Schouten, 2003), and CP is known to be present during the earliest stages of language acquisition (e.g. Eimas, Siqueland, Juszcyk & Vigorito 1971).

CP experiments regarding lexical tones have been conducted with both native and non-native listeners of tone languages. Wang (1976) found that when presented with a continuum changing between high-rising (T2) and high-level (T1) tones, the peak of the discrimination curve corresponded to the identification boundary, which served as evidence for the categorical perception of lexical tones. Francis, Ciocca & Brenda (2003) created three continua of Cantonese tones, changing from low level to high level, from high rising to high level, and from low falling to high rising, and divided each continuum into 10 equal steps. The endpoint tokens were synthesized according to the acoustical properties in the natural production of a native Cantonese speaker, and in-between steps with an equal distance were made by dividing the continuum between the two endpoints into equal steps. The participants' identification for both level tone and inflectional tone continua changed sharply and consistently from one tonal category to the other at the category boundary, and a slight peak could be observed for the discrimination of across-boundary stimuli. The abrupt change in identification at the category boundary together with the mild discrimination peak served as evidence that tone perception was categorical, and their result was consistent with Wang (1976) in showing an evident boundary effect.

Hallé, Chang & Best (2004) extended the CP paradigm to non-native tone language listeners to explore the question whether or not lexical tone processing is influenced by linguistic experience. Both Mandarin and French listeners participated in this Mandarin tone CP study. Three 8-step continua were generated, ranging from high-level to low-rising (T1-T2), from low-rising to high-falling (T2-T4), and from low-falling to high-falling (T3-T4) respectively. The in-between tokens were created by interpolating between the pitch contours of the endpoint tokens, providing a natural and smooth change along the pitch contour. The native Mandarin listeners perceived one tone changing into another at a certain step consistently, indicated by a steep slope along the identification curves. Again, a slight peak was observed for across-boundary stimuli pairs in the discrimination curves. Statistical analysis demonstrated that the slopes along the identification curve were significantly steeper for Mandarin than for French listeners, which indicated that French listeners failed to perceive the tonal contrasts in a categorical manner like native Mandarin listeners do. In the discrimination task, the Mandarin listeners' reaction times for the critical ambiguous tokens close to the boundary were significantly longer than for more extreme tokens near the peripheries of the continua, whereas the French listeners only displayed a slight increase in reaction time for the ambiguous tokens along the continuum. Moreover, the French listeners showed no clear discrimination peak for tokens across the boundary in the discrimination task. The performance of the French listeners demonstrated that they treated the in-between step differences equally, without being benefitted by the boundary information. These results served as evidence that the categorical perception of lexical tones is a result of linguistic experience, and the linguistic function of pitch change determined whether lexical tones are treated in a phonemic manner or on a psycho-acoustic basis.

Xu, Gandour & Francis (2006) conducted a CP experiment using high-level and low-rising tones (T1-T2) with native Mandarin and native English listeners. They adopted a domain-general perspective by presenting speech and non-speech stimuli (harmonic tones) to both groups of listeners. The results indicated that only Mandarin listeners perceived the lexical tones as well as the homologous harmonic non-speech stimuli categorically. The perception of the English listeners of the non-speech tonal stimuli was more categorical than their perception of the speech stimuli. Based on these results, Xu and his colleagues claimed that categorization in auditory perception, although influenced by long-term language experience, is a domain-general, rather than language-specific, mechanism.

All the three studies agree that, first, native listeners displayed a consistent and sharp shift between two different tones, indicating categorical perception. Second, non-native listeners, on the other hand, showed a smoother change along the identification curve, and no cross-boundary discrimination peak could be observed, which indicates that non-native listeners are able to track pitch change, but they perceive lexical tones in a psycho-acoustical manner. Moreover, categorical perception seems to extend to a wider auditory domain than speech as a result of long-term memory, while exposure to tonal language enhances categorical perception.

Informative as these studies are, it is somewhat surprising that no previous studies have examined the CP of T1-T4 and T2-T3 contrasts in Mandarin. As T2 and T3 are claimed to be only "partially contrastive" due to the alternations caused by T3 sandhi (Hume & Johnson 2003), and considering the relatively low accuracy in discriminating the T2-T3 contrast (Kiriloff 1969, Chuang & Hiki 1972, Gandour 1978, Li & Thompson 1978, Hume & Johnson 2003), it is natural to ask whether this contrast is perceived categorically by native listeners, or rather, whether their ability to accurately categorize T2 and T3 is hindered by T3 sandhi.

At this juncture it is important to note that T2-T3 is a (partly) neutralizable contrast in the Mandarin phonological grammar, while T1-T4 is not. Hence, comparing the perception of T2-T3 with that of T1-T4 may inform us about the specificity in perceiving the tones that are involved in the tonal phonological grammar, and reveal possible biases for why specifically these two tones undergo phonological contextual variation. In addition, in CP studies, an issue that has been largely neglected is how the order of presentation affects the discrimination of tones. In the case of the T2-T3 contrast, T3 sandhi is unidirectional and positional, and hence discriminating a T2-T3 pair can be expected to be more difficult than discriminating a T3-T2 pair, as the former could either be an underlying T2-T3 pair or a sandhied T3-T3 pair, while the latter is not confusing. It is highly plausible that the order of presentation in the discrimination task will influence listeners' responses. Concerning CP discrimination tasks, looking at how the different order of stimuli presentation interferes with the discrimination accuracy could provide us with a better understanding of how the directionality of T3 sandhi is processed. In contrast, for the T1-T4 pair, the phonological grammar does not render a specific order of tones more favorable for discrimination. Hence, to compare the order induced difference in response between discriminations along the T1-T4 and T2-T3 continua may help to reveal the perceptual knowledge of T3 sandhi.

Besides the crucial difference between the T1-T4 and T2-T3 contrast in terms of the phonological grammar, the unsettled issues about acoustical cues for tonal perception also calls for a closer examination of the perception of these tone pairs. The first pair, T1 and T4, shares the same high pitch onset, but whereas T1 stays high throughout its contour, the contour of T4 is falling. The second pair, T2 and T3, similarly shares the same (low pitch) onset, but whereas T2 has a slowly rising contour, T3 first falls slightly before rising. As mentioned in section 3.2, pitch dynamics are known to be an important cue for tone recognition in native listeners (Gandour & Harshman 1978, Moore & Jongman 1997, Massaro, Cohen & Tseng

1985), but debate still remains regarding what the most important cue for recognizing dynamic tones is: change of pitch direction (Abramson 1978, Wayland & Guion 2004), pitch level (Burnham, Kirkwood, Luksaneeyanawin & Pansottee 1992), or average pitch height (Gandour 1983).

In CP experiments, the acoustical difference between each step could be manipulated in a very subtle manner, which in turn provides us with the opportunity to look closely at the acoustical cues that contribute to the recognition of lexical tones. Meanwhile, CP allows for detecting the tonal boundary along the continuum when changing from one tone to the other, which in a way simulates the variations in one single tone that may occur in natural speech. Importantly, the location of the boundary could provide some clues about the perceptual space of each tone.

To understand how the innate perceptual biases, if any, may interact with the tonological grammar, it is not enough to only test native listeners. The performance of native listeners may be a result of language experience and may only reflect the frequency of tonal patterns present in the input language. To tease apart the innate perceptual biases from the knowledge of distribution of tonal patterns, it is necessary to test naïve listeners and compare their performance with that of native listeners. If non-native listeners without any knowledge of lexical tones are also consistent in displaying specific patterns for perceiving the T2-T3 contrast, such that a T3-T2 order is easier to discriminate than a T2-T3 order, this would give clear evidence for the perceptual bias of T3 sandhi.

The CP experiments involving the continuum between T1 and T4 have been carried out together with my colleague (see Liu 2010). However, results of this study will be reported in Chapter 2, together with the CP experiment involving the T2-T3 continuum, in order to allow for a direct comparison between the T2-T3 versus the T1-T4 contrast.

1.5 LEXICAL TONE PERCEPTION BY TONE LANGUAGE AND NON-TONE LANGUAGE LEARNING INFANTS

In order to study the perception of speech, and in particular how innate perceptual biases may influence the perception or even genesis of phonological regularities, the behavioral data of adults needs to be examined along with the performance of young infants. Infants of a very young age are not equipped with linguistic knowledge, and neither have they received much speech input; therefore, their response to speech stimuli would best reflect innate perceptual biases without the influence from language experience. In addition, as they grow older and accumulate exposure to ambient language, the developmental trajectory of perceiving speech sounds would help to explain how the initial state of human perception interacts and co-develops with knowledge acquired from input.

1.5.1 Knowledge of speech of preverbal infants

In the last a few decades, tremendous progress has been made in understanding the early development of speech perception. It is now widely accepted that newborn infants already have some knowledge about human speech as well as their native language. For example, neonates are able to discriminate speech of a non-native language from their own native language if the two languages are prosodically distinct (Mehler, Christophe 1995, Moon, Cooper, & Fifer 1993, Mehler & Dupoux 1994), and 3-day-old newborn infants are able to discriminate between sequences that differed in number of syllables (Bijeliac-Babic, Bertoncini, & Mehler 1993, Bertoncini, Floccia, Nazzi, & Mehler 1995). Taking into consideration that in the last trimester of pregnancy, infants already perceive prosodic information from the outside world inside the womb (e.g. DeCasper & Spence 1986, Granier-Deferre, Bassereau, Jacquet, & Lecanuet 1998, Querleu, Lefebvre, Titran, Renard, Morillion, & Crepin 1984), the perception of speech prosody in all likelihood precedes that of speech segments.

Regarding the perception of native language prosody, Jusczyk & Thompson (1978) found that 2-month-old infants responded to the difference between trochaic and iambic stress patterns. By 7.5 months, English infants were able to segment words with a strong/weak stress pattern from continuous speech, but they tended to mis-segment a word if it had a weak/strong stress pattern. Their preference for strong/weak words could be explained by the rhythmic pattern of English, whose dominant stress pattern is trochaic rather than iambic (Cutler & Carter 1987). By 10.5 months, however, they succeeded in segmenting words of both stress types (Jusczyk, Housten & Newsome 1999). Morgan & Saffran (1995) also found that by 9 months of age, native English infants were able to generalize across novel disyllabic words as being cohesive if they conform to the strong/weak stress pattern.

Besides knowledge of prosody, infants are aware of segmental information present in speech. For example, by 6 months infants have built up at least some native phonological vowel categories, are aware of “prototypes”, which are good exemplars of the category. Meanwhile, the not-so-well implemented instantiations of the same category are assimilated to the prototype and hence inhibit discrimination within the same category (Kuhl 1993, 1995). Infants are also aware of consonant categories, and between 6 and 12 months, perceptually, non-native consonants tend to be assimilated to native consonant categories (Best 1991, 1995) while the perception of native contrasts is enhanced (Kuhl, Stevens, Hayashi, Dequchi, Kiritani, & Iverson 2006). In particular, young infants are able to discriminate between some non-native sounds which are indiscriminable for their adult peers (e.g. Werker & Tees 1984, Best, McRoberts & Sithole 1988).

Based on the findings of previous research, it is agreed that, first, in terms of both the segment inventory and prosodic properties, presyllabic infants can have abundant passive knowledge of language. Second, young infants are supposed to be sensitive to sound contrasts in both their native language and non-native languages. How the initial sensitivity to all sounds changes to the sensitivity to only native contrasts, a process called perceptual reorganization, has received a lot of attention in recent years. In sections 1.5.2 and 1.5.3, I will go through the literature that deals with the perceptual reorganization of lexical tones, and proceed to how perceptual biases could be reflected in speech perception in early infancy in section 1.6.

1.5.2 Perceptual reorganization of lexical tones

Infants are able to discriminate both native and non-native sounds at birth, but towards the end of their first year, as they receive more exposure to their native language, they remain sensitive to only those sound contrasts that are relevant in their native language, while their discrimination of non-native sounds deteriorates (e.g. Werker & Tees 1984, Best, McRoberts & Sithole 1988). The process in which infants’ perception of speech sounds attunes to their native language is called perceptual reorganization. Numerous studies have explored the perceptual reorganization of consonants and vowels. For example, Werker & Tees (1984) tested native English infants between 6-7 months and between 10-12 months on their discrimination between glottalized velar and glottalized uvular consonants of Hindi. This sound pair is contrastive in Hindi while falling into the same phonological category in English. They found that English infants succeeded in discriminating the two sounds at 6-7 months old, but they performed poorly on the same task when they were 10-12 months. English children of four years of age as well as English adults also failed to discriminate between this Indian contrast. Hence, the authors argued that perceptual reorganization occurs in the second half of the first year of life. With regard to vowels, Kuhl, Williams, Lacerda, Stevens, & Lindblom (1992) found that for the 6-month-old infants, only the prototypical vowels in the infants' native language showed a strong magnet effect, in that the infants tended to accept that the variants of the prototypical vowels were identical. Polka & Werker (1994) found that 4-month-old English infants were able to discriminate between German vowels which were not contrastive in English, but they failed for the same task at 6 months. Their results were consistent with previous studies testing consonant perceptual reorganization, in that infants' sensitivity to non-native sounds declines within the first year of life. The different timeline of consonant versus vowel perceptual reorganization suggests that infants do not lose their sensitivity to non-native contrast at the same time. For individual contrasts, the age of successful discrimination needs to be determined specifically.

Compared to the perception of segments, the perception of lexical tones may form a different scenario. Burnham, Kim, Davis, Ciocca, Schoknecht, Kasisopa, & Luksaneeyanawin (2011) found that for Thai and Cantonese adults, though lexical tone plays a phonemic role in their native language, they still showed better phonological awareness than tonological awareness. Lexical tones can be both segmental and suprasegmental (Mattock & Burnham 2006). For native listeners, lexical tones are phonemic and therefore play a similar contrastive role as segments, but meanwhile they are suprasegmental, in that one tone may straddle multiple segments. Non-native listeners also encounter meaningful pitch variations in speech extensively, such as intonation, but intonational pitch variations do not change lexical meaning. In other words, the same physical property, fundamental frequency, needs to be processed differently by listeners with different language backgrounds. Moreover, for non-tone language adults, acquiring lexical tones seems to be extremely difficult (e.g. Kiriloff 1969). In the light of the difference between tone and non-tone languages, several interesting questions arise: is the perception difficulty surrounding lexical tones innate for both tone and non-tone infants, while only experience with tone languages can enhance the perception? Does the perception of lexical tones follow the same trajectory of perceptual reorganization as observed in consonants and vowels? Does the perception of different tonal contrasts follow the same timeline? Considering that Mandarin has a tonal grammar, in particular T3 sandhi, how does the perception of T3 sandhi develop? Are there biases that favor the infant to learn T3 sandhi? So far, there are no clear answers to these questions. Understanding how lexical tones are perceived in infancy, however, will help us to better understand the acquisition of language, and early development in speech prosody processing.

Among the limited studies on tone perception in early infancy, contradictory results have been found. Harrison (2000) tested native Yoruba infants and native English infants, aged between 6 and 7 months, on their perception of Yoruba tones. He also compared the perception pattern of the infants and that of native and non-native adults. Yoruba is a tone language spoken in west Nigeria and Benin in which three tonal contrasts are present, namely high-level, mid-level and low-level tones. The results, although preliminary and qualitative, suggested that only Yoruba infants were able to generalize across different tokens with the same pitch, and then discriminate the generalized tone category from another category that differed in F0. Moreover, Yoruba infants only responded to the pitch differences relevant for changing one tonal category into another while neglecting pitch differences of the same amplitude which correspond to within-category variations in Yoruba. The behavior of the Yoruba infants also matched the discrimination data of native Yoruba adults, who also discriminated between cross-boundary pitch differences while ignoring the within-category variations. English infants, on the other hand, failed to discriminate between the Yoruba tones. Therefore, Harrison argued that tonal perceptual reorganization follows the same pattern as segmental perceptual reorganization, which occurs around 6 months. The difference in responses towards the pitch differences of the same amplitude suggests that infants acquiring a tone language as young as 6 months already have some representation of the linguistic function of tonal categories and do not respond to F0 variations in a purely psycho-acoustical manner.

Later, Mattock & Burnham (2006) tested 6 and 9-month-old Chinese infants and native English infants on their perception of Thai tones. They examined two tonal contrasts, one was between two contour tones—a falling tone versus a rising tone, and the other contrast was between contour and level tones—a rising tone versus a low tone. They also tested the infants on their discrimination of non-speech noises which resembled the pitch contours of the two tonal contrasts. Cross-sectionally, both Chinese and English 6-month-old infants reached an above-chance accuracy rate in discrimination for both the contour-contour contrast and the level-contour contrast. Among the 9-month-old infants, however, only Chinese infants maintained an above-chance accuracy, while their English peers' performance dropped to below chance. These findings add more evidence that the perceptual reorganization of tones follows the same timeline as segments. In contrast, regarding the discrimination of non-speech stimuli, both language groups had an above-chance accuracy rate at both ages for both contrasts. Taking into consideration that the non-speech stimuli and the Thai lexical tone stimuli shared the same pitch contours, the authors claimed that the decline in detecting lexical tone changes by non-tone language learning infants is speech-specific.

Moreover, besides the common cross-sectional research recruiting different infants for different age groups, Mattock & Burnham (2006) also carried out a longitudinal study with the same group of infants when they were 6 months and 9 months respectively. Though practically more demanding compared to a cross-sectional study, a longitudinal study serves as a better method for examining developmental patterns by ruling out spurious differences between age groups due to individual variations. Ten Chinese infants and ten English infants participated in the longitudinal study. The Chinese infants' performance was consistent with the participants in the cross-sectional study, namely their discrimination of lexical tones was above chance at both ages and for both contour-contour and contour-level contrasts, and the same results also held for the discrimination of non-speech stimuli. For the ten English infants, although their performance in non-speech pitch discrimination corroborated the results in the cross-sectional study in the sense that a small decline in discrimination was found between 6 months and 9 months, somewhat unexpectedly, the declination in lexical tone discrimination across age did not reach significance.

Mattock, Molnar, Polka, & Burnham (2008) extended their research on younger infants with a different language background. They tested the discrimination of low and rising Thai tones by 4, 6, and 9-month-old native English and native French infants. Thai low and rising tones have very similar F0 contours, meaning that these two tones only start to diverge in the second half of the syllable, which makes them difficult for non-native listeners to discriminate. They found that both the 4-month-old and the 6-month-old groups succeeded in discriminating the two tones while the 9 month-olds failed in this task. They argued that the perceptual reorganization of lexical tones occurs between 6 and 9 months. Moreover, the fact that the 4 and 6-month-old infants, whether native English or native French, successfully discriminated the Thai tone pair, suggests that the perceptual reorganization of lexical tones for non-native infants occurs during the same period, regardless of the prosodic type (stress-timed or syllable-timed) of the input language. This result is informative in that it brings the very young 4-month-old infants in scope, and their success on the discrimination of Thai tones supports the claim that tonal perceptual reorganization does not begin until 6 months.

The results of Mattock & Burnham (2006, 2008) are consistent with the qualitative findings of Harrison, in that tone language learning infants are able to discriminate lexical tones from 6 months on, while non-native infants' sensitivity to lexical tones declines during the second half of their first year. Undoubtedly, Mattock & Burnham (2006, 2008) have provided valuable information about early lexical tone perception. However, several points in their study may need to be scrutinized. First, the Thai tones used in their studies are actually non-native for both Chinese and English/French infants, and though it might be that the Chinese infants perceived them as lexical tones, it is still not clear whether the infants perceived Thai tones as phonological categories or if they just perceived them as random acoustical patterns. Hence, it might be too straightforward to call the difference observed in the perceptual pattern between the infants they tested a perceptual reorganization of lexical tones.

Second, regarding the methodology used in Mattock & Burnham's studies, the Conditioned Head-Turn (CHT) paradigm was adopted in their research. In this paradigm, infants were first trained to turn their head to a visual reinforcer once they detected a sound change, and the percentage of correct responses, defined as head turns when there was a tone change together with non-turns when there was no tonal change divided by total number of trials, was used as the index for discrimination. It is a fact that the methodology that can be used in infant studies is very limited, and only indirect evidence, such as looking time or head turns can be used as an indicator of their understanding of the stimuli. Still, doubts may be raised regarding whether to consider the CHT a neutral procedure, which favors a positive response to a change of stimuli. In Mattock & Burnham's training phase, when a background tone changed to a target tone, the reinforcer popped up to encourage infants to give a headturn, while in the test phase, infants were also expected to turn their head towards the reinforcer once they detected a sound change. However, the auditory stimuli used in both the training phase and test phase were the same, and infants could only progress to the test phase once they turned their head to the reinforcer for three consecutive trials. Moreover, once the infants failed to turn their head for three consecutive trials in the test phase, the experiment went back to the training phase. Hence, this procedure was biased to favor head turns by the infants.

Gao, Shi, & Li (2011), using the visual fixation paradigm, found that 4-month-old Canadian French infants were able to discriminate between both the T2-T3 and T1-T4 contrast, which again seems to support the classical view regarding perceptual reorganization.

Recently, my colleagues Liu & Kager (2010) tested Dutch infants on their discrimination of Mandarin T1 and T4, preceded by a training phase that contained a statistical distribution of these two tones. They adopted the statistical learning paradigm of Maye (2002), in which infants were first familiarized by either a unimodal or a bimodal distribution of the target stimuli. To generate the distribution, an eight-step continuum was created, ranging from the endpoint T1 to endpoint T4. The unimodal distribution was made in such a way that step 4 and step 5 were presented with equal frequency, and formed the most frequently occurring tokens compared to the other steps. On the other hand, for the bimodal distribution, step 2 and step 7 occurred with equal frequency and occurred most frequently compared to the other steps. The unimodal training phase resembled the distribution of input with one single category, while the bimodal training phase supported the perception of two distinct categories in the input. As a result, the bimodal distribution was supposed to be beneficial for the infants to build two categories of the stimuli, while the unimodal distribution places all the tokens into a single category. Importantly, step 3 and step 6 occurred with equal frequency in both types of distributions. In the test phase following the familiarization, infants were tested on whether they were able to discriminate a step 3 from a step 6 using the habituation-dishabituation paradigm. Cross-sectionally, they found that 5-6 months old Dutch infants succeeded in discrimination, regardless of which distribution they were exposed to, 11-12 month-old Dutch infants succeeded in the same task only if they were familiarized with the bimodal distribution, while 14-15 months old Dutch infants failed in discrimination no matter on which distribution task they were trained. These different patterns suggest that computation of the statistical property of the input, which is assumed to be a general learning mechanism across domains (e.g. Maye et al. 2002, Saffran et al. 1996, Saffran et al. 2001, Saffran et al. 2003), may add to the effects of perceptual reorganization. From Liu & Kager's data, it seems that 5-6 month-old infants are fairly sensitive to the non-native tonal differences, and no extra statistical cue was needed for successful discrimination, while the 11-12 month-old infants, who were presumably going through perceptual reorganization, needed more information than the acoustical cues alone to maintain a successful discrimination of a non-native contrast. By 14-15 months, an age at which according to previous literature perceptual reorganization should have ceased, infants were unable to discriminate between the non-native contrast, even with extra support from the statistical information.

As steps 3 and 6 of the continuum used in Liu & Kager (2010) were not prototypical representations of Mandarin T1 and T4, for a clearer pitch of Mandarin tone perception by non-native infants, they tested Dutch infants again on their discrimination between endpoint T1 (step 1 on the continuum) and endpoint T4 (step 8 on the continuum). For this new task, the Dutch infants at all ages (5-6 months old, 11-12 months old, 14-15 months old) succeeded in discriminating a T1 from a T4 (Liu & Kager 2012). Yet, the discrimination effect of the 8-9 month-old infants failed to reach significance. If we compare the endpoint discrimination with the middle point discrimination, two patterns can be observed: first, acoustically, the endpoint T1 and the endpoint T4 were further apart and presumably more salient than the middle step 3-6 pair, and accordingly, infants performed better on the endpoint T1-T4 discrimination. Second, for the less salient step 3-step 6 contrast, the difference is initially perceivable, and as the infants grow older, it is not necessarily the case that they will lose the sensitivity completely, but rather, extra cues beyond pure acoustics are required to help them maintain their sensitivity. Thus, infants do react to differences in acoustical saliency, and perceptual reorganization is a flexible process restrained by a more general learning mechanism and perhaps influenced by additive information.

To summarize, though there are contradictory findings about the perceptual re-organization of lexical tones, it seems that young non-tone language infants are able to discriminate between the salient T1-T4 contrast, and this ability may remain until 14-15 months. Successful discrimination has also been found between the not-so-salient T2-T3 contrast at an early age among native French infants. Yet, these findings need to be tested in different paradigms and among infants with different native languages to see how generalizable the results are. Moreover, observing the infants across a larger age range may help us to have a more complete picture of early lexical tone perception.

1.5.3 Lexical tone perception by native Mandarin infants

Looking at the literature reviewed thus far, it seems that perceptual reorganization of lexical tones follows the trajectory proposed for segments; however, limited but contradictory evidence has been found among native Mandarin infants.

Tsao (2008), using the conditioned head-turn procedure, tested 10-12 month-old Mandarin infants on their discrimination of Mandarin T1-T3, T2-T3, and T2-T4 contrasts. While the accuracy of discrimination of all three contrasts are above chance, she demonstrated that the participants discriminated the T1-T3 contrast with the highest accuracy, while the discrimination accuracy between the T2-T3 contrast and the T2-T4 contrast did not greatly differ. The fact that T1 and T3 do not differ only in pitch height, but also in pitch contour probably makes this contrast salient in perception. Moreover, a discrimination asymmetry was only observed for the T1-T3 contrast, meaning that the change from a background T1 to a T3 was more noticeable compared to the change from a background T3 to a T1. This study provided evidence that by around one year old, native infants perceive native lexical tones with unequal ease. The more salient the tonal contrast is, the easier it was discriminated by native infants. One point of Tsao's study is worth mentioning is that her stimuli were different in duration as an effort to approach naturalness. Different tones do intrinsically differ in duration, yet in daily life, native listeners are able to recognize the same tone when it is produced with different durations. For example, even though the stimuli in Hallé et al. (2004) and Shen & Lin (1991) differed for about 100ms in terms of duration, yet the listeners in both studies reached an almost ceiling accuracy in identifying the well-realized T2 and T3. Therefore, it is still possible that the infants in her study discriminated on the basis of durational difference, or as a combination of both durational and tonal difference.

Using a visual fixation paradigm, Shi, Gao, & Li (2011) tested Mandarin infants on their discrimination of Mandarin T1-T4, T2-T3, and T1-T3 contrasts. Somewhat surprisingly, both 4-6 and 11-13 month-old native Mandarin infants failed to discriminate between Mandarin T1-T4, and 6-month-old infants also failed to discriminate between T2 and T3. Very little data has been collected for older infants on their discrimination of Mandarin T2-T3, and so far it has been shown that 7-11 month-old infants (N=10) were able to discriminate between T2 and T3. For the T1-T3 contrast, more data is needed for the 4-6 month-old group, and limited data from an 11-13 month-old group (N=8) suggest that infants at this age were able to discriminate T1-T3. To interpret these patterns, the authors considered T1-T4 as an acoustically similar contrast, T2-T3 as showing moderate acoustical difference, and T1-T3 as substantially differing acoustically. According to this division, native infants showed most difficulties in discriminating the acoustically similar contrast while being fairly successful in the discrimination of the contrast with a large acoustical difference. Leaving aside the fact that only a limited amount data has been collected for the analysis, their division of the not-so-salient acoustic contrast and acoustically salient contrast seems unconvincing. As mentioned earlier, much evidence has been accumulated from both native adults and native infants, as well as from non-native adults that it is the T2-T3, rather than T1-T4, that is perceptually most difficult. Even though acoustical saliency differs among Mandarin lexical tones, considering the phonemic role of lexical tones and the fact that there are only four citation tones, it is still unexpected that native Mandarin infants failed to discriminate between some of the tones by the age of 13 months. To make matters even more confusing, as we have seen, Dutch infants succeed in discriminating T1-T4 from 5 months on. It is a mystery why non-native infants would outperform native Mandarin infants in a T1-T4 discrimination. If we accept that compared to T1-T3, the T2-T3 contrast is acoustically more similar, and hence perceptually more difficult, the results of Shi et al. (2011) are somewhat consistent with Tsao (2008), in that less salient contrasts are learned later than the simple contrasts, and the acquisition of lexical tones does not follow a single timeline for all the contrasts.

While these recent studies have gathered the first pieces of information for understanding early lexical tone perception, infant lexical tone perception studies are still sparse. Moreover, some crucial questions regarding infant lexical tone perception remain unanswered. For example, what exactly is the trajectory of lexical tone perceptual reorganization? Does perceptual reorganization follow the same timeline for all the tones? For the less salient tonal contrasts, what cues may support their categorization? How is lexical tone discrimination assisted by cues present in the input? Do infants succeed in normalizing speaker variations as well as token variations of a lexical tone? If so, at what age do they succeed? Do infants display discrimination asymmetries among the tones, and if so, what can we infer from these asymmetries? Crucially, none of the above mentioned studies have dealt with the early perception of phonological regularities of lexical tones, for example T3 sandhi. However, considering the extensive presence of T3 sandhi, and the fact that native speakers apply T3 sandhi naturally to a novel T3T3 sequence, at least some traces of T3 sandhi knowledge or preference should be observed at an early age.

Answering these questions will provide valuable information for understanding early perception of speech prosody. Examining the course of development of lexical tone perception, and comparing it with the perception of stress or intonation will help to give a more complete picture about universality and language specificity in the acquisition of prosody. Specifically, regarding the perceptual knowledge of T3 sandhi, hardly any effort has been made from an acquisition perspective. Studying how infants process T3 sandhi, however, is crucial for understanding the genesis of this grammar. As the core interest of the current dissertation is to reveal possible innate biases in the perception of Mandarin lexical tones, it is necessary to bring the performance of non-native infants in scope, especially those young infants who have not yet been subjected to perceptual reorganization. Their performance can be considered as the initial state of language acquisition. Comparing the performance of non-native infants with the non-native adult listeners with the same native language, as well as with native Mandarin infants and adults, will help us tease apart the innate bias in tone perception from knowledge acquired from ambient input.

1.6 PERCEPTUAL BIAS FROM AN ACQUISITION PERSPECTIVE

1.6.1 Unequal development of contrast perception

It is generally believed that infants are born with the ability to discriminate between both native and non-native contrasts, while adults' perception of non-native contrasts is restrained by phonological categories in their native language. For adults, the non-native contrasts are filtered through the sound distribution of the native language, ending up with a perception different from the native listeners that do have this contrast (for reviews see Strange 1995, Sebastian-Gallés 2005). Though it is generally assumed that the perceptual reorganization occurs during the second half of the first year of life, this does not necessarily mean that infants lose their sensitivity to all non-native contrasts at the same time, and neither does it mean that all native contrasts are acquired with equal ease following an identical timeline. Instead, different patterns can be observed in their perception of different contrasts. In the early 1980s, Aslin & Pisoni (1980) proposed the idea that the discrimination of phonetic contrasts was influenced by both the experience with ambient input and the psycho-acoustical property of the contrast. For example, it has been demonstrated that the phonetically more similar contrasts are acquired later than the more distinct contrasts. Burnham (1986) proposed that the fragile contrasts needs more exposure to be learned. Narayan, Werker, & Beddor (2010) found that 4-5 month-old native English infants failed to discriminate the acoustically not-so-salient non-native initial /n/-/ŋ/ contrast, while they succeeded in discriminating the more salient /m/-/n/ contrast. At 4-6 months of age, which can be assumed to precede perceptual reorganization, the English infants were expected to be able to discriminate all non-native contrasts, yet they failed to discriminate the non-native /n/-/ŋ/ contrast. On the other hand, some contrasts, even non-native, remain discriminable from infancy until adulthood. It has been found that clicks with different place of articulation, which occur in some African languages, could be discriminated between each other by native English infants as well as native English adults (Best, McRoberts & Sithole 1988). The authors considered that the success in discriminating clicks was the result of the considerable acoustical difference between clicks and any other English consonants. The acoustical salience rendered clicks unassimilable to any native category, and hence the discrimination of clicks stayed fairly intact regardless of ambient language.

The findings that young infants failed to acoustically discriminate between a not-so-salient non-native contrast when they were supposed to be able to do so, while on the other hand older infants as well as adults succeeded in discriminating an acoustically salient non-native contrast, implies that phonetic properties of the contrast may have a substantial impact on the discrimination of phonological contrasts. The performance of native infants furthermore confirmed that more experience with less salient contrasts helped to improve their discrimination. For example, compared to 6-8 month-old infants, 10-12 month-old native Mandarin infants performed better regarding the discrimination of a native affricate-fricative contrast (Tsao, Liu, & Kuhl 2006). The perception of the English /r/-/l/ contrast by native English infants also improved between 6 and 12 months (Kuhl, Stevens, Hayashi, Deguchi, Kiritani, & Iverson 2006), and the discrimination of the /d/-/ð/ contrast has been found to improve from infancy to early childhood until adulthood (Polka, Colantonio, & Sundara 2001, Sundara, Polka, & Genesee 2006). Filipino-learning infants whose native language has the initial /n/-/ŋ/ contrast failed to discriminate the contrast at 6-8 months, but succeeded by 10-12 months (Nayaran et al. 2010). Narayan et al. (2010) contributed the different discrimination pattern observed between the /m/-/n/ and /n/-/ŋ/ contrasts to the less acoustic robustness of the place difference in articulation, namely the tongue tip/blade of /n/, and the tongue body of /ŋ/. The non-saliency of the /n/-/ŋ/ contrast is also consistent with its diminished frequency of occurrence across languages as compared to the /m/-/n/. In a way, their results also support the idea that the phonologies of human languages are shaped by certain perceptual biases, and the developmental pattern in infants' perception of speech sounds may give evidence for how the innate perceptual bias is integrated into the phonological grammar. The difficulty in perceiving specific contrasts for infants, such as observed in the discrimination of /n/ and /ŋ/, may be a reason why these contrasts do not occur frequently cross-linguistically.

The findings that native infants may not be able to discriminate some native contrasts until a later age, while some non-native contrasts are discriminable across all ages, are consistent with the attunement theory of Aslin & Pisoni (1980). It seems that infants are born with a sensitivity to acoustical differences, and the more robust the acoustical difference is, the higher the chance that infants would detect it with more ease at an early age. When acquiring a language, the experience with this specific language may help the learners to perceive and discriminate more subtle phonetic differences more accurately, but if the less salient contrast is absent in the ambient language, then its discrimination would stay poor until adulthood. In sum, acoustical properties may strongly interfere with early speech discrimination and perception, and the less salient a contrast is, the more input or training is needed for successful discrimination. The different timeline in discriminating acoustically salient versus not-so-salient contrasts may reflect the internal auditory limitation of a human being when perceiving speech sounds.

1.6.2 Innate biases reflected in early discrimination asymmetry

Not all sound contrasts are equally perceptible for infants, and moreover, the way in which a contrast is presented to infants may also influence their performance. Discrimination asymmetries have been widely observed for both consonants and vowels. In the relevant studies, for a same pair of sounds, the infants succeeded in discriminating between them only if one sound preceded the other, but not vice versa.

Fikkert (2006) found that in a word learning experiment, 14-month-old native Dutch infants were able to detect a switch from an initial labial consonant to an initial dorsal consonant /bIn/ to /dIn/, but not in the reversed order. Moreover, for the Dutch stop-fricative contrast /p/-/f/ in the word learning task, 14-month-old Dutch infants were able to discriminate this only if infants were habituated with syllables with a fricative onset, but not with a stop onset (Altvater-Mackensen & Fikkert 2010). The authors explained the observed asymmetries in terms of the specification of phonological features by the infants. For the /bIn/-/dIn/ contrast, the infants had established the specified feature [labial], whereas the feature [coronal] was underspecified. The specified feature was more securely retained in the mind, and hence, to notice a switch from a specified category to an underspecified one was easier than the reversed order. For the /p/-/f/ asymmetry, the authors similarly argued that infants had a specified representation of [continuant] regarding manner of articulation stored in their lexicons. The specified [continuant] helped the infants to detect a stop that was not part of it. As a result, a deviation from the specified feature could easily be detected, and as [stop] was not specified yet, the infants would fail to perceive the newly encountered token as falling into the same feature category.

Interestingly, the authors also mentioned in their article that native English adults less frequently detected the changes from a labial fricative to a stop than the change from a stop to a fricative (Cole et al. 1978). It thus seems that infants and adults have an opposite confusion pattern. Altvater-Mackensen & Fikkert (2010) argued that this opposite pattern could be due to infants and adults possibly giving different weight to the same acoustical cues. Taking into consideration that the study of Altvater-Mackensen & Fikkert (2010) was a word learning task, not a simple discrimination task, it could be that the cognitive demand of the word learning task diverted the perception of the infants. Direct comparisons for phoneme discrimination between adults and infants may present more informative data about the initial preference in speech perception, and how this initial bias could be reshaped by exposure to ambient language input.

Besides consonants, more consistent discrimination asymmetries have been observed in infants' vowel perception. Cross-linguistically, in vowel discrimination experiments, the order of presenting the referent vowel and the to-be-discriminated vowel has an evident influence on whether infants succeed in discrimination. Polka & Bohn (2003) offer a summary of the asymmetrical patterns found in infant vowel discrimination, together with an F1-F2 plot of the vowel space in these studies (Table 1.2).

Table 1.2 indicates that native English infants, after being habituated to /y/, were able to discriminate /u/ from /y/, however, if they were habituated on /u/, they were unable to discriminate /y/ from /u/ (Polka & Werker 1994, Polka & Bohn 1996). Taking a first glance at these results, one could assume that, as predicted by the perceptual magnet effect (Kuhl 1991, Kuhl et al. 1992), the native vowel, in this case /u/ would function as a perceptual magnet which attracts /y/ to its category, as /y/ could be a poor examplar of the /u/ category. However, the native German infants, who do have the category /y/ in their native language, displayed the same discrimination asymmetry. Another hypothesis would be that /y/ is a less common or a more marked vowel, and the change from a less common sound to a more common sound is more easily detectable. But again, this assumption is unsatisfactory if we see the results that English native infants discriminated /æ/ more easily from /e/ than the reverse order, in which the first is actually less frequent in English.

Studies of human infants and of animals that have reported asymmetries in vowel discrimination

Study	Infants' ambient language	Contrast
(a) Human infants		
Polka and Werker (1994)	English	y → u
		ʏ → ʊ
Polka and Bohn (1996)	English	y → u
		ɛ → æ
	German	y → u
		ɛ → æ
Bohn and Polka (2001)	German	ɪ → e
		(e → i)
		(ʊ → o)
Swoboda et al. (1976)	English	(ɪ → i)
Swoboda et al. (1978)	English	ɪ → i
Dejardins and Trainor (1998)	English	ɪ → i
Best et al. (1997)	English	ʉ → y
Best and Faber (2000)	English	i → y
	Species	
(b) Animals		
Hienz et al. (1981)	Blackbirds	ɔ → æ
		ɔ → ɑ
		ɔ → ɛ
		ɑ → æ
Hienz et al. (1996)	Cats	ɔ → æ
		ɔ → a
		ɔ → ɛ
		ɑ → æ
		ʊ → ɛ
		ʊ → æ
		ʊ → ɑ

Arrows show the direction in which discrimination was easier. Parentheses indicate non-significant trends for the contrasts listed.

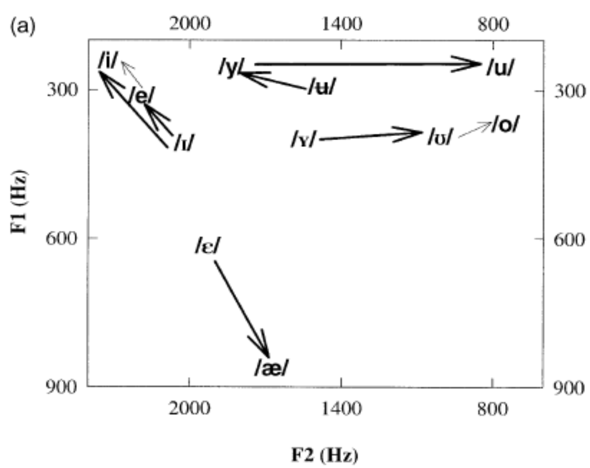


Table 1.2. Vowel perception asymmetries observed in infant perception data (cited from Polka & Bohn 2003: 223, 225). The lower panel gives the vowel space where asymmetries have been observed.

Polka & Bohn (2003) then established a new perspective to explain the perceptual asymmetry, namely that peripheral vowels such as /y/ and /æ/, which stretch to the extremes of the vowel space, serve as better referents in discrimination tasks. The age range of the infants who participated in these vowel perception studies ranged from 2 months to around one year, which straddles the pre-perceptual reorganization and the ongoing perceptual reorganization period. Hence, it does not seem likely that the asymmetry comes from a specific ambient language. Moreover, in adult AX and AXB vowel discrimination studies, adults also benefited from the peripheral-nonperipheral order of the stimuli presentation. Taken as a whole, the asymmetrical discrimination pattern was observed consistently in both adults and infants. Polka & Bohn (2011) proposed the Natural Referent Vowel (NRW) framework, in which they argued that the listeners were innately biased to favor peripheral over non-peripheral vowels, and these peripheral vowels naturally serve as a good referent for discrimination. Yet, the authors admitted that the exact role of these referent vowels is not yet clear. One of their speculations is that for infants, speaker normalization of these referent vowels is easier. They also proposed that at the initial state, the infants show biases that favor the referent vowels. Nevertheless, after being exposed to the native language, their vowel categories could be reorganized. Regarding the question as to why peripheral vowels are naturally preferred, the authors argued that the formants of the peripheral vowels converge, which produce high salience in the vowel spectra. I agree with the authors in that the consistent discrimination asymmetry across different languages shows innate perceptual biases in vowel discrimination, but the answer to the important question of what causes the bias is still quite vague. The authors speculated that peripheral vowels were more salient, but salience is difficult to define. Furthermore, even if we accept salience as a parameter to define a sound, it is not clear whether discriminating between a salient sound and not-so-salient sound will always be more successful if the salient sound precedes the not-so-salient sound. Regarding the question why asymmetry occurs, I tend to believe that specificity may play a role. In other words, the more specified a sound is, the easier to tell whether another sound is the same or not. Acoustic complexity may be a source of specificity, as the more features to be used to define a sound, the more specified the sound is. Yet, more efforts need to be made to see whether specificity correlates with distinctiveness naturally.

To investigate the discrimination asymmetry and especially the initial biases that favor such an asymmetry is of great importance, in that it gives us the chance to better understand how humans are auditorily equipped, and why some universal patterns are found cross-linguistically. To research what causes the biases, and how maintainable they are, would help answer the bigger question of how speech fits into the human perception system. To be able to shed light on these questions, it is worthwhile to test both native and non-native adults as well as native and non-native infants on their perception of the same phonological contrast or phonological regularity, and see whether we are able to find consistent perceptual patterns across different groups of participants. Any differences in linguistic knowledge between participants (e.g. native versus non-native; infants versus adults) would help us to tease apart the role of initial perceptual biases, acoustical properties of the sounds, and exposure to a certain language in human speech processing.

So far, only one study has reported a discrimination asymmetry in the field of early lexical tone perception (Tsao 2008). It is worthwhile to have a closer look at the early discrimination of lexical tones, and to see whether certain perceptual preferences can be observed. If similar asymmetrical patterns can be observed as in the discrimination of consonants and vowels, this will extend our knowledge about how lexical tone perception is innately restricted. Perceptual asymmetry is an interesting issue within the scope of the current dissertation. As I have mentioned earlier, T3 sandhi in Mandarin is unidirectional and asymmetrical in that T3T3 is only sandhied to T2T3, but not to T3T2 or other sequences. The question of why T3 sandhi occurs to the left-most T3 is largely unknown. The dissimilation process of T3 sandhi can not be explained by co-articulation, and neither is the OCP account satisfactory in explaining T3 sandhi (see section 1.3.3). As I have mentioned earlier, the specificity of the tones might play a role in deciding the unidirectionality of T3 sandhi. The currently available paradigms for testing the sound discrimination of infants actually provide us with a good opportunity to observe the natural distinctiveness of T2 and T3.

Due to intrinsic limitations in the methodologies that can be used in testing infants' speech perception, it is almost impossible to test discrimination directly such as the AX paradigm used to test adult listeners. Instead, infants are habituated on one category and once they lose interest in the habituated category, they are presented with a new category. The way to measure discrimination is that, if they are able to hear the difference between the two categories, then their listening time to the new category should be longer than their listening time to the habituated category due to the recovery of interest. Presumably, it could be that building up a representation for certain phonological categories is more difficult than for other categories when presented with the same amount of variations, which results in asymmetrical discrimination. As hinted by Polka & Bohn (2011), it is possible that certain natural referent sounds might be normalized more accurately. In other words, to discriminate a deviant token from a well-established category should be easier compared to discriminating a deviant token from a fuzzy category. Testing this hypothesis may reveal information as to the origin of the discrimination asymmetry.

Regarding the ease in categorizing different phonological categories, a natural question to ask is how the information in the input language may support or hinder the categorization of a speech sound. If a speech signal is to be processed efficiently, interference from other external phonetic variations, such as variations within the same phonological category and difference of speakers' voices, need to be compensated. Such normalization, to some extent, depends on the extraction of distributional information of the sounds in ambient speech input. In the following section, I will review the literature regarding the categorization of speech sounds, speaker normalization, and extraction of distributional patterns in early infancy.

1.6.3 Factors that interfere with early speech perception in infancy.

1.6.3.1 Categorization

When encountering a speech signal, listeners need to map the perceived acoustics onto linguistic categories. As mentioned earlier, in natural running speech, the same phoneme could be produced with substantial variations due to the linguistic context as well as different emotional status, and the voice quality of different speakers may also make the phonetic implementation of the same phoneme different. Thus, for efficient decoding of the speech signal, listeners need to compensate for these variations, and be able to map different realizations of the same phoneme onto a single category. Adult listeners are fairly capable of neglecting external information, such as gender of the speaker, pitch range of the speaker, rate of speech, etc., and are able to extract the internal features of speech sounds, which guarantees efficient recognition of phonemes as well as successful comprehension. However, categorization is a formidable task for infant listeners, based on their limited exposure to and knowledge about language (for a review of categorization from the perspective of speech perception, see Holt & Lotto 2010).

In recent studies regarding early lexical tone discrimination, researchers have introduced token variations in stimuli in order to make sure that the infants were discriminating between phonological categories rather than pure phonetics. For example, in Mattock & Burnham (2006, 2008), five tokens of each tone of one native speaker were used in the training phase in the Conditioned Head Turn procedure, while in Gao et al. (2010, 2011), the infants were presented with 12 different tokens of the same tone in the habituation phase. In Liu & Kager (2010), in the unimodal or bimodal statistical learning section of the experiment, infants were also presented with multiple tokens of one or two tone categories. The above mentioned studies seem to suggest that young infants are able to categorize speech sounds quite proficiently.

However, to what extent infants are able to categorize speech sounds is still under debate, and it has been proposed that infants tend to initially over listen to phonetic details (for a review, see Newman 2008). In the next section, I will discuss the research on another important aspect of categorization, namely speaker normalization.

1.6.3.2 Speaker normalization by young infants

Efficient speech processing calls for not only generalization of the same phoneme or words across different tokens of one speaker, but also the generalization across speakers with different voices. Kuhl (1983) tested 6-month-old infants on their discrimination between two vowels /a/ and /□/, and found that the infants were able to disregard the voice difference of male, female, and children, treating the same vowel produced by different voices as equivalent. Meanwhile, they remained sensitive to the change from one vowel category to the other. This finding suggests that infants as young as 6 months already possessed at least some of the categorization ability required for understanding speech. Though there has been evidence that infants are able to normalize speakers' voices (Kuhl 1979, 1983), it is still a much debated issue to what extent infants are able to compensate the inter-speaker variation and at what age they succeed in categorizing these acoustically dissimilar tokens. For example, at 7.5 months, after being familiarized with a word produced by a female voice, infants did not show preference of the same word produced by a male voice (Housten & Jucszyk 2000). Nevertheless, they recognized the words embedded in sentences with greater success if they were first familiarized with the word produced by multiple speakers in both genders. It has also been reported that 7.5-month-old infants were not able to generalize across the same word produced by a happy voice and that produced by a neutral voice (Singh, Morgan & White 2004). At 10.5 months, however, infants were able to ignore the vocal affect information attached to the word: after they were familiarized with a word produced with a happy voice, they were able to recognize it even if it was neutrally produced in a sentence, and vice versa (Singh, 2008). Moreover, Singh also found that infants tended to erroneously equalize phonetically similar words if they were presented with similar affect information, but if infants were familiarized with words produced by multiple speakers, they succeeded in telling apart sentences containing different words which were phonetically similar from those containing words that they were familiarized to.

For older infants who are at the stage of learning words, however, recent evidence has shown that they could succeed in discriminating new words only when presented with sufficient variation in the input. For example, it has been consistently demonstrated that 14-month-old infants have difficulties in tagging two words that differ minimally, such as *bih* and *dih*, to two different concepts, though they were able to discriminate the two words in a simple discrimination task (Stager & Werker 1997, Fennell & Werker 2003, Pater, & Stager, Werker 2004, Thiessen 2007, Werker, Cohen, Lloyd, Casasola, & Stager 1998, Swingley & Aslin 2007). For the same word learning task which involved a minimal pair, however, if extensive variations were introduced in the auditory stimuli such that each word was represented by 36 exemplars produced by different speakers at different pitch registers, infants did succeed in categorizing the auditory label of the new object that they learned, and were able to notice it if the labels of the objects were switched (Rost & McMurray 2009). Hollich, Jusczyk, & Brent (2002) have also found that on the one hand, 24-month-old infants stored specific information about the speaker, while on the other hand, exposure to a word produced by multiple speakers enhanced extraction of invariant information of the word, and hence helped word learning.

Based on these results, it seems that even infants as young as 7.5 months have some ability to categorize the phonetic representation of words, and more importantly, external information such as speaker and affect variation, form a beneficial factor for helping infants tease apart phonetic details from the core phonological representation of a word, and as a result assist both word learning and specification of phonological categories. But meanwhile, infants still store a large amount of specific phonetic information in their minds when learning words. One point that is worth mentioning is that, the above mentioned studies mainly involve word learning or recognition of words when embedded in sentences, which requires either relating a phonetic instantiation to a concept, or to segment a sequence from a larger phrase, and these tasks do not test the discrimination of phonological categories in a direct way. Regarding discrimination tasks, which simply test the perception of distinct phonological categories, whether speaker variation would assist or hinder categorization is still unknown. A second issue is that the age range of infants that have been tested for speaker or affect variation spanned from 7.5 months to 24 months, but for younger infants, especially those before the onset of perceptual reorganization, it is not clear whether they are overall sensitive to subtle phonetic differences, no matter whether those with a potentially phonological importance, or those unrelated to phonology such as a speaker's voice, or if they are only sensitive to differences that could potentially be related to meaning.

In the scope of early lexical tone perception, not much evidence has been collected regarding speaker normalization. Do speaker and token variations help the categorization of lexical tones and hence help the discrimination of tones? Or instead, do infants tend to memorize phonetic differences and are hindered in discrimination if there is too much variation? Do their compensation of variance and extraction of invariant information change with age? All these questions merit a closer look at the infant behavioral data. Moreover, taking into consideration that the high and low pitch range may have substantial overlap across different speakers, while high and low are contrastive features of lexical tones, it could well be that infants' speaker normalization of lexical tones is much more vulnerable compared to that of segments. As hinted by Polka & Bohn (2011), how tolerant the infants are in normalizing variability may give evidence for how well a tone could be a referent. Then the referent effect may be revealing for understanding innate perceptual biases regarding the discrimination asymmetry.

1.6.3.3 Normalization as abstraction of distributional information

If infants were to compensate variations of the same phonological category as presented in inter-speaker deviations and inter-token differences, at least to some extent, they need to establish an abstract representation of that category, and be able to detect whether an instantiation falls within or outside the category. If the infants succeeded in establishing the phonological category, then they presumably possess some knowledge of the distribution and statistical property of the phonological categories.

Indeed, numerous studies have demonstrated that infants do possess a powerful computational ability. For example, adults tend to perceive variant instantiations of one single vowel as belonging to the prototypical representation of that vowel, hence the phonetic prototype works as a perceptual magnet, which renders the non-prototypical tokens to be perceptually assimilated to the prototype rather than form a separate category (Grieser & Kuhl 1989, Kuhl 1991). As with adults, infants as young as 6 months also showed the perceptual magnet effect, but only the prototypes of a native vowel category worked as perceptual magnets, while the non-native prototypes failed to introduce an assimilatory effect. Crucially, the same instantiation of vowels was assimilated to different prototypes according to the vowel categories presented in the infants' ambient language. Therefore, with six months exposure to their native language, infants already built up some knowledge of the distribution of the vowel categories, and they are able to normalize the encountered phonetic differences accordingly (Kuhl et al. 1992).

Maye et al. (2002) tested the role of distribution in early phonological category learning by manipulating the statistical patterns presented to the infants preceding the test. They created an eight-step continuum changing from /da/ to /ta/, and the number of the occurrence of the eight steps formed either a unimodal or a bimodal distribution in the training phase. The unimodal distribution refers to the frequency pattern at which steps 4 and 5 occurred most frequently, and the frequency decreased bidirectionally along the continuum towards the endpoint tokens. The bimodal distribution refers to the pattern where step 2 and step 7 occurred with the highest frequency, and the frequency decreased towards the center and the extreme steps. Figure 1.2 gives the two types of distribution. Crucially, the unimodal distribution resembled the distributional pattern of one single category while the bimodal distribution is similar to the distributional pattern of two separate categories. As the figure demonstrates, step 3 and step 6 occurred with equal frequency in both unimodal and bimodal distributions. When presented with step 3 and step 6, both 6-month-old and 8-month-old infants succeeded in discriminating these tokens only if they were trained on the bimodal distribution, but not if they were trained on the unimodal distribution. The data suggests that infants are not only sensitive to the frequency distribution of speech sounds, but may actually rely on distributional information to form corresponding phonological categories. Later Maye, Weiss, & Aslin (2008) demonstrated that exposure to a bimodal distribution helped 8-month-old infants discriminate between categories that are initially difficult to perceive. Liu & Kager (2010), reviewed above, replicated Maye et al. (2002, 2008) by using a tonal continuum changing from Mandarin T1 to T4, and confirmed that the bimodal distribution was helpful for non-native infants to discriminate between the Mandarin T1 and T4.

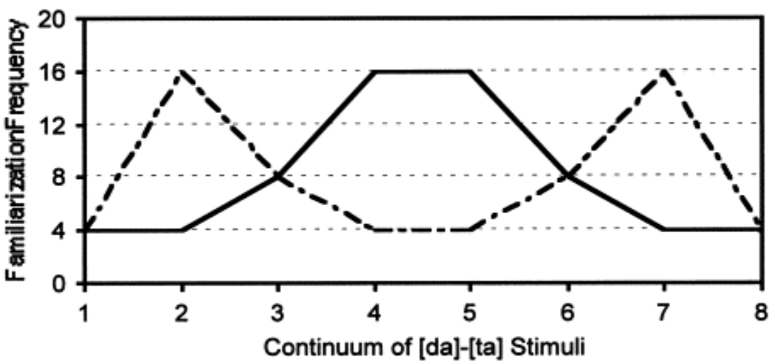


Figure 1.2. Bimonal v.s. Unimodal distributions of [da]-[ta] stimuli during familiarization. The continuum of speech sounds is shown on the abscissa, with Token 1 corresponding to the endpoint [da] stimulus, and Token 8 the endpoint [ta] stimulus. The ordinate axis plots the number of times each stimulus occurred during the familiarization phase. The presentation frequency for infants in the Bimodal group is shown by the dotted line, and for the Unimodal group by solid line (cited from Maye et al. 2002: B104).

Based on these findings, infants as young as 6 months are able to grasp at least some distributional information of the speech sounds as presented in ambient input, and are able to build up phonological categories based on the abstraction of the statistical distribution. However, to what extent they are able to normalize and generalize across the variant tokens of a single category is not yet clear. Bear in mind that young infants sometimes do fail in normalizing productions of different speakers (Houston & Jusczyk 2000). It could very well be that the abstraction of a phonological category based on tokens from multiple speakers with different voices is too demanding for young infants. Young infants may only be able to compensate the variations if the tokens vary along a dimension that helps the extraction of the invariant structure.

As mentioned earlier, so far the categorization of lexical tones, as well as inter-speaker and inter-token normalization of lexical tones has been largely neglected. Considering the habituation-dehabituation paradigm used to test the infants' discrimination, to manipulate the amount of variation in the input may help to reveal to what extent infants are able to normalize the tonal variations. If infants are detail-oriented listeners, then a pure phonetic difference discrimination task which involves only a single token of each tone category would be easiest for infants, but if variations are necessary for discriminating between phonological categories, then presenting infants with multiple tokens of a tone would improve their performance on tone discrimination tasks. Moreover, if tonal categorization is vulnerable to inter-speaker variation as assumed, then to train infants on productions from variant speakers may hinder the discrimination of lexical tones. Tackling the question about how much variation is needed for successful categorization and discrimination of lexical tone would provide us with evidence about the innate restrictions in perceiving specific lexical tones. Moreover, the variation needed for categorizing a tone would, in a way, reflect the perceptual space of different tones.

So far, we have discussed the Mandarin tones and T3 sandhi from multiple perspectives, and as discussed in previous sections, the lexical tones are mainly pitch variations. Yet, another interesting property of pitch variations is that they not only convey meaning in speech, but also in other domains, such as music. Considering the multifunctionality of pitch, studying pitch perception from a domain-general perspective may give us a more complete understanding of lexical tone processing.

Therefore, in the following section 1.7, I will discuss the literatures that study the pitch perception from a cross-domain perspective.

1.7. PERCEPTION OF LEXICAL TONES FROM A DOMAIN-GENERAL PERSPECTIVE

1.7.1 Entangled music and lexical tone perception in adults

As discussed in the previous sections, pitch variations realized in human speech bear plentiful information for conveying meaning. However, speech is not the only human communication that incorporates meaningful pitch variations, with another common usage of pitch variation being in music. The topic of interaction between music and speech perception has received much attention in recent years. It is still debated whether music and language work in distinctive domains in a bimodular way (Nettl 1983, Pinker 1997), or if both are dominated by universal dispositions and constraints (Trehub 2003, Zatorre, Belin, & Penhune 2002). Though the dissociation between language and music is to some extent supported by the observation of individuals who show disorder or impairment in one domain while processing in the other domain stays largely intact (Ullman, Corkin, Coppola, Hickok, Growdon, & Koroshetz 1997, Perets & Coltheart 2003, Peretz, Champod, & Hyde 2003, Hebert, Racette, Gagnon, & Peretz 2003), considering the fundamental role of pitch variations in both speech and music, it is at least conceivable that the perception of music would to some extent correlate with the ability of perceiving speech prosody.

It has been shown that musical aptitude and training enhances phonological awareness, such as the consistency or oddity in rhyme and consonants (Anvari, Trainor, Woodside, & Levy 2002), the accuracy in perceiving speech prosody (Thompson, Schellenber, & Husain 2004, Magne, Schön, & Besson. 2006), and music training also enhances the accuracy in perceiving an L2 segment (Slevc & Miyake 2006). Evidence has also been found supporting the correlation between music and speech processing (Patel, Peretz, Tramo, & Labreque 1998, Peretz, Ayotte, Zatorre, Mehler, Ahad, Penhune, & Jutras 2002, Patel, Foxton, & Griffths, 2005, Nan, Sun, & Peretz 2010). In other words, as music melody and speech prosody resort to the same acoustics, namely the fundamental frequency, the ability to process pitch would presumably affect both the music perception and speech perception domains.

Considering the highly contrastive status of lexical tones, which are mainly realized by pitch variations, the relation between musical pitch perception and lexical tone perception forms a solid base for studying the domain-general restrictions on pitch processing. Taking into consideration the issue of possible lexical tone discrimination asymmetry, the domain-general sensitivity to pitch needs to be considered if we are to draw any conclusion about innate perceptual biases. The main method for testing perception asymmetry is the discrimination task, in which the listeners are required to indicate whether or not two tones are identical. Naturally, we expect the listeners who have acute pitch perception to be more accurate in detecting a difference. It is possible that for those sensitive listeners, the order of stimuli presentation does not influence their judgment, while the way of presenting the stimuli may have an impact on the responses of the not-so-sensitive listeners. If we are to claim the universality of the biases in lexical tone perception, it is necessary to observe the same perception pattern across populations that may differ in general sensitivity to pitch. Considering that music perception calls for accurate recognition of subtle pitch differences, it may serve as a good index for general pitch processing. Furthermore, a correlation between lexical tones and musical pitch perception has been observed in multiple ways, which suggests that, an increased musical pitch processing ability indeed induces a higher accuracy in lexical tone perception. The correlation between music and lexical tone perception has been observed in the following aspects.

First, perception of lexical tones and the absolute pitch (AP), which refers to the ability of labeling or producing a note in isolation without references from other notes, has been found to be mutually beneficial. For example, though AP is rare and an indicator of musical talent, it has been observed that AP occurs more frequently among tone language speakers (Letivin & Rogers 2005, Deutsch, Henthorn, Dolson 1999, Deutsch 2002, Deutsch, Henthorn, & Dolson 2004). English musicians with AP also out-performed non-musicians as well as musicians without AP in the discrimination of Thai tones, filtered speech of the same pitch contour, or violin sounds of the same pitch contour. Moreover, musicians without AP and non-musicians found the discrimination of lexical tones more difficult than the other two tasks, while the musicians with AP did not show such difficulty (Burnham &

Brooker 2002).

Second, musical training helps the perceptual acquisition of lexical tones. Alexander, Wong, & Bradlow (2005) found that native English musicians out-performed non-musicians significantly on the discrimination of Mandarin lexical tones. Moreover, it has been demonstrated that after a short tutorial on identifying lexical tones, musicians reached a higher accuracy in identification than the non-musicians when encountering various lexical tone tokens produced by multiple speakers whose F0 range overlaps considerably (Lee & Hung 2008). Non-tone language speakers with a high melodic ability were better in detecting tonal variations compared to ordinary non-tone language speakers (Delogu, Lampis, Belardinelli 2006, Marie, Delogu, Lampis, Belardinelli, & Besson 2011). Wong & Perrachione (2007) trained native English listeners to use pitch patterns to differentiate meanings of pseudo English words, which is similar to learning lexical tones, and they found that learning success was associated with participants' sensitivity to pitch in a non-linguistic context and their previous musical experience.

Third, deficiency in music processing at least partially predicts inaccuracy in lexical tone processing. Congenital amusia is a neurogenetic disorder that affects the processing of musical pitch (Nan, Sun, & Peretz 2010: 2635), both in perception and production, which mainly manifests itself in having difficulties in processing subtle pitch differences. Cross-culturally, musical melodies are mainly composed of one semitone (corresponding to the pitch difference between adjacent keys on the keyboard) or two semitone units, and amusia is reflected in having difficulties in perceiving these one semitone pitch differences (Peretz et al. 2003, Hyde & Peretz , Foxton, Dean, Gee, Peretz, & Griffiths 2004). Congenital amusia is reported to be limited to impairment in music processing, while auditory, intellectual, memory and language skills are fairly intact (Peretz et al. 2002, Peretz et al. 2003). When native French speakers without prior experience to tone languages were asked to discriminate between Mandarin tones, though the scores of the amusic group and control group overlap largely, the amusic groups performed significantly worse compared to the control groups (Nguyen, Tillmann, Gosseln, & Peretz 2009). When congenital amusics were tested on a same/different judgment task for their discrimination of Mandarin and Thai tones, their performance was inferior to their matched controls, and in addition, when they were tested on a musical analog of the

Thai tones, the same inferiority was observed for the amusics (Tillmann, Burnham, Nguyen, Grimault, Gosselin, & Peretz 2011). The recurrently observed inferior performance on discriminating lexical tones by non-tone language amusics serve as convincing evidence that, at least to some extent, the perception of lexical tones is limited by general sensitivity to pitch across different domains.

Native tone language amusics' deficit in speech processing may extend to the perception of intonation; amusic native tone language speakers were impaired in identifying and discriminating native intonation contours between question and statement, as well as non-linguistic analogs of native intonation (Jiang, Hamm, Lim, Krik, & Yang 2010). The amusic Mandarin listeners may also differ from the Mandarin control group in lexical tone perception. Nan, Sun, & Peretz (2010) tested native Mandarin listeners on their identification and discrimination of Mandarin tones, which were superimposed on either monosyllables or disyllabic sequences. The participants were divided into an amusic group and a control group, which were balanced regarding musical background, years of education and age. Importantly, none of the amusic participants reported having difficulty regarding lexical tones in everyday communication. While the control group showed an almost ceiling accuracy in the lexical tone identification task (95%), the amusic group performed significantly worse than the control group. In the disyllabic lexical tone discrimination task, the amusic participants performed significantly worse than the control group only when the tone-bearing syllables were different. Interestingly, not all amusic participants were impaired in lexical tone perception. Among the 22 amusic participants, 9 were 3SD below the mean of the control group for the identification task, and 6 were 3SD below the mean of the control group for the discrimination task. These 6 participants, who performed significantly worse than the control group both in the identification and discrimination of native lexical tones, were defined as "lexical tone agnosia". Specifically, among the four Mandarin tones, the T2 seemed to be difficult for all the listeners, regardless of the control group, the amusics without lexical tone agnosia, or the amusics with lexical tone agnosia. In particular, the amusics with lexical tone agnosia performed at chance level for the identification of T2. In terms of production, however, no deficit in tone production has been observed for the 6 participants with lexical tone agnosia. Cross-rater identification of their tone productions reached an accuracy rate of 98.6%. Yet, for normal native Mandarin listeners, to what extent the perception of music is correlated to the perception of lexical tones is still not clear. Also, the question why T2 causes the most confusion among listeners needs further investigation.

Taken together, regardless of language background, both music and lexical tone perception are restricted by some general constraints in pitch perception. On the one hand, exposure to a tone language enhances the probability of the occurrence of musical talent (Deustch et al. 2004, 2006). On the other hand, musical training enhances lexical tone perception (Wang et al. 2007). It is important to note that although a lexical tone functions linguistically and calls for precise control of pitch, native tone language speakers are not immune to amusia, and native Mandarin amusics perceived the native lexical tones less accurately compared to control groups. So far, the mentioned studies that worked on the perception of music and lexical tones among native Mandarin listeners targeted the amusic group, while not much attention has been paid to the normal population. Considering the phonemic function of lexical tones, it is largely unknown whether for native listeners the lexical tone perception correlates with the music sensitivity, or if it has been consolidated to a linguistic property that is no longer restrained by general sensitivity to pitch. Comparing the lexical tone-music correlation between tone language and non-tone language listeners will help us understand how language experience may shape the processing of pitch variations.

Another essential-yet-neglected issue is how the domain-general sensitivity to pitch may be entangled with the perception of the phonological grammar regarding lexical tones. As mentioned earlier, the perceptual knowledge of T3 sandhi is largely unknown. As T3 sandhi mainly restricts the co-occurrence of pitch patterns, it is expected that the general sensitivity to pitch may play a role in the perception of T3 sandhi. If we desire to find out whether there is an innate perceptual bias that may have shaped the formation of T3 sandhi, listeners' general sensitivity to pitch needs to be taken into account. If we were to claim the universality of biases, which restricts the perception of T2 and T3, then listeners that differ in general sensitivity to pitch should display the same perceptual pattern. Specifically, as T3 sandhi occurs asymmetrically, it needs to be examined whether all listeners, regardless of general pitch sensitivity, demonstrate the same discrimination ease for the T3T2 order, rather than the reversed T2T3 order. It is also possible that those listeners who are sensitive to pitch in general, regardless of how the T2 and T3 are presented, will reach equally high accuracy rates.

In order to have a complete picture of the perception of T3 sandhi, in terms of the interaction between universal biases, language experiences, and cross-domain sensitivity to pitch, testing both native and non-native listeners on both language and music tasks is necessary. Based on the findings of previous CP studies (Hallé et al 2004, Francis et al. 2003, Xu et al. 2006), we can speculate that non-native listeners would perceive lexical tones psycho-acoustically. Then, presumably, non-native listeners serve as a baseline for evaluating the acoustical processing of lexical tones and music restrained by general auditory limitations, and their perception of T3 sandhi may well reflect innate biases. Native listeners, who perceive the lexical tones in a phonemic manner, may help us to understand how the linguistic function of lexical tones may influence the tonal perception and shape the musical pitch perception. Moreover, the perception of native listeners may demonstrate how the knowledge of tonal grammar interacts with the innate biases.

1.7.2 Pitch variations in music and speech processing

In the current dissertation, I will focus on two aspects of pitch realization, namely the relative and the absolute pitch, which play a crucial role in both speech and music perception. A clear example concerning relative and absolute pitch change has been given in Trehub & Hannon (2006). As is evident from Figure 1.3, if we define the original melody accompanying the lyric "happy birthday to you" as the "original pitch contour", the transposition refers to the exact same contour realized on a higher (or lower) pitch range, and the difference between the transposed contour and the original contour is counted as the absolute pitch difference. If we compare the phrase labeled as "same contour" to the original contour, we see that although the two phrases share the same pitch direction, the amplitude of the direction change is not the same. The "changed contour" phrase, on the other hand, has a different pitch direction compared to the original contour. Both the same contour and changed contour are instantiations of relative pitch change.

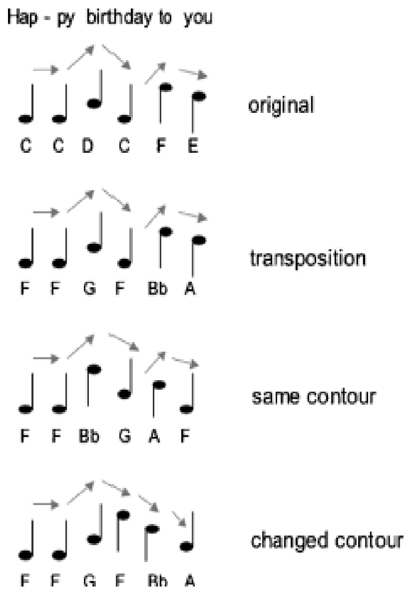


Figure 1.3. Relative pitch structure in melodies. Individual notes are assigned conventional pitch names (A B C), and arrows indicate the melodic direction. The tune, "Happy Birthday" (original) can be presented at a new pitch level (transposition), or one or more intervals can be changed in a way that preserves the original contour (same contour) or changes it (changed contour) (cited from Trehub & Hannon 2006: 76).

Most people, not only musicians, are able to recognize familiar melodies played at a different pitch height, and most adults are able to recognize a sour note in a familiar melody (Schellenberg & Trehub 2003). It is largely agreed upon that enjoying music primarily depends on the ability to perceive relative pitch (see Trehub & Hannon 2006 for a review). In terms of speech prosody, as mentioned earlier, it could well be that a high tone of a speaker with a low pitch range could acoustically be very similar to a low tone produced by a high pitch range speaker (Moore & Jongman 1997). As a result, relying on absolute pitch may be confusing. Thus, to efficiently track relative pitch in speech processing presumably outweighs the need to track absolute pitch. Specifically regarding lexical tones, as mentioned earlier, so far no language has been found to have more than five contrastive tones that only differ in absolute pitch level (Blevins 2006), and Wang (2012) even indicated that no language allows for more than four level tone contrasts. Then, plausibly, tracking relative pitch is more important than perceiving the absolute pitch height in the perception of lexical tones. Considering the more pervasive function of relative pitch difference in both speech and music, the relative pitch difference is presumably more salient in auditory processing than the absolute pitch difference.

One crucial difference between musical pitch variations and lexical tone contrasts is that the pitch contour of musical melody is mainly discrete, and the abrupt change occurs between different adjacent notes while lexical tone contour is continuous and smooth (Liu, Jiang, Thompson, Xu, Yang, Stewart 2012, Bidelman, Gandour, Krishnan 2011, Trehub 2000). Previous studies have found that the amusic listeners may achieve a high accuracy in the identification and discrimination of intonations realized on naturally produced speech stimuli, but their discrimination of discrete tones as well as gliding analogues of the intonation are frequently impaired (Ayotte, Peretz, & Hyde. 2002, Patel et al. 2005, Liu et al. 2012). Again, these studies focus on the amusic group, but it has not been previously discussed whether the normal population finds the discrete notes more difficult to perceive than continuous pitch. From the perspective of innate constraints of pitch perception, it would be informative to test young infants on their perception of discrete versus continuous pitch, and their accuracy in perceiving these two types of pitch patterns will shed light on the innate limitations in pitch perception. Moreover, observing the developmental pattern in pitch perception from infancy to adulthood, for both tone language listeners and non-tone language listeners, would give us hints as to how innate perceptual biases are shaped by language experience.

1.7.3 Music perception in early infancy

Considering the relative ease and proficiency in processing relative pitch in both speech and music, we would expect infants to show a similar preference in pitch processing. However, contradictory patterns have been found regarding whether infants tend to rely more on absolute pitch or more on relative pitch when processing music melody. On the one hand, infants tend to rely on absolute pitch in music processing. Saffran & Griepentrog (2001, 2003) tested 8-month-old English infants on a statistical learning task in which infants were first familiarized with 3 minutes of continuous melody consisting of four three-tone-sequences concatenated randomly, and the test trials were composed of either tone words—the original three-tone-sequences—or tone part-words, which were the concatenation of the last note of one original sequence and the first two notes of another sequence. They found that after the exposure infants succeeded in discriminating between tone words and tone part-words sharing the same relative pitch but differing in absolute pitch. The same pattern holds for both sequences created using random tones or generated from the C-major scale. Volkova (2004) has also shown that, after being familiarized to a novel lullaby song in a foreign language for two weeks, 7-month-old infants displayed a novelty effect when the same lullaby was sung on a different pitch level as compared to the original, suggesting that infants have a (mental) representation of absolute pitch. On the other hand, other studies indicated that infants responded to relative pitch differences more easily. For example, it has been found that 6-8 month-old infants more easily discriminated between melodies that differed in relative pitch than those differing in absolute height (Trehub, Bull, & Thorpe 1984). This result suggested that absolute pitch by itself was not enough to guarantee the successful discrimination of melodies by infants. Moreover, conducting the same statistical learning task as in Saffran et al. (2001, 2003), when the absolute pitch of the test words versus non-words was held constant, infants were able to track relative pitch (Saffran, Reeck, Niebuh & Wilson 2005).

Regarding the memory of music, infants tend to memorize relative pitch rather than absolute pitch. For example, Trainor, Wu & Tsang (2004) exposed 6-month-old infants to an unfamiliar melody for one week, and after the one week of exposure, they tested the infants on whether they preferred this now familiar melody, or whether they preferred a new, unfamiliar melody. Infants demonstrated novelty preference regarding the new melody, demonstrating that they were able to memorize a musical melody as defined by relative pitch difference between the notes. As the authors did not aim to test whether infants memorize absolute pitch or relative pitch, they did not test the infants on whether they would treat the transposed familiarized melody as novel or the same. Later, Plantinga & Trainor (2005) tested 6-month-old infants in order to add this missing piece of information. They also familiarized the infants to an initially unfamiliar folksong for one week, and the infants were tested on the eighth day. They found that the infants showed no novelty preference to the same melody transposed to different pitch levels, but they did show novelty preference to a new, unfamiliar melody. These two studies, taken together, imply that as young as 6 months, infants perceive and memorize real music melodies in the same way as adults. In other words, they tend to memorize the pitch contour of a melody, which is mainly defined by relative pitch. If we compare the results of Plantinga & Trainor (2005) and Volkova (2004), it seems that more time is needed for infants to memorize the absolute pitch level of a melody, and at least younger infants benefited more from relative pitch information.

Research has also indicated that infants are able to detect occasional note changes. Trainor & Trehub (1992) carried out a note change detection experiment with 8-month-old infants. The infants were presented with 10-note melodies that were identical in terms of relative pitch but differed in absolute pitch, and these melodies were called standard melodies. The authors found that the infants were able to discriminate between standard melodies and deviant melodies, the latter of which were also composed of 10 notes, with one note differing from the standard melodies. This note was an occasional change in the relative pitch of the melody. Later, using the ERP technique, Tew, Fujioka, He & Trainor (2009) tested 6-month-old infants' ability to detect the occasional note change of 4-note melodies, and they also found a discrimination effect between the standard melodies and the deviant melodies. Nevertheless, these two studies did not target whether the infants were able to discriminate between melodies that differ in pitch patterns, but attempted to find out whether the infants had knowledge of musical structures. They manipulated the melodies in a way that the occasional change in note could either fall within or outside the same key of the standard melodies, with the standard melodies differing in absolute pitch. Hence, though their studies hinted that infants were able to detect relative pitch differences, more direct evidence is needed to better understand the online discrimination of relative musical pitch difference versus absolute musical pitch differences.

So far, to my knowledge, no previous study has dealt with infants' online processing of discrete pitch versus continuous pitch, and their ability to discriminate between relative pitch differences versus absolute pitch differences. The aforementioned studies of Saffran et al. (2001, 2003) were searching for evidence of statistical learning from a domain other than language, and they did not directly test the discrimination of absolute or relative pitch. The method used in Volkova (2004), Trainor et al. (2004), and Plantinga & Trainor (2005) tested the long-term memory of a melody rather than the online processing of relative or absolute pitch. Trainor & Trehub (1992) and Tew et al. (2009) focused on whether infants had knowledge of musical structures. Therefore, more clear evidence is needed about how infants process musical pitch. How capable they are in dealing with relative and absolute pitch difference is important not only in the domain of music processing, but it will also help us interpret how the human auditory system is initially equipped, and may provide a domain-general view of possible limitations in early prosody processing. As the pitch contour of speech sound is generally continuous, while those in music are discrete (e.g. Bidelman et al. 2011), comparing how infants perceive musical pitch with how they perceive speech pitch may help to reveal the innate structure preference in pitch processing.

Besides the question of how infants perceive the musical relative pitch and musical absolute pitch, it is also intriguing to see whether infants show a correlation between music and speech processing. As we have seen in previous sections, numerous studies have found a correlation between music and lexical perception among non-tone language listeners (e.g. Wong et al. 2007, Deutsch et al. 2004 2006). However, the origin of this correlation is not clear. On the one hand, it could be that humans show such lexical tone-music correlation from early infancy on, which suggests innate domain-general restrictions on pitch processing. On the other hand, it could also be that such a correlation is a result of language experience, meaning that the lexical tones are not a linguistic function for non-tone language listeners, and hence they perceive the tones psycho-acoustically. It is the psycho-acoustical processing of lexical tones that correlates the lexical tone perception with music processing. To test these hypotheses, we need to have a closer look at how the perception of lexical tones and the perception of music patterns interact at the initial stage, and how they develop from infancy to adulthood. It is not enough to only look at non-tone language infants and adults, but we also have to look at the developmental course of tone language listeners. Trying to answer the question of whether tone language listeners differ from non-tone language listeners in pitch processing from infancy on, or whether they begin as identical listeners but later bifurcate due to language experience, will help us to tease apart the universal biases from language influence in pitch processing.

1.8 RESEARCH QUESTIONS AND OUTLINE OF THE DISSERTATION

Based on the body of literature reviewed above and the aforementioned questions that remain unanswered, the current dissertation addresses the following research questions:

1. Do native Mandarin listeners have well established T2 and T3 categories? Do native listeners and non-native non-tone-language listeners perceive the T2-T3 contrast in the same way?

2. Do native and non-native listeners display asymmetrical discrimination patterns for the T2-T3 contrast in tone sequences? If so, are the asymmetries language specific (a result of exposure to Mandarin) or universal?

3. How do native and non-native infants categorize Mandarin T2 and T3? Do they categorize T2 and T3 with equal ease?

4. For infants who have not finished perceptual reorganization yet, how can token variations assist them to categorize lexical tones?

5. Do infants show perceptual biases favoring the occurrence of T3 sandhi?

6. How does the perception of Mandarin lexical tones correlate with the processing of musical pitch, in native and non-native, infant and adult listeners? Do the correlations differ between these populations as a result of language experience?

To address these questions, this dissertation presents the results of a series of experiments, structured as follows. In Chapter 2, adult native Mandarin listeners and adult native Dutch listeners participate in categorical perception experiments involving T2 and T3. In Chapter 3, the same two populations are tested in a bisyllabic tonal sequence discrimination task and a monosyllabic tone discrimination task. In Chapter 4, native Dutch infants (6, 9, 11, and 14-month-old) were tested for their discrimination of Mandarin T2 and T3. Chapter 5 simulates a real language acquisition situation, where the infants are first familiarized with naturally produced Mandarin speech, and after that they are tested for the discrimination between T2T3-T3T3 and between T3T2-T3T3. With regard to the correlation between the perception of musical pitch and lexical tones, in Chapter 3, besides the lexical tone discrimination tasks mentioned above, the listeners are also tested with the Montreal

Battery of Evaluation of Amusia. Last in Chapter 6, Dutch and Mandarin young infants are tested for their discrimination of Mandarin T2-T3 and musical melodies that either differ in relative pitch or absolute pitch.

1.9 METHODOLOGY IN TESTING AUDITORY PERCEPTION

Considering that both adults and infants, and both native listeners and non-native listeners, will be tested in the current dissertation, and multiple test methods will be used, I will give a quick summary of the experimental paradigms to be used for a clear overview.

1.9.1 Methods for testing adults' auditory perception

1.9.1.1 Categorical perception experiment

Categorical perception (CP) has been extensively used to test how listeners categorize speech sounds into different phonological categories. In the CP experiment, a continuum is created between two endpoint tokens, which are both well realized exemplars of two separate phonological categories. Along the continuum, different steps are manipulated to form in-between tokens, which gradually change from one endpoint token to the other. By testing the identification of the in-between steps and the discrimination of the (nearly) adjacent steps, CP gives us the opportunity to measure how the listeners perceive the endpoint categories. As mentioned earlier, categorical perception is necessary for an efficient processing of speech. In order to perceive phonological categories contrastively, listeners have to ignore phonetic variations within the same category while still being able to detect the differences that distinguish one category from another. Therefore, CP forms a standard method for testing the representation of phonological categories by speakers of different language backgrounds. Three patterns will reflect categorical perception in CP: first, a steep change from one category to the other should be observed in the identification curve, and the location of the shift is the category boundary; second, when discriminating adjacent or nearly adjacent tokens, a discrimination accuracy peak should be observed for the steps that straddle the category boundary; third, the identification boundary predicts the discrimination peak (Gerrits & Schouten 2003). CP has been used for testing categorization of consonants (e.g. Liberman 1967, Pisoni 1973), vowels (e.g. Pisoni

1973, Repp 1984), and lexical tones (e.g. Francis et al. 2003, Hallé et al. 2004).

In the current dissertation, in Chapter 2, both native listeners of Mandarin, non-native adult listeners of Mandarin were teste using CP for their categorization of Mandarin lexical tone tokens. In particular, for testing the perception of T3 sandhi, I presented the compared tokens to the listeners in different orders.

1.9.1.2 Montreal Battery of Evaluation for Amusia

The Montreal Battery of Evaluation for Amusia (MBEA) is an instrument for screening amusics, and it is widely used in measuring musical ability of both the normal population and patients with brain injuries (for a review of the MBEA, see Peretz et al. 2003). The MBEA is open resource, and it is available online (http://www.brams.umontreal.ca/plab/research). The test consists of six parts, with each part aimed at one aspect in musical processing. The six parts always run in the same order: Part 1, melodies that differ in scale; Part 2, melodies that differ in relative pitch; Part 3, melodies that differ in absolute pitch; Part 4, melodies that differ in rhythm; Part 5, two melodies that differ in meter; Part 6, memory of melodies. Participants were asked to discriminate between two melodies. For Part 1 to Part 4, participants are asked to discriminate between two melodies, and for Part 5, participants are asked to judge whether a single melody is used for marching or for the waltz. In Part 6, participants are presented with single melodies, and they need to judge whether or not they have previously heard these melodies in the forgoing parts of the experiment. There are 30 trials in each part. For each task, the listeners receive a score which is the total number of the correct answers. Each listener may also receive a global score, which is the mean of the scores of each individual task.

Peretz et al. (2003) indicated that the global score of 160 listeners across different ages conformed to a normal distribution pattern. Only 3% of the listeners reached 100% accuracy, while at the low extreme, about 2% of the listeners got a global score two standard deviations lower than the average. The outliers at the low extreme were detected as amusics. Though the MBEA aims at screening the amusic individuals, the fact that it relies on the normal distribution of the whole population assures that it is sensitive in revealing musical aptitude among the normal population as well. Hence, the score of the MBEA could work as an index of sensitivity to musical pitch, and serves as a good comparison with the speech prosody perception.

By comparing the listeners' performance on the MBEA and lexical tone tasks, we may better understand the domain-general restrictions on pitch processing.

In the current dissertation, in Chapter 3, adult participants are tested with the MBEA for measuring their sensitivity to musical pitch, and their MBEA scores are used as an index of musical sensitivity for the correlation with their performance on the lexical tone perception tasks.

1.9.2 Methods for testing infants' perception

1.9.2.1 Visual habituation procedure

Since infants cannot speak and say whether they hear a difference between two stimuli, an indirect measure is needed to measure their discrimination. A widely used paradigm to test infants' discrimination between sounds is the visual habituation paradigm (for a review of the technique, see Colombo & Mitchell 2009, see also Golinkoff & Hirsh-pasek 2012). In this paradigm, visual responses of the infants are used as an indicator to reflect the perception of speech sounds. The procedure always starts with presenting one type of auditory stimulus to the infants simultaneously with the visual stimuli until their looking time to the visual stimulus drops below a preset criterion, which indicates the loss of interest. Then the auditory stimulus changes to another sound, which is also presented simultaneously with the same visual stimulus. If the infants are able to detect the difference, then their looking time to the visual stimuli should increase due to the recovery of interest. The novelty effect serves as an index for their discrimination of the two types of sounds presented to them.

In the current dissertation, using this procedure, I test infants on their discrimination of Mandarin lexical tones (Chapter 4), as well as their discrimination of musical melodies (Chapter 6).

1.9.2.2 Head-turn Preference Procedure

Another widely used procedure for testing preference of speech sounds by preverbal infants is the head-turn preference procedure (HPP) (Nelson, Jusczyk, Mandel, Myers, Turk, & Gerken 1995, see also Golinkoff & Hirsh-pasek 2012). Similar to the visual habituation procedure, the HPP also uses visual responses as an indicator for auditory perception of speech stimuli. The procedure has two phases, beginning with a familiarization phase, and immediately following the familiarization, there is a test phase. In the familiarization phase, infants are trained with auditory stimuli, and in the test phase, they are tested with trials in which stimuli that occurred in the familiarization phase and new stimuli alternate. Three blinking lights located in front, on the left and on the right of the infants serve as visual stimuli to keep the infants interested in the task. Importantly, in the test phase, the light at one side is always combined with one type of auditory stimuli. Differing from the visual habituation procedure, the HPP may provide a more natural language learning environment, and hence gives stronger evidence for generalization. For example, in the familiarization phase, infants can be familiarized with words, and in the test phase they can be tested with sentences that either contains the word they have heard in the familiarization phase or sentences that contain an unheard new word. If infants show a novelty effect to sentences containing a new word, then we could infer that they detect the difference between the familiarized word and the unheard word, and that they are able to generalize learned words in a new context (Bortfeld & Morgan 2010). By varying the stimuli in the familiarization phase, the HPP provides us with an opportunity to test what information in the speech input enhances discrimination between different phonological patterns, such as stress or probabilistic phonotactics (e.g. Saffran et al. 1996, Saffran & Grien. 2001, Saffran et al. 2003, Bortfeld, Morgan, Golinkoff, & Rathbun 2005).

In the current dissertation, in Chapter 5, Dutch infants are tested using the HPP for their bias towards asymmetrical patterns in T3 sandhi.

Chapter 2 Categorical perception of Mandarin T2 and T3 with native and non-native adult listeners

2.1 INTRODUCTION

As mentioned in Chapter 1, among the four Mandarin lexical tones, namely high level tone (T1), rising tone (T2), low dipping tone (T3), and falling tone (T4), T2 versus T3 is the only contrast that undergoes a tonal sandhi process. T3 sandhi restricts the co-occurrence of T3s, and requires the first T3 to change to a T2. Chapter 1 also mentions that, due to T3 sandhi, the T2-T3 contrast is only partially contrastive (Hume & Johnson 2003). Accordingly, there is a significant body of literature devoted to understanding the phonetic realization of the T3 sandhi (e.g. Xu 1997, among others) and the perception of the derived T2 (e.g., Chen, Shen, & Schiller 2011, Shen 1991, among others). Yet, in order for T3 sandhi to occur, it requires at least two syllables. However, nearly all of the aforementioned literature focuses on the perception of the derived T3 in isolation, without adequately investigating how the bisyllabic derived T2T3 sequence and the underlying T3T3 sequence are perceived. T3 sandhi is positional and asymmetrical, which means that only the first T3 syllable is neutralized to a T2 while the second T3 remains intact. Hence, a crucial question regarding T3 sandhi is that, if T2 and T3 are mutually confusable, why does the sandhi process occur in such an asymmetrical way? Furthermore, why is the second syllable in the bisyllabic sequence not neutralized?

In order to better understand the mental representation of the T2-T3 contrast, and possible traces of bias in perceiving these tones, this chapter focuses on the categorical perception (CP) experiments I ran with native Mandarin and native Dutch adult listeners. Two aspects in perceiving T2 and T3 were closely examined. First, I asked how the partially contrastive T2-T3 is perceived. Do native listeners perceive this contrast categorically or do they have more fuzzy perceptual categories of T2 and T3? Second, I queried whether, during the discrimination tasks, the asymmetrical pattern of T3 sandhi is reflected in online processing. In other words, I asked if the order of stimuli presentation influences the responses of the listeners.

Classical CP experiments create an acoustical continuum between two endpoint tokens, representing two well-formed contrastive sound categories. Along the continuum, the in-between tokens gradually change from one category to the other. Listeners are asked to identify the in-between tokens as either one or the other endpoint category and to discriminate the in-between tokens. If the listeners perceive the endpoint categories contrastively, they need to have separate representations of the two categories. Moreover, these listeners need to ignore the phonetic variation within the same phonetic category while remaining highly sensitive to cross-category boundary variations. Assuming they perceive two endpoint representations categorically, their identification of the in-between tokens will change abruptly, from one category to the other, at a certain point along the continuum. Also, the discrimination of pairs formed by adjacent tokens should be easiest for pairs falling across the identification boundary. Previous studies on the CP of lexical tones (Wang 1976, Francis et al. 2003, Hallé et al. 2004, Xu et al. 2006) have demonstrated that native tone language listeners tend to perceive lexical tone contrasts categorically, while non-tone language listeners perceive the lexical tones psycho-acoustically (these studies have been discussed in detail in section 1.4.3). Still, somewhat surprisingly, no earlier study has examined the CP of T2-T3 pairs in Mandarin.

When it comes to phonological grammars, as discussed in Chapter 1, the idea that phonologies of natural languages may be grounded in perceptual biases that are inherent to humans is not new. Many studies suggest that universal perceptual limitations may play a role in shaping phonological patterns (Steriade 1995, 1997, Liljencrants & Lindblom 1972; Lindblom 1986, Lindblom & Maddieson 1988). While much attention has been paid to universal biases affecting the perception of segments and segmental phonology, universal biases affecting the perception of lexical tones have been largely neglected. The primary purpose of this dissertation is to explore how innate perceptual biases may influence the grammars regarding lexical tones. Lexical tones are of great interest in that they form an important way of realizing supra-segmental information. To understand how the tonal grammar may be influenced by perceptual biases will help us to gain more knowledge about the formation of phonology.

In Mandarin, tones T2 and T3 are involved in T3 sandhi while T1 and T4 are spared from the phonological grammar. The perception of lexical tones of native

Mandarin listeners could very well result from language experience, and from the structure frequency in the speech input. Thus, if one aims to study the innate auditory biases that may have shaped the T3 sandhi, then one needs to undertake an examination of listeners who are naïve to lexical tones. Furthermore, such a study needs to include comparisons between: a) the perception of the contrasts that undergo the phonological grammar, namely the T2-T3 contrast and b) the contrasts that are not involved in the phonological grammar, namely the T1-T4 contrast. Comparing the performance of native Mandarin listeners to non-tone language listeners helps one to tease apart the knowledge introduced by exposure to Mandarin from the innate auditory biases. Moreover, comparing listeners' performance upon encountering a T2-T3 continuum and a T1-T4 continuum assists in illustrating the specificity in perceiving the tones that are involved in tonal grammar.

In the research present in this chapter, native Mandarin and Dutch listeners participated in two CP experiments investigating T1-T4 and T2-T3 continua respectively. For each continuum, an identification task, an AX discrimination task, and an AXB discrimination task were conducted. Dutch is a lexical stress language in which pitch variation is not lexically contrastive. As a result, Dutch listeners serve as naïve listeners to lexical tones in comparison to native Mandarin listeners. Given that T3 sandhi is unidirectional, meaning that only the T2T3 order is ambiguous while the T3T2 order is not, this study accounts for the order of stimuli presentation in the discrimination tasks. In the AX and AXB tasks, A could be either T2 or T3 with equal chance.

To summarize, the study presented here compares the performance of Mandarin and Dutch listeners. Both groups participated CP experiments involving both T2-T3 continuum and T1-T4 continuum. I attempt to answer the following questions:

1. Do native Mandarin listeners perceive the T2-T3 contrast categorically? How do native Mandarin listeners and native Dutch listeners differ in terms of tonal categorization?

2. Do specific discrimination patterns for the T2-T3 continuum hold true for listeners regardless of their difference in language background? If so, do they relate to the T3 sandhi pattern?

3. If biases exist, how do these biases interact with the T3 sandhi grammar?

CP experiments allow for the subtle manipulation of a linguistic contrast, the results of which can serve as an indicator of linguistic or non-linguistic processing of the contrast. The following two questions serve as the foundation for further research exploring the biases in perceiving T3 sandhi. First, how are T2 and T3 represented linguistically in native listeners' minds? Second, how are T2 and T3 perceived non-linguistically by non-native listeners?

2.2 EXPERIMENT 1 CATEGORICAL PERCEPTION OF THE MANDARIN T2-T3 CONTRAST

Participants

20 Mandarin and 20 Dutch participants were recruited (mean age: 22 years). All the Mandarin listeners were raised with and educated in Mandarin in China. All the Mandarin listeners had come to the Netherlands in order to obtain a higher-education University degree. The Dutch participants were recruited from a pool of volunteers who were willing to participate in the experiment. None of the Dutch participants had prior knowledge of Mandarin, nor had they been exposed to a Mandarin speaking environment. Both Mandarin and Dutch participants were students of Utrecht University at the time of this experiment. All participants also reported to have normal hearing and speech. They received 5 euros for their participation.

Stimuli

A T2-T3 continuum was created for the experiment. I chose the syllable /ma/ as the tone-carrying syllable, because of its nasal onset which ensures a continuous pitch contour. Furthermore, both /ma2/ (means hemp) and /ma3/ (means horse) are frequently used syllables in Mandarin. A female Mandarin speaker recorded /ma2/ and /ma3/ in isolation. To generate the 6 in-between steps, the pitch contours of naturally produced /ma2/ and /ma3/ were extracted by the software PRAAT. After normalizing the duration of these two contours (450 ms), of each tone, 100 points with equal temporal distance were obtained from the F_0 contour. The other 6 in-between contours were interpolated at each time point between the normalized endpoint contours of the /ma2/ and /ma3/. It was ensured that the pitch difference between the same points along the 8 pitch contour (e.g., point 2 at each of the 8 contours) was equal. All 8 pitch contours were re-synthesized to the original syllable using the PSOLA method (Moulines & Laroche 1995). Given that the underlying pitch contours of T2 and T3 were smooth curves connected by turning points, the in-between steps were acquired by interpolation and the PSOLA re-synthesis made the stimuli sound as natural as possible (Hallé et al. 2004). Figure 2.1 depicts the pitch contours of the 8-step continuum generated between T2 and T3. The time-normalized stimuli ruled out the possibility of interference from duration as a potential confounding factor in the experiment. Five native Mandarin speakers listened to the stimuli and were all in agreement that all the stimuli sounded like natural, normal speech.

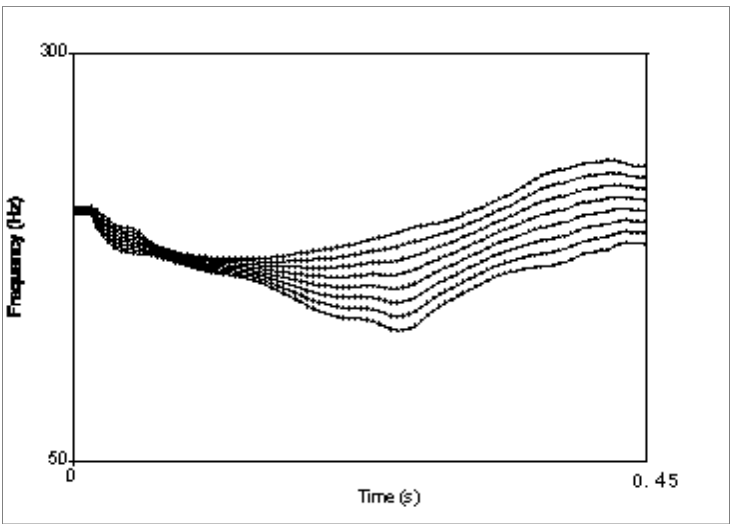


Figure 2.1. The 8-step continuum generated between T2 and T3, on the tone-carrying syllable /ma/. The stimuli change from step 1, T2 (top line) to step 8, T3 (bottom line).

Previous research has argued that the turning point and the F_0 difference between the onset and the turning point are cues for differentiating the two tones (Moore & Jongman 1997). In the current experiment, the in-between stimuli were generated by interpolations rather than by fine-grained calculations of the turning points along the time axis and F_0 difference along the frequency axis. This method was used because the current experiment does not aim to identify the exact perceptual boundaries between tones. Instead, this experiment attempts to explore whether CP for the T2-T3 contrast could be observed and whether the order of stimuli representation influences the listeners' responses.

2.2.1 Identification tasks

Procedure

Mandarin listeners took part in a forced choice identification task, in which they were presented with each step along the continuum and asked to identify whether the sound they heard was /ma/ with T2 or /ma/ with T3. Two characters 麻 and 马 with the same segmental information /ma/, but each with a different tone, were used as visual stimuli. The participants were asked to identify the syllable by clicking on the corresponding character on the screen. Each step occurred four times in this task, and was randomized across the experiment for each participant.

Given that Dutch listeners were unlikely to have had any categorical representation of T2 and T3 in Mandarin, identifying the two tones as "low dipping" and "low rising" may have been confusing for these listeners. Hence, an AXB identification paradigm was used for the Dutch listeners where A always served as the endpoint T2 token, which appeared first. On the other hand, B always served as the endpoint T3, which appeared last. The middle stimulus X varied along the 8 steps of the continuum. Participants were asked to identify whether X sounded the same as or closer to A (first syllable) or B (third syllable) by clicking on two buttons labeled "first" or "third" on the screen.[2]

[2] Here the AXB identification is different from the AXB discrimination in 2.1.3 and 2.2.3. In the discrimination task, A and B were always the to-be-discriminated steps, such as between step 1 and step 3 along the continuum, while X is either the same as the A or the same as the

Results and discussion

The mean proportion of correct T3 identifications at each step by Mandarin and Dutch listeners is depicted in Figure 2.2. As the figures demonstrate, there is an abrupt shift along the identification curve of the Mandarin listeners between step 5 and step 6, which indicates that native Mandarin listeners have formed two separate categories for T2 and T3. On the other hand, the identification curve of the Dutch listeners is much smoother and no abrupt shift could be observed. The smooth identification curve of Dutch listeners demonstrates that Dutch listeners lack two separate representations of T2 and T3, and they perceive the different steps along the continuum in a psycho-acoustical manner.

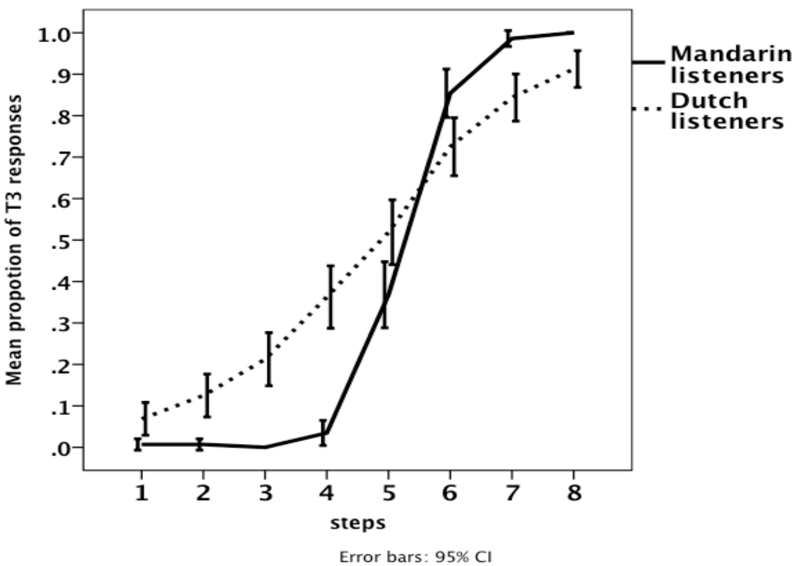


Figure 2.2 Mean proportion of T3 responses by native Mandarin and Dutch listeners.

Moreover, Mandarin listeners demonstrated minor variation among the steps that were close to the endpoints, indicating that native listeners were consistent in identifying well-realized tones. Conversely, the identification became vulnerable

B. In the identification task, A and B are the endpoint tokens, in this case Mandarin T2 and T3, while X varies along the continuum. The identification tasks aimed to examine in which category the in-between steps were categorized.

for ambiguous tokens close to the identification boundary. In comparison, Dutch listeners showed equal variation across the steps. A univariate ANOVA was conducted taking the mean proportion of the T3 responses from each participant at each step as a dependent variable, while taking steps (step 1 to step 8) and language background (native Mandarin or native Dutch) as fixed factors. A significant effect of steps was found, F_{steps} (7, 288) = 304.54, p<0.001, as well as a significant effect of language background, $F_{language\ background}$ (1, 288) = 16.11, p<0.001. There was also a significant interaction between steps and language background: $F_{interaction}$ (7, 288) = 14.41, p<0.001. These results demonstrate that native Mandarin listeners and native Dutch listeners responded differently to the stimuli, and their responses differed significantly for each step. Meanwhile, for the same step, Mandarin listeners and Dutch listeners also responded differently.

PROBIT analysis was used to calculate category boundary locations and the slopes of the Gaussian distribution functions (Hallé et al. 2004; Wu & Lin 2008). In this analysis, the three of the eight points along the continuum closest to the 50% mean response value were taken to fit into an Ogive curve. The intercept (the estimated 50% response position) represented the potential categorical boundary location. The slope (the estimated steepness value by 1/SD of the Gaussian function fitted to the data) indicated the degree or sharpness of a potential categorical perception. Table 2.1 presents the intercept and slope of the PROBIT analysis. The figures from the table show that is likely that the Mandarin listeners categorized steps 1 to 5 as T2 while steps 6 to 8 were categorized as T3. For Dutch listeners, as they do not have representations of Mandarin T2 and T3, it was hard to “identify” a tone. Hence their data was not submitted to the PROBIT analysis. Yet, it can still be seen that the Dutch listeners tended to perceive steps 6, 7, 8 to be the same as or similar to T3 while they perceived steps 1 to 5 to be the same as or similar to T2.

If we compare the pitch contours of T2 and T3, it can be seen that the main characteristic of T2 is a rising direction while T3 rises towards the offset of the pitch contour and drops at the beginning to have a low pitch. Therefore, a T3 carries more acoustical information than a T2, hence presumably T3 is more specified in nature. In other words, considering the specificity of T3, one can expect T3 to form a more distinctive tonal category compared to T2. What this study has found for the boundary position of the identification curve is consistent with the expectation, namely that in a continuum changing from T2 to T3, for both Mandarin listeners and Dutch listeners, T2 most likely forms a larger category than T3.

Experiment 2.1.1	Intercept	Slope
/ma2/-/ma3/	5.146	1.835

Table 2.1 The position of the category boundary for native Mandarin listeners along the T2-T3 continuum.

2.2.2 AX discrimination task

Procedure

Participants were presented with pairs of stimuli and were asked to make forced choices on whether the stimuli were the same or different by clicking on one of the two buttons on the computer screen labeled “same” and “different”. The paired stimuli in the test were separated by 2 steps (steps 1-3, 2-4, 3-5, 4-6, 5-7 and 6-8, named as pair 1, 2, 3, 4, 5 and 6 hereafter). Note that both ascending step (step 1-3, 2-4, etc.) and descending step (step 3-1, 4-2, etc.) orders occurred in these pairs. To prevent the introduction of potential experimental bias, 32 extra trials of same-step pairs (e.g., steps 1-1, steps 3-3) were used as filler trials. Altogether there were 96 trials (6 “different pairs” * 4 repetitions * 2 orders + 6 “same pairs” * 8 repetitions). All trials were randomized, and the inter-stimulus interval (ISI) was set at 200 ms.

Results and discussion

The responses of the “different” pairs with a two-step difference (i.e., pair 1, 2, 3, 4, 5, 6) were submitted for analysis. A univariate ANOVA was carried out taking the accuracy rate from each participant at each pair as a dependent variable, and using the language group and the stimuli pair as fixed factors. The results demonstrate that both pairs and language background are significant factors for the accuracy rate: F_{pairs} (5, 228) = 5.89, $p<0.001$, $F_{language\ background}$ (1, 228) = 26.83, $p<0.001$, and there is significant interaction between pairs and language background: $F_{interaction}$ (5, 228) = 4.895, $p<0.001$. Figure 2.3 provides the mean accuracy rate of discrimination at each step by both language groups. As the figure illustrates, there is a clear discrimination peak for Mandarin listeners at pair 4. According to the identification boundary found in 2.1.1, step 4 and step 6 formed pairs that straddled the boundary of the T2 and T3 categories along the continuum. When faced with cross-boundary pairs (pairs 4, 5), Mandarin listeners benefited from their tonal category knowledge and outperformed Dutch listeners. Moreover, Mandarin listeners' native knowledge of tonal categories appeared to inhibit within-category discrimination which also suggests the presence of tonal categories. For native Dutch listeners, however, no clear cross-boundary discrimination peak could be observed. Again, the results indicate that Mandarin listeners possess categorical representations for T2 and T3 while Dutch listeners perceive the T2-T3 continuum psycho-acoustically.

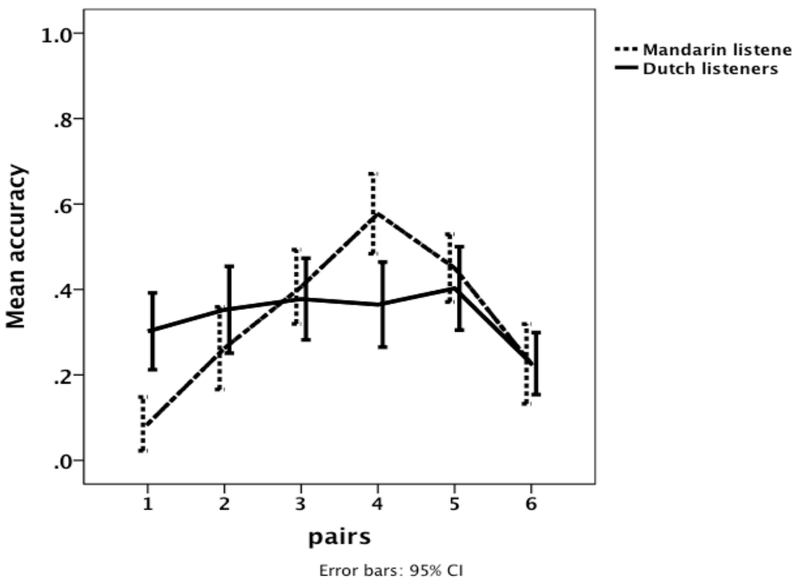


Figure 2.3 Mean accuracy rate for the discrimination of each pair by both native Mandarin listeners (solid line) and native Dutch listeners (dashed line).

To explore any order effects, pairs ranging from T2 to T3 (e.g., steps 1-3, 2-4) were defined as an "ascending order" while pairs ranging from T3 to T2 (e.g., steps 3-1, 4-2) were defined as a "descending order". In this case, the ascending order resembles the sandhied T3T3 sequence in the sense that the first syllable is closer to a T2 while the second syllable is closer to a T3. The T3 sandhi knowledge may encourage the Mandarin listeners to accept the two tones in a ascending order as identical. Hence the hypothesis is that, for native Mandarin listeners, the ascending order is presumably more confusing and difficult to discriminate than the descending order. The data was submitted to a repeated measures ANOVA for each language group separately. The mean accuracy rate of each participant was the dependent variable, and pairs (pair 1, 2, 3, 4, 5, 6) and order (ascending or descending) were used as within-subject factors. The analysis of native Mandarin listeners illustrated that pairs had a significant effect on response accuracy, $F_{Mandarin}$ (2.99, 56.74) = 19.07, $p < .001$[3], while for Dutch native listeners pairs showed marginal significance, F_{Dutch} (5, 15) = 2.88, p = 0.051. The analysis also showed that order constitutes a significant factor for native Mandarin listeners, where: $F_{Mandarin}$ (1, 19) = 87.49, $p < .001$. For native Dutch listeners, order demonstrated a tendency to be a significant factor, where: F_{Dutch} (1, 19) = 3.90, p = 0.06. Pair and order only significantly interacted for Mandarin listeners: F (5, 15) = 8.82, $p < .001$, but not for Dutch listeners: F (5, 15) = 1.82, $p > 0.1$. A mixed effect model that took the accuracy rate of responses as the dependent variable, the pairs together with order as the within-subject variable, while using language background as the between-subject variable revealed a significant main effect of pairs F_{pairs}(3.79, 144.14) = 16.958, $p<0.001$[4] and order F_{order}(1, 38) = 64.331, $p<0.001$. Yet, language background failed to yield a significant main effect, $F_{language}$ (1, 38) = 0.04, $p>0.1$. Nonetheless, both pairs and order significantly interacted with language background: $F_{pairs*language\ background}$ (5, 34) = 7.407, $p<0.001$; $F_{order*language\ background}$ (1, 38) = 27.30, $p<0.001$. The three-way interaction between pairs, order, and language background was also significant: F (5, 34) = 8.22, $p<0.001$. The results of the mixed effect model suggest that although the overall accuracy did not differ greatly between native Mandarin

---

[3] Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of pairs, χ^2(14) = 29.59. Therefore, the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity (ε= 0.60 for the main effect of pairs).

[4] Mauchly's test indicated that the assumption of sphericity had been violated for the main effects of pairs, χ^2(14) = 28.43. Therefore, the degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity (ε= 0.46 for the main effect of pairs).

listeners and native Dutch listeners, native Mandarin listeners and native Dutch listeners responded differently to the same pair of stimuli. The listeners also responded differently to the same pair of stimuli when presented in different orders. Figure 2.4 and Figure 2.5 depict the accuracy rate of each pair in both ascending order and descending order, separated by language background.

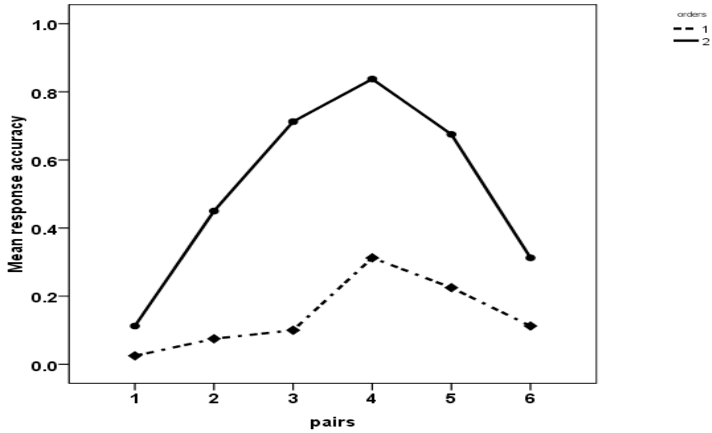


Figure 2.4 The mean accuracy of the responses of each pair in ascending order (dashed line) and in descending order (solid line) of native Mandarin listeners

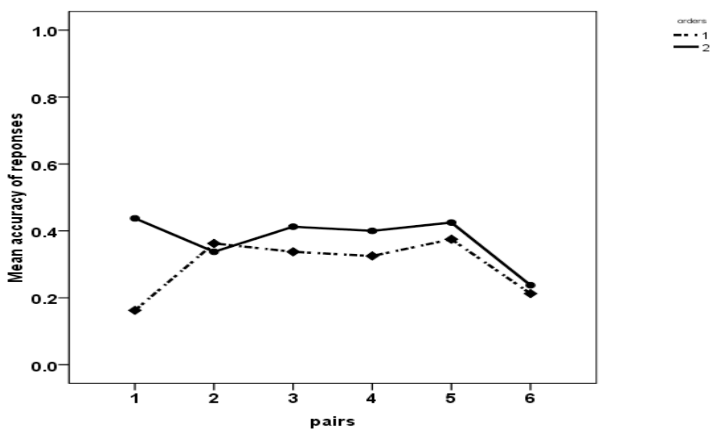


Figure 2.5 The mean accuracy of the responses of each pair in ascending order (dashed line) and in descending order (solid line) of native Dutch listeners.

From the figures, it can be seen that Mandarin listeners performed better when stimuli were presented in a descending order, (i.e., when the first token heard was closer to a T3 and the second token was closer to a T2). Moreover, the order effect becomes stronger when the stimuli pairs are closer to the T2-T3 boundary (e.g., pair 5). Interestingly, the same facilitation effect in the descending order is observed for Dutch participants too, albeit to a lesser extent. For Dutch listeners, the distances between the ascending and descending order response curve, however, did not become larger when approaching the tonal boundary. This provides evidence that Dutch listeners lack the representation of two separate tonal categories, T2 and T3, and that they fail to benefit from the cross-boundary variation.

The facilitating effect of the descending order could be explained by native Mandarin listeners' grammatical knowledge of the T3 sandhi rule. In this instance, a surface T2-T3 sequence may correspond to either an underlying T2-T3 sequence or a sandhied T3-T3 sequence. Consequently, when presented with a pair of tones in an ascending order, (i.e., step 1-3, step 2-4, step 3-5, step 4-6, step 5-7, and step 6-8) native listeners may have confused this with a sequence in which the second token is repeated. The surprising fact is that Dutch listeners display a similar pattern in discrimination without having any prior knowledge of Mandarin. This facilitating pattern occurs with Dutch listeners despite their lack of representations of the T2 and the T3 categories. The order facilitation observed in Dutch participants suggests that in an AX discrimination task, a T2-T3 sequence is intrinsically more difficult to distinguish than a T3-T2 sequence. Nevertheless, the figures demonstrate that for native Mandarin listeners: a) the accuracy difference between the ascending and descending order is larger than that of Dutch listeners; and b) the closer one gets to the boundary, the larger the accuracy difference between ascending and descending pairs. Hence, it seems that acquaintance with T3 sandhi strengthens the order effect for native listeners, while the categorical representation of lexical tones amplifies the order effect on the tonal boundary.

2.2.3 AXB discrimination task

Procedure

Participants heard three stimuli per trial along one continuum and were required to make a forced choice as to whether the second stimulus (X) sounded the same as or similar to the first stimulus (A) or the third stimulus (B) by clicking on the two choices labeled as “first” and “third” on the screen. All four possible orders (AAB, ABB, BAA and BBA) were presented. A and B always occurred at a two-step distance (e.g., 1-1-3, 1-3-3, 3-1-1, and 3-3-1 respectively). Six two-step triplets * 4 possible orders were repeated four times and resulted in a total of 96 test trials. To ensure the prevention of experimental bias, 48 extra trials with evidently distinct steps were used as filler trials. These filler trials used stimuli where A and B were more than 3 steps apart (e.g., triplet 2-2-6). All trials were randomized, and the ISI within the trials was set at 200 ms.

Results and discussion

A repeated measures ANOVA was conducted. In the model, the accuracy for each triplet (combinations collapsed) was the dependent variable, and triplets was a within-subject variable, while language background was a between-subject variable. The result reveals a significant main effect of triplets $F_{triplet}(5, 34) = 5.22$, $p<0.01$ and no significant main effect of language $F_{language}(1, 38) = 0.307$, $p>0.1$. The ANOVA also reveals that the two factors interact significantly $F_{interaction}(1, 38) = 4.731$, $p<0.05$. Figure 2.6 illustrates the mean accuracy rate of discrimination for each triplet by both language groups.

With the current task, native Mandarin listeners and native Dutch listeners do not differ significantly in terms of overall accuracy. In the AX discrimination task, in which only one referent was available, Mandarin listeners outperformed Dutch listeners. What this task reveals is that when Dutch listeners were supplied with two referents (i.e., both A and B), which offered more acoustical information, they were able to perform at a fairly-high or nearly native-like accuracy in lexical tone discrimination. For native Mandarin listeners, although their discrimination in the AXB task also improved considerably in comparison to the AX task, they did not perform significantly better than the tone naïve listeners.

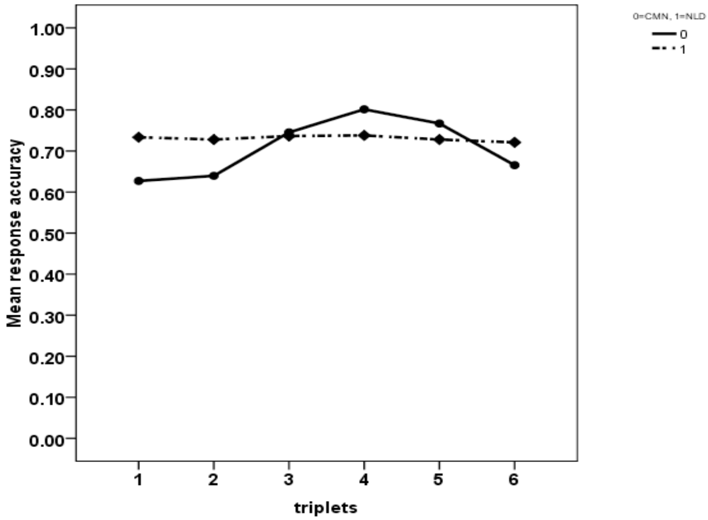


Figure 2.6 Overall mean response accuracy of each triplet (combinations collapsed) by native Mandarin listeners (solid line) and native Dutch listeners (dashed line).

Consistent with the AX discrimination results, Mandarin participants discriminated the triplets with the highest accuracy when A and B belonged to different tonal categories as shown in the identification task, (i.e., the triplet that involved steps 4 and 6). The correspondence between the discrimination peak and the identification boundary offers evidence for tonal CP. In comparison, Dutch listeners discriminated all the triplets with almost equal accuracy, and no cross-boundary facilitation could be observed. In terms of accuracy, Dutch listeners outperformed Mandarin listeners when discriminating triplets 1, 2 and 6. These three triplets involved the steps that were identified as belonging to a single tonal category by native Mandarin listeners. Native Mandarin listeners' lower accuracy suggests that within-category discrimination is inhibited and that non-tone language listeners outperform native tone language listeners in tracing the subtle acoustical differences between variations within the same category.

After reviewing the overall response accuracy of each triplet, the influence of combinations on the response accuracy was examined within each separate language group. Here, the response accuracy of each participant was the dependent variable, triplets and combinations were the within-subject variables, and the data was submitted to a repeated-measures ANOVA. For Mandarin listeners, both triplets and combinations have a significant main effect on the response accuracy, where: $F_{triplets}$ (5, 95) = 9.603, p<0.001; $F_{combinations}$ (2.05, 38.86) = 10.83[5], p<0.001. There was also a significant interaction between triplets and combinations, where: $F_{interaction}$ (15, 285) = 4.03, p<0.001. Figure 2.8 provides the mean response accuracy rate for each triplet in each combination by native Mandarin listeners. Bonferroni post hoc test indicated that the responses to triplets 1, 2 and 6 were significantly different from those to triplets 2, 3, and 4, where p<0.05. Furthermore, responses to the combination ABB differed significantly from the responses to combination BAA, and responses to the combination ABB differed significantly from the responses to the combination BBA, where p<0.05. Figure 2.7 illustrates the accuracy of native Mandarin listeners in the AXB experiments, where combinations are separated.

The results show that, in general, the Mandarin listeners discriminated more easily between the triplets in the BAA and BBA orders than in the ABB and AAB orders. In other words, they discriminated the triplets better if they first encountered a step closer to T3. Meanwhile, the triplets 3, 4, and 5 were easier to discriminate than triplets 1, 2, and 6. The least distinguishable triplets were the ABB combination at triplets 1 and 2. In both cases, a closer to T2 step preceded a closer to T3 step, and the last two syllables in the triplets were the same. In terms of the memory load, the listeners had to remember all three syllables in order to relay a correct judgment. In terms of T3 sandhi, the ABB combination is confusing in that the first syllable is closer to T2 and the second syllable is closer to T3, and thus a T2T3 sequence could be misconstrued as identical. These discrimination patterns demonstrate that at first, native listeners benefited from their knowledge of tonal categories (i.e., triplets that include two steps straddling the tonal boundary were easier to discriminate). Second, native listeners' knowledge of T3 sandhi influenced the discrimination, given that they discriminated more poorly

---

[5] Mauchly's test indicated that the assumption of sphericity had been violated for the the main effects of triplets, χ^2 (5) = 11.85. Therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε= 0.51) for the main effect of pairs.

between the triplets if A preceded B, as a T2T3 sequence is confusing due to T3 sandhi.

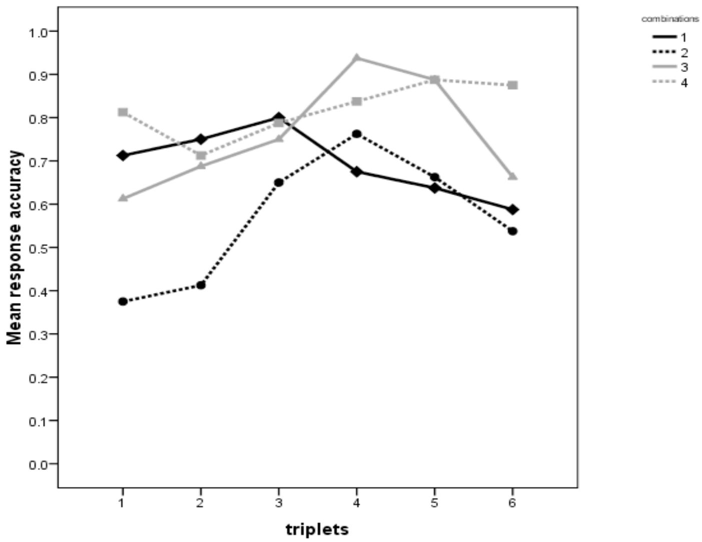


Figure 2.7 Mean response accuracy rate for each triplet in each combination by native Mandarin listeners. The solid black line represents the responses in AAB order, the dashed black line represents the responses in ABB order, the grey solid line represents the responses in BAA order, and the grey dashed line represents the responses in BBA order.

The same type of statistical analysis was also conducted for Dutch listeners' data. The repeated measures ANOVA revealed a significant main effect of combination, where: $F(3, 17) = 5.976$, $p<0.01$. However, this data revealed no main effect of triplets, where: $F(5, 15) = 0.60$, $p>0.1$. There was also no interaction between triplets and combinations, where: $F(15, 5) = 1.23$, $p>0.1$. Figure 2.8 depicts the mean response accuracy for each triplet in each combination by native Dutch listeners.

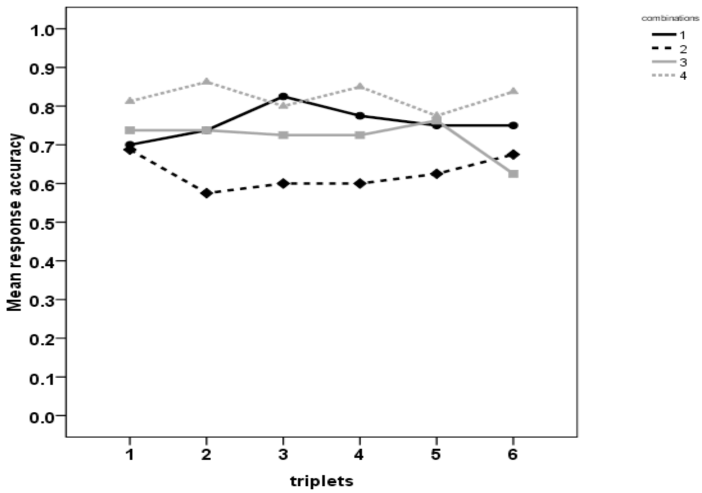


Figure 2.8 Mean response accuracy rate for each triplet in each combination by native Dutch listeners. The solid black line represents the responses in AAB order, the dashed black line represents the responses in ABB order, the grey solid line represents the responses in BAA order, and the grey dashed line represents the responses in BBA order.

For Dutch listeners, the response accuracy to different combinations can be ranked from high to low as BBA > AAB > BAA > ABB. From this, three patterns can be observed. First, due to the lower demand of memory load, discrimination was easier if the first two syllables, rather than the last two, were identical. Second, given that the grey lines generally falls above the black lines, it seems that judgments were easier to make if B preceded A rather than vice versa. This results in a pattern similar to that of Mandarin listeners. To have a clear view of the order effect, I collapsed the AAB order and ABB order into the sequence “AB” and collapsed the BAA and BBA order into the sequence “BA”. Then the response accuracy rate of each participant was submitted to a repeated measures ANOVA as dependent variable, and triplets and sequences as within-subject variables. A significant main effect of sequences was found, $F(5, 125) = 2.67$, $p<0.05$ while neither triplets, nor the interaction between sequences and triplets gave a significant effect: $F_{triplets}(5, 195) = 0.084$, $p>0.1$, $F_{interaction}(5, 195) = 0.84$, $p>0.1$. These results demonstrate that Dutch listeners discriminated significantly better when presented with “BA” sequences than with “AB” sequences, in the same fashion as the native Mandarin listeners. The third and final observation relates to the higher discrimination accuracy of the AAB combination than that of the BAA combination. Hence the advantage of lower memory burden surpasses that of the order of stimuli presentation.

Taking these results together, one can see that Mandarin listeners perceived T2-T3 categorically and that their discrimination of the triplets benefited if the A and B comparison straddled the tonal boundary. T3 sandhi influenced the native listeners’ discrimination in that the accuracy was higher if a closer to T3 step preceded a closer to T2 step. In this scenario, no confusion was introduced by T3 sandhi. On the other hand, Dutch listeners had no knowledge of tonal boundary and no facilitation could be observed for cross-boundary discrimination. More importantly, although unaware of Mandarin tones and the T3 sandhi rule, Dutch listeners demonstrated consistency with the Mandarin listeners in showing higher accuracy in discrimination when a closer to T3 step preceded a closer to T2 step. The discrimination asymmetry of Dutch listeners cannot be accounted for by language experience or structure frequency in the speech input, and therefore must reflect universal perceptual biases. Nevertheless, T3 sandhi in Mandarin amplified the order effect in the AXB discrimination task in the sense that the accuracy of Mandarin listeners was largely enhanced by the absence of ambiguity introduced by T3 sandhi. Moreover, the facilitation of order outweighed the memory burden.

2.2.4 General discussion

Referring back to the research questions posed above, the results of the CP experiments using a T2-T3 continuum are discussed with a focus on categorization and possible biases in the perception of T3 sandhi. Regarding categorization, Mandarin listeners did perceive the T2-T3 contrast categorically. Listeners were consistent in identifying the variations generated along the T2-T3 continuum, and when Mandarin listeners were asked to identify tones in isolation, they relied on the surface form of the tone. Even though it is claimed that this pair is only “partially contrastive” in Mandarin Chinese (Hume & Johnson 2003), native listeners were able to construct two contrasting representations of each tone separately. The confusion between the T2 and the T3 is only observed when the proper context for T3 sandhi to occur was present (i.e., when the listeners discriminated bisyllabic or trysillabic sequences). In order to trigger the perception of T3 sandhi, at least a bisyllabic sequence, where a T2 and a T3 co-occur, is needed. The categorical perception of the T2-T3 contrast is consistent with the findings of Francis et al. (2003) and Hallé et al. (2004).

When the tones occur together, however, the asymmetrical sequential phonological rule of T3 sandhi hampers the discrimination of sandhied T3-T3 sequences, which manifests as T2-T3. Most importantly, although non-native listeners lack the categorical representation of the T2 and the T3, they reveal a discrimination pattern similar to native listeners. In other words, they achieved a higher discrimination accuracy if a token closer to the T3 preceded a token closer to the T2. This did not hold true for the reverse, where the token closer to the T2 preceded a token closer to the T3. The same asymmetrical pattern is displayed in the discrimination of the T2-T3 pairs regardless of the native language of the listeners. Hence, this difference is maintained regardless of whether such a pitch difference is used phonemically. This finding suggests that sequential biases in the perception of tone contrasts can be universal in the sense of being intrinsic to human perception rather than dependent on linguistic experience.

Yet, when comparing the performance of native Mandarin and Dutch listeners, it becomes evident that native phonological knowledge of Mandarin magnifies the discrimination difference between the 'confusing' T2-T3 and the 'non-confusing' T3-T2 sequences. Compared to Dutch listeners, native Mandarin listeners expressed greater accuracy difference between the ascending order and the descending order in the AX task. In accordance with Mandarin listeners' knowledge of the boundary between the T2 and the T3, this study found that the closer a stimulus pair lies to this boundary, the stronger the sequential difference becomes. For Dutch listeners, who lack the category boundary, no such effect was found. Therefore, only for Mandarin listeners the phonological rule of T3 sandhi in the input language strengthens the order effect, as the accuracy difference between the ascending order and the descending order is more evident. Moreover, if it were not due to the input from Mandarin, then one would expect native listeners to express the same discrimination peak for the cross-boundary pairs and no amplification of the order effect towards the boundary. In other words, phonological grammar in Mandarin strengthens the natural bias in discrimination.

Another difference between native and non-native listeners relates to the strength of the sequential bias versus memory burden in discrimination tasks. It was found that the performance of non-native listeners was severely limited by general memory load, especially for the AXB discrimination task, which was the most cognitively demanding task of all the experimental tasks undertaken in this study. In brief, the more items non-native listeners had to hold in their memory for judgment, the higher the error rates became. Only when the memory burden was equal across different conditions, the natural bias of sequential perception of the T2-T3 became apparent. In contrast, the tonal processing of native listeners was driven largely by linguistic knowledge so that the difference in memory load between conditions became practically irrelevant to the discrimination accuracy. For instance, native listeners discriminated tonal triplets better than Dutch listeners as long as the initial token was T3. This result did not vary if the first two tokens or the last two tokens were the same. Despite the fact that native listeners might be less cognitively burdened than non-native listeners due to their familiarity with the lexical tones, this study observed that the discrimination is facilitated if a token closer to the T3 occurs first.

The order effect observed among Dutch listeners provides evidence that the perception of tonal sequences is, to some extent, universally biased. Meanwhile, the stronger order effect observed among Mandarin listeners provides evidence that the phonological rule, namely T3 sandhi, amplifies the universal bias in discriminating the T2 and the T3. It appears that Mandarin listeners have phonologized the bias in discriminating the T2 and the T3 into a sandhi rule. In return, the grammar of T3 sandhi consolidates the natural bias in linguistic processing. In other words, the T3 sandhi rule in Mandarin is likely to originate from a perceptual bias, which neutralizes the T2-T3 contrast when presented in a T2T3 order. Based on perception, the T2 and the T3 are the most similar tones and a T2T3 share the same structure with T3T3 in the sense that the two tones in the sequence are identical in perception. Hence, it is plausible that the T3 sandhi's surface and the underlying form are the two bisyllabic sequences that are perceived as the most similar. Yet, the origin or cause of the bias that hampers the discrimination in the T2T3 sequence is not clear. One possibility is that T3 has an acoustically more complex contour than T2. In this case, the processing of T3 in the second position of a tonal sequence draws more heavily upon memory load.

The second possible explanation is that for a discrimination task involving two sounds, the first works as a referent. In this scenario, the T3 is a better referent than T2. Given that the T3 is more complex, possibly it is also more specified. Hence, the specificity makes T3 distinctive in perception, and to exclude a different tone from T3 is fairly easy. The T2, in comparison, is less specific and may occupy a larger perceptual space. As a result, T2 tolerates more variation, and a T3 that follows a T2 could be misperceived as a peripheral realization of the T2.

Putting aside what may cause the bias, two questions linger and must be addressed before one can conclude an innate asymmetry in the discrimination of the T2-T3 and biases for T3 sandhi. First, although the order effect has been clearly observed in the CP experiment involving the T2-T3 continuum, it is not certain whether other tonal contrasts, which do not involve the T2 and the T3, also show a similar perceptual asymmetry. Second, thus far this study has only tested the discrimination between monosyllabic variants of the Mandarin T2 and T3. In order to have a more direct view of the perceptual distance between the underlying bisyllabic sequence T3T3 and the surface form T2T3, in comparison to other bisyllabic sequences such as T3T2, one needs a bisyllabic discrimination task. In the next section, section 2.3, the results found in Liu's study (2010) [6] are presented. This study tested the categorical perception of the T1 and the T4 by both native Mandarin listeners and native Dutch listeners. Liu's study provides a good comparison with the CP experiments involving the T2-T3 contrast discussed above. Later in Chapter 3, the results of both native Mandarin and native Dutch listeners participating in bisyllabic sequences discrimination tasks will be discussed.

2.3 RESULTS OF THE CATEGORICAL PERCEPTION EXPERIMENT INVOLVING T1 AND T4

[6] The results of the categorical perception experiment involving Mandarin T1 and T4 have previously been reported in Liu (2010). These results are also reported here in order to facilitate comparison with the categorical perception experiment involving Mandarin T2 and T3. The cross-continuum comparison is original work of the current dissertation.

Participants

Another 20 Mandarin and 20 Dutch participants were recruited for the the current experiment, which involved a continuum shifting from the Mandarin T1 to the Mandarin T4, with /ta/ being the tone bearing syllable (Liu 2010). All participants reported to have normal hearing and speech. Another three native Mandarin participants were tested, but excluded from the analysis for reasons of experiment bias (N = 1) and dialect interference (N = 2). One other native Dutch participant was tested, but excluded from the analysis due to his/her previous experience with Mandarin. The remaining 20 Dutch participants had no prior knowledge of Mandarin, nor had they been exposed to a Mandarin speaking environment.

Description of the tasks

The native Mandarin listeners and Dutch listeners participated in four tasks: a forced choice identification task, an AX discrimination task, a free choice discrimination task, and an AXB discrimination, respectively. As for the T2-T3 continuum, the listeners did not participate in the free choice identification task, the result of this task regarding T1-T4 continuum will not be reported here. In the remaining three tasks, the procedures and amount of stimuli were the same as previously described (i.e., the identification task previously described in section 2.1, the AX discrimination task reported in section 2.2, and the AXB task reported in section 2.3). The only difference is that native Dutch listeners were asked to identify the tone as "level" or "falling" by clicking on the buttons labeled as "level" or "falling" on the screen. Given that native Dutch listeners did not participate in a free choice identification task in this study, the results of the free choice identification task in Liu (2010) will not be presented here. Compared to the aforementioned experiment involving the T2-T3 continuum, the only difference with respect to Liu (2010) is that all the stimuli were generated between the endpoint T1 and the endpoint T4. In Liu (2010), the endpoint tokens were arranged along four continua each with four different tokens of the T1 and four different tokens of the T4. Yet, these different continua did not have a significant impact on the responses (see Liu 2010). For this reason, the four continua were collapsed in the following analysis. Figure 2.9 illustrates the F0 continuum created between the Mandarin T1 and T4.

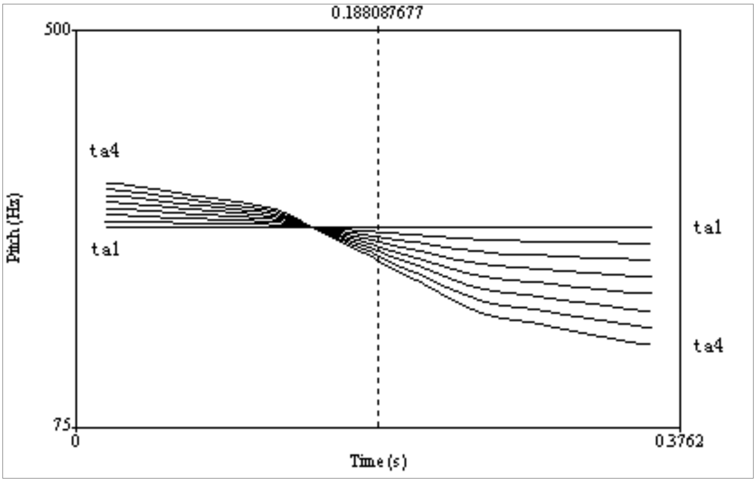


Figure 2.9 The continuum created between the Mandarin T1 and T4 conducted on a /ta/ syllable.

2.3.1 Results of the identification task

Before comparing the performance of the groups, the proportion of the T4 responses was calculated for each participant. Figure 2.10 depicts the proportion of the T4 responses at each step according to language background.

For each language group, a PROBIT analysis (Best & Strange 1992) was used to calculate the category boundary locations and the degree of a potential categorical perception (Table 2.2).

Experiment1	Intercept		Slope	
	Chinese	Dutch	Chinese	Dutch
/ta1/-/ta4/	2.927	4.117	2.276	0.811

Table 2.2 The position of the category boundary and the slope shift between the T1 and the T4 along the identification curve.

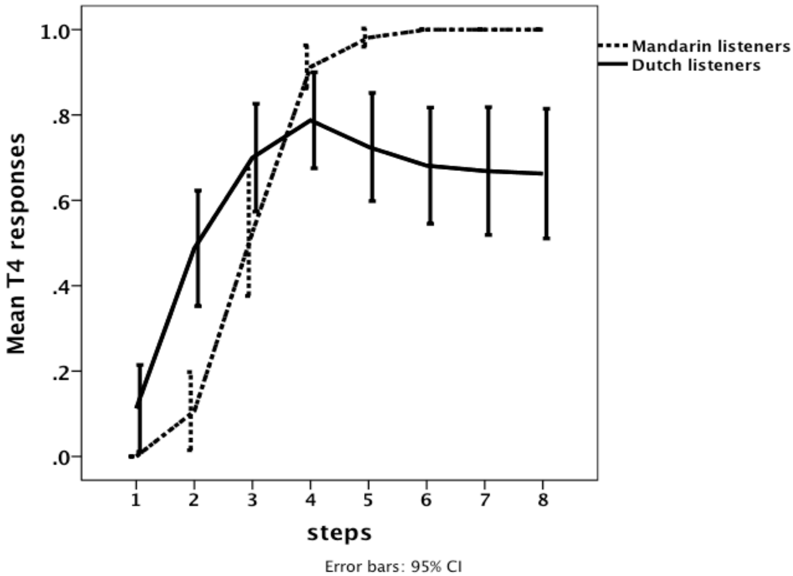


Figure 2.10 Mean proportion of the T4 responses at each step by native Mandarin listeners (solid line) and native Dutch listeners (dashed line).

The intercept differed significantly between the two language groups, where: $F (1, 22) = 22.61$, $p = 0.02$. Mandarin listeners perceived steps 1, 2, 3 as the T1, while the remaining five steps were perceived as the T4. The boundary of Dutch listeners, who do not have representations of the tones, fell in the middle of the continuum. The slope was significantly sharper for the Mandarin listeners, where: $F (1, 38) = 2.8$, $p < 0.001$. A repeated measures ANOVA was conducted. This test used the proportion of the T4 responses at each step of each participant as the dependent variable, steps as the within-subject variable, and the language background as the between-subject variable. Results reveal a significant main effect of steps, where: F_{steps} (7, 32) = 58.16, $p<0.001$. Results also illustrate a significant main effect of language background, where: $F_{language\ background}$ (1, 38) = 4.20, $p<0.05$. Lastly, steps and language background also interacted significantly, where: $F_{interaction}$ (7, 32) = 9.35, $p<0.001$.

Figure 2.10, in combination with the results of the PROBIT analysis, suggests that Mandarin listeners perceived steps 1 and 2 as the T1, and steps 4 through 8 as the T4. Step 3 remained ambiguous for native listeners. Dutch listeners, on the other hand, identified steps 5 through 8 as having a "falling" pitch contour and steps 1 through 4 as having a "level" contour. Hence, "level" and "falling" occupied equal perceptual space. The responses of Dutch listeners can be considered natural identification of "level" and "falling" without interference of linguistic knowledge of level and falling tones. In comparison, in the identification task involving the T2-T3 continuum, Dutch listeners showed slightly more frequent identification of the T2 than the T3. Regarding the variance observed in the responses at each step of the T1-T4 continuum, similar to the identification of the T2-T3 continuum, it was found that Mandarin listeners are quite consistent when identifying the close-to-endpoint tokens while some variation occurs for the ambiguous tokens. Dutch listeners, however, demonstrated fairly equal amounts of variation for all the steps along the continuum.

To summarize, only the Mandarin listeners formed two tonal categories for the T1 and the T4. The sharp shift along the identification curve, together with the high level of consistency in identifying the close to endpoint tokens, is evidence that these listeners process lexical tones in a phonemic manner. Based on the evidence discussed thus far, regarding the T2-T3 and the T1-T4 continuum, Mandarin listeners demonstrated a categorical perception whereas native Dutch listeners perceived both contrasts in a psycho-acoustical fashion. It is important to mention that even though the Mandarin T2 and T3 sequences undergo the T3 sandhi process, and T2 and T3 are neutralized in certain contexts, native listeners have two well-established categories for the T2 and the T3 when they are presented in isolation. In other words, the perception of the T2 and the T3 in isolation is not hindered by T3 sandhi. Furthermore, T3 sandhi only begins to play a role in the perception in a bisyllabic context.

2.3.2 Results of the AX discrimination task

A repeated measures ANOVA was undertaken. Here, the mean accuracy rate of each participant at each step was the dependent variable, pairs (steps 1-3, 2-4, 3-5, 4-6, 5-7, 6-8) were the within-subject variables, and the language background was the between-subject variable. The results revealed a significant main effect of pairs, where: F_{pairs} (5, 34) = 2.04, $p<0.001$. Yet, there was no significant main effect of language background, where: $F_{language\ background}(1, 38) = 0.07$, $p>0.1$. The pairs and the language background, however, interacted significantly, where:

$F_{interaction}$ (5, 190) = 9.90, $p<0.001$. Figure 2.11 depicts the mean response accuracy rate of each pair for both the Mandarin listeners and the Dutch listeners.

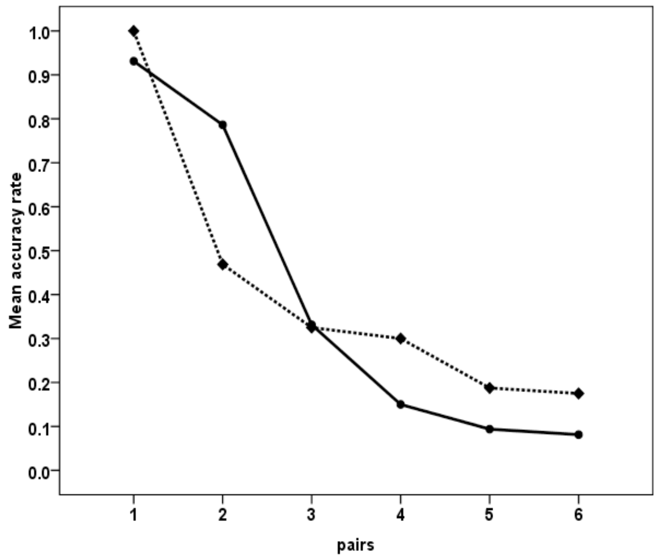


Figure 2.11 Mean response accuracy rate of each pair of the native Mandarin listeners (solid line) and the native Dutch listeners (dashed line).

The ease of discrimination of Mandarin listeners matches the identification boundary observed along the identification curve. Mandarin listeners were most accurate in discriminating pairs 1 and 2, which straddled the tonal boundary. This provides additional evidence for the categorical perception of the Mandarin T1 and T4. Meanwhile, Dutch listeners discriminated pair 1 with 100% accuracy, which suggests that they were able to detect the difference between a level pitch and a falling pitch very easily, even when the change resided in a subtle drop of pitch contour slope. However, for all the other pairs Dutch listeners' accuracy rates dropped to below the level of chance. Yet, Dutch listeners discriminated pairs 3, 4, 5, 6 (which are all categorized as the T4 by Mandarin listeners) better than Mandarin listeners. The distinctive role of lexical tones in Mandarin causes native listeners to perceive tones in a categorical (phonemic) way (i.e., where within-category discrimination is inhibited and cross-boundary discrimination is facilitated). In contrast, Dutch listeners perceive the pitch difference in a psycho-acoustical way and are capable of tracking the subtle change of the pitch contours due to a lack of experience with lexical tones. These results demonstrate that adult listeners are able to keep track of the realization of subtle pitch differences in syllables, but the phonemic function of lexical tones inhibit native Mandarin listeners from perceiving these acoustical variations of the same tone (Francis et al. 2003, Hallé et al. 2004).

Given that a clear order effect was observed for native Mandarin listeners in the AX discrimination task involving the T2-T3 continuum, the same task involving the T1-T4 continuum will be discussed with distinct orders. For the ease of comparison, the pairs going from the T1 to the T4 (e.g., steps 1-3, 2-4) were defined as an "ascending order" and pairs going from the T4 to the T1 (e.g., steps 3-1, 4-2) were defined as a "descending order". The accuracy rates for both pairs and order were compared separately for each language group. A repeated measures ANOVA was used. The mean accuracy rate of each pair in each order was the dependent variable, and the pairs (1, 2, 3, 4, 5, 6) and the order (ascending or descending) were the within-subject variables. Results demonstrate that both pairs and order have a significant main effect on the response accuracy, where: F_{pairs} (5, 15) = 85.73, $p<0.001$, F_{order} (1, 19) = 7.24, $p<0.05$. There was also a significant interaction between pairs and order, where: $F_{interaction}$ (5, 15) = 3.78, $p<0.05$. For native Dutch listeners, however, only the pairs have a significant main effect on response accuracy, where: F_{pairs} (5, 15) = 90.61, $p<0.001$. The order, however, failed to be a significant factor, where: F_{order}(1, 19) = 0.55, $p>0.1$. Moreover, the interaction between pairs and order showed no significance, where: $F_{interaction}$ (5, 15) = 1.75, $p>0.1$. Figure 2.12 and Figure 2.13 illustrate the mean accuracy in the AX discrimination task and depict the responses of native Mandarin listeners and native Dutch listeners in different orders respectively.

Only for Mandarin listeners, the descending order appeared easier to discriminate. One possible interpretation of this pattern relates to the fact that the pitch decreases towards the end of a phrase due to the down-step in intonation in running speech. If a step closer to the T1 preceded another step closer to the T4, then the latter could be misperceived as carrying the same tone as the first syllable. The Dutch listeners did not show any order effect in discrimination, but they were extremely sensitive to the difference between a level tone (step1) and a slightly falling tone (step3). When both tones had falling contours, they failed to distinguish how much a pitch contour dropped.

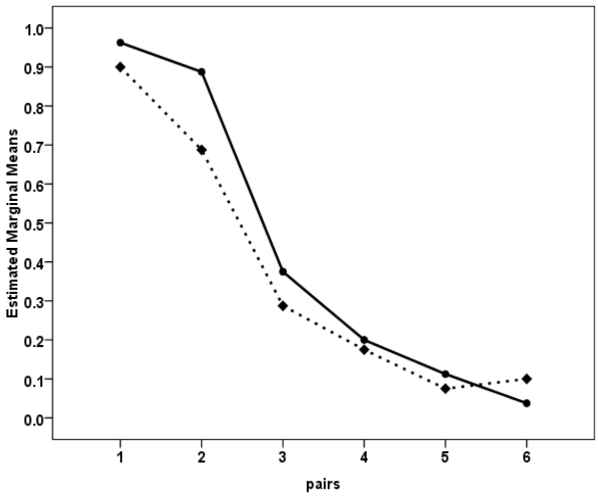


Figure 2.12 Mean response accuracy at each step in ascending order (dashed line) and descending order (solid line) of Mandarin listeners

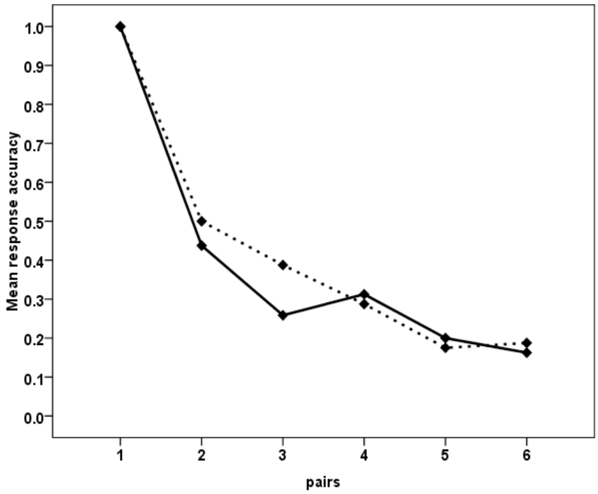


Figure 2.13 Mean response accuracy at each step in ascending order (dashed line) and descending order (solid line) of Dutch listeners.

If one compares the performance of Mandarin listeners on the AX task involving the T2-T3 continuum to the task involving the T1-T4 continuum, it becomes clear that order of stimuli presentation has a much more dramatic influence on the discrimination accuracy along the T2-T3 continuum. Moreover, the enlargement of the order effect at the pairs that crossed the tonal boundary was more obvious for the T2-T3 continuum. For native Dutch listeners, on the other hand, there was a small order effect only in the T2-T3 continuum. The different discrimination patterns of the Dutch listeners with the T1-T4 continuum and the T2-T3 continuum suggests that T2 and T3 sounds more similar in a T2-T3 order than in a T3-T2 sequence while T1 and T4 are spared of order related confusion. Therefore the T3 sandhi rule may reside on this natural perceptual bias that is specific to the tones involved in the sandhi process.

2.3.3 Results of AXB discrimination

In this task, the A and the B were again always two steps apart while the X could be either the same as A or the same as B. The AXB triplets were defined as triplet 1 if the A and the B involved step 1 and step 3, as triplet 2 if the A and the B involved step 2 and step 4, and so on.

In order to investigate how language background influenced the accuracy in discrimination, a repeated measures ANOVA was carried out. This test used the accuracy rate of each participant at each triplet as the dependent variable, the triplet as the within-subject variable, and the language background as the between-subject variable. Results demonstrate that triplets have a significant main effect on response accuracy, where: $F_{triplets}$ (5, 34) = 29.45, p<0.001. There is also a demonstrated tendency for language background to have a main significant effect, where: $F_{language\ background}$ (1, 38) = 3.53, p<0.07. Triplets and language background also interacted significantly, where: $F_{interaction}$ (5, 34) = 11.09, p<0.001. Bonferroni post hoc analysis indicated that for the triplets, responses to triplets 1, 2 and 3 were significantly different from one another. In contrast, the responses to triplet 4, 5, 6 were not significantly different from one another, and the responses to triplets 1, 2, 3 were significantly different from the responses to triplets 4, 5, 6. Figure 2.14 illustrates the mean response accuracy for each triplet for each language group.

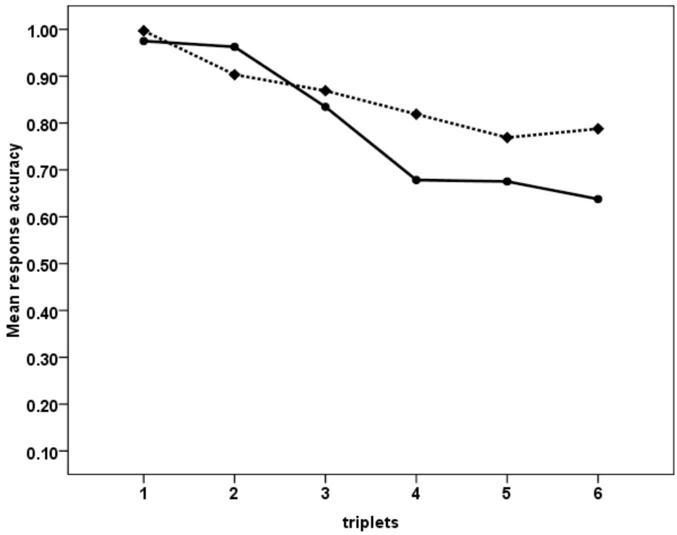


Figure 2.14 Mean AXB response accuracy for each triplet involving the T1-T4 contrast of native Mandarin listeners (solid line) and native Dutch listeners (dashed line).

Consistent with the AX discrimination results, Mandarin participants discriminated the A-B pairs best when the A and the B belonged to different tonal categories as shown in the identification tasks. A discrimination peak was predicted by the category boundary, offering evidence for tonal CP. Without the facilitation from the knowledge of the tonal categories, Dutch listeners performed more poorly than Mandarin listeners in discriminating the triplet 2, in which the A and the B straddled the tonal boundary according to Mandarin listeners' identification. However, Dutch listeners outperformed Mandarin listeners in tracing the subtle differences for the within-category triplets. Similar to the AX discrimination task, Dutch listeners discriminated pair 1 perfectly, showing high sensitivity to changes in pitch direction.

In order to find out whether the combinations of the stimuli also affected the response accuracy in the T1-T4 continuum discrimination within each language group as they did in the T2-T3 continuum discrimination, a repeated measures ANOVA was conducted. Here, the independent variable was the response accuracy rate of each participant for each triplet in each combination, with the within-subject variables being defined by triplets (1, 2, 3, 4, 5, 6) and combinations (AAB, ABB, BAA, BBA).

For Mandarin listeners, the results demonstrate a significant main effect of triplets as well as combinations, where: $F_{triplet}(5, 15) = 28.00$, $p<0.001$, and $F_{combination}(3, 17) = 6.49$, $p<0.005$, respectively. There was also a significant interaction between triplets and combinations, $F(15, 5) = 9.49$, $p<0.05$. Bonferroni post hoc tests indicate that the responses of triplet 3 is significantly different from all other triplets ($p<0.005$). In contrast, the responses to triplets 1 and 2 do not differ significantly from one another ($p>0.1$), and the responses to triplets 4, 5, 6 do not differ significantly from one another ($p>0.1$). However, the responses to triplets 4, 5, 6 are significantly different from that to triplets 1, 2 ($p<0.001$). Concerning the combinations, the Bonferroni post hoc test indicates that the combination BAA tends to be significantly different from all the other combinations ($p_{BAA\text{-}AAB}<0.05$, $p_{BAA\text{-}ABB}<0.06$, $p_{BAA\text{-}BBA}<0.005$), while responses to the remaining three combinations do not significantly differ from one another ($p>0.05$). Figure 2.15 illustrates the mean accuracy of Mandarin listeners of each triplet in each combination.

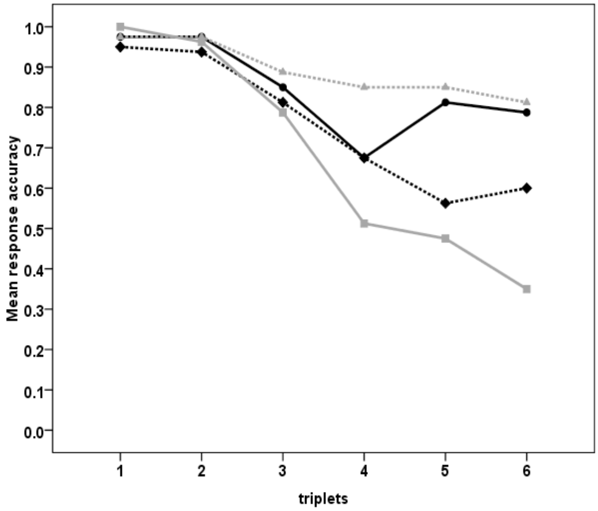


Figure 2.15 Mean response accuracy of Mandarin listeners of each triplet in combination AAB (black solid line), ABB (black dashed line), BAA (grey solid line), BBA (grey dashed line).

The accurate response proportions were higher for orders BBA and AAB than for ABB and BAA. A possible explanation for this discrepancy would be a higher memory load for the latter orders: for orders AAB and BBA, listeners could already decide after hearing the first two syllables, but for the other two orders, they must hear all three syllables before deciding. Collapsing the AAB together with BBA, and ABB together the BAA combination using the position of the identical steps (either the first two syllables were identical steps, or the last two syllables were identical steps) as a within-subject variable, with triplets being the other within-subject variable, and the mean response accuracy of each triplet in each position was submitted to a repeated measures ANOVA. A significant main effect of position was found $F_{position}$ (1, 19) = 10.77, p<0.005, as well as of triplets: $F_{triplets}$ (5, 15) = 27.64, p<0.001. However, there was no significant interaction between triplets and position: $F_{interaction}$ (5, 15) = 2.15, p>0.1. Figure 2.16 depicts the mean response accuracy of native Mandarin listeners of each triplet according to the position of the identical syllable.

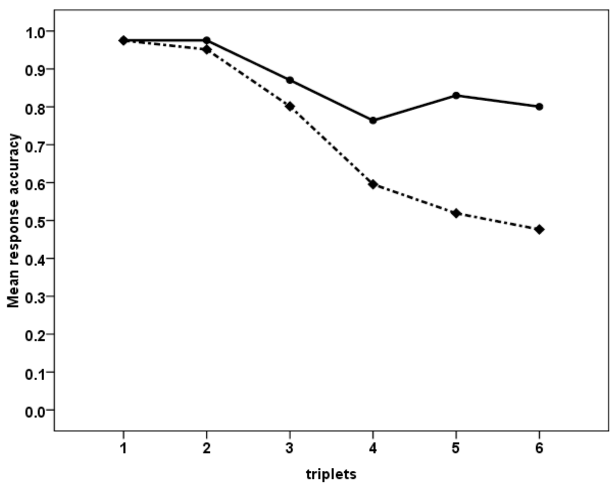


Figure 2.16 Mean accuracy rate of Mandarin listeners of each triplet while either the first two syllables (solid line), or the last two syllables (dashed line), were identical steps.

Mandarin listeners' proportions of correct responses were higher than 90% for pairs 1 and 2, no matter in which order the stimuli were presented. These two triplets were where A and B straddled the category boundary. The order effect only became prominent when the pairs moved towards the T4 endpoint. This pattern suggests that for Mandarin listeners, the cross-boundary facilitation due to their knowledge of tonal categories was strong enough to overcome the disadvantage of a higher memory load needed for a successful judgment. The memory burden only started to affect discrimination when the stimuli belonged to the same tonal category.

If we carry out the same statistical tests with the responses of Dutch listeners, it can be seen that triplets is a significant main factor for response accuracy, $F_{triplets}(5, 15) = 6.81$, $p<0.005$. But both combinations and the interaction between triplets and combinations failed to have a significant effect on the response accuracy: $F_{combinations}$ $(3, 17) = 1.93$, $p>0.1$, $F_{interaction}$ $(15, 5) = 1.59$, $p>0.1$. Figure 2.17 illustrates the mean response accuracy of the Dutch listeners of each triplet in each combination.

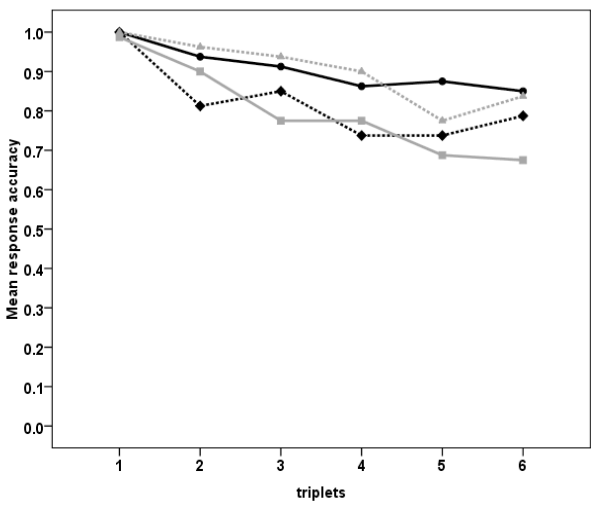


Figure 2.17 Mean accuracy response of Dutch listeners of each triplet in the AAB (black solid line), ABB (black dashed line), BAA (grey solid line), and BBA combinations (grey dashed line).

As was the case for native Mandarin listeners, native Dutch listeners also showed higher accurate response proportions for orders BBA and AAB than for

ABB and BAA. Again, this pattern could be explained by the lower memory load when processing the combinations in which the first two syllables are identical steps. Again, the responses of the AAB and BBA combination as well as the responses of the ABB and BAA combination were aggregated to introduce the factor of position (the first two syllables were identical steps or the last two syllables were identical steps). A repeated measures ANOVA was carried out taking response accuracy of each participant for each triplet as dependent variable, taking triplets (triplet 1, 2, 3, 4, 5, 6) as well as position as within-subject factors. It can be seen that both triplets and positions have a significant main effect on the response accuracy of Dutch listeners: $F_{triplets}$ (5, 15) = 6.75, $p<0.005$, $F_{position}$ (1, 19) = 5.02, $p<0.05$, but there was no significant interaction between triplets and positions, $F_{interaction}$ (5, 15) = 1.411, $p>0.1$. Figure 2.18 illustrates the mean response accuracy of each triplet according to the position of the identical syllable.

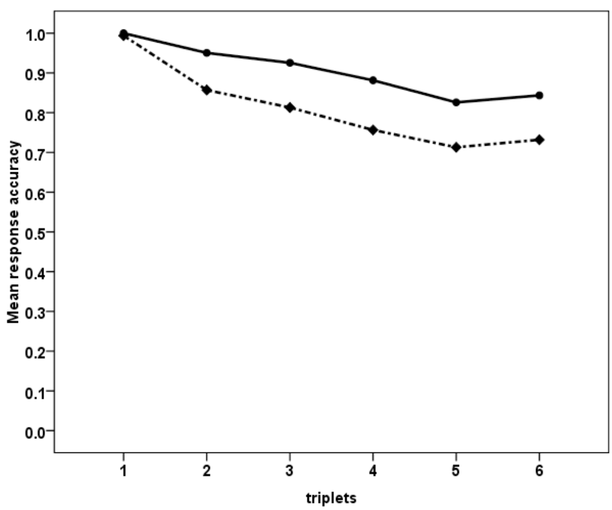


Figure 2.18 Mean response accuracy of Dutch listeners of each triplet while either the first two syllables (solid line), or the last two syllables (dashed line) were identical steps.

For Dutch listeners, although identification tasks predicted steps 4 and 5 as cross-boundary tokens, no peak could be observed in discrimination tasks for triplet 3 and triplet 5. The lack of a discrimination peak indicates that although Dutch listeners perceive the acoustic difference between level and falling tones, they do not have a representation of two tonal categories. Moreover, regardless of the stimuli presentation order, Dutch listeners discriminated pair 1 successfully. This suggests that without interference of linguistic knowledge, the direction of a pitch contour remains a salient and reliable psycho-acoustic cue for pitch discrimination in non-native listeners, regardless of a difference in memory load.

2.4 CROSS-CONTINUUM ANALYSIS

2.4.1 Cross-continuum analysis of degree of CP

CP can be seen as the relationship between identification and discrimination tasks, with the core of CP lying not only in how listeners categorize sounds, but also whether listeners are able to tell the difference between sounds that straddle the category boundary (Gandour 1978). If participants perceive stimuli in a categorical manner, their performance in identification tasks should determine their discrimination task performance. In principle, the stronger the correlation between the identification and discrimination functions, the higher the degree of CP. The degree of CP can be calculated using the "categorical-perception index" (Van Hessen & Schouten 1999):

$$CP = [r / (1 + 2 * | p(obt) – p(pred) |)] * 100$$

In this formula, CP refers to the degree of categorical perception. By definition, the index value refers to how categorically the target is perceived. The CP value may vary from 0 to 100. The numerator r represents the correlation coefficient between an identification function and a discrimination function. The denominator includes "p(obt) – p(pred)," representing the difference between the discrimination and identification function across data.[7] The degree of CP can be observed in Table 2.3.

Regardless of the task combination between the identification and discrimination, the degree of CP is always higher for the Mandarin listeners than for the Dutch listeners. The mean difference between the two groups is fairly large. Hence, it is likely that the Mandarin listeners' performance in identification tasks

[7] For more information about the formula and algorithm, please refer to Van Hessen & Schouten, 1999.

predicts their performance in discrimination tasks, but for Dutch listeners the predictive power of the identification curve for the discrimination tasks is much lower. To summarize, the results of the CP index analysis confirm the hypothesis that Mandarin listeners are more likely to perceive the lexical tonal stimuli categorically than Dutch listeners.

Tones	Task correlation		CP index	
	Identification	Discrimination	Mandarin	Dutch
T1-T4	Forced choice	AX	41.144	35.974
T1-T4	Forced choice	AXB	44.382	20.212
T1-T4 Mean			48.075	35.019
T2-T3	Forced choice	AX	38.543	18.973
T2-T3	Forced choice	AXB	16.720	1.907
T2-T3 Mean			29.046	10.440
Grand Mean			38.561	22.730

Table 2.3 Degree of CP of Mandarin and Dutch listeners for the T2-T3 continuum and T1-T4 continuum.

2.4.2 Cross-continuum analysis for the order effect

The focus of this chapter is to find the similarities and differences between the T1-T4 and T2-T3 contrasts and to see whether there are perceptual biases that favor T3 sandhi to occur in the way as it does now. This issue can be analyzed using cross-continuum analysis. As order information interferes most in discrimination tasks within each language group, the results of the two continua (T1-T4 and T2-T3) are analyzed for the AX and the AXB discrimination separately to test the order effect.

For the AX task for Mandarin listeners, a repeated measures ANOVA was conducted, where the accuracy rate of each participant was the dependent variable and order (ascending or descending) was the within-subject factor. Continua (T1-T4 or T2-T3) was the between-subject factor. Results show that both order and continua work as a significant factor: F_{order} (1, 38) = 22.05, p<0.001, $F_{continua}$ (1,

38) = 156.40, p<0.001. Results also show significant interaction between order and continua, $F_{interaction}$ (1, 38) = 26.51, p<0.001. Figure 2.17 illustrates the accuracy rate of the participants when presented with either the T1-T4 or T2-T3 continuum, with the order separated. For the AXB task, again I carried out a repeated measures ANOVA, using the accuracy rate of each participant as dependent variable and order (A occurred first or B occurred first) as the within-subject factor. Here, continua (T1-T4 or T2-T3) was the between-subject factor and it was found that both order and continua work as significant factors: F_{order} (1, 38) = 31.44, p<0.001, $F_{continua}$ (1, 38) = 5.90, p<0.02. There was also significant interaction between order and continua: $F_{interaction}$ (1, 38) = 45.59, p<0.001. Figures 2.19 and 2.20 depict the accuracy rate of Mandarin listeners in AX and AXB tasks respectively with the continua and order separated.

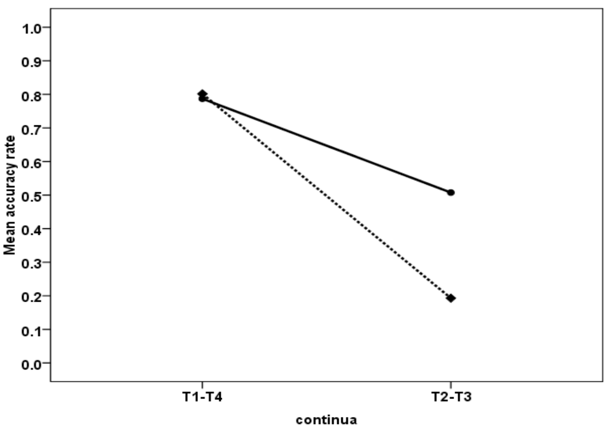


Figure 2.19 Accuracy rate of Mandarin participants for the AX discrimination task for the T1-T4 continuum and T2-T3 continuum in ascending order (dashed line), and descending order (solid line).

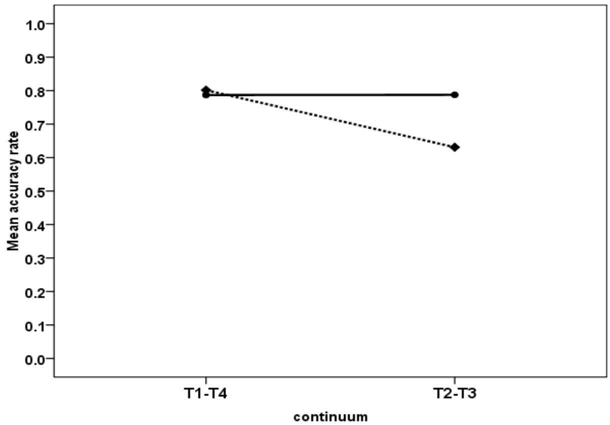


Figure 2.20 Accuracy rate of Mandarin participants for the AXB discrimination task for the T1-T4 continuum and T2-T3 continuum when A occurred first (dashed line), and B occurred first (solid line).

Figures 2.19 and 2.20 can be discussed from several perspectives. First, for the T1-T4 continuum, the native listeners reached similar accuracy in both the AX and AXB discrimination tasks. Compared to the AX experiment, the AXB experiment presents the listeners with more reference, and hence in theory is more favorable for successful discrimination. Yet, for the T1-T4 continuum, native listeners reached fairly high accuracy, regardless of in which task they participated. Second, no clear order effect could be found for the discrimination between the stimuli along the T1-T4 continuum.

However, for the T2-T3 continuum, we can see that compared with the AX task, the discrimination accuracy largely improved in the AXB task. Moreover, contrasting the T1-T4 continuum, an evident order effect was observed in the T2-T3 continuum; listeners more effectively discriminated the stimuli when a token closer to T3 preceded a token closer to T2. These patterns suggest that even for native listeners, the discrimination of a T2-T3 contrast is more difficult than the discrimination of a T1-T4 contrast, and hence only the discrimination between T2-T3 benefited from the richer context in the AXB task. It is important to note that the order effect in discrimination is particular to the T2-T3 contrast, and regardless of the task, both Mandarin and Dutch listeners more effectively discriminated between T2 and T3 tokens if a token closer to T3 preceded a token closer to T2.

For Mandarin listeners, the order effect specific to the T2-T3 continuum can be explained by the T3 sandhi rule, which neutralizes a T3T3 into a T2T3. As a result, when native listeners were presented with a T2T3, they may have confused the sequence with the underlying T3T3 form, demonstrating a tendency to misperceive the two syllables as carrying the same tone. However, the question whether the order effect comes from knowledge of phonological regularities present in ambient input, or whether it reflects universal biases in perception cannot be answered by only looking at the performance of native listeners. Evidence of innate perceptual biases needs to be gathered from naïve listeners without knowledge of Mandarin tones.

Regarding Dutch listeners, Liu and I analyzed the effect of the continuum in the same way as for native Mandarin listeners. For the AX task, no significant effect was found for continua, $F_{continua}$ (1, 38) = 1.75, $p>0.05$, and order showed a marginal significant effect: F_{order} (1, 38) = 3.82, $p= 0.058$. There was no significant interaction between continua and order: $F_{interaction}$ (1, 38) = 0.85, $p>0.05$. Figure 2.22 provides the mean accuracy rate of Dutch listeners in the AX task when presented with either the T1-T4 or T2-T3 continuum. For the AXB task, on the other hand, a significant effect was found for both order and continua: F_{order} (1, 38) = 5.26, $p<0.05$, $F_{continua}$ (1, 38) = 9.53, $p<0.01$. There was also a significant interaction between order and continua, $F_{interaction}$ (1, 38) = 9.56, $p<0.01$. Figure 2.23 depicts the mean accuracy rate of Dutch listeners in AXB task when presented with the T1-T4 continuum and when presented with the T2-T3 continuum.

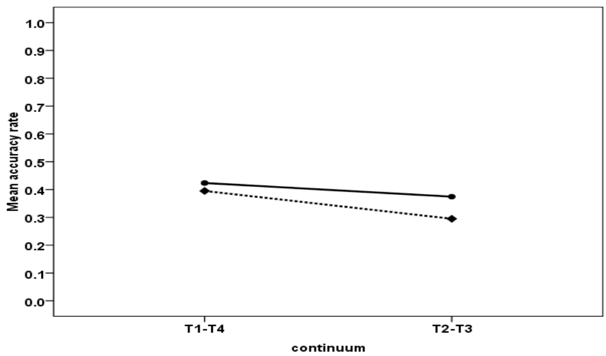


Figure 2.22 Mean accuracy rate of Dutch listeners in the AX task when presented with either the T1-T4 continuum or the T2-T3 continuum in ascending order (dashed line) and descending order (solid line).

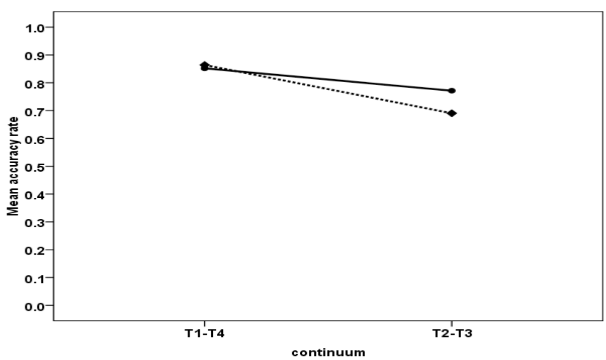


Figure 2.23 Mean accuracy rate of Dutch listeners in the AXB task when presented with the T1-T4 continuum and the T2-T3 continuum when A occurred first (dashed line), and when B occurred first (solid line).

As can be seen from Figures 2.22 and 2.23, native Dutch listeners performed quite poorly in the AX task, and neither the discrimination of the T1-T4 continuum nor that of the T2-T3 continuum reached an overall accuracy higher than chance level. The listeners, however, benefited from the richer information in the AXB discrimination, reaching a fairly high accuracy. In both the AX and AXB tasks, the order effect was more visible for the T2-T3 continuum. In the AXB task, the significant interaction between continua and order together with Figure 2.24 indicate that the order effect was specific for the T2-T3 continuum, while the discrimination of the T1-T4 continuum was not considerably affected by the order of stimuli presentation. The absence of significant interaction between continua and order in the AX task could be a result of generally low discrimination accuracy in both continua. Nevertheless, it can still be seen that the accuracy difference between the ascending and descending order became larger for the T2-T3 continua.

As native Dutch listeners did not have either exposure to or knowledge of Mandarin tones, the order effect observed among Dutch listeners cannot be a result of language experience; rather, it must reflect certain perceptual biases innate to the human auditory system. To put it differently, it was easier for naïve listeners to discriminate between tokens along the T2-T3 continuum if a token closer to T3 preceded a token closer to T2 than vice versa. If we compare the performance of native Mandarin listeners to that of native Dutch listeners, it can be seen that both groups demonstrated a similar discrimination pattern, meaning that the discrimination for the T1-T4 continuum was more accurate than that for the T2-T3 continuum, and the order effect was only found in the T2-T3 continuum. Nevertheless, the order of stimuli presentation has a stronger effect on Mandarin listeners than on Dutch listeners. A possible explanation for this phenomenon could be that T3 sandhi in the input amplifies the order effect among Mandarin listeners. Therefore, it is likely that T3 sandhi occurs on the basis of natural perceptual biases, which tends to misperceive the two syllables in a T2T3 sequence as being identical. The phonologization of this bias, the T3 sandhi rule, requires a T3T3 to change to a T2T3 obligatorily. In other words, the phonological grammar residing in a natural bias, and the grammar in return encourages the native Mandarin listeners to equalize T2T3 as T3T3.

2.5 GENERAL DISCUSSION

The participants in the experiments in this chapter were native Dutch and native Mandarin listeners. Two identification tasks and two discrimination tasks were conducted for both T1-T4 and T2-T3 perception. For both the T1-T4 and T2-T3 continua, Mandarin native listeners were consistent in identifying a tonal token as falling within one category or another, as shown by the abrupt change along the identification curves. Moreover, their discrimination peak was correctly predicted by the identification boundary. These findings provide strong evidence that lexical tones are processed categorically in a phonemic way by native listeners. On the contrary, although Dutch listeners were able to track subtle pitch differences in the stimuli, they showed no trace of categorical tonal representation in these multiple tasks, and they processed the stimuli in a psycho-acoustical way. The findings show congruence with previous studies in that native speakers of a non-tonal language do not perceive linguistic tones categorically, regardless of which non-tonal language the participants speak (Hallé et al. 2004; Xu et al. 2006). Therefore, it seems safe to conclude that language experience is a crucial factor in the development of tonal categories.

The discrimination results of the T1-T4 (level versus falling) contrast demonstrate a clear difference between native and non-native listeners. Non-native listeners were sensitive to phonetic details in the stimuli, and they discriminated level tones from contour tones by relying on the change of direction of the pitch contour. Yet, this sensitivity was constrained by the acoustical property of the tokens, as Dutch listeners failed to discriminate tonal pairs that contained falling contours. To clarify, a level and a falling contour are very distinctive, but two falling contours that differ in the degree of the falling are difficult to perceive. The perception of the native Mandarin listeners was tuned by the phonology of lexical tones. The Mandarin listeners were no longer sensitive to phonetic detail within a tonal category and perceived the speech sounds categorically due to the phonemic function of lexical tones. The acoustical property of the stimuli seemed to be more influential for non-native listeners than for native listeners, whereas categorical perception influenced only native listeners. For native Mandarin listeners, the discrimination of tokens belonging to different linguistic categories was facilitated while discrimination of within category difference was inhibited. What is more, in the AX and AXB discrimination tasks involving the T1-T4 continuum, no consistent order effect could be observed among either native or non-native listeners. Whether a T1 preceded a T4, or a T4 preceded a T1 did not markedly influence the accuracy. Therefore, the listeners were not biased to discriminate the contrast better in one order compared to the other.

The findings of our set of experiments using a T2-T3 continuum are three-fold. Firstly, though the T2-T3 contrast is to some extent neutralized by T3 sandhi and is claimed to be only “partially contrastive” (Hume & Johnson 2003), native listeners were able to build up two contrastive representations of each tone separately. Only when the tones occurred together did the asymmetrical sequential phonological rule of T3 sandhi hamper the discrimination between T2-T3. Due to T3 sandhi, the syllables in a T2-T3 sequence can be an underlying T3-T3 sequence, and the two syllables have the same underlying tone. Therefore, Mandarin native listeners' knowledge of the phonological rule of T3 sandhi only has an impact on their perception of co-occurrences of T2 and T3, rather than on these tones in isolation.

Secondly, although non-native listeners lacked the categorical representation of the two different tones, in the discrimination tasks they revealed a pattern similar to native listeners in terms of the order of stimuli presentation. An asymmetrical pattern occurred in the discrimination of T2-T3 pairs regardless of the native language of the listeners; both Mandarin and Dutch listeners always discriminated the steps along the T2-T3 continuum better if a token closer to T3 preceded a token closer to T2. This discrimination asymmetry stayed regardless of whether the continuum was perceived in a categorical way or in a psycho-acoustical manner. This finding suggests that sequential biases in the perception of tone contrasts could be universal, in the sense of being intrinsic to human perception rather than dependent on linguistic experience.

Yet, when comparing the performance of native Mandarin and Dutch listeners, it becomes evident that phonological regularities of Mandarin magnify the difference in discrimination between 'confusing' T2-T3 and 'non-confusing' T3-T2 sequences. In accordance with Mandarin listeners' knowledge of the boundary between T2 and T3, it is found that the closer a stimulus pair is to this boundary, the stronger the sequential difference becomes. This increase in order effect towards the cross-boundary tokens was absent among Dutch listeners.

Thirdly, another difference between native and non-native listeners relates to the strength of the sequential bias versus memory burden in discrimination tasks. It was found that the performance of non-native listeners was severely limited by general memory load, especially for the AXB discrimination. In brief, the more items non-native listeners had to hold in memory for judgment, the higher the error rates became. Only when the memory burden was equal across different conditions, the natural bias of sequential perception in T2-T3 became apparent. In contrast, the tonal processing of native listeners was so driven by linguistic knowledge that the difference in memory load between conditions became irrelevant. Mandarin listeners discriminated tonal triplets better as long as the initial token was T3, regardless of whether the first two tokens or the last two tokens were the same. That is, although the perception of tonal sequences was to some extent universally biased, the Mandarin grammar seems to have phonologized this bias into a sandhi rule of categorical status. Hence, native speech processing is directly dependent on grammatical knowledge.

In the current study, a homogeneous perceptual pattern holding for both native and non-native tone language listeners was found; both groups found it harder to discriminate contrasts containing a T2-T3 sequence compared to a T3-T2 sequence. Here it is important to note that the T2-T3 contrast is the only tonal contrast affected by a phonological sandhi rule in Mandarin. It was suggested above that for native listeners, this sandhi rule may be ultimately grounded in the same asymmetrical perceptual bias that has been evidenced by non-native listeners. Comparing T1-T4 and T2-T3 contrasts, the latter contrast is intrinsically less salient than the former contrast, and it is plausible that the low salience of the T2-T3 contrast makes it more vulnerable to sequential confusion. Moreover, the non-native asymmetrical perceptual pattern leads to the idea that the genesis of the sandhi tone, at least in Mandarin Chinese, may reside in an innate and universal limit of pitch processing. The results of the current paper suggest that, in the domain of speech prosody, humans can be pre-equipped with intrinsic biases that facilitate the emergence of phonological rules in a language-specific manner, which is consistent with the findings regarding segmental grammar (e.g. Ohala 1990, 1993, Berent et al. 2007, 2008). As we learn language through ambient spoken input, we are able to extract, amplify and consolidate regularities that resonate with the innate biases, store them in long-term representations in the mind (a grammar), and use them in the processing of spoken language whenever we encounter tasks where these regularities are involved.

At this point, however, we are not able to draw a firm conclusion about where the bias comes from. Recall that in the AXB identification task, Dutch listeners were asked to identify the X as similar to A, an endpoint of T2, or B, an endpoint of T3. Without any knowledge of Mandarin tones, they tended to identify a larger number of variations as T2 (step 1 to step 5), while T3 tolerated smaller variations (step 6 to step 8 are identified more as T3 than as T2). On the other hand, for the T1-T4 continuum, the boundary fell almost at the middle point. Based on this finding, one speculation is that the complex contour of T3 makes it more specified and more distinctive in perception, while the rising contour of T2 is a more general pattern that tolerates more variation. As a result, in the discrimination task, when listeners were first presented with a distinctive token, it was easy for them to detect a token that deviates from this distinctive one. Conversely, if they were first presented with a token that carried little specific information, it was difficult for the listeners to tell whether another token was different from the first token. To support this speculation, we need to accumulate more information about the strictness of the perceptual space of Mandarin T2 and T3. For solving this issue, a valuable clue might be how infants build up the Mandarin T2 and T3 categories; if T3 forms a stricter perceptual space and is more distinctive, then with the same amount of variations, we would expect infants to build up a representation of T3 more easily. Likewise, in order to gather more convincing evidence concerning the innateness of the discrimination asymmetry of T2T3, it would be necessary to test infants who have no knowledge of any language yet.

In the current chapter, we observed order biases in the discrimination tasks, yet the stimuli pairs consisted of two-step differences, which is substantially more subtle when compared to the underlying T2-T3 contrast. In addition, some of the pairs, such as step 2- 4 or step 5-7, in fact involved steps that belong to the same tonal category according to the identification boundary of Mandarin listeners. If we want to obtain direct evidence about the asymmetry in discriminating Mandarin T2 and T3, we need to test listeners with prototypical T2s and T3s. Moreover, all the stimuli were realized on one single tone-carrying syllable, namely /ma/, and to rule out possible biases introduced by one single tone-carrying syllable, we need to test listeners on multiple syllables carrying Mandarin T2 and T3.

The last remark regarding the aforementioned CP experiments is that the general sensitivity to pitch may have become tangled with the identification and discrimination of Mandarin lexical tones, particularly for non-native Mandarin listeners. Non-native Mandarin listeners, as we have seen, perceive both the T1-T4 contrast and T2-T3 contrast in a psycho-acoustical manner. Hence, it is highly probable that their accuracy in Mandarin tone discrimination is restrained by the general sensitivity in detecting pitch change. If this is the case, then based on the current paradigm we cannot rule out the possibility that the discrimination difficulty observed in the ascending order may be caused by a poor ability to perceive pitch variation in general for some listeners. For those listeners who are sensitive to pitch, they may reach equal accuracy in discrimination regardless of the order in which stimuli are presented. If we are to have a more thorough understanding of the psycho-acoustical constraints on lexical tone perception, we need to discern a measure that is able to indicate the general sensitivity to pitch variations, and then look into the performance of listeners that differ regarding this general sensitivity.

Chapter 3 Universal bias in the perception of Mandarin T2 and T3 from a domain-general perspective

3.1 INTRODUCTION

In chapter 2, we saw that native Mandarin listeners perceived Mandarin lexical tones categorically and Dutch listeners psycho-acoustically. We also saw that native Mandarin listeners more easily discriminated the steps along the T2-T3 continuum if a token closer to T3 preceded a token closer toT2, and not vice versa. Discrimination of steps along the T1-T4 continuum did not demonstrate such an order-induced asymmetry. An explanation for this asymmetry is T3 sandhi in Mandarin: the co-occurrence of two T3s is required to change a T2T3, and as a result, when encountering a T2T3, native listeners may perceive it as a sandhied T3T3, and accept that the two syllables in the sequence carry the same tone. Furthermore, Dutch listeners also demonstrated a similar discrimination asymmetry as their Mandarin counterparts along the T2-T3 continuum, as if they had knowledge of the T3 sandhi rule. However, the Dutch listeners had no familiarity with either Mandarin tones or the T3 sandhi rule, hence their performance must have reflected innate perceptual biases that were not introduced by language experience. Therefore, I argued that T3 sandhi may reside in the innate perceptual biases of human beings, and the phonologization of these biases in return has strengthened the bias in speech perception.

The purpose of the CP experiments was to reveal how the tonal variations are processed by listeners with different language backgrounds. In the CP experiments mentioned in Chapter 2, the discrimination of two-step differences was examined. The acoustical difference in the to-be-discriminated pair is subtle, and the stimuli in each trial could be ambiguous tokens according to the Mandarin tonal boundary. Therefore, although the results of the CP experiments suggested that there might exist universal biases favoring the discrimination between T2 and T3 in a T3T2 order rather than in a T2T3 order, more direct evidence is needed for demonstrating the presence of an innate asymmetry in the T2-T3 discrimination.

Previous studies dealing with T3 sandhi have focused on the perception of the derived T2 in isolation (e.g. Chen 2011, Wang & Li 1963, Peng 1996), without questioning how the sandhi surface form and the underlying form are perceived.

Regarding the relation between possible innate biases and T3 sandhi, I mentioned in Chapter 1 that the domain of T3 sandhi is at least a bisyllabic sequence, and therefore, besides the perception of monosyllabic tones, we need to examine how the bisyllabic underlying T3T3 form and its surface T2T3 form are processed. Though the CP experiments in Chapter 2 revealed that the listeners, regardless of language background, consistently discriminated the T2-T3 contrast better in the bisyllabic T3T2 sequence than in the T2T3 sequence, they did not present direct evidence concerning the perceptual distance between the derived T2T3 and the underlying T3T3 form. In order to gain a direct view about the perceptual knowledge of T3 sandhi, I tested listeners on their discrimination of the underlying T3T3 form and of the T2T3 surface form, comparing it with the discrimination between T3T3 and other bisyllabic sequences that involve T2 and T3. More importantly, to test whether there are biases that favor T3 sandhi to occur as it does now, it is necessary to compare the discrimination of bisyllabic tones that undergo T3 sandhi, namely T2 and T3, and those tones that stay intact from phonological grammar, namely T1 and T4. By comparing listeners' performances when they encounter bisyllabic sequences involving either T2 and T3 or T1 and T4, we may be able to discern T2 and T3-specific discrimination patterns.

The core interest of this dissertation is to find out whether there are any universal auditory biases that may have shaped the T3 sandhi rule. To tease apart the universality of the biases from the possible influences from language experiences, it is necessary to test listeners who are naïve to Mandarin lexical tones as well as native Mandarin listeners. Despite the fact that native Mandarin listeners have knowledge of T3 sandhi while non-native listeners do not have knowledge of such a grammar, Mandarin listeners never encounter a T3T3 sequence realized in its underlying form within one single prosodic word. As a result, a T3T3 sequence in its underlying form could be perceptually salient due to its uncommonness. If this were the case, we would expect Mandarin listeners to be highly and equally accurate when they are asked to discriminate between a common sequence, such as T2T3 or T3T2, and an uncommon T3T3 realized in its underlying form. However, if the native listeners still perceive a specific sequence, hypothetically a T2T3, as perceptually more similar to the underlying T3T3, then it serves as evidence that the confusion asymmetry stays regardless of the frequency distribution in native language input, which is even stronger evidence for a perceptual bias. If similar perceptual patterns are consistently observed among non-native listeners, it serves as strong evidence that there are auditory biases that tend to equalize the T2T3 and T3T3, and that these biases may have shaped the genesis of T3 sandhi. By comparing the performances of native Mandarin listeners to non-native listeners, we will have a clearer picture of the role of universal biases versus that of language input with regard to the perception of a tonological grammar.

In Chapter 2, the listeners participated in monosyllable discrimination tasks in an AX experiment and an AXB experiment, while the compared tokens were not prototypical exemplars of the Mandarin T2 and T3 category. To make the results in the current chapter more comparable with those in Chapter 2, besides carrying out the lexical tone discrimination tasks which involve bisyllabic sequences, I also included a discrimination task which involved the prototypical exemplars of the four Mandarin lexical tones. Another reason to introduce the monosyllabic tone discrimination tasks is that, as mentioned previously, for native Mandarin listeners, discriminating a bisyllabic sequence from the underlying T3T3 form may be influenced by the unequal frequency of the compared pairs. The discrimination of monosyllables, however, only involved natural tones, while the order of presenting the monosyllabic tone may still play a role in the discrimination.

As argued in Chapter 1, there has been extensive evidence showing the correlation between music and lexical tone processing (e.g. Duetsch 2004, 2006, Li & Hung 2008). In addition, numerous studies have found that, for non-native listeners, musicians outperform non-musicians in the perception of lexical tones (e.g. Musacchia, Sams, Skoe, & Kraus 2007, Wang et al. 2007, Deustch et al. 2004, 2006). Looking back at the results of the CP experiments in Chapter 2, the non-native listeners tend to perceive the Mandarin tones in a psycho-acoustic manner. The acoustics for realizing pitch are the same, regardless in speech or in music, namely the F0 variations. Hence, for non-native listeners, it is highly possible that there are domain-general constraints on the perception of pitch. The fact that music perception correlates with perception of lexical tone among non-native adults allows us to use the listeners' performance on a music processing task as an index of general pitch sensitivity, while at the same time giving us a chance to look at the processing of lexical tones by listeners that differ in general pitch sensitivity.

In this dissertation, I seek to find out direct evidence whether there are possible biases that may favor T3 sandhi to occur in its present form, such that a T3T3 is sandhied to T2T3. A discrimination task is the main method to test this hypothesis. A discrimination task targets at listeners' accuracy in detecting pitch differences between two tones (or tonal sequences). Naturally, it is expected that listeners who are sensitive to pitch variations will display a higher accuracy in these discrimination tasks. Moreover, it is likely that, for those sensitive listeners, their discrimination of Mandarin lexical tones is not affected by stimuli presentation order. If this were the case, then the hypothesis of universal biases becomes less convincing, because it is the general sensitivity to pitch that restricts the accuracy in T2-T3 discrimination, and not the order of presenting T2 and T3.

In order to gather an idea of the general pitch sensitivity of the listeners they were, in addition to the lexical tone discrimination tasks, also asked to finish the Montreal Battery of Evaluation of Amusia (MBEA) (Peretz, Champod, & Hyde 2003). The MBEA evaluated their accuracy in detecting a single note change in musical melodies. Taking their score on the MBEA as an index of general sensitivity to pitch variations, I divided the listeners into two groups—a group displaying a general high sensitivity and a group displaying a general low sensitivity to pitch—and I looked at their performance on lexical tone discrimination tasks separately.

For native listeners, the question whether the perception of lexical tones is correlated to the perception of music has remained unanswered. Nan et al. (2010) have found that among native Mandarin music listeners who have deficiency in musical pitch processing, some may also have tone agnosia. However, among the tone language listeners with a normal sensitivity to pitch, it is unknown whether the perception of music pitch still correlates with the perception of lexical tones. Regarding this issue, two possibilities exist. For native Mandarin listeners, the absence of a correlation between perception of musical pitch and perception of Mandarin lexical tones would suggest that native listeners process lexical tones as a linguistic function which is no longer influenced by general auditory constraints. On the other hand, if a correlation is present, then it would mean that, even for native tone language listeners, lexical tones are not perceived in a completely phonemic manner, but immersed in a more general acoustical processing of pitch. Based on the findings in Chapter 2, namely that native Mandarin listeners have established categorical representations of Mandarin lexical tones, which suggests linguistic processing of lexical tones, I tend to believe in the first hypothesis that there will be no correlation between the perception of music and speech. By comparing native Mandarin listeners and native Dutch listeners in terms of the correlation between lexical tones and discrimination of musical melodies, we will better understand how tonal language experience may shape the processing of pitch across different domains.

In order to obtain a direct view of the discrimination between the sandhied T3T3 sequence and other bisyllabic sequences that involve T2 and T3, a bisyllabic discrimination experiment was carried out. Specificity in the processing of T2 and T3 was observed by comparing the bisyllabic discrimination of sequences made up by T2 and T3 and those made up by T1 and T4. With the intention of directly comparing results with the CP experiments in Chapter 2, the listeners also participated in a monosyllabic lexical tone discrimination task. Finally, in order to understand how general sensitivity to pitch may influence accuracy in lexical tone discrimination, listeners were also tested for their accuracy in music perception.

The experiments were organized in such a way that a bisyllabic discrimination task always preceded a monosyllabic discrimination task. Due to the fact that the bisyllabic discrimination task gives direct evidence of the perceptual distance between the sandhied T3T3 sequence and other bisyllabic tonal sequences, the listeners first participated in this task. The reasoning behind this is that if listeners, particularly Dutch listeners, first participate in the monosyllabic tone discrimination tasks, then the manner in which they experience Mandarin tones from the monosyllabic task may shift their perception in the bisyllabic task. Moreover, for the monosyllabic tone discrimination task, if the listeners still discriminate T2 and T3 more easily in one order but not in the other after the experience with the bisyllabic tones discrimination task, then it serves as even stronger evidence for the auditory biases in perceiving the Mandarin T2-T3 contrast. After the tasks involving lexical tones, the listeners were tested with the MBEA as an evaluation of their sensitivity to pitch variations in music.

Considering the interest in innate biases in perceiving Mandarin lexical tones and T3 sandhi, I began with the native Dutch listeners. The Dutch listeners are naïve to Mandarin tones, and their performance cannot be a result of language experience and could serve as a baseline for the natural perception of lexical tones. After discussing the perceptual patterns of the naïve listeners to better understand the role of native language input in shaping the perception of lexical tones, I will proceed to native Mandarin listeners, who participated in the same tasks in the same order.

3.2 EXPERIMENTS WITH DUTCH PARTICIPANTS

Participants

48 native Dutch speakers participated in the experiment. At the time of the experiment, three participants were working, and they had finished at least 15 years of education. The remainder of the participants (aged 19-46 years) were Utrecht University students. Among the 48 participants, 38 had music lessons at some point in their lives (average length of having music lessons: 6.86 years; range of having music lessons: 1.5-19 years). None of the participants reported speech or hearing problems. The participants took part in three experiments: a Mandarin bisyllabic sequence discrimination task, a Mandarin monosyllable discrimination task, and a music perception task. An additional three participants were tested but excluded for the following reasons: one for not finishing all the three tasks, and two for reporting not to have understood the instructions after being tested.

3.2.1 Experiment 1A bisyllabic tonal sequence discrimination

Stimuli

Bisyllabic sequences were generated for a Mandarin lexical tone discrimination task. A female native speaker of Mandarin recorded multiple tokens of monosyllables /ba/, /bou/, /bi/, /da/, /dou/, /di/, /la/, /lou/, /li/, /ma/, /mou/, /mi/, /na/, /nou/, /ni/ with all possible tones in isolation, together with other syllables and sentences. The recording took place in a sound-proof phonetic lab equipped with a DAT Tascam DA-40 recorder and a Sennheiser ME-64 microphone. The sounds were recorded with PRAAT software (Boersma & Weenick 1997). After recording, one token of each syllable with each tone were chosen for further manipulation.

All the chosen syllables were divided into two groups—those involving T1 and T4, and those involving T2 and T3. Within each group, in order to ensure that the other acoustical properties such as duration and intensity were identical across different tones, the original syllables with the same segments were first converted into identical duration by using the “lengthening” function in PRAAT. After manipulation, the duration of the syllables had a range between 380 and 604 milliseconds.

Secondly, to get the stimuli carrying T1 and T4, for each syllable, I extracted the pitch contour of T1, and then replaced the original pitch contour of the T4 syllables with the extracted T1 contour. By manipulating the syllables in this way, I got a T1 syllable which has exactly the same segmental as well as intensity information as the original T4 syllable. To get the stimuli carrying T2 and T3, for each syllable, I extracted the pitch contours of the original T3, and replaced the original pitch contour of the T2 syllable with the extracted T3 contour and hence got a T3 syllable which has exactly the same segmental as well as intensity information as the original T2 syllable.

Thirdly, to generate the bisyllabic sequences for the same syllable, I concatenated T1 and T4 to get a T1T4 sequence, and using the same method, I generated T4T1, T4T4, T1T1, as well as T2T3, T3T2, T3T3, T2T2 sequences. These bisyllabic sequences were used as stimuli. By manipulating the original recordings in this manner, it is ensured that the pitch contour, i.e. the lexical tones, is the only difference between the syllables with the same segments. Therefore, if the listeners were to discriminate between the sequences, the only information that they could rely on was the lexical tone.

Procedure

The participants were asked to discriminate between two sequences as fast as possible by pressing a button box. Within the same trial, the two sequences had the same segmental information. Taking T3T3 sequences as referents, participants were asked to indicate whether or not the T2T3, T3T2, and T2T2 bisyllabic sequences were identical to the referents. Similarly, taking T4T4 sequences as referents, the participants were asked to indicate whether or not T1T4, T4T1, T1T1 were identical to the referents. The referent had an equal chance of being either the first sequence or the last sequence in a test pair. Besides these “different” pairs, in which one of the to-be-compared sequences occurred together with a referent, I also constructed “same” pairs by repeating the afore-mentioned bisyllabic sequences within one trial.

The stimuli were randomized across the participants. The inter-stimulus interval between the two sequences in a trial was 1500ms, and the response duration was one second (i.e. if the participants did not give their response within one second, the next trial began, and they missed one trial).

Results and discussion

Before running the statistical analysis, the answers of the participants were converted into an accuracy rate for each tonal pair. The accuracy rate equals the correct answers divided by the number of total trials. Correct answers refer to “same” responses to the same pairs, and “different” responses to the different pairs. If the participants failed to give a response within the response window, that trial was counted as a missing trial. As the main interest of the dissertation is to find out how listeners discriminate between bisyllabic sequences carring different tones, in the following analysis, I will focus on the accuracy in responding to the different pairs. In order to see whether the listeners were more accurate in discriminating between T3T2 and T3T3 compared to discriminating between T2T3 and T3T3, I submitted the accuracy rate of T3T2-T3T3 pairs and that of T2T3-T3T3 pairs to a paired t test. In the t test, for the trials involving T2T3, those where T3T3 occurred first and those where T3T3 occurred last were collapsed; similarly, for the trials involving T3T2, those where T3T3 occurred first and those where T3T3 occurred last were collapsed. The result showed that the accuracy rate of T3T2-T3T3 pairs was significantly higher than that of the T2T3-T3T3 pairs: $T_{T2T3\text{-}T3T2}(47) = -4.74$, $p<0.001$ (2-tailed). In order to explore whether the position of a referent also affected the judgment accuracy, I submitted the data to a repeated measures ANOVA. Here, the accuracy rate was the dependent variable, and sequences (T2T3 or T3T2 to be discriminated with T3T3) and order of referent (T3T3 occurred first or last) were the within-subject factors. The results showed that both sequences and order have a main effect $F_{sequences}$ (1, 47) = 18.035, $p<0.001$, F_{order} (1, 47) = 53.987, $p<0.001$, and there was no interaction between sequences and order, $F_{sequences*order}$ (1, 47) = 0.228, $p>0.1$. Figure 3.1 plots the accuracy rate of native Dutch listeners when discriminating sequences involving T2 and T3 in each order.

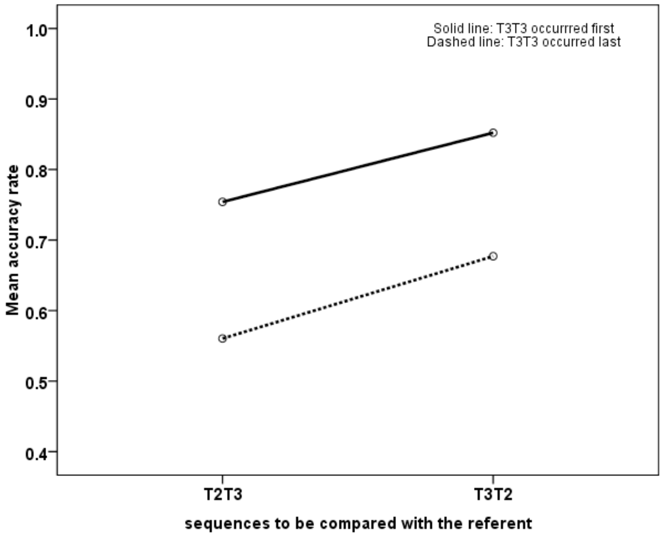


Figure 3.1 Mean accuracy rate of native Dutch listeners when discriminating the sequences involving T2 and T3 in both orders.

As can be seen from figure 3.1, participants discriminated T3T2 from T3T3 more accurately than T2T3 from T3T3. Moreover, their discrimination accuracy was always higher when the referent T3T3 occurred first than if it occurred last, regardless of which sequence was to be compared with the referent. These perception patterns suggest that, first, compared to a T3T2 sequence, a T2T3 is perceptually more similar to a T3T3 sequence; second, a T3T3 sequence seems to be a good referent because it was fairly easy to discriminate T2T3 and T3T2 sequences from it, while discriminating a T3T3 sequence from a T3T2 sequence, and especially from a T2T3 sequence, was much more difficult. The fact that native Dutch participants discriminated more easily between T3T2-T3T3 than between T2T3-T3T3 confirms the hypothesis that a T2T3 is perceptually more similar to a T3T3 sequence than a T3T2 sequence, and hence it is plausible that T3 sandhi occurs as a way to keep the smallest perceptual distance between the sandhied form and the underlying T3T3 form.

A paired t test did not demonstrate a significant difference in accuracy between the discrimination of T1T4-T4T4 and T4T1-T4T4 (with the position of the referent T4T4 collapsed): T (95) = 0.527, p>0.1. A repeated measures ANOVA with sequences (T1T4 or T4T1) and order (referent T4T4 occurred first or last in the trial) as within-subject variables did not result in a significant effect of either sequence or order, but there was a significant interaction between sequence and order, $F_{sequence}$ (1, 47) = 0.32, p>0.01, F_{order} (1, 47) = 0.75, p>0.1, $F_{sequence*order}$ (1, 47) = 14.128, p<0.001. Figure 3.2 plots the accuracy rate of Dutch listeners of each sequence in each order.

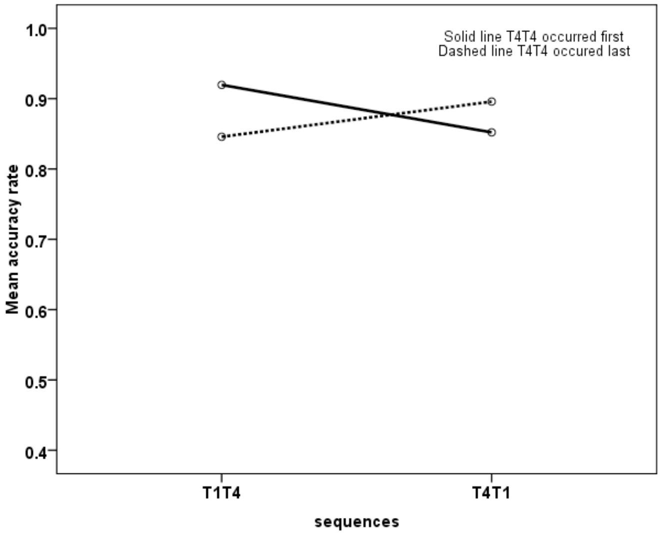


Figure 3.2 The mean accuracy of Dutch listeners when discriminating bisyllabic sequences involving T1 and T4 in both orders.

In the above figure, the participants discriminated the bisyllabic sequences involving T1 and T4 fairly accurately, with an average accuracy rate above 80%. Moreover, contrasting to the discrimination pattern observed in sequences involving T2 and T3, the participants neither exhibited a facilitation effect for specific sequences, nor for a fixed order of stimuli presentation. Comparing the perceptual patterns of the participants regarding sequences involving T2 and T3, and T1 and T4, it is clear that the discrimination asymmetry is specific to the sequences involving T2 and T3, which means that a T3T2 is perceptually more distinguishable from a T3T3 than a T2T3. Additionally, the discrimination between T3T3 and other sequences was facilitated if T3T3 occurred first in the to-be-compared pair. For sequences involving T1 and T4, however, there was no consistent effect regarding either sequences or order.

After collapsing the orders, the accuracy rate was submitted to a repeated measures ANOVA, with tones (T1T4, T4T1, T1T1, T2T3, T3T2, and T2T2) as a factor. Tones was revealed to be a significant factor for the discrimination accuracy $F(4, 44) = 16.2$, $p<0.001$. Post hoc tests indicated that the accuracy of T2T3 sequences was significantly different from all other sequences ($p<0.05$), and the accuracy rate of T3T2 sequences was not significantly different from T1T4 and T4T1 sequences ($p>0.05$). In addition, the accuracy rate of T2T2 sequences was significantly different from both T2T3 and T3T2 sequences ($p<0.05$). Figure 3.3 depicts the accuracy rate of Dutch listeners when discriminating bisyllabic sequences with orders collapsed.

One can see from figure 3.3 that the participants made the most errors when discriminating between T2T3 and T3T3 sequences. More importantly, there was no significant difference in terms of error rate between the discrimination of T3T2-T3T3 pairs and the discrimination of other pairs involving T1 and T4. The fairly high accuracy in discriminating between T3T2 and T3T3 suggests that the difficulty in discriminating T2 and T3 greatly depends on the order of presenting the stimuli. In addition, to discriminate a T2T2 sequence from a T3T3 sequence was fairly easy for the listeners, though structure wise the sequences were similar, in that both were composed of identical tones. Therefore, we could draw the conclusion that T2T3 has the smallest perceptual distance to T3T3.

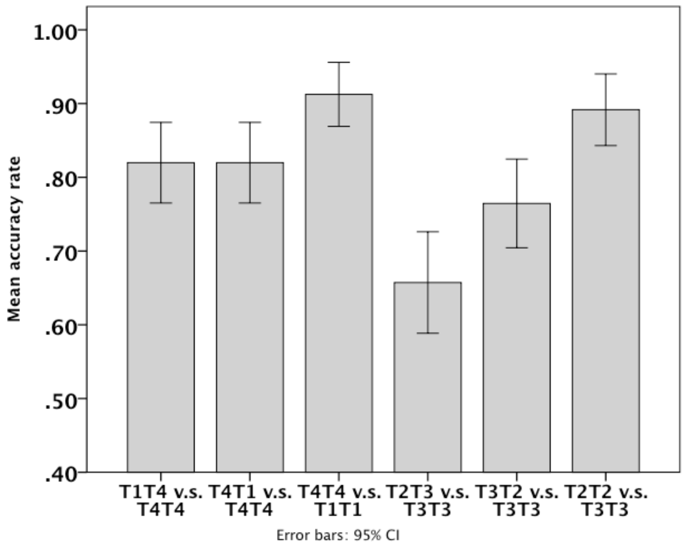


Figure 3.3 Mean accuracy of Dutch listeners when discriminating the bisyllabic sequences with orders collapsed.

To summarize the above, first, for Dutch listeners who are naïve to Mandarin lexical tones, the discrimination between bisyllabic tonal sequences involving T1 and T4 was, overall, easier than the discrimination between bisyllabic tonal sequences involving T2 and T3. Second, a clear discrimination asymmetry was observed for the sequences involving T2 and T3, meaning significantly more errors were observed for the discrimination between T2T3 and T3T3, compared to the discrimination between T3T2 and T3T3. Furthermore, the discrimination asymmetry is limited to the sequences involving T2 and T3. Third, only for sequences involving T2 and T3, the position of the referent, in this case the T3T3, had a significant effect on discrimination accuracy. For the sequence involving T1 and T4, a position effect of the referent was absent.

In Chapter 2, we saw that Dutch listeners listened to the acoustical details of the lexical tones, and perceived the tones in a psycho-acoustical fashion. Therefore, their discrimination pattern reflects how the acoustics of lexical tones are processed without linguistic interference. The results demonstrate that a T2T3 sequence naturally tends to be frequently misperceived as a T3T3 sequence, and perceptually a T2T3 is the most similar sequence to the underlying T3T3 sequence. By comparing the responses to the sequences involving T1 and T4 to those involving T2 and T3, it can be easily seen that T1 and T4 are exempted from discrimination asymmetry. Referring to the fact that T3 sandhi occurs in a way that the consecutive T3T3 sequence is sandhied to a T2T3, it seems highly possible that T3 sandhi neutralizes between the perceptually most similar sequences.

One point that is worth mentioning is that, in this bisyllabic tonal sequence discrimination task, for those sequences involving T2 and T3, it was clearly observed that the discrimination accuracy was significantly higher if a T3T3 occurred first rather than last. Again, this pattern did not apply for the T4T4 referent. Hence, it seems that the referent effect is specific to the T3T3 sequence. At this point, it is difficult to discern why presenting T3T3 first makes discrimination easier. In Chapter 2, the identification task in CP found that along the T2-T3 continuum, both native Mandarin listeners and non-native listeners categorized a token more often as belonging to the T2 category than belonging to the T3 category, and I have argued that this is a hint of a smaller perceptual space of T3 compared to that of a T2. Following this speculation, a possible reason for the referent effect of T3T3 could be that T3 has a complex tonal contour which makes it more specified, and therefore occupies a small perceptual space. Similarly, a combination of two T3s may also occupy a small perceptual space. As a result, it would be easy to discriminate a tonal sequence that falls outside the perceptual space of T3T3. On the contrary, the rising contour of a T2 is more general and tolerates much variation.

One thing that I would like to emphasize here is that the listeners were all Dutch adults without either exposure to or knowledge of Mandarin, and their performance reflects natural ease in processing the tonal acoustics. Also, multiple tone-carrying syllables have ruled out possible biases introduced by segmental information. Therefore, it is highly possible that there are innate auditory biases specific to Mandarin T2 and T3, and that the phonological grammar, namely T3 sandhi, may have occurred based on these innate perceptual biases.

So far, in the aforementioned experiments, we have seen the discrimination of bisyllabic sequences composed by well-realized tones, and we have looked at the two types of tones separately, namely those that undergo the phonological grammar, the T2 and T3, and those that are exempted from the grammar, the T1 and T4. In chapter 2, we have seen the discrimination tasks that involved monosyllables while these monosyllables were not prototypical tokens of either T2 or T3. In the CP experiments, the tokens of the discrimination task were monosyllabic tones generated between two well-realized tones. Hence, one piece of information is still missing for a complete picture of the perception of these two groups of tones: the discrimination of well-realized monosyllabic tones. Testing Dutch listeners on their discrimination of monosyllabic well-realized tones will give us the opportunity to get a direct view of the referent effect of T3, if any. In addition, comparing the discrimination between T2-T3 and between T1-T4 would give more information about the specificity in perceiving the tones that are targeted by the phonological grammar. Therefore, to better understand the natural processing of the monosyllabic tones without the interference from language knowledge, I tested the Dutch listeners on their discrimination of Mandarin monosyllabic tones.

3.2.2 Experiment 2A monosyllabic tone discrimination

Stimuli

The manipulated monosyllables with the four tones were used as test stimuli. Within each trial, two syllables with the same segments, but different tones, were presented to the participants with an inter-stimulus interval of 1500ms. Again, the four tones were divided into two groups: one group went through the tonal grammar, the T2 and T3, and another group that stayed intact from the tonal grammar, T1 and T4. Within a single trial, the two syllables carried either T1 and T4 or T2 and T3, and for each contrast, both orders, such as T1-T4 and T4-T1, occurred with equal chance. Besides these “different” pairs, “same” pairs were constructed by repeating one syllable twice within a trial.

Procedure

The procedure here was the same as in Experiment 1. Participants were asked to indicate as fast as possible whether the two sounds in one trial were the same or different by pressing a button box labeled with “same” and “different”. The order of the stimuli was randomized across the participants. The response window for the participants was 1000ms.

Results and discussion

As was the case in experiment one, the responses of each participant for each tonal pair were converted into an accuracy rate. A response is counted as “correct” if the listener gave a “same” response for a “same” pair and a “different” response for a different pair. If the participants failed to give a response within the response window, that trial was counted as a missing trial. The accuracy rate equals the number of correct responses divided by the total number of trials.

First, the accuracy rate of the pairs involving T2 and T3 and those involving T1 and T4 were compared in a paired t test. The t test showed that there was a significant difference in discrimination accuracy for the T2-T3 contrast and T1-T4 contrast, $T_{T2T3\text{-}T1T4}$ (95) = -7.783, p<0.001 (2 tailed). Second, in order to explore how the order of presentation may have affected the discrimination of these two tonal contrasts, the accuracy rate of each participant was submitted to a repeated measures ANOVA, where contrasts (T1-T4 or T2-T3) and order (T3 occurred first or last, T4 occurred first or last) were within-subject factors. Both contrasts and order had a significant main effect on the accuracy in discrimination, $F_{contrasts}$ (1, 47) = 51.946, p<0.001, F_{order} (1, 47) = 15.067, p<0.001. There was also a significant interaction between contrasts and order, $F_{contrasts*order}$(1,47) = 25.580, p<0.001. Figure 3.4 plots the mean accuracy of each contrast in each order.

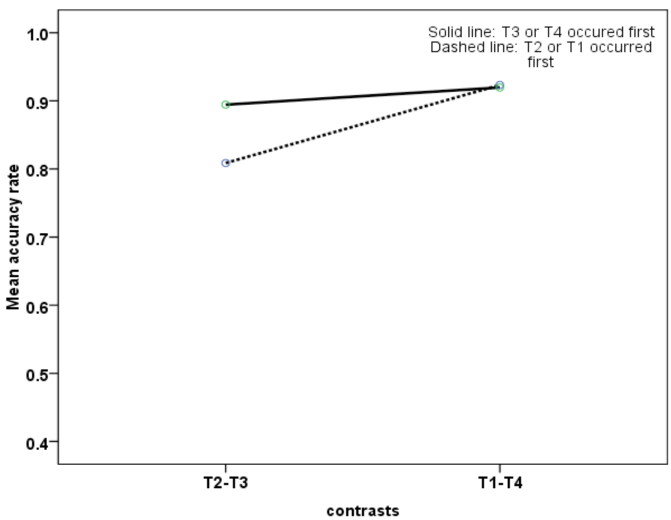


Figure 3.4 Mean response accuracy of native Dutch listeners when discriminating monosyllabic Mandarin lexical tones.

In Figure 3.4, the Dutch participants demonstrated a fairly high accuracy in the monosyllabic tone discrimination task. Yet, the Dutch participants were still more accurate in discriminating the T1-T4 contrast than in discriminating the T2-T3 contrast. Moreover, the main effect of order was again found in the trials involving T2-T3 contrasts: the discrimination accuracy was higher when a T3 preceded a T2 than vice versa. For the T1-T4 contrast, on the other hand, no obvious accuracy difference could be observed for the different orders.

Based on these findings we can see that, as in Experiment 1A, the Dutch listeners showed an evident referent effect for T3, meaning that they were more able to discriminate between a T2 and a T3 if the T3 occurred first in the to-be-discriminated pair. Again, this referent effect was absent for the T1-T4 contrast. T3 sandhi requires a T3T3 to change to a T2T3, so if a listener has knowledge of T3 sandhi, he or she may be tempted to accept the T2 in the T2T3 sequence as a sandhied T3, and hence judge the T2 as the same as the T3. Our Dutch listeners, who obviously had no knowledge of Mandarin, behaved in a way as if they were influenced by T3 sandhi. Moreover, taking together the findings in Experiment 2A and those in Chapter 2, it can be seen that the order effect was specific to T2 and T3, and the effect stayed consistent across prototypical tone discrimination and the subtle tonal variations discriminations. Therefore, the asymmetrical discrimination pattern of T2 and T3 is quite stable and strong, regardless of how the tones are phonetically implemented. Therefore, tone sandhi could have originated from this innate bias which equalizes T2 and T3 in a T2T3 as T3T3. Coming back to the question of why such asymmetry occurs, my speculation is that the complex contour of T3 makes it highly specified perceptually and tolerates little variation. Yet, more evidence is needed to prove this speculation. Infants serve as a suitable population to test the natural perceptual space of the T3 category. If we assume that T3 is more specified and more distinctive than T2 in perception, then with the same amount of variation, the infants are expected to establish a T3 category more easily than a T2 category (the infants experiments will be discussed in Chapter 4). In addition, it is worth the effort of conducting future research to test whether the discrimination asymmetry between a specified category and a less specified category also holds for other sounds.

As argued in the introduction of this chapter, considering that Dutch listeners perceive the Mandarin lexical tones in a psycho-acoustical fashion, it is plausible that their accuracy in tone discrimination is constrained by general sensitivity to pitch. Naturally, we would expect those who are sensitive to pitch in general to be more accurate in perceiving Mandarin tones. Considering the influence of the general pitch sensitivity, one possibility is that those listeners who are generally sensitive to pitch variations are able to pick up the difference between Mandarin T2 and T3 regardless in which order the tones are presented, while only those who are not so sensitive to pitch variations benefit from a fixed order of presenting the tones. If this were the case, then it is arguable that the asymmetry pattern found in the discrimination tasks is a result of a difference in general sensitivity to pitch rather than a universal preference of certain order. The second possibility is that the listeners exhibit the same asymmetrical pattern despite of differences in general sensitivity to pitch. If this is the case, it serves as strong evidence for the universality of the perception asymmetry of the Mandarin T2 and T3.

3.2.3 Experiment 3A Music sensitivity evaluation of native Dutch listeners

In Experiment 3A, the participants took part in a music perception task, and their accuracy in detecting a change of one single note in a music melody was measured. Considering that the core of music processing is pitch perception, the participants' performance on the music task was taken as an index of their general sensitivity to pitch.

After participating in the previous two tasks, the participants finished the Montreal Battery Evaluation of Amusia (MBEA)[8]. Though the focus of the test is mainly to screen people with amusia, which is a deficiency in processing music, it has been validated for evaluating overall musical ability and correlates with the *Musical Aptitude Profile* tests of Gordon (Peretz et al. 2003). The MBEA targets the ability of discriminating and recognizing musical melodies. The battery consists of six parts, referring to scale discrimination, contour discrimination, interval discrimination, rhythm discrimination, meter discrimination, and memory, respectively. There are 30 trials in each part. In each trial, two melodies that are

---

[8] For a complete introduction, please see Peretz et al. 2003. The scores of the MBEA is also available online: http://www.brams.umontreal.ca/plab/research

either the same or only differ in one single note are presented to the listeners, and they are asked to indicate whether the two melodies are the same or different by ticking either “same” or “different” on an answer sheet. The melodies have a range of duration between 3.8 and 6.4 seconds, with a mean of 5.1 seconds. Across the first three parts, the average pitch interval changes are equivalent, with a mean of 4.3, 4.3, and 4.2 semitones respectively.

I first calculated the mean score of the participants on each pitch task and compared them with the norms provided by Peretz online (see Table 4.2 for a complete comparison of the Dutch participants, the Mandarin listeners, and the norms provided by Peretz online). As the first three tasks in the MBEA are meant to test the sensitivity to musical pitch, and considering that pitch is the most relevant feature for lexical tone perception, for later analysis, I will focus on the performance of the participants on the first three tasks. The mean score of all the participants on the three pitch tasks were normally distributed. Among the participants, none had reached 100% accuracy in the three pitch tasks. There was one outlying participant who received a score two standard deviations lower than the mean of all the participants. The scores of this participant on the first three tasks were 16, 19, and 18, respectively. This participant also reported that she had neither interest in nor aptitude for music.

For the ease of comparison, the following scores are listed in Table 2 in section 3.2.3: the scores of each individual on the MBEA task, the mean score of the three pitch perception tasks in the MBEA, and the overall average score on all six MBEA tasks. These scores were listed separately for the native Dutch participants, the native Mandarin participants, and for the norms provided by Peretz online. As can be read from the table, neither Dutch listeners nor Mandarin listeners differed significantly from the norms. In other words, Dutch listeners and Mandarin listeners adhere to a normal distribution of musical sensitivity within a population.

3.2.4 Cross-Experiment analysis

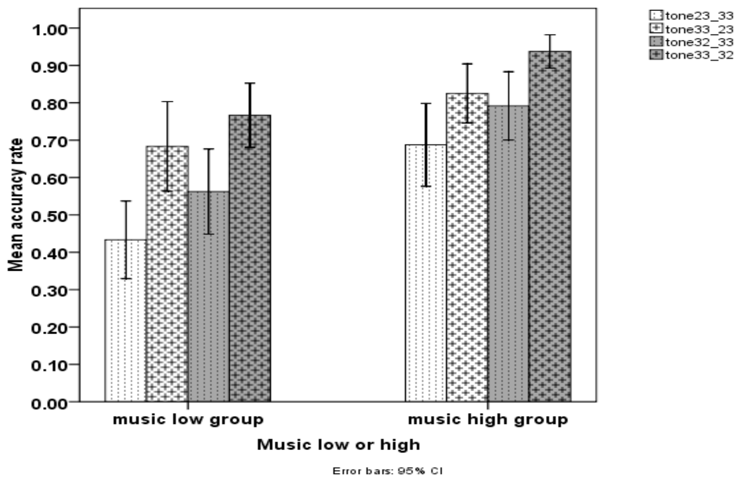
a. Musical ability and the discrimination of bisyllabic sequences

In order to find out how the performance of the musical pitch discrimination predicts the accuracy in lexical tone discrimination, linear regression models were carried out. One model took the average score of the three pitch tasks in the MBEA as the predictor, and the accuracy rate of the bisyllabic tonal sequences discrimination as the dependent variable, and another model took the average score of the three pitch tasks in the MBEA as predictor while taking the accuracy rate of the monosyllabic tone discrimination as the dependent variable. The results showed that the average score of the three pitch tasks is a significant predictor for both the accuracy of bisyllabic tonal sequence discrimination and the accuracy of monosyllabic tone discrimination: $R^2_{\text{music-bisyllabic}} = 0.236$, $B = 0.021$, $SE = 0.006$, $\beta = 0.486$, $p<0.001$; $R^2_{\text{music-monosyllable}} = 0.271$, $B = 0.019$, $SE = 0.005$, $\beta = 0.521$, $p<0.001$. The results showed that there was positive correlation between the average score of the three music pitch tasks in the MBEA and the accuracy of lexical tone discrimination. In other words, the better a participant performed on music pitch tasks, the better he or she discriminated the Mandarin lexical tones. However, even though the score of music pitch tasks is a significant predictor, it is only able to account for less than 30% of the observed variance in the discrimination of Mandarin lexical tones. Hence, it seems that the perception of lexical tones is restricted, but not completely determined, by the sensitivity to musical pitch. Therefore, even though the stimuli of the MBEA are very different from lexical tones, either in terms of timbre or in terms of pitch structure and duration, still a significant positive correlation was found between music discrimination and lexical tone discrimination. The correlation serves as strong evidence for cross-domain perception of pitch in adulthood among non-tone language listeners. Also, it is justified to use the average score on the MEBA pitch tasks as an index of general accuracy in lexical tone perception. It needs to be acknowledged that the MBEA may have measurement errors, and if the music stimuli were constructed in the way that is more similar to lexical tones, stronger predictive power may have been observed.

As mentioned earlier in this chapter, one possibility that may hamper my claim of universal biases in the T2-T3 discrimination is that the order effect is valid only for the listeners who are not so accurate in tracking pitch variations, while for the sensitive listeners the order of stimulus presentation may not affect their accuracy. If this were the case, then the asymmetry that we observed in the previous experiments would be the result of low accuracy in perceiving pitch of some participants, but not of universal biases favoring the discrimination of T2 and T3 in a fixed order. In order to answer the question of whether the discrimination pattern that we have observed in Experiment 1 and Experiment 2 stays regardless of general sensitivity to pitch, I divided the participants into two subgroups according to their average score in the three musical pitch tasks. Taking the median of the average pitch score as the cutting point, the 24 participants who had a score lower or equal to the median form the musical “low” group, and the other 24 participants who had a score higher or equal to the median form the musical “high” group.

Again, for the bisyllabic sequence discrimination task, for the trials involving T2 and T3, a repeated measures ANOVA was carried out. In the test, the accuracy rate of the participants was the dependent variable, and sequences (T2T3 or T3T2) and order of presentation (referent T3T3 occurred first or last) were within-subject factors. Music pitch ability (music low or high) was put into the model as a between-subject factor. The result showed that music pitch ability has a significant main effect on the accuracy of bisyllabic sequences discrimination, F_{music} (1, 46) = 14.642, p<0.001. Sequences and order of presentation again show a significant main effect on the bisyllabic discrimination accuracy $F_{sequences}$ (1, 46) = 17.652, p<0.001, F_{order} (1, 46) = 56.229, p<0.001, while no significant interaction between sequences and order was found, $F_{interaction}$ (1, 46) = 0.226, p>0.1. Importantly, no significant interaction was found between music ability and sequences, between music ability and order, or between sequences, order and music ability: $F_{music*sequence}$ (1, 46) = 0.002, p>0.01; $F_{music*order}$ (1, 46) = 3.021, p>0.05; $F_{music*sequence*order}$ (1, 46) = 0.471, p>0.1.

For the sequences involving T1 and T4, the music sensitivity (low or high) was also introduced in a repeated measures ANOVA as a between-subject variable. The model took the response accuracy of each participant as the dependent variable, and sequences (T1T4 or T4T1 to be compared with the referent T4T4) and order (referent occurred first or last) as the within-subject variables. It was found that music sensitivity had a marginally significant effect on the response accuracy, F_{music} (1, 46) = 3.23, p=0.079, while neither sequences nor order had a significant main effect: $F_{sequence}$ (1, 46) = 0.31, p>0.1, and F_{order} (1, 46) = 0.73, p>0.1. Figure 3.5 illustrates the accuracy rate of each sequence in each order of the music low and high groups separately.



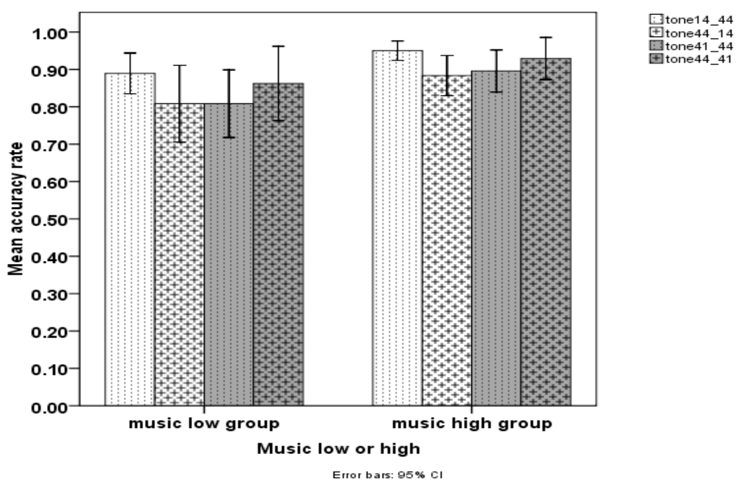


Figure 3.5 Mean accuracy of Dutch music low and high groups in the discrimination of bisyllabic sequences involving T2 and T3 sequences (upper panel) and bisyllabic sequences involving T1 and T4 sequences (lower panel) in both orders.

As seen in figure 3.5, for the bisyllabic sequences discrimination that involved T2 and T3, the music high group outperformed the music low group on each sequence in each order. Nevertheless, both groups exemplified the same discrimination pattern. Both the music low group and the music high group discriminated the bisyllabic sequence pair T3T2-T3T3 more accurately than the bisyllabic sequence pair T2T3-T3T3, and both groups discriminated with increased accuracy if the T3T3 occurred first rather than last in the to-be-discriminated pair. Hence, the discrimination asymmetry for the sequences involving T2 and T3 was consistent among listeners, regardless of their sensitivity to pitch in general. Moreover, after excluding the outlier of the MBEA mentioned earlier, if we look at the accuracy rate of the music low group when discriminating T2T3-T3T3 pairs, in which the T3T3 occurred last, the accuracy rate was not significantly different from chance level (0.5), $t(23) = -1.327$, $p>0.1$. This result suggests that for a substantial population, in this case half of the participants, under certain circumstances the difference between T2T3, which is the sandhied T3T3, and the original T3T3 is unhearable. Bear in mind that the participants in the current experiments were native Dutch speakers with neither exposure to nor knowledge of Mandarin lexical tones which rule out the influence from the frequency distribution in spoken input. Hence, our results serve as strong evidence that, perceptually, the listeners are biased to perceive T2T3 and T3T3 as similar or equal, and T2T3 and T3T3 are perceptually the most similar sequences.

Another point that is worth mentioning is that for naïve listeners without knowledge of lexical tones, the discrimination of sequences involving T2 and T3 is more restricted by the domain-general sensitivity to pitch than the sequences involving T1 and T4. The accuracy difference between the music low and music high group is much more evident for sequences involving T2 and T3 while both groups were fairly accurate in discriminating sequences involving T1 and T4. This pattern suggests that the discrimination of Mandarin tones are not equally restricted by cross-domain sensitivity to pitch; rather, the not-so-sensitive listeners have the most difficulties in discriminating between T2 and T3, while their perception of the salient T1 and T4 is as good as their music-sensitive peers. The different performance of the native Dutch listeners on the discrimination of sequences involving T2 and T3 and those involving T1 and T4 adds more evidence to the claim that the T2-T3 contrast is acoustically less salient than the T1-T4 contrast (e.g. Huang 2001, Johnson & Hume 2003). Therefore, the misperceptions tend to occur between non-salient contrasts, and T3 sandhi neutralizes these easily misperceived tones.

If we collapse the order and submit the discrimination accuracy rate to a repeated measures ANOVA, defining all the bisyllabic sequences (T2T3, T3T2, T1T4, or T4T1) as within-subject factors, and music ability (low or high) as the between-subject factor, music ability still had a significant main effect on discrimination accuracy, $F(1) = 9.773$, $p<0.01$. Moreover, the music ability and sequences show a tendency of a significant interaction, $F_{music\text{-}sequence}(1, 46) = 5.510$, $p=0.06$. Figure 3.6 depicts the discrimination accuracy of the sequences and order of the music low and music high groups separately.

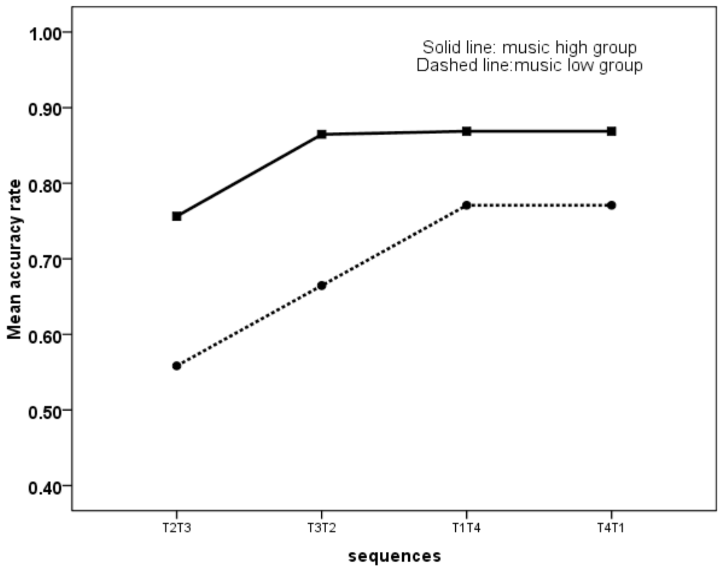


Figure 3.6 Mean accuracy of native Dutch listeners divided into a music low and music high group for the discrimination of bisyllabic sequences with the orders (i.e. the position of the referent) collapsed.

As figure 3.6 depicts, the music high group outperformed the music low group for all the sequences. For the music low group, the discrimination between sequences involving T2 and T3 was always more difficult than the discrimination between sequences involving T1 and T4. However, if we look at the music high group, their accuracy in discriminating a T3T2 from T3T3 did not differ significantly from the accuracy in the discriminating between T1T4-T4T4 and T4T1-T4T4 (post hoc $p>0.05$), which were all higher than 85%. Therefore, for listeners who were sensitive to pitch variation in general, the discrimination between

T3T2 and T3T3 was actually not so problematic. The pattern we have found suggests that the perceptual difficulty for discriminating the Mandarin T2-T3 contrast found in previous studies (e.g. Huang 2001) may be partly due to the order of presentation. In other words, presenting the stimuli in a T3-T2 order might have improved the accuracy of discrimination.

To summarize, for native Dutch listeners overall, the discrimination of bisyllabic tones involving T2 and T3 was more difficult than those involving T1 and T4, and the music ability of the listeners correlated positively with their accuracy in bisyllabic tone discrimination. In other words, the discrimination of Mandarin lexical tones is constrained by domain-general sensitivity to pitch. Next and most importantly, regardless whether the listeners were sensitive to musical pitch variations or not, they all consistently demonstrated the same discrimination asymmetry, namely a T2T3 was less discriminable from a T3T3 than a T3T2. For all listeners, the discrimination was easier if T3T3 occurred first rather than last in the to-be-discriminated pair. For the not-so-sensitive listeners, a T2T3 was indistinguishable from a T3T3 if the two sequences were presented in a T2T3-T3T3 order. Additionally, among all the listeners, this discrimination asymmetry was T2/T3 specific and was absent for the sequences involving T1 and T4.

At this point, we have a clearer picture regarding T3 sandhi: the T2-T3 contrast is less salient compared to the T1-T4 contrast, and the perception of it is more vulnerable. As a result, phonological neutralization occurs between the tones that are most likely to be misperceived. Moreover, the discrimination asymmetry is present among all listeners, and is specific to the T2-T3 contrast; namely, it is hard to discriminate between T2T3 and T3T3, but not between T3T2 and T3T3. Therefore, there are innate biases for T3 sandhi to occur in the way as it does now; the underlying form of T3T3 is neutralized to the most similar perceptual candidate, the T2T3. It is possible that the asymmetry in T2-T3 discrimination may be partly due to the complex contour of T3 that renders it perceptually distinctive, and the distinctiveness of T3 facilitates the discrimination of other tones from itself.

b. Music ability and monosyllable discrimination

In order to explore whether music ability correlates with accuracy in discriminating monosyllabic tones in isolation, and to see whether there is a consistent discrimination pattern regardless of general sensitivity to pitch, again, I took into consideration musical sensitivity in the analysis of the monosyllabic tone discrimination. I ran a repeated measures ANOVA, where accuracy rate in discrimination was the dependent variable, and tones (T2-T3 contrast or T1-T4 contrast) and order (T3 occurred first or last, T4 occurred first or last) were the within-subject variables. Music ability (low or high) was the between-subject variable. The analysis yielded a significant main effect of tones: $F_{tones}(1, 46) = 53.449$, $p<0.001$. There was also a significant main effect of order: $F_{order}(1, 46) = 15.004$, $p<0.001$. Tones and order also showed significant interaction: $F_{tones*order}(1, 46) = 26.374$, $p<0.000$. Music ability also had a main effect on the accuracy of lexical tone discrimination: $F_{music}(1, 46) = 7.817$, $p<0.01$. However, neither tones nor order showed a significant interaction with music ability ($p>0.05$), and there was no significant interaction between tones, order, and music ability ($p>0.05$). Figure 3.7 provides the discrimination accuracy rate of each contrast in each order of both the music low and music high groups separately.

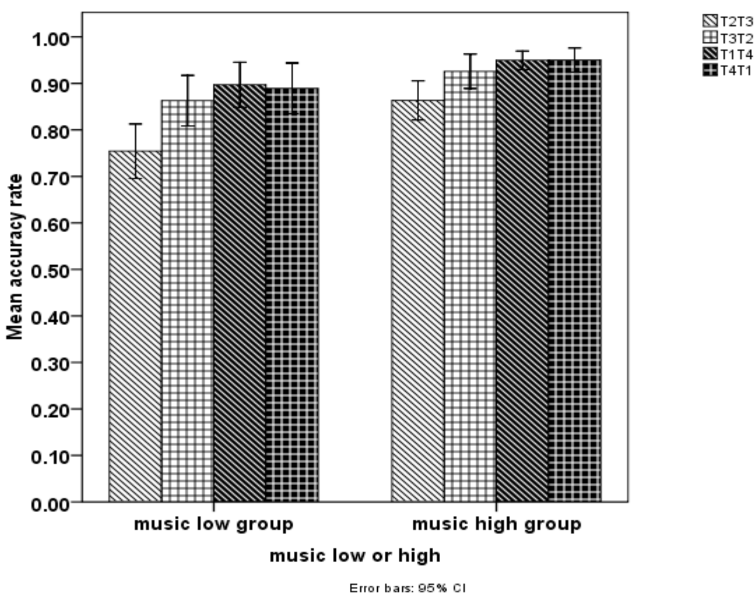


Figure 3.7 Mean accuracy of the discrimination of monosyllabic Mandarin lexical tones in both orders separated by Dutch music low and high groups.

From the figure, we see that the main effect of orders mainly came from the discrimination of the T2-T3 contrast, meaning that both groups discriminated between T2 and T3 better if a T3 preceded a T2, rather than in the reversed order. For the T1 and T4 contrast, however, no clear order facilitation was found. Moreover, similar to what has been found in the bisyllabic sequence discrimination, the music high group outperformed the music low group in all the contrasts in all the orders. Yet, the music low and music high groups showed a consistent discrimination pattern, where discriminating between T1 and T4 was easier than between T2 and T3. Moreover, both music low group and music high group discriminated a T2 from a T3 more accurately than they discriminated a T3 from a T2.

Taking together what we have found so far, it can be seen that for non-tone language listeners, overall, those who were more accurate in perceiving music melodies were also more accurate in lexical tone discrimination. Nevertheless, the fact that discriminating between T2 and T3 is more difficult than discriminating between T1 and T4 holds for both groups in both monosyllabic tone discrimination and bisyllabic tone discrimination. Also, for both groups, the discrimination asymmetry is specific to T2 and T3. The result of the monosyllabic tone discrimination gives more direct evidence of the referent effect of T3, namely to distinguish a tone from T3 when the T3 comes first is fairly easy. Moreover, the tendency to confuse between T2 and T3 in a T2T3 sequence renders the T2T3 the same as T3T3 in terms of the structure, which means that both sequences are composed of identical tones. In this way, the T2T3 and T3T3 become the perceptually most similar sequences both in terms of acoustics and in terms of structure. Hence, it is highly likely that T3 sandhi merges the perceptually most similar sequences.

In order to find out the predictive relationship between monosyllabic tone discrimination, bisyllabic tone discrimination, and music pitch sensitivity, I carried out a linear regression model (backward stepwise). In this linear regression model, the accuracy rate of bisyllabic sequences discrimination was the dependent variable, and the average pitch score in music and the accuracy of monosyllabic tone discrimination were defined as predictors. The results of the models are summarized in Table 3.1.

Step1	B	SE B	β
Constant	-0.021	0.108	
Monosyllabic	0.874	0.12	0.755***
Music pitch score	0.004	0.004	0.373
Step2	B	SE B	β
Constant	0.031	0.091	
Monosyllabic	0.93	0.102	0.803***

R^2 = 0.645 for step 2, ΔR^2 = 0.006 for step 1, ***p<0.001

Table 3.1 Results of the linear regression model taking the response accuracy of the bisyllabic discrimination as the dependent variable, and taking the response accuracy of monosyllabic discrimination and the average score on the three pitch perception tasks in the MBEA as predictors.

As illustrated in table 1, the accuracy rate in monosyllabic tone discrimination was a significant predictor of the accuracy of bisyllabic sequences tone discrimination, and monosyllabic tone discrimination accuracy alone accounted for 64.5% of the variations observed in bisyllabic tonal sequence discrimination. In comparison, the regression model shows that the average score on music pitch tasks of the MBEA failed to be a significant predictor for the accuracy of bisyllabic tonal sequence discrimination, and the average score of music sensitivity accounted for only 0.6% of the variations observed in the discrimination of bisyllabic tonal sequences. Therefore, though the sensitivity to music pitch correlated positively to accuracy in both monosyllabic and bisyllabic lexical tone discrimination, the performance of the bisyllabic lexical tone discrimination was predicted better by the performance of monosyllabic lexical tone perception tasks than by the accuracy in detecting musical pitch change. The two lexical tone discrimination tasks had more in common while both of the lexical tone tasks were very different from the musical tasks, so if the three tasks were more similar, then musical sensitivity might have been a stronger predictor.

If we examine the results of the aforementioned experiments, the following conclusions can be drawn. First, for non-native listeners, accuracy in lexical tone perception correlates significantly with accuracy in detecting differences in musical melody, suggesting that general sensitivity to pitch restricts the perception of lexical tones. A second issue is that the listeners who differed in sensitivity to pitch revealed a significant accuracy difference in discriminating T2-T3, while both groups reached fairly high accuracy rates in discriminating T1-T4. This finding adds more evidence to T2-T3 being perceptually less salient and hence more vulnerable in perception. Third, and most importantly, regardless of the difference in general sensitivity to pitch, all the listeners showed a consistent discrimination asymmetry for the T2-T3 contrast. Those listeners who were generally sensitive to pitch were not spared from the discrimination asymmetry, because they still perceived the difference between T2 and T3 more easily in one direction that the other. Viewing these findings as a whole, I argue that T3 sandhi may originate from perceptual biases. Since the perceptually less salient tones, the T2 and T3, are more vulnerable in perception, they are more likely to be involved in contextual variation, as such variation may not be perceptible. The natural perceptual biases consistently cause confusion between a monosyllabic T2 and T3 in a T2T3 sequence, and confusion between the bisyllabic T2T3 and T3T3. As a result, T2T3, the derived form from T3 sandhi, and the underlying T3T3 are the two sequences that are, in terms of perception, naturally the most similar sequences. To put in another way, T3 sandhi neutralizes the two most similar sequences due to perceptual restriction.

The performance of native Dutch listeners in the aforementioned experiments reflects the innate perceptual bias present when perceiving Mandarin lexical tones psycho-acoustically. Next, a natural question to ask is how the experience with lexical tones would influence such natural biases. It is only for native Mandarin listeners that the lexical tones have a phonemic function, and as a result, as seen in Chapter 2, native Mandarin listeners perceived the lexical tones categorically. It is intriguing to inquire into how linguistic perception of lexical tones may shift the natural discrimination of T2 and T3. Moreover, Mandarin listeners have knowledge of T3 sandhi, which encourages them to neutralize the T2T3 and T3T3. How they would discriminate between T2 and T3 will help us to understand the interaction between natural biases and phonological grammar.

Regarding the stimuli, as the Dutch listeners had no experience with Mandarin, all the test sequences were equally new to them. Native Mandarin listeners, however, had frequently encountered all the test sequences except the T3T3, which is illegal within a prosodic word. As a result, it might be the case that T3T3 is perceptually salient for Mandarin listeners due to its uncommonness. If this were the case, then we would expect Mandarin listeners to discriminate T2T3, T3T2, and T2T2 from T3T3 equally well, as all the pairs consist of one common sequence with another never-heard sequence. However, if we observe the accuracy difference among these pairs, especially if a T2T3 is perceived as most similar to T3T3, then the frequency in ambient language input cannot explain the different patterns observed in discrimination, which is even stronger evidence for universal biases. For the monosyllabic discrimination, based on what we have seen in the CP experiments, Mandarin listeners are expected to make more errors when encountering a T2-T3 pair than their Dutch peers, as the underlying tones of the two syllables in a derived T2T3 are actually the same, namely two T3s.

Moreover, as mentioned in the introduction of this chapter, considering that Mandarin listeners perceive the native lexical tones phonemically in a linguistic way, it is not clear whether the perception of lexical tones still correlates with the perception of music. To compare the lexical tone-music correlation between Mandarin and Dutch listeners is of great importance in helping us to understand whether linguistic experience would make the perception of one type of pitch, namely the lexical tone, independent from general sensitivity to pitch. A secondary goal of testing Mandarin listeners with the MBEA is that, by comparing the performance of Dutch listeners to Mandarin listeners on each individual task, we would have a clue whether or not the experience with lexical tone enhances the accuracy in perceiving musical melody. Therefore, in the following experiments, I tested native Mandarin listeners on exactly the same tasks in exactly the same order as with the native Dutch listeners.

3.3 EXPERIMENTS WITH MANDARIN LISTENERS

Participants

48 native Mandarin listeners participated in the experiments. All the participants were born and raised in China in a Mandarin-speaking environment. At

the time of the experiment, all the participants were either enrolled in a bachelor or a post graduate program at Utrecht University (mean age: 26.96, age range 21-35 years). 28 of them had received musical instruction at some point in their lives (average length of having music instructions: 4.28 years, range of length of music instructions: 0.5-20 years). None of the participants reported hearing or speech problems. Four additional participants were tested, but excluded from analysis after reporting to have misunderstood the instructions after the experiment was over (N=3), and for missing pages of the MBEA (N=1).

3.3.1 Experiment 1B bisyllabic tonal sequence discrimination

Stimuli and procedure

This experiment used the same stimuli and procedure as experiment 1A.

Results and discussion

Before running the statistical analysis, the answers of the participants were converted into an accuracy rate for each tonal pair. The accuracy rate equals the correct answers divided by the number of total trials, in which a correct answer refers to a "same" response for the same pairs and a "different" response for the different pairs. If the participants failed to give a response within the response window, that trial was counted as a missing trial. In order to compare whether it was easier for the native listeners to discriminate between T2T3 and T3T3 than to discriminate between T3T2 and T3T3, I submitted the accuracy rate of T3T2-T3T3 pairs and that of T2T3-T3T3 pairs to a paired t test, regardless of whether T3T3 occurred first or last in the stimuli pair. The result demonstrated that the accuracy rate of discriminating the T3T2-T3T3 pairs was significantly higher than that of the T2T3-T3T3 pairs: $T_{T2T3\text{-}T3T2}(47) = -6.45$, $p<0.001$ (2-tailed). In order to explore whether the position of the T3T3 also affected the accuracy in judgment, the response accuracy of each participant was submitted to a repeated measures ANOVA, taking sequences (T2T3 or T3T2) and order (T3T3 occurred first or last) as within-subject factors. This demonstrated that both sequences and order had a main effect, where $F_{sequences}$ $(1, 47) = 41.65$, $p<0.001$, F_{order} $(1, 47) = 22.49$, $p<0.000$, but there was no interaction between sequences and order, $F_{sequences*order}$ $(1, 47) = 0.67$, $p>0.1$. Figure 3.8 plots the accuracy rate of each sequence in each order.

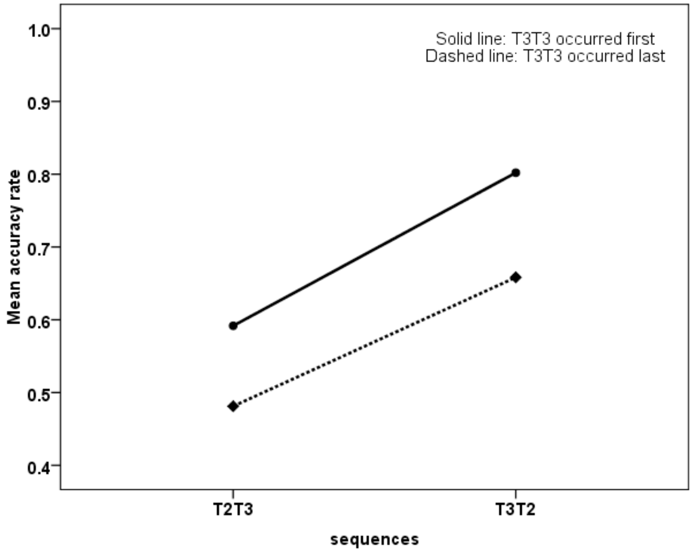


Figure 3.8 The mean accuracy of native Mandarin listeners in discriminating between T3T3-T2T3 and T3T3-T3T2. The solid line represents the accuracy of discrimination when the referent T3T3 occurred first, and the dashed line represents the accuracy of discrimination when the referent T3T3 occurred last.

Figure 3.8 shows that participants always discriminated more accurately between T3T2 and T3T3 than between T2T3 and T3T3. Meanwhile, their discrimination accuracy was always higher when the T3T3 occurred first than when it occurred last, regardless of which sequence was compared with it. First, these perception patterns suggest that, compared to a T3T2 sequence, a T2T3 is perceptually more similar to a T3T3 sequence. Second, as was the case for Dutch listeners, the position of T3T3 had a significant effect on discrimination accuracy. For Mandarin listeners, the facilitation effect obtained by presenting the T3T3 sequence first was somewhat unexpected. If the unfamiliarity with T3 makes it a solid referent, then whether it occurred first or last should not have made a difference. Taking together the discrimination patterns found among native Mandarin listeners and that found among Dutch listeners, it seems that T3T3 served as a good referent in discrimination, no matter whether the listeners had experience with Mandarin tones or not.

The accuracy rate of each participant for the discrimination of the T2T3-T3T3 pair was submitted to a one sample t test, which showed that the mean accuracy rate was not significantly different from chance level: $t\ (47) = 1.48$, $p>0.1$ (2-tailed). As we have seen earlier, Dutch listeners of the music low group also performed at chance level for the discrimination of T2T3-T3T3. Next, in order to obtain a direct view of how language background may affect discrimination, I submitted the accuracy rate of both the Mandarin participants and the Dutch participants to a repeated measure ANOVA. This used sequences (T2T3 or T3T2 to be compared with T3T3) and order of presentation (T3T3 occurred first or last) as within-subject factors, while language background was the between-subject factor. Results show that language background only had a marginally significant effect on discrimination accuracy: $F_{\text{language background}}\ (1,\ 94) = 3.80$, $p=0.054$. Yet, both sequences and order have a significant main effect on the accuracy of discrimination: $F_{\text{sequences}}\ (1,\ 94) = 58.87$, $p<0.001$, $F_{\text{order}}\ (1,\ 94) = 71.98$, $p<0.001$. A significant interaction was observed between language background and sequences: $F_{\text{interaction}}\ (1,\ 94) = 4.85$, $p<0.05$. Importantly, there was no significant interaction between language background and order: $F_{\text{interaction}}\ (1, 94) = 2.44$, $p>0.1$.

If we compare Figure 3.8 and Figure 3.1, we can see that native Mandarin listeners did not outperform Dutch listeners in the discrimination of sequences involving T2 and T3. Moreover, native Mandarin listeners' discrimination between T2T3 and T3T3, no matter in which order the two sequences were presented, was always poorer than the native Dutch listeners. It seems that T3 sandhi in Mandarin strengthens natural perceptual biases. The fact that T3 sandhi neutralizes T2T3 and T3T3 encourages the native listeners to equalize a T2T3 to a T3T3. This explanation is corroborated by the fact that the accuracy difference between T2T3-T3T3 and T3T2-T3T3 among native Mandarin listeners is more substantial than for native Dutch listeners. A noteworthy finding is that both Dutch and Mandarin listeners obtained a higher accuracy if a T3T3 sequence was presented first rather than last. Dutch listeners perceive lexical tones in a psycho-acoustical manner while native Mandarin listeners process the lexical tones categorically, but despite the different ways of perceiving lexical tones, the fact that T3T3 excluded other sequences easily does not change. In other words, the perceptual biases are independent from language experience.

In order to find out whether for native listeners, this discrimination asymmetry is specific to sequences involving T2 and T3, or if it is widely observed among other contrasts, I ran the same statistical analysis as aforementioned with the stimuli that involved the T1 and T4 contrast. First, the Mandarin participants' accuracy in discriminating between T1T4 and T4T4 and the accuracy in discriminating between T4T1 and T4T4, regardless of the position of T4T4, was submitted to a paired t test. No significant difference was found: $t\ (47) = 0.17$, $p>0.01$. Second, a repeated ANOVA with sequences (T1T4 or T4T1) and order (referent T4T4 occurred first or last in the trial) did not yield a significant effect of sequence, $F_{sequence}\ (1, 47) = 0.1$, $p>0.1$. The main effect of order was marginally significant, $F_{order}\ (1, 47) = 3.57$, $p=0.065$. There was no significant interaction between sequence and order: $F_{sequence*order}\ (1, 47) = 0.085$, $p>0.1$. Figure 3.9 plots the accuracy rate of each sequence in each order.

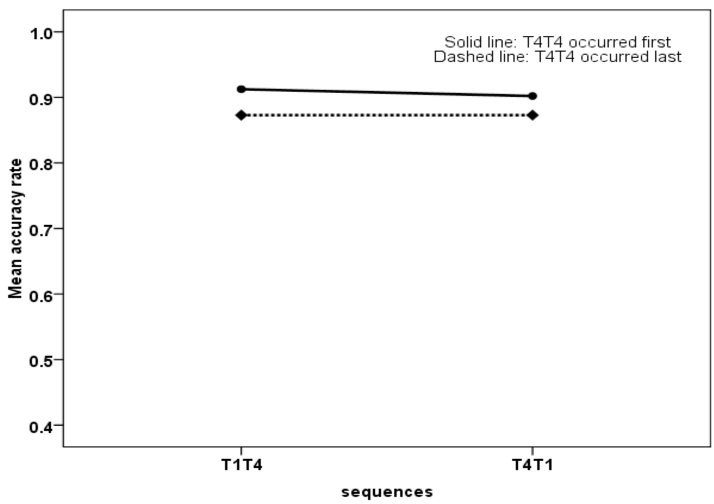


Figure 3.9 The mean accuracy of native Mandarin listeners of their discrimination between T4T4-T1T4 and T4T4-T4T1. The solid line represents the accuracy of discrimination when the T4T4 referent occurred first, and the dashed line represents the accuracy of discrimination when the referent T4T4 occurred last.

From the figure, one can extrapolate that unlike the native Dutch listeners, who did not show a consistent order effect when discriminating between T1T4 and T4T4, Mandarin listeners showed a tendency to discriminate the T4T4-T1T4 pairs more easily compared to the T1T4-T4T4 sequence. This pattern is to some extent similar to our findings in the categorical perception experiments in Chapter 2. The CP experiments demonstrated that when native Mandarin listeners were first presented with the T4, they discriminated between T1 and T4 more easily but not vice versa. Yet, this discrimination asymmetry introduced by order of presentation was quite marginal in both the current experiment and the CP experiments in Chapter 2. This slight facilitation effect in the discrimination when the T4 tokens occurred first might be the result of a downstep in intonation. The pitch naturally goes down in a sentence or in a phrase, and when a T4 follows a T1, it could be that the pitch downshift is perceived as a natural result of energy decrease. Hence, the T4 could be misperceived as a T1.

For the purposes of directly comparing the performance of native Mandarin listeners with native Dutch listeners, I submitted the accuracy rate of each participant to a repeated measures ANOVA, where sequences (T1T4 or T4T1 to be compared to T4T4), and order (referent occurred first or last) were within-subject factors, and language background was the between-subject factor. Results showed that neither sequences nor order had a significant main effect on accuracy: $F_{sequences}$ (1, 94) = 0.37, p>0.1, F_{order} (1, 94) = 0.24, p>0.1. Language background failed to leave a significant main effect for response accuracy too, $F_{language}$ (1, 94) = 0.172, p>0.1. A marginally significant interaction could be observed between order and language, $F_{interaction}$ (1, 94) = 3.86, p=0.053. There was a significant interaction between sequences and order $F_{interaction}$ (1, 94) = 5.08, p<0.05. The three-way interaction between sequences, order, and languages was also significant, F (1, 94) = 7.25, p<0.01. Looking at figure 3.9 and Figure 3.2, we can see that the interaction between sequences and order mainly came from native Dutch listeners. The significant three-way interaction demonstrated that the responses of Dutch listeners and Mandarin listeners were significantly different for different sequences in different orders. The distinct role of order among the Dutch native listeners and the Mandarin native listeners is likely to be driven by different language experiences. It is important to note that for the acoustically more salient T1-T4 contrast, native Dutch listeners performed equally well as native Mandarin listeners, though they lacked the linguistic representation of the tones.

For native Mandarin listeners I also collapsed the orders, and submitted the accuracy rate of each participant of each sequence to a repeated measures ANOVA. Here, sequences (T1T4, T4T1, T1T1, T2T3, T3T2, T2T2) were the within-subject factor, and a significant effect of sequences for the discrimination accuracy was found, $F_{sequences}(5, 43) = 33.97$, $p<0.001$. Figure 3.10 depicts the mean accuracy for each pair by native Mandarin listeners. A Bonferroni post hoc test found a significant difference between the accuracy of the pairs that involved T1 and T4, and those that involved T2 and T3, $p<0.001$. Among the three pairs that involved T2 and T3, the accuracy rate of the T2T3-T3T3 pair was significantly different from the accuracy rate of T3T2-T3T3 and T2T2-T3T3 sequences, both $p<0.001$.

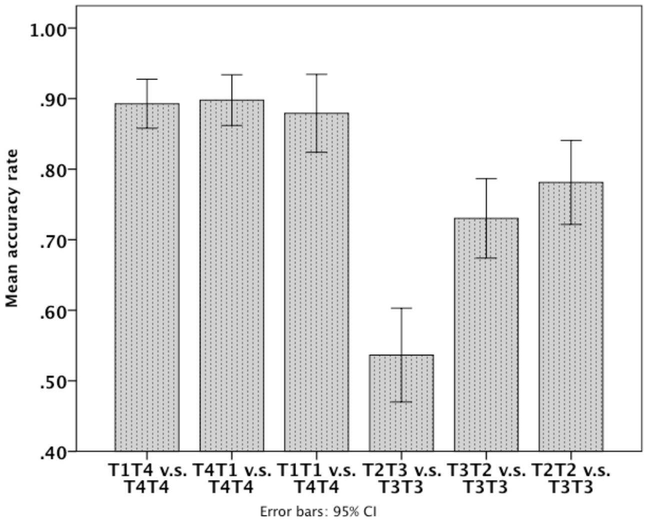


Figure 3.10 Mean response accuracy of native Mandarin listeners in discriminating each pair composed of bisyllabic lexical tone sequences.

As is evident from the figure, native Mandarin listeners discriminated between bisyllabic sequences that involve T1 and T4 with fairly high accuracy. Compared to other pairs, they showed discrimination difficulty if the to-be-discriminated pair involved T2T3 and T3T3. If we look at the discrimination of the pairs that involved T2 and T3, though T2T2 and T3T3 were discriminated more accurately compared to other T2 and T3 pairs, the accuracy rate of the discrimination between T2T2 and T3T3 was still lower than any of the pairs that involved the T1 and T4. Even though T2T2 had the same structure as T3T3 in that both sequences were composed of identical syllables, compared to T2T3 and T3T2, the T2T2 was perceptually less similar to T3T3.

3.3.2 Experiment 2B monosyllabic tone discrimination

Stimuli and procedure

The 48 native Mandarin listeners participated in precisely the same task and procedure as in Experiment 2A.

Results and discussion

Before running the statistical analysis, the answers of the participants were converted into an accuracy rate for each tonal pair. The accuracy rate equals the correct answers divided by the number of total trials, in which a correct answer refers to a "same" response to the same pairs, and a "different" response to the different pairs. If the participants failed to give a response within the response window, that trial was counted as missing. The accuracy rate equals the number of correct responses divided by the total number of trials.

Similar to the previous experiment with the Dutch participants, a paired t test compared the discrimination accuracy rate of the pairs involving T2 and T3 and those involving T1 and T4. The t test showed that there was a significant difference in the discrimination accuracy of the T2-T3 contrast and T1-T4 contrast: $T_{T2T3\text{-}T1T4}$ (47) = -6.747, $p<0.001$. To explore how the order of stimuli presentation may affect the response accuracy, the accuracy rate of each contrast of each participant was submitted to a repeated measures ANOVA. Contrasts (T1-T4 or T2-T3) and order (referent T3 occurred first or last, and referent T4 occurred first or last) were within-subject factors. Results illustrated that both contrasts and order had significant main effect on the response accuracy: $F_{contrasts}$ (1, 47) = 65. 41, $p<0.001$, F_{order} (1, 47) = 85.02, $p<0.001$. The results also indicated that there was a significant interaction between contrasts and order, $F_{interaction}$ (1, 47) = 64.79, $p<0.001$. Figure 3.11 depicts the mean accuracy of the Mandarin listeners when discriminating the T2-T3 and T1-T4 contrasts.

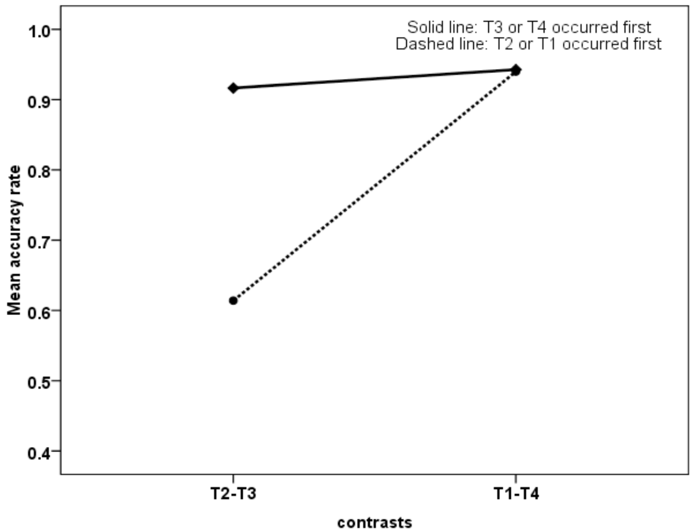


Figure 3.11 Mean accuracy rate of the Mandarin listeners for the discrimination of T2-T3 and T1-T4 contrasts realized on monosyllables. The solid line represents the pairs in which T3 or T4 occurred first, and the dashed line represents the pairs in which T2 or T1 occurred first.

Figure 3.11 demonstrates that native Mandarin listeners reached almost 100% accuracy for the discrimination between T1 and T4, no matter in which order the tones were presented. On the other hand, their discrimination of T2 and T3 was heavily influenced by the order of presentation: they discriminated the T2-T3 contrast much more accurately if the stimuli were presented in a T3-T2 order than in a T2-T3 order. This discrimination asymmetry was expected considering T3 sandhi, which changes a T3T3 to a T2T3. As a result, when encountering a T2-T3 pair, the listeners may have perceived it as a sandhied T3T3, hence considering the two tones in the pair as identical.

If we compare Figure 3.3 and Figure 3.11, it is apparent that native Mandarin listeners and native Dutch listeners shared the same perceptual patterns in the monosyllabic discrimination in several ways. First, both groups reached a higher accuracy when discriminating Mandarin T1-T4 contrasts than when discriminating Mandarin T2-T3 contrasts. Second, the order of stimuli presentation only had an effect on the discrimination of the Mandarin T2-T3 contrast. Third, it was always easier to discriminate the T2-T3 contrast if a T3 preceded a T2 rather than vice versa. Yet, compared to native Dutch listeners, native Mandarin listeners showed a stronger order effect for the discrimination of the T2-T3 contrast.

As discussed earlier, the fact that Mandarin listeners discriminated a T2-T3 pair significantly better than a T3-T2 pair was expected. The Dutch listeners also showed a similar discrimination asymmetry, not differing from Mandarin listeners in terms of having or lacking the asymmetry. However, the Dutch and Mandarin listeners did differ in terms of the magnitude of the asymmetry. Regardless of whether the lexical tones are perceived psycho-acoustically (Dutch listeners) or categorically (Mandarin listeners), the tendency to confuse a T2 and T3 in a T2T3 order as well as between T2T3 and T3T3 stays present. Therefore, T3 sandhi in Mandarin may be generated from universal biases that are independent of specific languages. T3 sandhi in Mandarin has consolidated this bias into a grammar, and a T3T3 is required to be neutralized into a T2T3. As a result, the knowledge of the T3 sandhi rule, in return, has amplified the natural biases for Mandarin listeners.

3.3.3 Experiment 3C music sensitivity evaluation of native Mandarin listeners.

The 48 native Mandarin listeners, just as their Dutch counterparts, participated in the MBEA for the evaluation of their sensitivity to music. The average score of all the tasks of each participant and the average score of the three pitch tasks of all the participants was submitted to a K-S test. The results from the K-S test showed that both of the average scores did not significantly deviate from a normal distribution: $D_{\text{all tasks}}(48) = 0.12$, $p>0.05$, $D_{\text{all pitch tasks}}(48) = 0.12$, $p>0.05$. No outlier was detected for both average scores. The average score of native Mandarin listeners and native Dutch listeners of each individual task, the overall mean score of all the tasks, and the mean score of the three pitch tasks are listed in Table 3.2

Tasks	Dutch listeners (N=48)	Mandarin listeners (N=48)	Norms (N=285)
Task 1 scale difference discrimination	25.12 (2.95)	26.21 (2.87)	26 (2.63)
Task 2 different pitch contour discrimination	25.75 (2.73)	26.62 (2.34)	26 (2.64)
Task 3 same pitch contour, different absolute pitch discrimination	24.98 (3.21)	25.48 (3.51)	26 (2.80)
Task 4 rhythm discrimination	25.85 (3.09)	26.98 (3.47)	27 (2.60)
Task 5 meters discrimination	26.83 (3.21)	27.27 (3.74)	26 (4.12)
Task 6 short term memory of the melodies	28.54 (1.85)	28.38 (1.50)	27 (2.43)
Overall score all tasks	26.22 (2.09)	26.78 (1.94)	26 (1.88)
Overall score all pitch tasks	25.28 (2.52)	26.10 (2.51)	26 (2.00)

Table 3.2 The average score of each individual task, of all six tasks, and of three pitch discrimination tasks of native Mandarin listeners and Dutch listeners and the norms provided by Peretz online.

All the above-listed scores were submitted to an independent sample t test, dividing the participants into two groups according to their language background (native Mandarin or native Dutch). No significant difference was found between the two groups of listeners for any of the tasks: T_{scale} (94) = 1.82, p=0.072, $T_{different\ contour}$ (94) = 1.69, p-=0.095, $T_{same\ contour}$ (94) = 0.73, p>0.1, T_{rhythm} (94) = 1.68, p=0.097, T_{meters} (94) = -0.615, p>0.1, T_{memory} (94) = 0.49, p>0.1. The results demonstrated that the Mandarin listeners and the Dutch listeners did not significantly differ in terms of accuracy in detecting change of musical notes. If we compare the scores of our participants to those provided by Peretz, there was also no significant difference. Therefore, both Dutch listeners and Mandarin listeners conform to the normal distribution in processing musical melodies. Importantly, our Mandarin listeners did not show any advantage in the MBEA tasks. In other words, the accuracy in music pitch processing was not enhanced by familiarity with lexical tones. It is possible that the MBEA may have measurement errors, and if the two groups of listeners are tested on a more demanding music task, maybe a language effect would surface. Hence, this may be an interesting topic for future linguistic research.

3.3.4 Cross-experiment analysis of Mandarin listeners

a. music ability and the discrimination of bisyllabic tones

In order to find out whether the performance on the music tasks correlated with accuracy in native lexical tones discrimination for native listeners, linear regression models were carried out. The predictor was the average score of the three pitch tasks of the MBEA, and one model took the accuracy rate of the bisyllabic tonal sequences discrimination as the dependent variable while another model took the monosyllabic tone discrimination as the dependent variable. The average score of the pitch tasks in the MBEA failed to be a significant predictor for either the accuracy of bisyllabic tonal sequence discrimination, or for the accuracy of monosyllabic tone discrimination: $R^2_{\text{music-bisyllabic}} = 0.01$, $B = -0.01$, $SE = 0.006$, $\beta = -0.106$, $p>0.1$; $R^2_{\text{music-monosyllabic}} = 0.003$, $B = 0.001$, $SE = 0.003$, $\beta = 0.053$, $p>0.1$. Unlike native Dutch listeners, for native Mandarin listeners, there was no significant correlation between the performance on music pitch tasks and the performance on lexical tone discrimination tasks. Moreover, Mandarin listeners' performance on the music pitch perception tasks did not predict their performance on the lexical tone discrimination tasks. Among a normal population of native Mandarin listeners, the lack of correlation between music pitch perception and lexical tone perception is expected. Bear in mind that the lexical tones function phonemically, as we have seen in Chapter 2, and native tone language listeners need to neglect subtle pitch variations to ensure efficient processing. Hence, though both lexical tones and music are mainly realized by pitch variations, native listeners tend to perceive the tonal variations in a linguistic way. The linguistic processing of lexical tones makes the tone perception of native Mandarin listeners independent from general sensitivity to pitch processing.

Though performance on music pitch perception failed to predict the performance on lexical tone discrimination, it could still be that listeners who were more adept at catching music pitch difference were also better in perceiving lexical tones, and thus would not be thrown off by the order of stimuli presentation. Therefore, as was done with native Dutch listeners, I divided the participants into two subgroups according to their average score on the three musical pitch tasks. Taking the median of the average pitch score as the cutting point, the 24 participants who had a score lower or equal to the median formed a musical "low" group, and the remaining 24 participants, who had a score higher or equal to the median, formed the musical "high" group. As I have failed to find a significant correlation between the accuracy in musical pitch perception and the accuracy in lexical tones discrimination, I did not expect that the music high and music low group would differ significantly.

For the discrimination of bisyllabic sequences that involve T2 and T3, a repeated measures ANOVA was conducted. The dependent variable was the accuracy of each participant, and the within-subject factors were the sequences (T2T3 or T3T3 to be compared with the T3T3) and the order (T3T3 occurred first or last). The music pitch sensitivity (high or low) was the between-subject factor. There was a significant main effect of sequences and order for the response accuracy of the native Mandarin listeners: $F_{sequences}$ (1, 46) = 42.70, $p<0.001$, F_{order} (1, 46) = 22.04, $p<0.001$. However, music sensitivity failed to have a significant effect on the response accuracy, and it did not significantly interact with either sequences or order: F_{music} (1, 46) = 0.00, p=1, $F_{music*sequences}$ (1, 46) = 2.18, $p>0.1$, $F_{music*order}$ (1, 46) = 0.05, $p>0.1$, $F_{music*sequences*order}$ (1, 46) = 0.04, $p>0.1$.

I also ran a repeated measures ANOVA with the responses of the sequences involving T1 and T4, where the response accuracy of each participant was the dependent variable, and sequences (T1T4 or T4T1 to be compared with the referent T4T4) and order (referent occurred first or last) were the within-subject variables. Music sensitivity (low or high) was the between-subject variable. I found that none of the variables showed a significant main effect: $F_{sequences}$ (1, 46) = 0.09, $p>0.1$; F_{order} (1, 46) = 3.52, $p>0.05$, F_{music} (1, 46) = 2.22, $p>0.1$. There was no significant interaction between any of the factors, all p>0.1. Figure 3.12 provides the mean accuracy of both the Mandarin music low and Mandarin music high group for the sequences involving T2 and T3 in both orders, and those involving T1 and T4 in both orders, respectively.

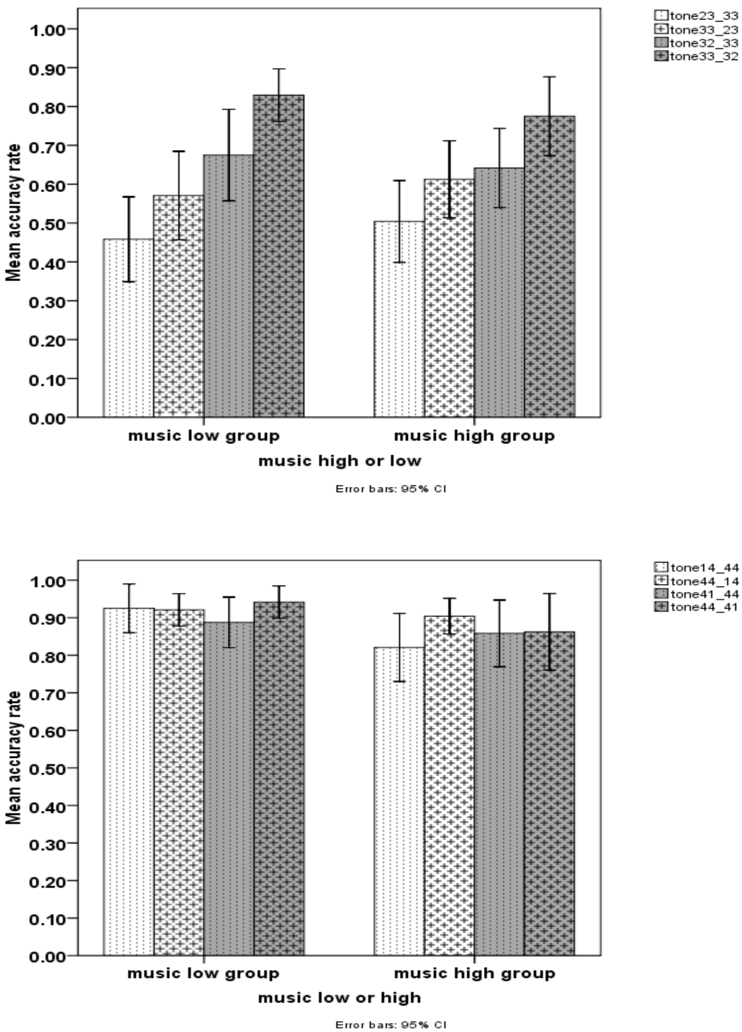


Figure 3.12 The mean response accuracy of native Mandarin listeners for the discrimination of sequences involving T2 and T3 (upper panel) and sequences involving T1 and T4 (lower panel).

After collapsing the orders, the stimuli were divided into four groups according to the bisyllabic sequence that was compared to the two referents, namely T2T3, T3T2, T1T4, T4T1. A repeated measures ANOVA was run. The dependent variable was the response accuracy of each participant, and sequences (T2T3, T3T2, T1T4, T4T1) were the within-subject variable. Music sensitivity (low or high) was used as the between-subject variable. Music sensitivity failed to leave a significant effect on the response accuracy: F_{music} (1, 46) = 0.374, p>0.1. Figure 3.13 depicts the mean accuracy of the music low and music high group of native Mandarin listeners of each bisyllabic sequence.

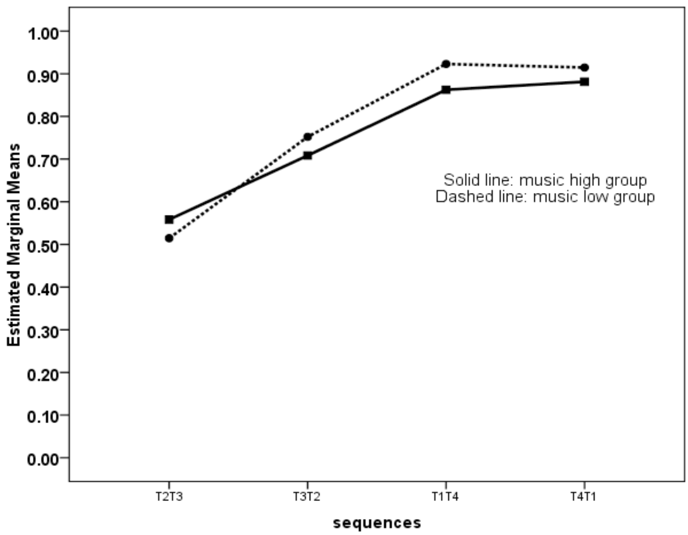


Figure 3.13 Mean response accuracy of music low and music high native Mandarin listeners when discriminating the bisyllabic sequences from the referent.

Examining the results from the statistical analysis together with figure 3.13, it can be seen that native Mandarin listeners, regardless of whether they were sensitive to musical melodies, discriminated Mandarin lexical tones with equal accuracy. For both groups, again, discriminating between bisyllabic sequences involving T2 and T3 was more difficult than discriminating between bisyllabic sequences involving T1 and T4. Moreover, the order-induced discrimination difference was only observed for the sequences involving T2 and T3. Hence, for native listeners, though their acquaintance with lexical tones makes their perception of lexical tones independent from general constraints of pitch perception, the discrimination asymmetry between T2 and T3 persisted and is T2/T3 specific.

In Chapter 2, Dutch listeners listened to the acoustics of lexical tones, while native listeners have been diverted from the tonal acoustics but processed the tonal variations phonemically. In this chapter, we saw that for non-native listeners, the acoustic processing of the lexical tones was restrained by general sensitivity to pitch, while for native listeners, the categorical perception of lexical tones became independent from general constraints. Nevertheless, no matter in what manner the listeners perceived the tones, linguistically or acoustically, the asymmetry in the discrimination between T2 and T3 stayed and remained to be T2/T3 specific. Moreover, for the Dutch native listeners, the discrimination of monosyllabic or bisyllabic sequences involving T2 and T3 was still impaired in one specific direction even when the listeners were generally sensitive to pitch. The performance of Dutch listeners strongly suggests that there are innate biases that favor T2T3 being perceptually neutralized to T3T3. Therefore, T3 sandhi may have originated from these innate biases. For native listeners, on top of the categorization of lexical tones, which is the result of the phonemic function of lexical tones, the grammar of T3 sandhi in the input phonologizes the bias and the neutralization between T3T3 and T2T3 is obligatory. The categorical status of the grammar, in return, strengthened the natural discrimination asymmetry.

b. Music ability and monosyllable discrimination among native Mandarin listeners

For the monosyllable discrimination, I also introduced the sensitivity (low or high) as a between-subject variable to the original repeated measures ANOVA. In the model, tones (the T2-T3 contrast or the T1-T4 contrast) and order (T3 occurred first or last, T4 occurred first or last) were within-subject variables. Again, I found that music sensitivity failed to have a significant impact on the response accuracy, F_{music} (1, 46) = 2.65, $p>0.1$. Both tones and order had a significant main effect: F_{tones} (1, 46) = 64.10, $p<0.001$, F_{order} (1, 46) = 83.21, $p<0.001$. In addition, tones and order also interacted significantly: $F_{tones*order}$ (1, 46) = 63.64, $p<0.001$. With the exception of tones and order, all the interactions failed to reach a significance level (all $p>0.1$). Figure 3.14 depicts the mean accuracy of Mandarin music low group and music high group when discriminating between the T2-T3 and T1-T4 contrast in both orders.

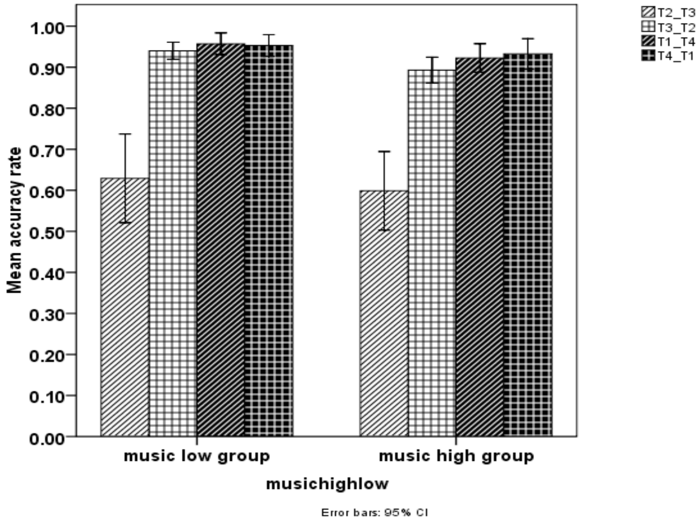


Figure 3.14 Mean accuracy of Mandarin music low and high group when discriminating between T2-T3 and T1-T4 contrasts in both orders.

In the above figure, the native Mandarin listeners discriminated the monosyllabic tones with almost equal accuracy regardless of whether their music score was low or high, and for both the music low group and the music high group, perceptual difficulty was observed only for the discrimination of the T2-T3 pair. As discussed earlier, this bias was probably due to the knowledge of T3 sandhi.

3.4 GENERAL DISCUSSION

In these experiments, I tried to shed light on the possible auditory biases that cause T3 sandhi to occur in the way as it does now, and how this innate bias, if any, interacted with general constraints on pitch perception and language experience. The results found in the experiments can be discussed from several perspectives. First, a significant correlation was found between musical melody discrimination and lexical tone discrimination for non-native listeners, while for native listeners such a correlation was absent. This result corroborated the patterns found in Chapter 2. In the CP experiments in Chapter 2, we saw that non-native listeners perceived lexical tones psycho-acoustically. It is likely that this psycho-acoustical manner in the processing of lexical tones that made the lexical tone perception correlate with music perception. Native listeners, on the other hand, perceived the Mandarin tones in a categorical manner, and the linguistic processing of the lexical tones made it independent from the perception of pitch acoustics. Second, Dutch listeners did not perform significantly worse than their Mandarin peers in discriminating monosyllabic T1 and T4 or bisyllabic sequences that involved T1 and T4. On the other hand, they were less accurate than Mandarin listeners in discriminating between T2 and T3. This finding adds more evidence to the claim that T2 and T3 are the most similar tonal contrast in Mandarin. Moreover, it also suggests that Mandarin input and knowledge is important in improving the perception of less salient tones. Third and most crucially, regardless of how the listeners perceived the lexical tones, some patterns were specific to the T2 and T3 perception. To be more specific, T2 and T3 were discriminated better if a T3 preceded a T2 rather than vice versa. Regarding the bisyllabic tonal sequences, T2T3 was most frequently confused with T3T3. Moreover, the discrimination was always more accurate if T3T3 preceded T2T3. These patterns held for the non-native listeners regardless of the general sensitivity to pitch. Last, the aforementioned asymmetry was specific to the discrimination between T2 and T3, since such an asymmetry was absent for the T1-T4 contrast.

Based on these findings, we can safely draw the conclusion that there are innate perceptual biases that favor the occurrence of T3 sandhi. To start with, the sandhi rule targets at the perceptually most vulnerable contrast, and as there are innate biases that tend to confuse the T2T3 and T3T3, the sandhi rule neutralizes these tonal contrasts. The phonologization of this bias consolidates the innate bias into a categorical grammar, namely a T3T3 has to obligatorily change to a T2T3. As a result, the knowledge of such a grammar enlarged the discrimination asymmetry among native listeners. That a T3T3 is not sandhied to other sequences, as we have seen in this chapter, is probably due to the fact that other alternations are easily noticeable. Perceptually, T3T3 is easily confusable with T2T3, and precisely the fact that T2T3 already occurs in the language is the factor that undermines T3T3, and ultimately triggers sandhi.

Up till now, there are still several questions that remain unanswered. The first is why such a confusion asymmetry occurs between T2 and T3. As I have mentioned earlier, it could be that the complex contour of the T3 makes it perceptually specified, and hence more distinctive from other tones. The distinctiveness of the T3 only allows a strict perceptual space. As a result, when listeners first encountered a distinctive tone, it is fairly easy to tell whether another tone is the same. On the contrary, the rising contour of the T2 is quite general and tolerates more variation. Hence, when listeners are first presented with a rising tone, they tend to accept that a T3 is a different phonetic implementation of a rising contour. The speculation that the discrimination asymmetry occurs between a distinctive sound and a more general sound still needs more cross-linguistic evidence. Second, regarding how general pitch sensitivity restrains lexical tone perception, by now it seems plausible that tone language listeners and non-tone language listeners listen to the lexical tones in different ways, and the psycho-acoustical processing of the latter group makes the perception restricted due to general acoustical sensitivity. However, I am not able to affirm that the perception of lexical tones by native listeners remains completely independent from their general pitch sensitivity or from the accuracy in music processing. More perception experiments using more demanding tasks need to be carried out in order to have a more accurate picture of the interaction between the linguistic processing of lexical tones and the processing of pitch realized in other domains. Moreover, it is not known whether the pitch perception in lexical tones and the pitch perception in music initially correlate and separate later in life due to the input of a tonal language, or if they do not initially correlate in the first place. In this case, it might be that non-tone language listeners do not encounter the phonemic function of lexical tones, and they perceive the lexical tones in a psycho-acoustical manner in adulthood, and hence show the correlation between music and speech. This issue is of crucial importance in that it will allow us to have a clearer picture of how human beings are "equipped" initially to perceive prosody, and will help us to better understand how language shapes our mind.

In the following chapters, I will take up the topic of infant studies. In chapter 4, I will report on the experiments that deal with the discrimination of Mandarin T2 and T3 among Dutch infants across different ages, which will serve as a basis for the following experiments which examine the possible bias from an acquisitional perspective as well as the early (un)correlation between the perception of lexical tones and the perception of music melody.

Chapter 4 Mandarin lexical tone perception by non-native infants

4.1 INTRODUCTION

Thus far, we have examined the perception of Mandarin lexical tones by both native and non-native adults. The findings could be summarized from the following perspectives. First, as shown by the CP experiments in Chapter 2, due to the phonemic function of the lexical tones, native listeners perceive the lexical tones categorically. Non-native listeners, on the other hand, tend to process the acoustics of the tones and perceive them psycho-acoustically. Second, for both Mandarin listeners and Dutch listeners, the discrimination accuracy between T2 and T3 was always lower than that of T1 and T4, both in the monosyllabic discrimination tasks and in the bisyllabic discrimination tasks. For the T1-T4 contrast, the discrimination accuracy of the non-native adult listeners was not significantly lower than that of the Mandarin listeners. Third, and most importantly, regardless of whether the listeners perceived Mandarin lexical tones linguistically or psycho-acoustically, a discrimination asymmetry was consistently observed for T2 and T3. Both Mandarin listeners and Dutch listeners discriminated the monosyllabic T2 and T3 more accurately if a T3 preceded a T2 than if a T2 preceded a T3, and both groups made most errors when discriminating between T3T3 and T2T3 sequences. On the other hand, Mandarin and Dutch listeners were fairly successful in discriminating between T3T3 and T3T2, and between T3T3 and T2T2. Moreover, It is also found that in the bisyllabic sequence discrimination task, both groups showed benefits if T3T3 occurred first rather than last in the to-be-compared pair. Fourth, in the case of the non-native listeners, lexical tone discrimination correlated positively and significantly with the discrimination of musical melodies. It seems that the psycho-acoustical perception of lexical tones of the non-native listeners made their processing of lexical tones restrained by general sensitivity to pitch, while the phonemic function of lexical tones has caused native listeners to perceive them linguistically.

The perception of the Dutch listeners, who were naïve to Mandarin lexical tones, has demonstrated that there are innate biases that favor the occurrence of T3 sandhi: the sandhi process targets the perceptually not-so-salient contrasts, and it neutralizes the two bisyllabic sequences that are most similar and sometimes indistinguishable. Moreover, it seems that the discrimination is facilitated if T3 occurs first in the discriminated pair. However, as I have mentioned in Chapter 2 and Chapter 3, one question remains open: why does this asymmetry occur? Considering that T3 is a complex tone, embodying features of falling, dipping, and rising at the same time, I have assumed that the complex contour of the T3 may make it more specified and hence more distinctive in perception. In comparison, the rising contour of T2 is less specified in perception and tolerates more variation. As a result, first encountering a distinctive tone may help the listeners to discriminate another tone from it. However, if a more general tone is presented first, it is more difficult to tell whether another tone is the same. The result of the identification task in CP using the T2-T3 continuum has shown that, when presented with a continuum changing between T2 and T3, Dutch listeners without any knowledge of lexical tones tended to categorize more tokens into the T2 category than into the T3 category. This pattern hinted that T3 may naturally occupy a smaller perceptual space than T2.

How infants build up phonological categories is very revealing for understanding the innate biases in perceiving speech sounds. It has been largely agreed that the infants' speech acquisition pattern reflects a natural ease in perceiving and categorizing speech sounds (Jakobson 1941, Stampe 1969, Trubetzkoy 1931, 1939). An advantage of looking at the performance of infants, especially those who have not finished perceptual reorganization (Werker & Tees 1984, Polka & Werker 1994), is that they have not yet established substantial knowledge of any language, and as a result, their perception pattern reflects innate ease in perceiving the speech sounds without interference of language knowledge. Referring to the unsolved question regarding the perceptual space of T2 and T3, to observe how easy it is for the infants to categorize these two tones may shed light on the natural distinctiveness of these tones. Due to the complexity of the T3 contour, it is presumably also more specified in perception. If so, assuming that the distinctiveness of T3 makes it occupy a narrow perception space, it is expected that for infant language learners, it is easier to establish a category of T3 than a category of T2.

The available methodology for testing young infants actually gives us the chance to infer how easy it is to categorize a lexical tone. A widely used paradigm for testing early discrimination of speech sounds is the habituation-dishabituation paradigm. The paradigm starts with a habituation phase, in which the infants are familiarized with one phonological category. Gradually the infants get used to the repetition of the same category, and once their listening time drops below a prefixed criterion, the habituation phase ends and the test phase starts. In the test phase, they are presented with a trial presenting a new sound (together with another trial presenting the category of the habituation phase). If the infants are able to discriminate between the two categories, it is expected that their listening time will increase in the novel trial, due to the interest recovery when hearing something new (e.g. Best et al. 1988, Burnham & Dodd 2004, Pegg, Werker & McLeod 1992).

It has to be acknowledged that the habituation-dishabituation paradigm only gives indirect evidence of discrimination. Yet, it does provide us with the opportunity to observe the categorization process. In the habituation phase, variations could be introduced to the tokens representing one single phonological category. Taking lexical tones as an example, in the habituation phase, tokens of the same tone could be produced by different speakers, having different durations, or not precisely displaying the same pitch contours. In the test phase, if the infants were to discriminate a new tone from the habituated tone, they first need to normalize the multiple tokens occurred in the habituation phase into one single category, and then compare the one exemplar of the habituated category with another one exemplar of a new category. Therefore, by controlling the variations in the habituation phase, we have the chance to observe how infants establish a tonal category and then succeed in discriminating two lexical tones. In the scope of the current study, assuming that T3 forms a perceptually stricter category, it is expected that, with the same amount of variations in the habituation phase, the young infant learners will establish a T3 category more easily. Accordingly, those infants who are habituated on T3 are expected to discriminate the T2-T3 contrast more easily than those habituated on T2.

If we were to observe the natural ease in categorizing specific tones, it is necessary to test infants without any input from a tonal language. So far, it has been largely agreed that the perceptual reorganization of lexical tones occurs later than 6 months (Mattock & Burnham 2006, Mattock et al. 2008, Shi et al. 2011, Gao et al. 2011). Hence for non-tone language infants younger than or equal to 6 months, the influence of a native language on lexical tone perception is still marginal. Therefore, these young non-tone language infants are an appropriate population for testing the innate perceptual ease in perceiving Mandarin tones.

Another reason to test tonal discrimination of young infants is that it allows us to observe the developmental patterns in lexical tone perception. As mentioned in Chapter 1, a lot of discrepancy still remains regarding the perceptual development of lexical tone perception (Mattock & Burnham 2006, 2008, Shi et al. 2011, Tsao 2008, Liu & Kager 2012), which contradicts the classical view of perceptual reorganization (Werker & Tees 1984, Kuhl et al. 1992). The results of the mentioned studies could be summarized as follows. First, non-native infants before six months of age were able to discriminate between Mandarin T1 and T4 (Canadian French infants: Gao et al. 2011, Dutch infants: Liu & Kager 2012), and regardless of the rhythmic type of the infants' native language (Dutch is stress-timed while French is syllable-timed). Second, Dutch infants show sensitivity to the T1-T4 contrast until 14-15 months, which is far beyond the assumed perceptual reorganization window (Liu & Kager 2012). Third, Mandarin infants discriminated the T1-T3 contrast better than the T2-T3 and T2-T4 contrasts at 10-12 months (Tsao 2008). Yet, the failure of 6-month-old Mandarin infants in discriminating the T1-T4 contrast in Shi et al. (2011) is unexplainable. Taken together, it seems that non-native infants were fairly good at discriminating a salient Mandarin contrast, namely the T1-T4. Native infants also more easily discriminated the lexical tones that have highly distinctive pitch contours. However, how the not-so-salient contrast, namely the T2-T3 contrast, is perceived by non-native infants across different ages is not clear. To compare the early discrimination of not-so-salient tones with the discrimination of salient tones will help to reveal the initial capacity of the human auditory system, and will give more insight into the composition of acoustic salience. To compare the discrimination pattern of salient versus not-so-salient contrasts across different ages will also give us a more comprehensive picture of the perceptual reorganization of lexical tones.

In order to shed light on the natural perceptual space of Mandarin T2 and T3, and to explore how the perception of acoustically similar tones develops in non-tone language learning infants, I tested Dutch infants on their discrimination of Mandarin T2 and T3 at different ages (5-6, 8-9, 11-12, and 13-14 months old). At 5-6 months of age, according to previous literatures, the infants should not have finished (or even started) the perceptual reorganization yet, so infants at this age are supposed to be universal listeners, and their perception of T2-T3 contrast will reflect the initial status in T2 and T3 perception. By 13-14 months of age, the infants have finished perceptual reorganization, and they are expected to lose sensitivity to lexical tone contrast. By observing the process in which infants change from language-universal listeners to language-specific listeners, we will have the chance to understand how initial biases are shaped by language experiences.

4.2 EXPERIMENTS

Participants

All participants were full-term monolingual Dutch infants. None of the infants were reported to have had an ear infection in the three weeks before the experiment. In total, there were 28 5-6 month-old infants (5:04-6:14), 30 8-9 month-old infants (7:25-9:28), 30 11-12 month-old infants (11:1-12:10), and 28 13-14 month-old infants (13:02-15:01). Another eleven 6-month-old infants were tested and excluded from analysis (three for crying during the experiment, three for fussiness, three for not meeting the habituation criteria, one for equipment failure, and one for parental interference). An additional four 9-month-old infants were tested but excluded from analysis (two for crying during the experiment, one for not passing the post-test, and one for failing to meet the habituation criteria). Another twelve 11-month-old infants were tested but excluded from analysis (four for crying, four for equipment failure, three for failing to meet the habituation criteria, one for delay in motoric development as reported by parents after the experiment). Lastly, another ten 13-14 month-old infants were tested but excluded from analysis (one for equipment failure, three for crying, two for parental interference, four for failing to meet the habituation criteria).

Stimuli

Two female Mandarin native speakers produced multiple tokens of syllable /ma/ bearing T2 and T3 respectively as well as several other monosyllabic words with different tones. Two pairs of tokens of /ma/, with both tones from each speaker, were chosen as stimuli in the habituation phase. Another pair of tokens with both tones from one speaker was used as stimuli in the test phase. The duration of stimuli was between 517 and 814 ms. The stimuli were balanced in intensity. The time-normalized average pitch contours of the stimuli of each tone as produced by the two speakers are depicted in Figure 4.1. During the habituation and the test phase, the visual stimulus was a colorful infant-friendly picture. Between each trial, a smiling baby face was used as attention getter.

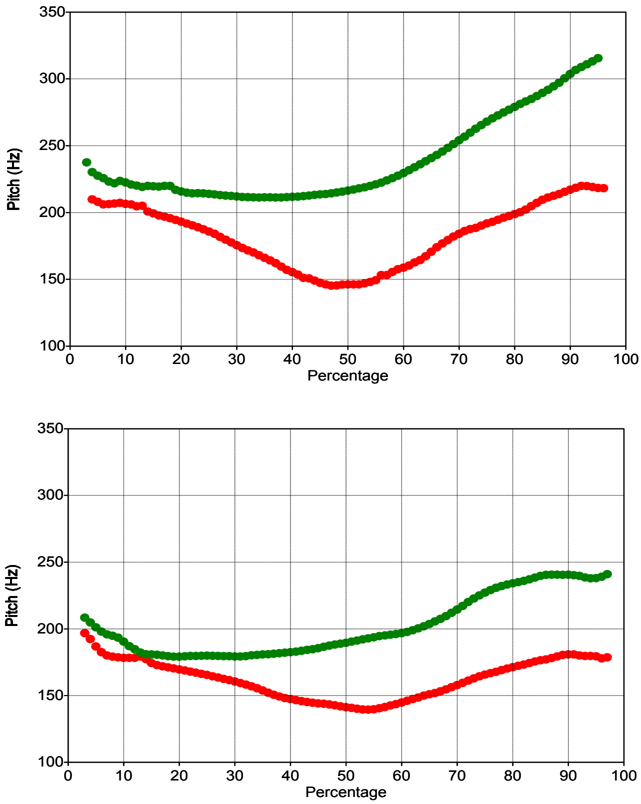


Figure 4.1 the average f0 contours of test stimuli T2 (upper line) and T3 (lower line) produced by the two speakers.

Experiment settings and procedure

The visual fixation paradigm was selected for the current study. The lab consisted of a test cabin and a separate control room for the experimenter. During the experiment, infants sat on their parent's lap in the test cabin, and in front there was a 14 inch computer screen about one meter away from the baby displaying the visual stimuli. The auditory stimuli were presented at a comfortable volume through a hidden speaker in front of the baby. The parent listened to background music through headphones to prevent possible interaction with the infants. There was a hidden camera above the screen recording the behavior of the infants, and the live video stream was transferred to experimenter's computer in the control room. The experimenter observed the video and recorded the visual fixation of the infants by pressing the "looking" and "non-looking" button on a button box connected to the control computer.

The experiment included a habituation phase and a following test phase. A pretest and a posttest were used to measure general attention of the participants, in which the stimuli were moving infant-friendly pictures accompanied by beeps. When the pretest finished, and once the infant focused on the screen, the experimenter initiated the first habituation trial by pressing the "looking" button, and once the infant looked away, the experimenter pressed the "non-looking" button. The looking and non-looking of the infants were always recorded by these two buttons. If the infant looked away and then looked back at the screen within two seconds, the same trial continued, while if the infant looked away for more than two seconds, the current trial ended and the attention getter appeared on the screen to get the attention of the participant back. After the attention getter was activated and the infant looked back at the screen, the experimenter started the next trial by pressing the "looking" again. All the trials were started and finished as described above. The looking time of each look and the total looking time of one single trial were recorded automatically by the experimenter's control computer. In this way, the infants' looking time to the visual stimuli was used as an indicator of their attention to the auditory stimuli.

The total looking time of the first three trials in the habituation phase was used as a baseline for measuring habituation. After the first three habituation trials, once the total looking time of three consecutive habituation trials was less than 65% of the total looking time of the first three habituation trials, the habituation criterion was met, and the test phase started automatically. In the test phase, the infants were presented with one "old" trial, which was another token of /ma/ with the same tone as they had heard in the habituation phase, and another "novel" trial which was /ma/ carrying the other tone that they had not previously heard in the habituation phase. One single token of each tone was used in the test phase. The tones that were used in the habituation phase and the order of the "old" and "novel" trials in the test phase were counter-balanced among the participants.

Prediction

In the test phase, if the infants were able to detect the difference between the two tones, then upon hearing the novel trial, their listening time should be recovered due to hearing something new. In other words, they would have a longer looking time to the novel trial than to the old trial. Specifically regarding this procedure, there is no reason to suspect that the looking time of the novel trial would be significantly shorter than the looking time to the old trial. In other words, only increase in looking time can serve as an indicator of successful discrimination. While equal or decrease of listening time may just be the result of loosing attention in general. Regarding the transition from habituation phase to test phase, I do not expect significant looking time difference between the last habituation trial and the old trial in the test phase. Moreover, if there is no significant decrease from last habituation trial to the old trial in test phase in terms of looking time, then infants reaction to the old trial is stable, which gives extra evidence of successful habituation.

Results and discussion

The videos of all the participants were recoded offline after the experiment before being submitted to analysis. After recoding, the raw looking time of the "old" and "novel" trials was logarithmically converted to correct the skewness of the distribution of the raw data. The log transformed looking times of all the age groups fitted a normal distribution, and the statistics hereafter are based on the log transformed looking time (LGLT): $D_{\text{6-months old trial}}(28) = 0.09$, $p>0.1$; $D_{\text{6-months novel trial}}(28) = 0.10$, $p>0.1$; $D_{\text{9-months old trial}}(30) = 0.08$, $p>0.1$; $D_{\text{9-months novel trial}}(30) = 0.13$, $p>0.1$; $D_{\text{11-months old trial}}(30) = 0.15$, $p>0.1$; $D_{\text{11-months novel trial}}(30) = 0.12$, $p>0.1$; $D_{\text{14-months old trial}}(28) = 0.11$, $p>0.1$; $D_{\text{14-months novel trial}}(28) = 0.1$, $p>0.1$.

a. Results of the 6-month-old Dutch infants

First, a 1-tailed paired t test was carried out between the LGLT of the “old” and “novel” trials, between the LGLT of the “old” trial and the last trial in habituation (lasthab), and between the LGLT of the “lasthab” and the “novel” trial. None of the pairs revealed a significant difference: $T_{old\text{-}novel}(27) = -0.391$, $p>0.05$; $T_{lasthab\text{-}old}(27) = -0.036$, $p>0.05$; $T_{lasthab\text{-}novel}(27) = -0.328$, $p>0.05$.

Second, the LGLTs of the participants habituated on T2 and T3 were analyzed separately. As the previous analysis had shown that the LGLTs of the lasthab trial and the novel trial did not differ significantly, only the LGLTs of the old trial and the novel trial were submitted for statistical testing. A paired t test showed that, regardless of the tone that infants had been habituated on, there was no significant difference between the LGLT of the old trials and the novel trials: $T_{habT2}(13) = -1.25$, $p>0.05$; $T_{habT3}(13) = 0.563$, $p>0.05$. Figure 4.2 gives the overall mean LGLTs of the three trial types of the 6 month-old infants, and the mean LGLT of the old trial and the novel trial separated by the habituation tone.

As can be read from the figures, although infants tended to look longer to the visual stimuli in the novel trials, the LGLT difference failed to reach statistical significance. The non-significant result implies that in the current task, 6-month-old Dutch infants did not demonstrate a discrimination effect of the two tones. Moreover, this pattern remains the same regardless on which tone they were habituated. These results suggest that the 6-month-old Dutch infants failed to discriminate between Mandarin T2 and T3. It is difficult to draw a conclusion about the failure in T2-T3 discrimination based on a null result. More infants need to be tested to examine the validity of the null results.

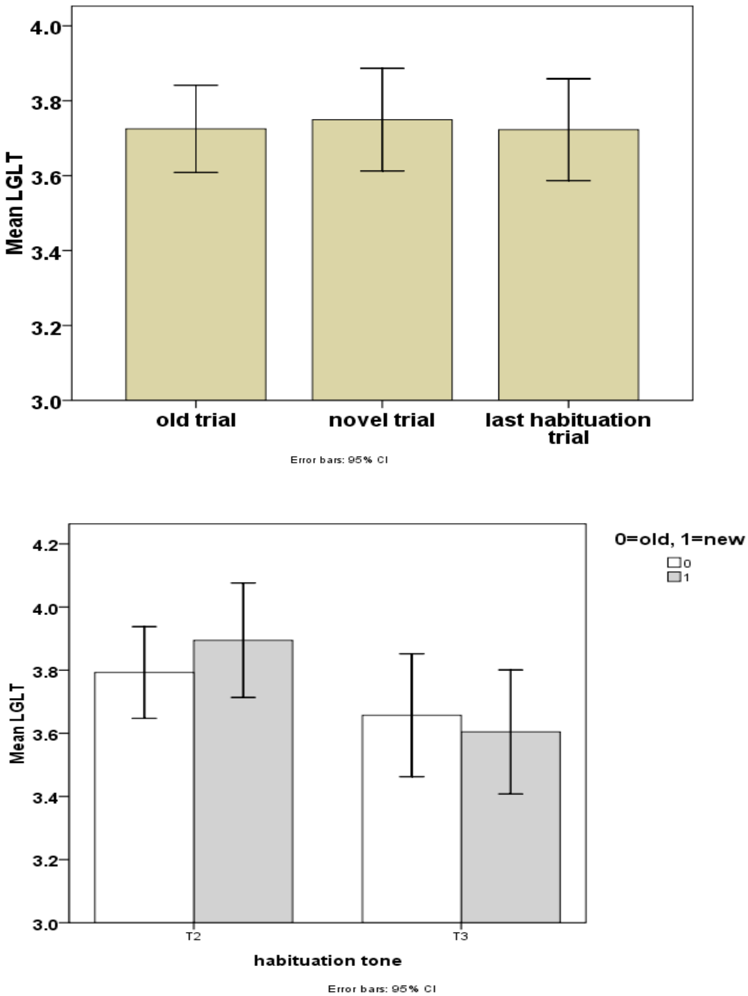


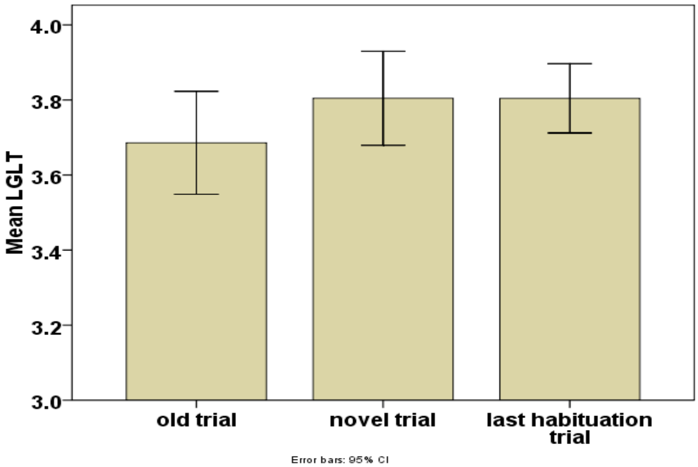
Figure 4.2 Overall mean LGLT of the three trial types (upper panel), and the mean LGLT of the old trial and the novel trial separated by the habituation tone (lower panel) of 6-month-old Dutch infants.

b. Results of the 9-month-old infants and discussion

The same statistical tests used for the 6-month-old infants were used for the 9-month-old infants. The LGLT difference between the last habituation trial and the old trial, as well as between the last habituation trial and the novel trial, failed to reach significance: $T_{lasthab\text{-}old}(29) = 1.50$, $p>0.05$; $T_{lasthab\text{-}novel}(29) =$

-0.01, $p>0.05$. However, the difference in LGLT between the novel trial and the old trial did reach significance: $T_{old\text{-}novel}(29) = -1.70$, $p<0.05$ (1-tailed). When the participants were separated into two sub-groups based on the tone that they had been habituated on, only those habituated on T3 showed significantly longer LGLTs to the novel trial: T_{habT2} (15) = 0.792, $p>0.05$; T_{habT3} (13) = −3.597 $p<0.005$. Figure 4.3 depicts the overall mean LGLT of the three trial types, and the mean LGLTs of the old trial and the novel trial separated by the habituation tone.

Overall, the 9-month-old Dutch infants succeeded in discriminating between Mandarin T2 and T3. Yet, it is evident that the overall significant difference between the LGLT of the old trial and the LGLT of the novel trial was caused by the highly significant LGLT difference of those infants who were habituated on T3. On the other hand, those infants who were habituated on T2 failed to show a discrimination effect between T2 and T3. Considering the rationale of the paradigm, if the infants were to discriminate between T2 and T3, they had to build up a tonal representation based on the various tokens of the tone presented to them in the habituation phase, and they then had to tell whether the tokens in the old trial and the novel trial in the test phase were the same as the built-up tonal representation. Therefore, it seems that with the same amount of variation in the habituation phase, to build up a representation of T3 in the mind was somehow easier than to build up a representation of T2. As a result, discriminating a T2 from the T3 category was easier than the other way around. The performance of the 9-month-old infants was consistent with the postulation that I made in earlier chapters. It seems that T3 is perceptually more distinctive than T2, and hence it forms a stricter tonal category than T2. Moreover, in a way, the discrimination pattern of the 9-month-old Dutch infants resembled the discrimination of Dutch adults, meaning that discriminating a T2 from a T3 was easier than discriminating a T3 from a T2.



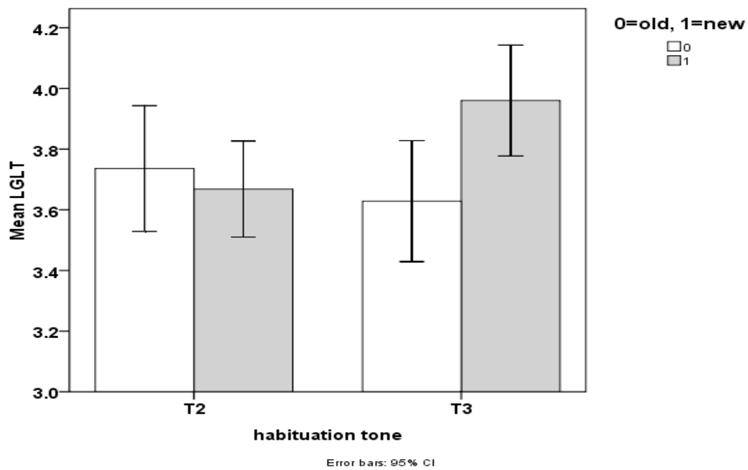
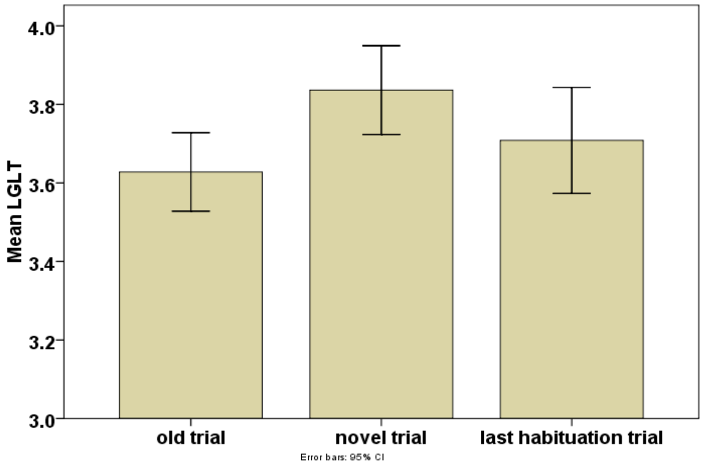


Figure 4.3: Overall mean LGLT of the three trial types (upper panel), and the mean LGLT of the old trial and the novel trial separated by the habituation tone (lower panel) of 9-month-old Dutch infants.

c. Results of the 11-month-old Dutch infants

The same statistical tests as used for the 6-month-old infants were carried out for the 11-month-old infants. As was the case with the 9-month-old group, the LGLTs of the novel trial were significantly longer than those of the old trials: $T_{old\text{-}novel}$ (29) = -2.661, $p<0.05$ (2 tailed). Again, the difference in the LGLT between the last habituation trial and the old trial failed to reach significance: $T_{lasthab\text{-}old}(29) = 1.448$, $p>0.05$. However, differing from the 9-month-old group, the LGLT of the novel trial was significantly longer than the last habituation trial: $T_{lasthab\text{-}novel}(29) = -1.648$, $p<0.05$. The reason for the significant difference between the last habituation trial and the novel trial could be that by 11 months of age, infants are quite mobile, and can be bored more easily. As a result, by the end of the habituation phase, their looking time decreased significantly. When the participants were separated into two sub-groups based on the tone that they were habituated on, again, only for those infants who were habituated on T3, the LGLT difference between the old trial and novel trial reached significance: T_{habT2} (15) = -1.04, $p>0.05$ (2 tailed); T_{habT3} (15) = -3.70 $p<0.05$ (2 tailed). Once again, the overall significant LGLT difference between the old trial and novel trial was caused by those who were habituated on T3, as their LGLT to the novel trial was much longer than that to the old trial. Figure 4.4 gives the overall mean LGLT of the three trial types, and the mean LGLTs of the old trial and the novel trial separated by the habituation tone.

The Dutch 11-month-old infants, just as their 9-month-old counterparts, showed a similar discrimination asymmetry. Again, it seems that for 11-month-old Dutch infants, with the same amount of variation, it was easier to establish a T3 category than to establish a T2 category. Accordingly, the infants discriminated a T2 from a T3 category more easily than a T3 from a T2 category. Hence, the results support the speculation that T3 is more specified and more distinctive, which facilitates the exclusion of a token that falls outside its category. The complex tonal contour of T3 might be a reason for its perceptual distinctiveness, as the co-occurrence of multiple features (low, dipping, rising) could easily specify a tone.



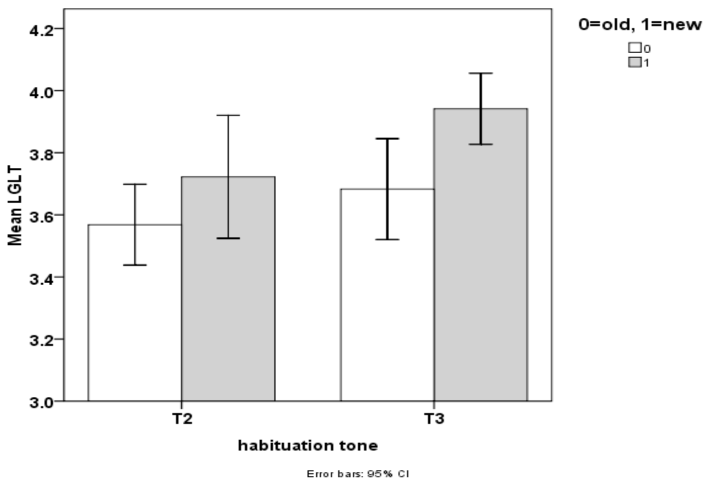


Figure 4.4 the overall mean LGLT of the three trial types (upper panel), and the mean LGLT of the old trial and the novel trial separated by the habituation tone (lower panel) of 11-month-old Dutch infants.

d. Results of the 14-month-old Dutch infants

The same statistical tests as used for the 6-month-old infants were carried out for the 14-month-old infants. The t test demonstrated a marginal significant difference between the LGLT of the old trial and the LGLT of the novel trial: $T_{old\text{-}novel}(27) = -1.62$, $p=0.059$ (1 tailed). The difference in LGLT between the last habituation trial and the old trial failed to reach significance: $T_{lasthab-old}(27) = -0.326$, $p>0.05$. The LGLT of the novel trial was significantly longer than the last habituation trial: $T_{lasthab-novel}$ $(27) = -1.827$, $p<0.05$ (1 tailed). When the participants were separated into two sub-groups based on the tone that they were habituated on, no matter on which tone they were habituated, the LGLT difference between the old trial and the novel trial failed to reach significance: T_{habT2} $(13) = -1.226$, $p>0.05$ (1 tailed); T_{habT3} $(13) = -1.087$, $p>0.05$ (1 tailed). Figure 4.5 depicts the overall mean LGLT of the three trial types, and the mean LGLTs of the old trial and the novel trial separated by the habituation tone.

Overall, for the 14-month-old Dutch infants, there was a tendency for successful discrimination of the Mandarin T2-T3 contrast. Yet, when the participants were divided into two groups, neither group showed a significant novelty effect. The results suggest that under the current paradigm, by 14 months of age, the Dutch infants do not show evident sensitivity to the T2-T3 contrast. When interpreting this null result, one thing is worth mentioning: at 14 months of age, the infants were already quite mobile, and thus grew bored much more easily than the younger groups. Hence, it is possible that the habituation-dishabituation is not entertaining enough to maintain the interest of the older infants. Whether other procedures would show a discrimination effect of the T2-T3 contrast needs to be tested in the future.

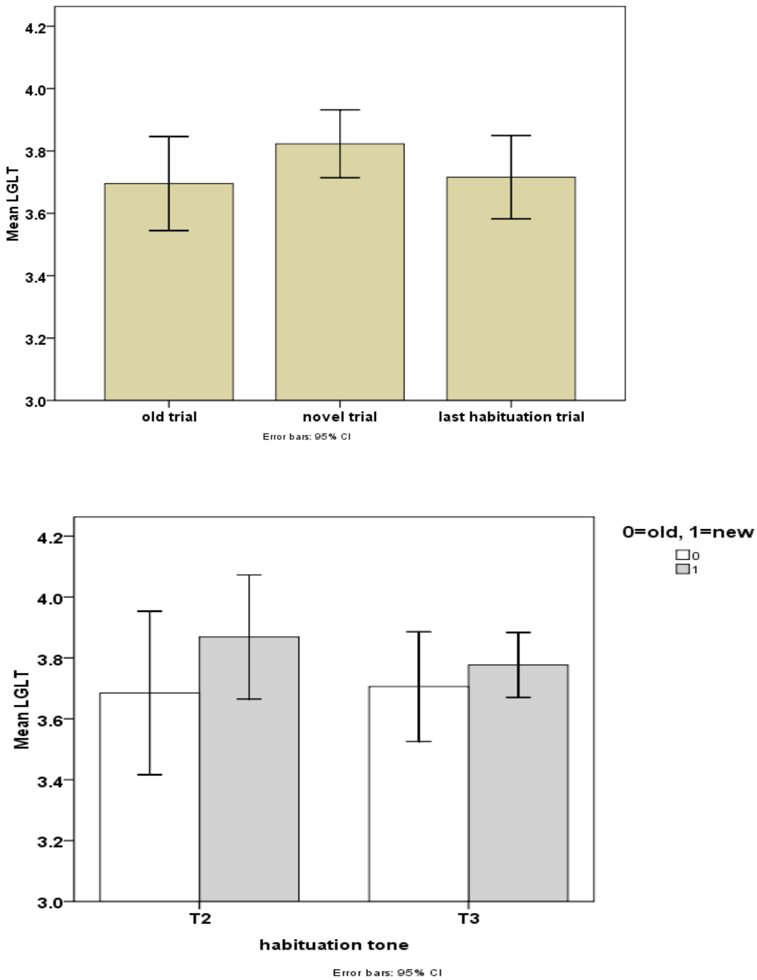


Figure 4.5 the overall mean LGLT of the three trial types, and the mean LGLT of the old trial and the novel trial separated by the habituation tone of 14-months-old Dutch infants.

4.3 CROSS-AGE ANALYSIS OF THE MANDARIN T2-T3 DISCRIMINATION BY DUTCH INFANTS

In order to have a direct view of the developmental pattern of the perception of this tonal contrast, the performance of all the four age groups was compared. Taking into consideration that the LGLT of the old trial and that of the last habituation trial did not differ significantly, only the LGLT of the old trial and the LGLT of the novel trial were submitted to a mixed ANOVA. The trial types (old or novel) was the within-subject variable, and habituation tone (T2 or T3) and age (6, 9, 11, 14) were the between-subject variable. A significant main effect of trial types was found— $F_{\text{trial type}}$ (1, 103) = 13.53, p<0.001. However, neither age nor the habituation tone turned out to have a significant main effect on the LGLT of the infants: F_{age} (3, 103) = 0.07, p>0.1, $F_{\text{habituation tone}}$ (1, 103) = 0.003, p>0.1. A significant interaction was found between age and habituation tone, F (3, 103) = 2.72, p<0.05. There was also a significant interaction between trial types, age and habituation tone, F (3, 103) = 3.82, p<0.05. From the results, we could see that, overall, the infants tended to look longer at the novel trial, and infants of different age behave differently when habituated on different tones. Bearing in mind the discrimination pattern of each individual age group, it can be seen that the main effect of trial type mainly came from the successful discrimination of the 9-and 11-month-old infants. The significant interaction could be interpreted in the way that 9 and 11 month-old infants succeeded in discrimination only if they were habituated on T3, while the other two age groups failed to discriminate the tonal contrast regardless of the habituation tone.

4.4 GENERAL DISCUSSION

As we can see from the results above, native Dutch infants were unable to discriminate between Mandarin T2 and T3 at 6 months, but unexpectedly, their ability to discriminate this tonal contrast improved by 9 months, and the same perception pattern could be observed among the 11 month-old infants. By 14 months, there was a tendency for successful discrimination, but the discrimination effect was not significant. If we look at the infants habituated on different tones separately, at 9 and 11 months, they succeeded in discrimination only if they were habituated on T3.

These findings can be discussed from several perspectives. First, contradictory to previous research, in the current study, the 6-month-old infants were unable to discriminate between non-native tonal contrasts. At the age of 6 months, it is assumed that the perceptual reorganization of lexical tones has not yet completed, and infants of this age were previously found to be able to discriminate non-native tones (Harrison 2000, Mattock & Burnham 2006, 2008, Shi et al. 2011, Gao et al. 2011). It is plausible that the failure was due to the acoustical similarity between the two tones. Dutch infants did succeed in discriminating the Mandarin T1 versus T4 contrast at 5-6, 8-9, 11-12, and 14-15 months, and once the T1 and T4 were manipulated so that the acoustical difference between the two tones shrank, Dutch infants started showing a weaker discrimination effect (Liu & Kager 2012). The failure to discriminate between not-so-salient non-native contrasts by young infants has also been observed in early segmental perception (Narayan et al. 2010). As shown in the previous two chapters, the T2-T3 contrast is the most difficult contrast for both native Mandarin adults and non-native adult listeners. In addition, previous research has also shown that T2 and T3 form the most difficult perceptual contrast for native Mandarin infants and children (Li & Thompson 1977, Zhu 2000). If we look at the F_0 contours of the two tones, we can see that they start from almost the same height, while in the second half of the F_0, both tones rise in parallel. The difference is that the T3 is overall lower than T2, and there is an evident dipping along the T3 contour. Thus, it could be that the acoustical difference between the tones is too small, and infants of 6 months of age were not mature enough auditorially, so they failed to sufficiently track the pitch contour of the two tones.

By 9 months, in the current experiments, the Dutch infants became more capable of hearing finer-grained acoustical differences and began displaying some discrimination of the Mandarin T2 and T3. Viewing the results from the 6-month-old Dutch infants together with the findings in Narayan et al. (2010), it seems that infants are not endowed to perceive all sounds contrastively. For the not-so-salient contrasts, they need some time to develop their auditory system in order to process them sufficiently. Yet, different from Narayan et al. (2010), who observed that English infants failed to discriminate between the not-so-salient Filipino onset /n/-/ŋ/ contrast consistently at both 6-8 months and 10-12 months, the current experiments found that the perception of a non-native not-so-salient contrast improved between 6 and 9 months of age. It might be that the perception of prosody and the perception of segments develop in different ways. To better understand what causes these different developmental patterns, the perception of speech prosody in early infancy needs to be studied more closely. Future studies may need to take a closer look at the "acoustical salience"— what kind of acoustics makes a salient contrast salient? Furthermore, how salient is salient enough for successful discrimination in early infancy?

Second, besides the possible acoustical property of the T2-T3 contrast that may have influenced the discrimination of the infants, the manner in which the stimuli were presented may have also influenced the responses. As mentioned in Chapter 1, it has been observed in various experiments that young infants (7.5 months old) were unable to generalize across the same words produced by different genders or in a different vocal affect status (Houston & Jusczyk 2000, Singh et al. 2004). Their ability of generalization has been observed to improve by the age of 10 months (Sing 2008). On the other hand, however, Kuhl (1983) found that 6-month-old infants were already able to treat the same word produced in different voices as equivalent. In word learning tasks, it has been found that sufficient token variations were helpful for linking a sound with a concept (Rost & MucMurray 2009, Hollich et al. 2002). In the field of early lexical tone perception, multiple tokens of one single speaker (five tokens of the Thai tones in Mattock & Burnham 2006, 2008, 12 tokens in Shi et al. 2011, four tokens in Liu & Kager 2011) have been used as stimuli in the studies that reported successful discrimination of lexical tones by non-native infants. In the current experiments, I used two tokens of two different females as stimuli in the habituation phase, and the timbre of the two voices did differ substantially. As a result, it is possible that the failure of discrimination of the 6-month-old infants was not due to them being unable to hear the acoustical difference between Mandarin T2 and T3 *per se*, but due to the failure in normalizing the different tokens of different voices. To test this hypothesis in Chapter 6, I will reduce the number of voices in the experiment by using only one speaker in the habituation phase, while extending the token variations within the productions of the speaker. From a different angle, if we were to test whether it is the acoustical property of the two tones that hinders the infants from successful discrimination, the infants need to be tested with the pure acoustics of the tones, such that they are presented with one single token of the T2 or T3 in the habituation phase, and then tested with the same habituation token with another single token of the unheard tone. In this way, the cross-token variation is maximally reduced and the infants are encouraged to perceive the acoustics of the tones.

Third, and most importantly, we have observed that habituating 9- and 11-month-old Dutch infants on T3 assisted them in detecting a change from T3 to

T2, but habituating them on T2 failed to trigger the detection of the change. It seems that with the same amount of variations, i.e. 2 tokens of two speakers, it was easier for the infants to establish a representation of T3; hence, to discriminate a T2 that fell outside the space of T3 was successful. But with the same variation of four tokens, the infants failed to build up a category of T2, and tended to accept a T3 as similar to a T2. This pattern gives positive evidence for my earlier hypothesis: T3 is perceptually more specified and more distinctive, hence excludes other tones easily. On the other hand, T2 is less specified and more general. Thus, a T3 could be perceived as a peripheral instantiation of T2, but not vice versa.

To put our findings in a bigger picture, discrimination asymmetry among infants regarding segments is not rare. As mentioned in Chapter 1, coronal consonants have been claimed to be underspecified as compared to labial consonants, which hinders the discrimination of labial from coronal consonants but not vice versa (Altvater-Mackensen & Fikkert 2010). Various studies have found that infants succeeded in detecting a vowel change unidirectionally (Polka & Werker 1994, 1996, Bohn & Polka 2001, Swoboda et al. 1976, 1978, Best et al. 1997, Best & Faber 2000). The discrimination asymmetry observed in vowels has been interpreted in the way that the marginal vowels of the vowel space serve as better referents for discrimination than the inner vowels (Polka & Bohn 2003). In 2011, Polka and Bohn proposed the Natural Referent Vowel framework (NRV), in which they argued that the extreme vowels which occupy the periphery of the F1-F2 space, such as /i/, /a/, /u/, occur in all the languages, and naturally serve as distinctive perceptual referents. Young infants are biased to perceive the referent vowels. Polka and Bohn (2011) explained the referent effect from the perspective of focalization, namely the formants convergence is strongest among the peripheral vowels, which makes the peripheral vowels perceptually most salient. Therefore, a change from more central vowels to more salient vowels could be easily detected. When encountering a change of the reversed direction, the infants tended to equalize the central vowels as instantiations of the referent vowels (Bohn 2007, Polka & Bohn 2011). This framework looks at early vowel perception from a new angle, and though it is inspiring, the authors pointed out that the precise role of the NRV has not yet been fully understood. Their explanation of why the peripheral vowels naturally serve as good referents is also quite vague. Nevertheless, they hinted that "NRVs may facilitate the recognition of equivalence across vowels produced by different talkers"

(Polka & Bohn 2011: 475). This speculation seems to be consistent with what I have found with the infants: being habituated with either four tokens of T3 or four tokens of T2, to form a T3 category was easier. However, in Polka & Bohn (2011), to discriminate central vowels from NRVs is more difficult, and the authors speculated that the speaker normalization would be easier for NRVs. In my case however, it seems that to discriminate T3 from T2 is more difficult, while the speaker normalization of T3 is easier. Hence, what mechanism stands behind the observed discrimination asymmetries still needs to be examined more closely in the future.

Despite that perception asymmetries have been extensively found in various studies, there is still a long way to go to answer the question what causes the asymmetry. My speculation is that, the complex pitch contour of T3 makes it acoustically specified and hence more distinctive in perception. As a result, it is easier to distinguish the exemplar of another category from it. The labials in Altvater-Mackensen & Fikkert (2010) or T3s in the current experiments, might be perceptually distinctive and form a strict perceptual space, which favors the exclusion of other tokens. Considering that a perceptual asymmetry is extensively observable among infants for both segments and tones, for future studies, it is worth the effort to test whether the unequal degree of specification could explain the discrimination asymmetry for both segments and speech prosody.

Fourth, the youngest group in the current study is 6 months old, and we know that the vowel perceptual reorganization may start earlier than 6 months (Polka & Werker 1994). Though previous studies did not find positive evidence of lexical tone perceptual reorganization before 6 months (Mattock & Burnham 2006, Mattock et al. 2008, Liu & Kager 2011, Gao et al. 2011), it is still possible that native Dutch infants' sensitivity to Mandarin T2 and T3 decreases at an earlier age. It might be the case that at a younger age, the Dutch infants are able to discriminate between T2 and T3. To test whether the perceptual reorganization happens at a younger age than discovered in the aforementioned studies, in Chapter 6, I will test young infants that are very unlikely to have lost their sensitivity to non-native contrasts.

Last, regarding the failure of the 14 month-old infants, it has to be acknowledged that the methods that can be used to test infants are very limited, and these methods only give a very indirect measurement. Hence, when interpreting the null results, the limitation of the experiment paradigm has to be considered. The

14-month-old infants were already able to stand and walk, and had had quite some experience with the surrounding world; hence, the visual fixation paradigm was probably too boring for them. As could be seen from Figure 4.5, there were substantial variations in the looking times of the infants, and quite some infants became fussy and agitated at the end of the experiment. Therefore, it is possible that the current paradigm is not entertaining enough to encourage successful discrimination. To obtain a clearer picture of whether the 14-month-old infants are able to discriminate between T2 and T3, it is useful to test the infants with other paradigms.

Chapter 5 Contextual learning of Mandarin T3 sandhi by non-native infants

5.1 INTRODUCTION

As demonstrated in Chapter 4, native Dutch infants failed to discriminate between Mandarin T2 and T3 at 6 months, but their perception of this contrast improved between 6 and 9 months. Nevertheless, at both 9 and 11 months, they showed a discrimination effect of the two tones only if they were habituated on Mandarin T3. The asymmetry observed in the aforementioned experiments suggests that, initially, establishing a representation of T3 might be easier than establishing a representation of T2. Regarding why this pattern occurred, I inferred that compared to T2, T3 has a more complex pitch contour which may make it more distinctive in perception. Due to the specificity of T3, it is relatively easy to detect a T2 token that falls outside the T3 category. The relative ease in categorizing a T3 may be an explanation of why the asymmetry has been consistently observed in the T2-T3 discrimination.

It needs to be acknowledged that this paradigm does not provide direct evidence of the order effect regarding T2 and T3 discrimination by infant listeners. Restricted by the paradigms that are usable for testing early discrimination of sounds, in the experiments discussed in Chapter 4, the infants were first trained by multiple instantiations of one tone, and then presented with two test trials, one presenting the habituated tone, and one presenting a new tone. Within each test trial, they were also presented with multiple repetitions of one tonal token. Hence, with the currently widely used habituation-dishabituation procedure it is not possible to test the order effect of the T2 and T3 discrimination among infants in a direct way. It is also impossible for the infants to participate in an AX discrimination task as the adults, and they cannot say whether the two tones are the same or different. With regard to the bisyllabic tonal sequences discrimination (as I did with the adults in Chapter 3), adopting the habituation-dishabituation paradigm may introduce confounding factors. For example, if we habituate the infants on the bisyllabic T3T3 sequence, and then in the test phase provide them with trials alternating between T3T3 and T3T2, even if they show the discrimination effect, it is very hard to infer whether they perceive a T3T2 as different from a T3T3, or whether they just detect the difference between a monosyllabic T3 and a monosyllabic T2 at the right edge of a word, or whether they are just discriminating between a T2 and T3 within the T3T2 sequence.

Therefore, we still need to rely on an indirect measurement to test the discrimination asymmetry between T2 and T3 among infant participants. In addition, the motivation behind testing infants is that they have not yet accumulated substantial knowledge of any language, and how they learn to perceive speech sounds gives an opportunity to observe the perceptual biases reflected in the language acquisition process. Therefore, to simulate a situation that is similar to real language acquisition would be the best way to reflect the innate biases in language acquisition.

In order to have a more direct view on the possible perceptual biases of T3 sandhi, as well as to reveal how such a perceptual bias may affect the early phonological acquisition of young infant learners, Dutch infants were first trained on naturally produced Mandarin speech with either the confusing T2T3 sequence or the non-confusing T3T2 sequence embedded, and later tested with the discrimination between the embedded sequence and a T3T3 sequence. As shown in Chapter 4 with the habituation-dishabituation procedure, only 9 and 11-month-old infants demonstrated sensitivity to the possible order effect for the discrimination between Mandarin T2 and T3. In the current experiments, I also tested infants of these two ages.

5.2 EXPERIMENT 1 WITH 9-MONTH-OLD DUTCH INFANTS

Participants

32 healthy full-term 8-9-month-old (age range 8:15- 9:16) Dutch infants participated in the research. All the infants were monolingual Dutch infants, and none of them had exposure to any tone language. Another four infants were tested but excluded from analysis because of crying (N =1), equipment failure (N = 1), and experimenter's error (N = 2).

Stimuli

For the training phase, I created a carrier sentence ta(T1) zai(T4) zhuo(T1) zi(T0) shang(T4) hua(T4) le(T0) zhi(T1)____ ("He drew a ___on the table"). The blank at the end of the sentence could either be /mama/ carrying a T2T3 (condition A), or /mama/ carrying a T3T2 (condition B). A female native speaker produced multiple sentences of both condition A and condition B with an infant friendly intonation. The /mama/ sequences in both conditions were naturally produced bisyllabic sequences embedded in the carrier sentence, rather than a concatenation of monosyllables produced in isolation. Among the multiple productions, five sentences were selected as stimuli for the training phase. The range of the duration of the sentences was between 3.10s and 3.15s. The /mama/ carrying a T2T3 had a duration between 914ms and 1009ms, and /mama/ carrying a T3T2 in the sentence had a duration between 873 and 1000ms. Half of the infants participated in Condition A while the other half participated in Condition B.

For the test phase, the same speaker produced the monosyllable /ma/ with either T2 or T3 in isolation, and then the monosyllables were concatenated to form three bisyllabic sequences, namely /mama/ carrying a T2T3, T3T2, and T3T3. One token of each sequence was used as a stimulus in the test phase. For condition A, the T2T3 and T3T3 sequences were used as stimuli, and for condition B, the T3T2 and T3T3 sequences were used as stimuli. All three sequences were manipulated to have a duration of 1008ms.

The pitch contour of a T3 preceding a non-T3 tone is realized differently than a T3 produced in isolation. Yet, it has been argued that "low" is the most important feature of T3, and T3 is the only low tone in Mandarin (Yip 1980, Milliken 1989, Wu 1988). Hence, it is expected that the T3s realized in the carrier sentences would not be assimilated to other tones in perception. As we have shown in Chapter 4, for native Dutch adults as well as for native Mandarin adults, a T2T2 sequence was quite distinguishable from a T3T3 sequence; hence I did not include the condition in which a naturally produced T2T2 sequence was embedded as a target word in the carrier sentence.

The stimuli were generated in this way for a number of reasons. First, in the training phase, the infants were presented with sentences in which all the four tones occurred. The presence of the neutral tone assured that all the tonal categories present in Mandarin were presented to the infants in the familiarization phase. Hence, the training phase was to a large extent similar to the natural input when acquiring Mandarin. Second, the target words /mama/ were located at the end of the sentences, and formed a prosodic word on its own. By doing so, it is assured that the target words were perceptually salient for the infants. I expected the infants to be able to link the sequences that they heard in the test phase with the target words in the training phase. Third, in the test phase, I used concatenated bisyllabic sequence rather than naturally produced sequences. The reasons for doing so are twofold: first, if I had presented the infants with naturally produced target words as in the carrier sentences, together with concatenated T3T3 sequences, then the infants would be diverted to pay attention to the difference between concatenated sounds and co-articulated sounds. If this were the case, we would not be able to know whether they were discriminating between lexical tones or discriminating between properties of the speech. Second, if a T3T3 sequence were produced naturally, T3 sandhi would have occurred, and it would not be possible to get a T3T3 with its underlying tonal form. Lastly, I am interested in finding evidence from an acquisitional perspective for the occurrence of T3 sandhi, asking why the underlying form of T3T3 is sandhied to a T2T3. Hence, it is necessary to test the perceptual distance between the sandhied form and the underlying form. In a natural language learning situation, the infants only encounter naturally produced tonal sequences, such as the co-articulated T2T3, and they need to build up the underlying form of this sequence, namely a T2 and a T3 in a row, and equalize the underlying form and the surface form. One step further, with regard to T3 sandhi, the Mandarin learners need to equalize the co-articulated T2T3 sequence to the underlying T3T3 sequence. The design of the current experiments, i.e. training the infants with naturally produced sequences while testing them on the concatenated underlying form of each tone, reflects the process of acquiring phonology as it occurs in the natural learning of a language.

Both the sentences and the concatenated sequences were presented to two native Mandarin phoneticians, and they all agreed that the tones were natural and unambiguous.

Prediction

Considering the asymmetrical and positional property of T3 sandhi, if the infant language learners tend to accept the naturally produced T2T3 as an instantiation of T3T3 from exposure to natural language, then we would expect no discrimination effect in the test phase in condition A. In condition B, however, if the infants are able to perceive the difference between T3T2 and T3T3, then we would expect a discrimination effect. One might argue that the infants were discriminating between concatenated sequences in the test phase and naturally produced sequences in the training phase, but if this were the case, then we would expect the infants to look equally long in the two types of test trials. Moreover, if the infants tend to consider the tokens in the "same" trials as a different category from the target words in the training phase, then again, when presented with two new sequences, we would not expect any discrimination effect in the test phase.

Procedure

During the experiment, the infants sat in the test cabin on the caregivers' lap and the caregiver listened to background music through headphones. At the infant's eye level, a green light was placed in front, while two red lights were placed to the left side and right side of the infant. The three lights were about one meter away from the infant. The experimenter sat in the control cabin outside the test cabin and recorded the looking time of the infants by pressing either the "looking" or "non-looking" button on a button box. A hidden camera in front of the infant took a video of the experiment for each child, and the video was transferred live to a computer in the control room.

To start the experiment, the experimenter pressed the "looking" button, and the green light in the middle began blinking. Once the infant looked at the green light, the experimenter pressed the "looking" button again, and one of the red sidelights would start blinking. Once the infant looked at the blinking light, the experimenter pressed the "looking button" again to start the stimuli, and the control computer started recording the looking time of the infant. Once the infant looked away, the experimenter pressed the "non-looking" button, and the control computer of the experimenter would start to count the non-looking time. If the infants looked back to the blinking light within two seconds, the experimenter pressed the "looking" button, and the light on the same side would continue blinking, and the same trial continued. If the infants looked away for more than two seconds, the side light would turn off, finishing the trial. Then the experimenter pressed the "looking" button again, and the green middle light would start blinking, and once the infant looked at the green light, the experimenter would press the "looking" button, and one of the sidelights would start blinking. Once the infant looked at the blinking light, the experimenter pressed the “looking” button to start the next trial and the calculation of the looking time. In the training phase, the blinking of the sidelights was always controlled by infants, but regardless of where they were looking, the sound stimuli kept playing. In the training phase, the red light on the same side could not blink for more than two consecutive trials. For each infant, the five carrier sentences were repeated for 10 times in a random order, which made the training phase last for around 3 minutes. Besides familiarizing the infants with the target words, the training phase also presented the infants with a chance to link their looking to the blinking of the sidelights.

Immediately after the training phase, the test phase began. The experimenter started each test trial by pressing the “looking” button to initiate the blinking of the middle green light, and once the infant looked at the green light, the experimenter pressed the “looking” button, and one of the sidelights would begin blinking. Once the infant looked at the sidelight, the experimenter pressed the “looking” button to initiate the auditory stimuli. There were speakers on both sides of the test cabin for the playing of the auditory stimuli, and within each trial it was always the speaker at the same side as the blinking light that played the auditory stimuli. Once the infant looked away from the blinking sidelight, the experimenter pressed the “non-looking” button, and the control computer began to calculate the duration of “non-looking”. If the infant looked back to the blinking light within two seconds, then the same trial continued, and if the infant looked away for more than two seconds, then the current trial stopped. To initiate the next trial, the experimenter pressed the “looking” button, and the middle green light would start blinking. The same procedures as mentioned above applied for each test trial. There were eight trials in the test phase consisting of two types of trials that alternated. The “same trial” was the presentation of the concatenated bisyllabic tonal sequence sharing the same underlying tone as the target words in the carrier sentence, and the “different trial” was the presentation of the T3T3 bisyllabic sequences. In other words, for condition A in the test phase, the trials with concatenated /mama/ carrying a T2T3 and the trials with /mama/ carrying a T3T3 alternated, and for condition B, it was the trial with concatenated T3T2 and the trial with T3T3 alternating. For each participant, the first trial could either be a “same trial” or a “different trial”, and the orders of these two trials were counter-balanced across the participants. There was a maximum of 15 tokens for a single test trial.

Results and discussion

Before running the statistical tests, the video of each participant was recoded offline to correct possible misrecordings of the experimenter. After recoding, a mean looking time to the “same” trials and a mean looking time to the “novel” trials was calculated for each infant. The following statistical analysis will be based on the mean looking time to each trial type of each infant.

In order to see whether a discrimination effect between the “same” trials and the “different” trials exists in the test phase, a paired t test was carried out between the mean looking time of the “same” trial and “different” trial of each participant for each condition. The mean looking time of the “same” trial and of the “different” trial in both conditions fitted a normal distribution: $D_{\text{same condition A}}(16) = 0.16$, $p>0.1$, $D_{\text{different condition A}}(16) = 0.16$, $p>0.1$, $D_{\text{same condition B}}(16) = 0.16$, $p>0.1$, $D_{\text{different condition B}} = 0.18$, $p>0.1$. There is one outlier in condition A whose looking time to the different trial lay more than two standard deviations higher than average. After excluding this one single participant from condition A, I only found a significant difference between the looking times in condition B, $T_{\text{same-different}}(15) = -2.801$, $p<0.05$, but not in condition A, $T_{\text{same-different}}(14) = 0.23$, $p>0.1$. Figure 5.1 depicts the average of the mean looking time of the “same trial” and that of the “different” trial in both condition A and condition B.

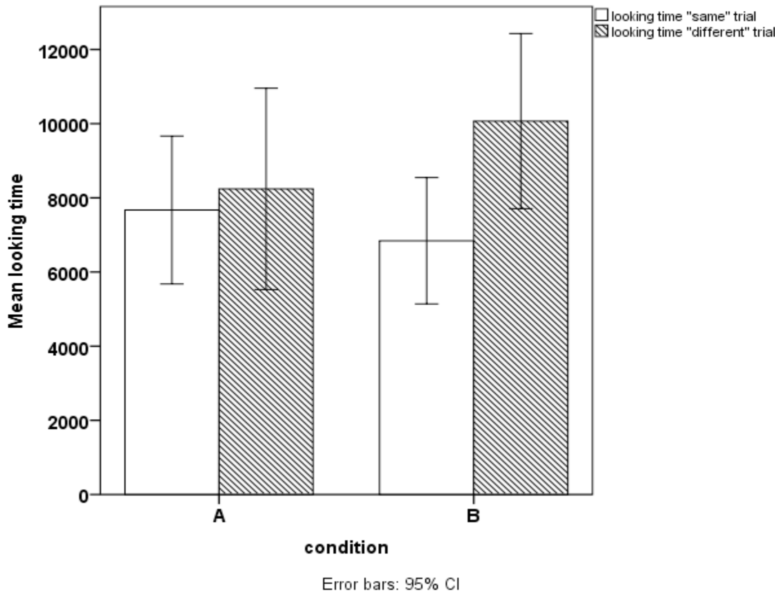


Figure 5.1 Mean looking time to the “same” trial and “different” trial by 9-month-old Dutch infants in condition A (N= 16) and condition B (N=16).

As demonstrated in the above figure, a significant discrimination effect of the 9-month-old Dutch infants could be observed only in condition B, in which they were trained with sentences with the naturally produced T3T2 embedded as the target word. Later in the test phase, they were able to discriminate between the underlying form of the heard target, namely a T3 concatenated with a T2, from the underlying form of a T3T3 sequence. For condition A, in which they were trained with sentences in which the naturally produced T2T3 target sequence was embedded, they were not able to discriminate the underlying form of the T2T3 form, namely the T2 concatenated with a T3 from the concatenated T3T3 sequence.

In a way, the current experiment simulated a language acquisition situation in which the infants were first presented with different repetitions of the same sentence, and then they were tested on their discrimination between underlying forms of the bisyllabic sequence that occurred in the sentence, and the underlying forms of another unheard bisyllabic tonal sequence. As in the real case of acquiring Mandarin, the underlying T3T3 sequence is sandhied to a surface form carrying T2T3, and the infants did not have the chance to encounter a naturally produced T3T3 sequence in its underlying form within a prosodic word. As a result, similar to the real language learning situation, the infants either heard a T2T3 within a prosodic word, or a T3T2 in a prosodic word, while the underlying form of the former sequence could be ambiguous. If the infants were to acquire T3 sandhi, they need to learn to link the surface T2T3 form to the underlying T3T3 form. In the current experiment, I also trained the infants with either the ambiguous condition, as in Condition A, or the unambiguous condition, as in Condition B. The Dutch infants were naïve listeners to Mandarin, and they had neither experience with nor knowledge of Mandarin lexical tones. Therefore, how they reacted to the “same” and “different” trials in the test phase, presumably reflect the acquisitional ease of discriminating between the underlying forms of different tonal sequences. We have seen that after exposure to natural Mandarin speech, the 9-month-old Dutch infants discriminated between the underlying T3T2 form and the underlying T3T3 form more easily than between the underlying T2T3 and T3T3 forms. The different discrimination pattern observed for T2T3-T3T3 and T3T2-T3T3 suggests that young language learners tend to accept the underlying form of T2T3 as equal to T3T3 while differentiating between T3T2 and T3T3. Hence, the pattern of T3 sandhi may also have an acquisitional motivation: if the T3T3 sequence were to be sandhied to a T3T2 rather than a T2T3, the learners may perceive the sandhied form and the underlying form as two different sequences.

5.3 EXPERIMENT 2 WITH 11-MONTH-OLD DUTCH INFANTS

Participants

32 healthy 10-11-month-old Dutch infants participated in the experiment (age range 10:16-11:29). All the infants were monolingual Dutch infants, and none of them had exposure to any tone language. Half of the infants were assigned to condition A and the other half were assigned to condition B. Another nine infants were tested but excluded from analysis due to crying (N=6), error of experimenter (N = 1), short attention (N = 1), and equipment failure (N = 1).

Prediction

Similar to the 9-month-old infants, I expect that the infants who participated in condition B will display a discrimination effect, while no clear discrimination effect will be observed for those participating in condition A.

Stimuli

Exactly the same stimuli as in Experiment 1 were used for the current experiment.

Procedure

Exactly the same procedure as in Experiment 1 was used in the current experiment.

Results and discussion

Before running the statistical test, the video of each participant was recoded offline to correct possible misrecordings of the experimenter. After recoding, a mean looking time to the "same" trials and a mean looking time to the "novel" trials was calculated for each infant. The statistical analysis will be based on the mean looking time to each trial type for each infant.

In order to see whether there was a discrimination effect between the "same" trials and the "different" trials in the test phase, a paired t test was conducted between the mean looking time of the "same" trial and the mean looking time of the "different" trial of each participant for each condition. The mean looking time of the "same" and "different" trials in condition A significantly deviated from a normal distribution: D_{same} (16) = 0.23, $p<0.05$, $D_{different}$ (16) = 0.22, $p<0.05$. However, the difference score between the mean looking time of the "old" trials and the mean looking time of the "novel" trials were normally distributed in both conditions: $D_{difference\ score\ condition\ A}$ (16) = 0.116, $p>0.1$, $D_{difference\ score\ condition\ B}$ (16) = 0.15, $p>0.1$. In Condition A, there was one participant whose difference score was more than two standard deviations away from the mean. After excluding the one single outlying participant from condition A, I carried out a one sample t test for each condition separately, taking zero as the test value. The difference score of neither condition showed a significant difference from 0: $T_{condition\ A}(14) = -1.43$, $p>0.1$, $T_{condition\ B}$ (15) = 1.11, p>0.1. Figure 5.2 depicts the difference score of condition A and condition B of the 11-month-old participants. For the ease of comparison with the 9-month-old infants, I also depict the mean looking time to the “same” trials and the mean looking time to the “different trials” of the 11-month-old infants in each condition in Figure 5.3.

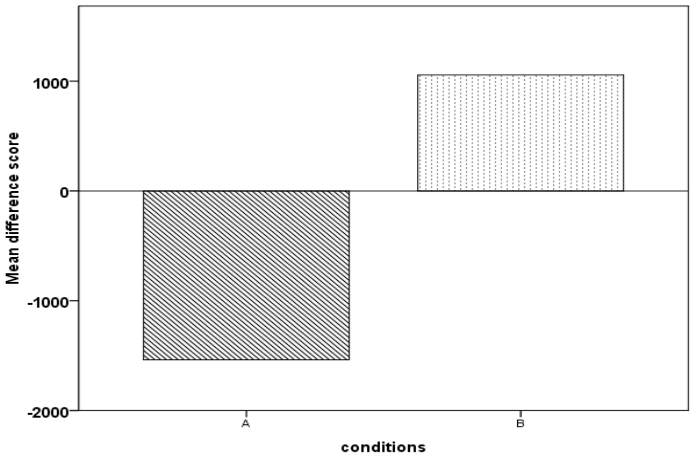


Figure 5.2 The difference score between the mean looking time of the “same” trial and the mean looking time of the “different” trial of the 11-month-old Dutch infants.

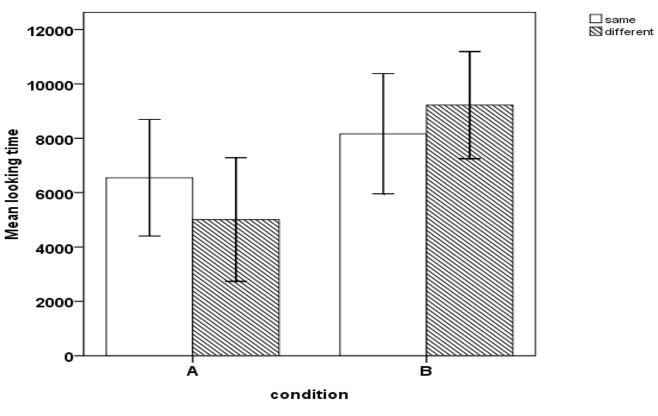


Figure 5.3 Mean looking time to the “same” trial and mean looking time to the “different” trial in condition A and condition B by 11-month-old Dutch infants.

Contrary to the expectations, the 11-month-old infants failed to show a discrimination effect in either condition A or condition B. In other words, for the discrimination between T3T3 and the other bisyllabic tonal sequences, they did not benefit from the preceding training phase, regardless of whether naturally produced T2T3 or T3T2 sequences were embedded. Yet, from Figure 5.3 and Figure 5.4, we could still see that the infants only listened longer to the unheard T3T3 sequence in the test phase in condition B. The mean looking time difference between the “same” trial and the “different” trial was 1056ms, meaning the infants listened for about one token more in the “different” trial, but this difference was not large enough to reach significance.

Two factors may have contributed to the failure observed with the 11-month-old infants. First, I tested only 16 participants in each condition, and though the difference score between the mean looking time to the “same” trials and that to the “different” trials fitted to a normal distribution, I acknowledge that the sample size was quite limited. Bear in mind that the mean looking time to the “same” trials and mean looking time to the “different” trials of the participants were still skewed, so it could be that a larger sample size is needed for generating more robust results.

Another possible factor is that I set up the experiment in the way that there was a three-minute training phase followed by a test phase consisting eight trials, which made the whole experiment last for about 6 minutes. It could be that the duration of the experiment was too long for the 11-month-old infants, and towards the end of the experiment, they became restless and hence failed to focus their attention on the stimuli. Yet, the head-turn preference procedure has been used to test 15-month-old (Boll-Avestisyan 2012) and 18-month-old (Gomez 2002) infants with a similar amount of training. Even though the 18-month-old infants are more mobile and get bored more easily compared to 11-month-old infants, the procedure still worked. Therefore, though it is a possibility that the infants were not focusing enough during the experiment, it is unlikely that the procedure alone was the main reason for the non-result we have found.

5.4 GENERAL DISCUSSION

In the current experiments, I tested 9 and 11-month-old Dutch infants with a setting that in a way simulated a natural language learning situation. These two age groups were chosen because in Chapter 3, they displayed a discrimination asymmetry of Mandarin T2 and T3 under the habituation-dishabituation paradigm. Importantly, both age groups only succeeded in the discrimination when they were habituated on T3 rather than T2, which suggests a possible order effect for the perception of T2 and T3 in early infancy, namely the first acquaintance with T3 enhanced the discrimination between T2 and T3.

In order to gain a more direct view of such an order effect, and to test the possible perceptual bias of Mandarin T3 sandhi reflected in early language acquisition in infancy, the training phase of the experiments in the current chapter were designed with two conditions. The target words in the two conditions were both composed of T2 and T3, and the only difference was that in condition A, the T2T3 sequence was the target word, while in condition B, the T3T2 sequence was the target word. In this case, one condition presented the infants with a confusing target word in terms of T3 sandhi (/mama/ carrying T2T3 in condition A), while the other condition presented the infants with non-confusing target word (/mama/ carrying T3T2 in condition B). In the test phase, the infants were tested on the underlying form of the target word presented in the training phase, together with the underlying form of the sandhied T3T3 sequence. Compared to the natural production of complete sentences in the training phase, in the test phase the infants were presented with the underlying forms of the tonal sequences generated by concatenating naturally produced citation tones. I expected that the training phase of the non-confusing condition would encourage the infants to discriminate between the underlying tonal form of the target word and the underlying tonal form of the T3T3, while the training phase of the confusing condition would lead the infants to erroneously accept the underlying form of the target word as identical to a T3T3 sequence. However, only the 9-month-old infants performed as I had expected, whereas the 11-month-old infants did not show any discrimination effect for either condition. Nevertheless, for the 11-month-old infants who were trained with sentences in condition B, they tended to looked longer to the unheard T3T3 form in the test phase, while no novelty effect whatsoever could be observed for those who participated in condition A.

The results, though with limited significance, suggest that when young language learners who have no knowledge of lexical tones or T3 sandhi are presented with naturally produced Mandarin speech, they are likely to accept a T2T3 as equal to a T3T3 whereas they tend to discriminate between a T3T2 and a T3T3. This pattern suggests that, when acquiring Mandarin, language learners may be biased to perceive T2T3 and T3T3 as identical, and the fact that T3 sandhi requires the first T3 in a T3T3 sequence to change to a T2 may have something to do with acquisitional ease. In other words, there is a natural tendency for infant language learners to accept the T2T3 as an implementation of T3T3. It is very likely that due to this natural perceptual bias, the T2T3 and T3T3 were merged together. On the contrary, T3T2 and T3T3 are distinctive from one another, and the alternation between the two sequences is easily noticeable. Hence, there is no motivation to merge T3T2 and T3T3. Moreover, if T3T3 were to be sandhied to other surface forms, such as T3T2, the learners may mistakenly perceive the surface form and the underlying form as two different representations.

Though very unlikely, there is still a chance that the discrimination effect that we have seen among the 9-month-old infants in condition B reflects a natural auditory perception preference, rather than being the result of the training phase. In other words, it could be that naturally, compared to a T3T2 sequence, the infants preferred to listen to a T3T3 sequence, and it is this natural preference that caused the longer listening times in the "different trials" in the test phase. If this were the case, then the training phase was irrelevant for the longer looking time of the T3T3 observed in the test phase. Yet, as I have argued in section 5.2, if the infants failed to link the concatenated T3T2 to the naturally produced T3T2 in the sentences presented in the training phase, then T3T2 and T3T3 were both new sound sequences to them. In this case, we would expect them to show an equal interest for the new sounds. Moreover, even if we accept that the looking time difference of the 9-month-old infants between the "same" trial and the "different" trial in condition B reflected a natural preference, the same interpretation should hold for condition A as well. Infants failed to display a preference difference between T2T3 and T3T3, which in a way also suggests that the infants did not consider T2T3 and T3T3 as different. To have a clearer picture of how the training phase may have influenced the perception of the infants, it is worth the efforts for future studies to test the auditory preference of T2T3 versus T3T3 and T3T2 versus T3T3 without the training phase, or with a training phase that familiarizes the infants with speech material that is irrelevant to Mandarin tones.

So far, we have seen the discrimination between T2 and T3 by native Dutch adults and native infants, as well as their discrimination between the underlying T3T3-T2T3 and T3T3-T3T2 sequences. For monosyllabic tone discrimination, both Dutch adults and Dutch infants benefited if they were first presented with T3(s) rather than first presented with T2(s). Dutch adults also discriminated more easily between T3T3 and T3T2 sequences than between T3T3 and T2T3 sequences. In the current chapter, I gathered more evidence for the possible bias for the perception of T3 sandhi from an acquisitional perspective, namely that after being familiarized with naturally produced Mandarin speech, the infants tend to equalize the underlying form of T2T3 and that of T3T3, while they stay sensitive to the difference between the underlying form T3T2 and that of T3T3. Until now, the results of all the tasks, despite that the procedure may differ substantially, quite consistently indicate that to some extent, T3 sandhi as it occurs now is a result of possible perceptual biases. In other words, the surface form of the sandhied T3T3, namely the T2T3, has the smallest perceptual distance to the underlying form T3T3, and for a substantial population the difference between T2T3 and T3T3 is not detectable. Therefore, to merge the T2T3 and T3T3 is a natural tendency.

In Chapter 4, we saw that under the habituation-dishabituation paradigm, Dutch infants discriminated a T2 from a T3 more easily than they discriminated a T3 from a T2. Considering the experiment procedure, it seems that with the same amount of variations, Dutch infants established a T3 category more easily than they established a T2 category. The ease of T3 categorization suggests that T3 is more specified and more distinctive in perception, and it might be that the distinctiveness of T3 excludes other tones easily. The difference in degree of specificity may be the reason why the discrimination asymmetry between T2 and T3 is triggered.

From the perspective of language evolution, it is proposed that the misperception of the acoustics in the transmission of speech contributed to the formation of phonological regularities (e.g. Ohala 1981, 1990, 1993, 1996, 2012).

Regarding Mandarin T2 and T3, even when there was a short pause between the two tones (see the experiments in Chapter 2 and Chapter 3), listeners tended to misperceive the T2-T3 contrast in a T2-T3 pair as identical, and they tended to accept the T2T3 as equal to the T3T3. Therefore, it is highly possible that in the process of language evolution, the misperception of the listeners have shaped the phonological rule. This means that if the listeners are not able to hear the difference between a T3T3 and a T2T3, in the long run, the misperception may lead the T2T3 and T3T3 to be merged into one single perceptual representation.

Chapter 6 cross-domain pitch perception in very early infancy

6.1 INTRODUCTION

In Chapter 5, I discussed the possible perceptual biases in the discrimination of Mandarin T2 and T3. In the current chapter, an acquisition perspective will be taken again. I will return to the unsolved issue of Chapter 4, namely why the 6-month-old Dutch infants failed to discriminate the Mandarin T2-T3 contrast. As demonstrated in recent research, at 6 months, the perceptual reorganization of lexical tones is not supposed to be complete, meaning that the infants should have been able to discriminate between non-native lexical tones (Harrison 2000, Mattock & Burnham 2006, 2008, Gao et al. 2011, Liu & Kager 2012). To find out why such an unexpected failure occurred will help us to better understand the innate limitations in the perception of the T2-T3 contrast. Moreover, by controlling the variations in the habituation phase, if we are able to observe the discrimination effect at an earlier age (than the 9 months shown in Chapter 4), we may accumulate stronger evidence about how T2 and T3 are perceived initially. As demonstrated in Narayan et al. (2010), the acoustical salience of a contrast may have a substantial influence on the early discrimination of speech sounds. Therefore, besides to look for more evidence about innate perceptual bias regarding T2 and T3, another focus of the current chapter is to understand to what extent the low salience of the T2-T3 contrast may hamper discrimination. Nevertheless, there are other factors that may influence the perception in early infancy. Before proceeding to the experiments adjusted to test the 6 month-old and younger infants on their discrimination of T2 and T3, I would like to discuss the possible reasons, as previously stated in Chapter 4, that may have contributed to the discrimination failure among Dutch 6-month-old infants.

First, the discrimination failure could be explained by 6-month-old infants' lack of auditory maturity, meaning that the acoustical difference between T2 and T3 is too subtle for the infants to distinguish. Dutch infants, however, are able to discriminate between Mandarin T1 and T4 (Liu & Kager 2012). With respect to the acoustical properties of Mandarin T2 and T3, compared to the T1-T4 contrast, two points are worth noting. First, the absolute pitch difference between T2 and T3 is smaller than that between T1 and T4, either at the onset, offset or the lowest point along the F0 contour. Second, T1 and T4 differ in pitch direction while T2 and T3 share a similar pitch direction. The difference between T2 and T3 resides in that T2 rises overall, while T3 dips before it rises. These properties of the T2-T3 contrast may have contributed to the failure of discrimination. If we were to accept that T2 and T3 are acoustically less salient, then either absolute pitch difference or the dynamic pitch direction may contribute to the lack of salience. Importantly, if we were to claim that it is the lack of acoustical saliency that hinders the infants from successful discrimination, we would not expect the infants to discriminate between sounds which demonstrate even more subtle acoustical differences.

Second, the observed failure may be due to the cognitive burden of normalizing between limited tokens of different speakers. Young infants may have limited ability to compensate the variations between different speakers' voices (see Housten & Jucszyk 2000, Singh, Morgan & White 2004, Rost & McMurray 2009). If the inability of speaker normalization was the explanation of the infants' failure to discriminate between the tones, then we would expect the infants to succeed in discrimination by reducing the habituation stimuli to productions of one single speaker. Regarding the normalization of the productions of one single speaker, two scenarios should be discussed separately. The first scenario is that the infants are habituated with one single token of one single speaker of one lexical tone, and are tested with one single token of another lexical tone. In this case, the infants are supposed to discriminate between the pure phonetic attributes of the Mandarin T2 and T3, with no normalization whatsoever being involved. The second scenario is that the infants are presented with multiple tokens of one single tone of one speaker, and are tested with one token of the habituated tone together with one token of the unheard tone. Compared to the first scenario, the second scenario may more closely resemble the perception of speech in real language acquisition situations. In real life, young infants encounter multiple productions of the same phonological category from the caregivers, and they need to learn to equate these different productions, and build up one single representation of the phonological category. In early infancy, though the infants may hear speech produced by more than one speaker, the speaker diversity may well be very limited. To summarize, if it was inability of speaker normalization that caused the 6-month-old Dutch infants' failure to discriminate between T2 and T3 in Chapter 4, then by reducing the habituation tokens to the productions of one single speaker, regardless of multiple tokens or repetitions of one single token, the infants should succeed in discrimination. If the infants are able to perceive the acoustical difference between Mandarin T2 and T3, they should by all means be able to display the discrimination effect under the single-token-habituation scenario. On the other hand, if the infants are only able to discriminate the lexical tones when the linguistic function of the lexical tones is highlighted, then they should fail to discriminate between T2 and T3 in the single-token-habituation scenario, but succeed in the multiple-token-habituation scenario.

Third, the discrimination failure could be accounted for by the possibility that the perceptual reorganization of lexical tones may occur earlier than 6 months, and the failure observed among the 6-month-old infants was due to the loss of sensitivity to specific tones. To test this hypothesis, we need to test younger infants who, based on the evidence accumulated so far, have certainly not lost their sensitivity to non-native contrasts yet.

To summarize, if we want (to be able) to pinpoint the factor that hinders the discrimination of the T2-T3 contrast, all the aforementioned possibilities need to be considered. Therefore, in the current chapter, 4-month-old Dutch infants are tested, who certainly have not completed the perceptual reorganization of lexical tones. To observe the role of acoustical salience, the infants were tested on stimuli that exhibit a much smaller acoustical difference. Moreover, to reduce the burden of speaker normalization, the stimuli in the habituation phase are limited to productions of one single speaker. In addition, to better link to the patterns observed among 6-month-old Dutch infants, the same infants were tested at both 4 months and 6 months using the same procedures and the same stimuli.

Considering that the phonetic attributes of lexical tones, namely pitch variations, are also the main phonetic attributes of music, to study the early perception of lexical tones from a domain-general perspective will give us a more complete picture of pitch perception in early infancy. The highly contrastive function of pitch in lexical tones makes it a suitable case to study how infants cope with subtle pitch differences across different domains. As we have seen in Chapter 3, native Dutch adults and native Mandarin adults demonstrated differences in the processing of music and lexical tones, and only Dutch listeners showed a significant correlation between the discrimination of musical melodies and the discrimination of lexical tones. To study how young infants perceive lexical tones and music may give us a more complete picture regarding the general auditory constraints in early lexical tone perception. Specifically, with regard to acoustical salience, music gives us the chance to test the infants with very subtle pitch differences, as the minimal pitch difference used in music is substantially more subtle than that used in speech. In music, one semitone—the pitch difference between adjacent piano keys—is the critical unit for pitch variations. In comparison, the offset difference between Mandarin T1 and T4 is more than ten semitones, and the difference between Mandarin T2 and T3 is five to six semitones. Thus, music allows us to generate pitch patterns that display much more subtle acoustical differences than lexical tones. In addition, it is possible to construct the musical melodies in a way that they resemble the pitch directions of T2 and T3—the rising and the dipping contour. By comparing the infants' discrimination of musical melodies and lexical tones with the same pitch direction, we have the chance to examine whether pitch direction *per se* is the reason that renders T2 and T3 not salient. Moreover, comparing the infants' performance on subtle musical pitch discrimination and Mandarin T2-T3 discrimination will allow us to examine the role of acoustical salience in early pitch processing. In sum, to test infants with musical stimuli allows us to study the early processing of subtle pitch differences, and to compare the infants' performance in speech tasks and their performance in music tasks will help reveal what acoustical cues might be critical for successful pitch discrimination in early infancy. Furthermore, how the infants perceive pitch across different domains will help us to better understand the initial status of human auditory perception.

Regarding the processing of music, previous research has demonstrated that infants show perception sensitivity in various experiments. It has been found that infants as young as 2 months old were able to distinguish a folksong played at home from another unheard folksong (Plantinga & Trainor 2009). At 6-8 months of age, infants also discriminate between music structures such as consonance and dissonance (e.g. Trainor & Schellenberg 1996), and the interruption of consonant intervals (e.g. Trainor 1997). 8-month-old infants were also able to detect an occasional note change in a ten-note melody while they treated the transposition of a melody as equal to the original melody (Trainor & Trehub 1992). More recently, in an ERP study, 6-month-old infants were tested with four-note melodies, and the infants were occasionally presented with another melody where the last note deviated from the original note. In correspondence to the deviant notes, a slow positive wave was found, indicating that infants were able to hear the difference between the original melody and the deviant melody (Tew et al. 2009). Tew et al.

(2009) also found that the neural mechanism involved in detecting the deviant note was different in adults and in 6-month-old infants. All these findings suggest that the young infant listeners possess substantial ability when it comes to processing the pitch patterns realized in music. Yet, the perception of music pitch *per se* has been largely neglected. Though in Trainor & Trehub (1992) and Tew et al. (2009) the infants showed a discrimination effect between different melodies, the authors intended to understand how the infants learn to process the combination of musical notes in a well structured way, but not how they discriminate musical pitches. Moreover, besides Plantinga & Trainor (2009) and Tew et al. (2009), the other studies tested infants older than 6 months; hence, how young infants process musical pitch variations remains unclear.

To my knowledge, except Mattock & Burnham who tested English and Chinese infants on their perception of Thai tones and non-speech analogues of these tones, so far no other studies have attempted to compare the early perception of speech pitch to the early perception of music. This topic, however, is of great importance to understanding how human beings are initially equipped to perceive pitch variations. The answers to some important questions are still vague. For example, are infants equally sensitive to the pitch patterns realized in music and in speech? Or do they display better pitch processing in one domain compared to the other? Is perception of musical pitch correlated to the perception of speech pitch, and if so, in what way? As I have shown in Chapter 3, for native Dutch adults, there was a significant positive correlation between their accuracy in detecting musical pitch difference and their accuracy in the discrimination of Mandarin lexical tones. Then, at least to some extent, for adult non-tone language listeners, the perception of lexical tones is restricted by a more domain-general sensitivity to pitch. For Mandarin listeners, on the other hand, no significant correlation was found between the discrimination of musical melodies and the discrimination of lexical tones. It seems that the phonemic function of lexical tones has made tone language listeners process them linguistically, with the perception of lexical tones no longer being restrained by general auditory sensitivity. However, as was mentioned in Chapter 3, it is not yet known how tone and non-tone language listeners develop their pitch processing. It is not clear whether the perception of lexical tone and the perception of musical pitch initially correlate—and it is the experience with tone languages that makes the perception of lexical tones independent from general auditory constraints—or if the perception of musical pitch and the perception of lexical tones fail to initially correlate. In this latter instance, it is the lack of the contrastive function of pitch that makes non-tone language listeners perceive lexical tones psycho-acoustically and hence become restricted by general sensitivity to pitch variations. To better understand the initial stage of pitch processing, it is necessary to test infants on their discrimination of lexical tones and musical melodies.

If we wish to observe when and how tone language listeners and non-tone language listeners begin to display differences in pitch perception, it is necessary to include native tone language infants in the scope of this study/dissertation. As mentioned earlier, infants up to six months of age do not display language-specific tonal perception. Therefore, testing infants younger than 6 months should delineate the initial state of pitch processing without influence from the native language. Therefore, by looking at the performance of Mandarin and Dutch infants, we may have the chance to examine whether the pitch perception of tone and non-tone language listeners differ initially, regardless of language experience. Comparing the music-lexical tone correlation of Dutch to Mandarin infants may help to clarify the issue of early domain-general pitch perception that was raised in the previous paragraph.

To shed light on the unsolved issues mentioned above, in the current chapter, native Dutch infants of 4 and 6 months are tested longitudinally on their discrimination of Mandarin T2 and T3, as well as their discrimination of musical melodies. The musical melodies were constructed to reflect relative and absolute pitch differences. Most importantly, in order to examine whether it was the pitch dynamics of T2 and T3 that rendered the perception difficult for young infants, the musical melodies that differed in relative pitch were constructed to have a similar pitch direction as Mandarin T2 and T3. Additionally, in the musical absolute pitch discrimination task, the infants were tested with melodies that only showed a very subtle absolute height difference. Regarding the lexical tone stimuli, in order to examine how the variations are normalized, each infant participated in both the single-speaker-single-token habituation scenario and the single-speaker-multiple-token habituation scenario. At 4 months of age, infants are assumed to be language-universal listeners; indeed, so far no evidence of language-specific listening has been found in terms of the phonological inventory at

4 months. In addition, the longitudinal design should maximally reduce possible individual differences, and observing the same infants at both 4 and 6 months should provide us the opportunity to delineate the developmental perception change, if any, in pitch perception in early infancy.

Mandarin infants were also tested on the single-speaker-single-token scenario of the Mandarin T2-T3 discrimination task and on the music absolute pitch discrimination task. Due to the limited access to Mandarin infants, only 6-month-old infants participated in the experiments.

A summary of the experiments is listed in Table 6.1.

Expt	N	Task	Tokens	Age	Language
1A	39	T2-T3	Single token	4m	Dutch
2A	37	T2-T3	Multiple tokens	4m	Dutch
3A	38	Musical relative pitch		4m	Dutch
4A	35	Musical absolute pitch		4m	Dutch
1B	33	T2-T3	Single token	6m	Dutch
2B	32	T2-T3	Multiple tokens	6m	Dutch
3B	32	Musical relative pitch		6m	Dutch
4B	31	Musical absolute pitch		6m	Dutch
1C	12	T2-T3	Single token	6m	Mandarin
4C	12	Musical absolute pitch		6m	Mandarin

Table 6.1 Summary of the experiments reported in the current chapter.

6.2 EXPERIMENT 1A discrimination of T2-T3 under the single-token-habituation scenario by 4-month-old Dutch infants

Participants

39 4-month-old, healthy full-term Dutch infants (age range 4:02-4:29) participated in this experiment. None of the infants had been diagnosed with hearing deficiencies, and no ear infections were reported for the tested infants. Among the 39 participants, 19 were female and 20 were male. Another two infant were tested, but were excluded from analysis due to crying.

Stimuli

I used the endpoint T2 and endpoint T3 tokens in the adult categorical perception experiment discussed in Chapter 2 as test stimuli. Presenting the infants with one single token of each tone provides us with the opportunity to test the infants on the pure acoustics of the Mandarin T2 and T3. Presumably, if the infants are able to perceive the acoustical difference between these two tones, then in the current scenario they should be able to demonstrate a discrimination effect in the test phase. For the ease of comparing the stimuli used in the musical melody discrimination tasks, I have converted the frequency of the stimuli into semitones, and Figure 6.1 depicts the pitch contours of the two tones in semitones.

The same visual stimuli used in the experiments discussed in Chapter 4 were used in the current experiment.

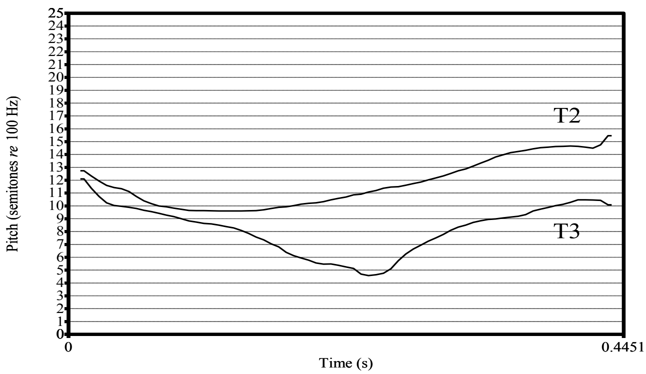


Figure 6.1 Pitch contours of the T2 and T3 stimuli used in the single-token-habituation scenario with frequency denoted in semitones.

Procedure

A procedure identical to the experiments discussed in Chapter 4 was used for the current experiment. The only difference is that in the habituation phase, infants were habituated with the repetition of only one single token of either T2 or T3. In the test phase, they were presented with one "old" trial, in which they heard the same token as played in the habituation phase, and one "novel" trial, in which they were presented with the other previously unheard tone. In theory, if the infants are able to hear the difference between the two tones in the test phase, their listening time to the novel trial should be longer than their listening time to the old trial due to the recovery of interest.

Results and discussion

The raw looking time for the "old" and "novel" trials was logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times conformed to a normal distribution: $D_{LGLT\ old}$ (39) = 0.112, p>0.1, $D_{LGLT\ novel}$ (39) = 0.071, p>0.1. The statistics hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was conducted between the LGLT of the "old" and "novel" trials. The pair failed to reveal a significant difference: $T_{old\text{-}novel}$ (38) = 0.387 p>0.1. In order to learn whether the infants performed differently when habituated on different tones, the infants were divided into two sub-groups according to the habituation tone. Within each sub-group I ran a paired t test with the LGLT of the old trial and the LGLT of the novel trial. Again, neither subgroup demonstrated a significant discrimination effect: T_{habT2} (17) = 0.435, p>0.1, T_{habT3} (20) = 0.19, p>0.1. Figure 6.2 gives the overall mean LGLT of the novel trial and mean LGLT of the old trial together with the mean LGLT of the novel trial and old trial separated by habituation tone. As can be seen in the Figure 6.2, compared to the looking time when presented with the old trial, the infants did not look longer at the visual stimuli when they were presented with the new trial, suggesting that they were not able to discriminate between Mandarin T2 and T3. The same result remained regardless of on which tone they were habituated.

The discrimination failure in the current experiment suggests that, for the 4-month-old infants, reducing the habituation tokens to one single token does not help the infants to pick up the acoustical differences between T2 and T3. Moreover, by 4 months of age, the infants are supposed to still be sensitive to non-native contrasts, hence the failure of these young infants could not be explained by the perceptual reorganization of lexical tones.

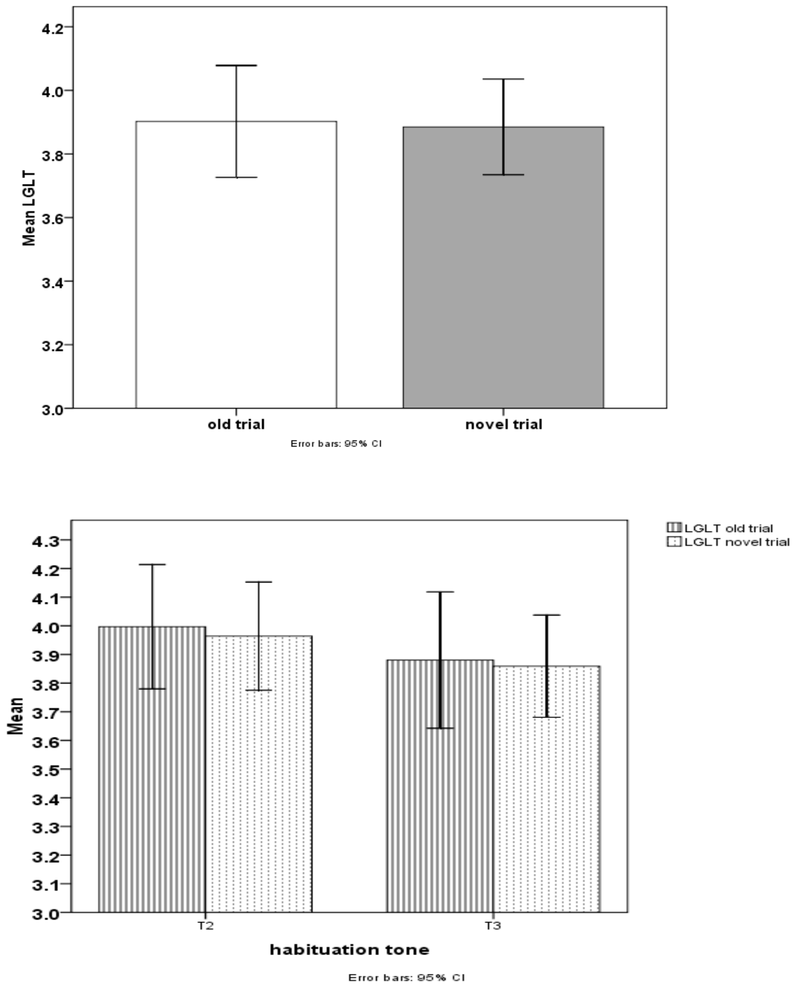


Figure 6.2 Under the single-token-habituation scenario, the overall mean LGLT of the old and of the novel trial (upper panel), and the mean LGLT of the old and novel trial separated by habituation tone (lower panel) of 4-month-old Dutch infants.

6.3 EXPERIMENT 2A discrimination of Mandarin T2-T3 under the multiple-token-habituation scenario by 4-month-old Dutch infants

Participants

37 4-month-old, healthy full term Dutch infants (age range 4:02-4:29) participated in this experiment. None of the infant had been diagnosed with hearing deficiencies, and no ear infections were reported for the tested infants. Among the 28 participants, 19 were female and 18 were male. Another four infants were tested but excluded from analysis due to crying.

Stimuli

In order to explore whether a multiple-token presentation in the habituation phase, which provides more linguistic hints concerning lexical tones, will help infants in building tonal categories, in the current experiment, multiple tokens of either T2 or T3 are used as habituation stimuli. A native Mandarin female speaker recorded multiple tokens of /ma/ carrying either T2 or T3, and 12 tokens that shared a similar intensity and duration were chosen as habituation stimuli. The mean duration of the syllable /ma/ carrying T2 was 431ms (SD = 20.5ms, range 396ms – 459ms), and the mean duration of /ma/ carrying T3 was 431 ms (SD = 20.7ms, range 400ms – 453ms).

For each tone, I generated four chains each containing three different tokens of the 12 tokens. Within each chain, the interstimulus interval was 1 second, which was the same as the interstimulus interval in the above experiment 1A. Within each trial of the habituation phase, the infants were presented with one of the chains, and all four chains occurred in the habituation phase. The occurrence order of the chains in the habituation phase was randomized for each infant. By manipulating the stimuli in this manner, I created a semi-random presentation of the multiple tokens of each tone in the habituation phase.

In the test phase, I selected another token of each tone produced by the same female speaker. These two tokens were manipulated in PRAAT to have an equal duration of 460ms. The visual stimuli were the same as in Experiment 1A.

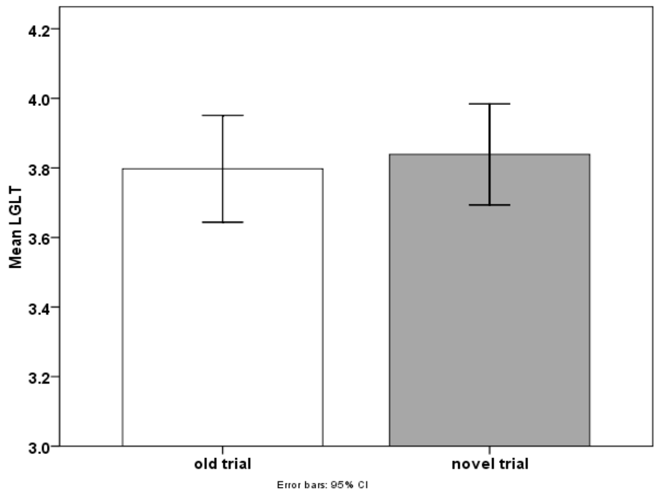
Procedure

The identical procedure as in Experiment 1A was used for the current experiment.

Results and discussion

After recoding, the raw looking times for the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times conformed to a normal distribution: $D_{LGLT\ old}$ (37) = 0.129, p>0.1, $D_{LGLT\ novel}$ (30) = 0.123, p>0.1. The statistics hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was carried out between the LGLT of the “old” and “novel” trials. The paired t test did not reveal a significant difference: $T_{old\text{-}novel}$ (36) = -0.547, p>0.1. In order to assess whether the infants performed differently when habituated on specific tones, I divided the infants into two sub-groups according to the habituation tone. Within each sub-group, I ran a paired t test with the LGLT of the old trial and the LGLT of the novel trial. Again, neither subgroup displayed a significant discrimination effect: T_{habT2} (15) = -1.09, p>0.1, T_{habT3} (20) = 0.43, p>0.1. Figure 6.3 illustrates the overall mean LGLT of the novel trial and the mean LGLT of the old trial, together with the mean LGLT of the novel trial and the old trial separated by habituation tone. As is evident in the figure, compared to the looking time when presented with the old trials, the infants failed to look longer at the visual stimuli when presented with new stimuli, suggesting that they were unable to discriminate between Mandarin T2 and T3. Moreover, the result stays the same regardless of on which tone they were habituated.



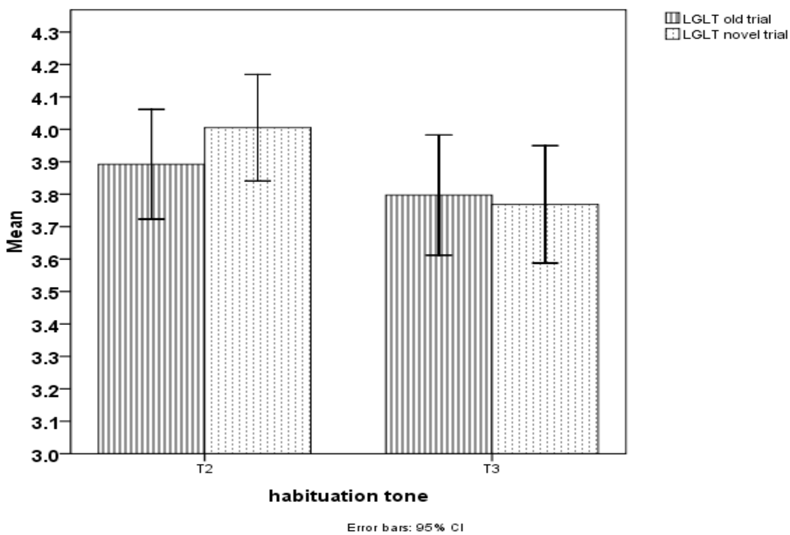


Figure 6.3 Under the multiple-token-habituation scenario, the overall mean LGLT of the old and of the novel trial (upper panel), and the mean LGLT of the old and novel trial separated by habituation tone (lower panel) of 4-month-old Dutch infants.

In the current experiment, when habituated on multiple tokens of Mandarin T2 and T3, the infants failed to show any discrimination effect of the contrast.

Moreover, the infants failed the discrimination regardless of the tone on which they were habituated. Compared to Experiment 1A, in the current experiment, the stimuli in the habituation phase provided more information of the linguistic function of the lexical tones. However, the infants were still unable to distinguish the two tones. Combining the results obtained in Experiment 1A and 2A, 4-month-old Dutch infants consistently failed to discriminate between Mandarin T2 and T3. No matter whether we encouraged them to process the pure acoustics of the two tones, or whether we offered them a more linguistic presentation of these two lexical tones, they were unable to perceive the difference between T2 and T3. At this age, the failure could not be accounted for by the perceptual reorganization of lexical tones. The consistent failure suggests that, most likely, the acoustic difference between T2 and T3 is not salient enough, either due to the dynamic pitch contours, or due to the absolute pitch difference between the two tones.

In the following two sections, I tested 4-month-old Dutch infants on their discrimination of musical melodies. Specifically, to find out whether the pitch direction *per se* is enough to account for the failure in the discrimination of the T2-T3 contrast, in Section 6.3 the infants were tested on two melodies that shared the same pitch direction with T2 and T3— a rising contour versus a dipping contour. In Section 6.4, in order to establish whether the 4-month-old infants are able to discriminate between acoustically very similar pitch patterns, they were tested with two melodies that shared exactly the same pitch contour but differed in one semitone in absolute height. By running these two tasks, we can examine what acoustical cues may hinder the successful discrimination of pitch patterns by young infants.

6.4 EXPERIMENT 3A Discrimination of musical relative by 4-month-old Dutch infants

Participants

38 healthy 4-month-olds (4:02-4:29) participated in the current experiment. None of the infant has been diagnosed with hearing deficiencies and no ear infections were reported for the tested infants. Among the 38 participants, 20 were female and 18 were male. Another two infants were tested but excluded from analysis due to crying.

Stimuli

16th notes of D4, E4, F4, and C4 were synthesized using a Nyquist script (for a description of Nyquist, see http://audacity.sourceforge.net/help/nyquist and http://www.cs.cmu.edu/~music/music.software.html). The notes were generated with a timbre of a piano, and all the notes were generated on the C4 (middle C) scale, along which the fundamental frequency of A4 equals 440Hz. Nyquist synthesized the notes with the default equal temperament, meaning that the pitch differences between adjacent notes, such as between D4 and #D4, and between E4 and F4, are equal to one semitone. After synthesizing the four single notes separately, D4, E4, F4 were concatenated to obtain a three-note melody— D4E4F4 (melody 1). D4, C4, and F4 were also concatenated to obtain another three-note melody— D4C4F4 (melody 2). Melody 1 and 2 were used as stimuli in this experiment. The pitch contours of melodies 1 and 2 are given in Figure 6.4. Comparing melodies 1 and 2, we can see that in terms of pitch contours, melody 1 has a rising pitch contour while melody 2 displays a dipping pitch contour. To put it differently, melodies 1 and 2 differ in terms of relative pitch. Importantly, melodies 1 and 2 were controlled to have identical initial and final pitch levels. Hence, if the infants were to discriminate between the melodies, they would not succeed by only paying attention to the initial or final portion of the melody, but they would have to perceive each melody of three notes as a whole and discriminate them based on the complete pitch contour. Melodies 1 and 2 both had durations of 830 ms. The visual stimuli were the same as in Experiment 1A.

The pitch direction of melody 1 and melody 2 were constructed in a way that made them similar to Mandarin T2 and T3 in terms of pitch direction. Comparing the infants' response to the Mandarin T2-T3 contrast and their response to melody 1 and melody 2 would help to reveal whether rising and dipping pitch directions in general hinder discrimination. If the infants succeed in discriminating between melody 1 and melody 2, then it serves as evidence that young infants are able to track the difference between a rising pitch contour and a dipping pitch contour. Then the failure that we have observed for the discrimination of T2 and T3 cannot be accounted for by pitch direction alone. If the infants also fail to discriminate between melody 1 and melody 2, then the difference between a rising contour and a dipping contour is very likely to be difficult in general for the infants to perceive.

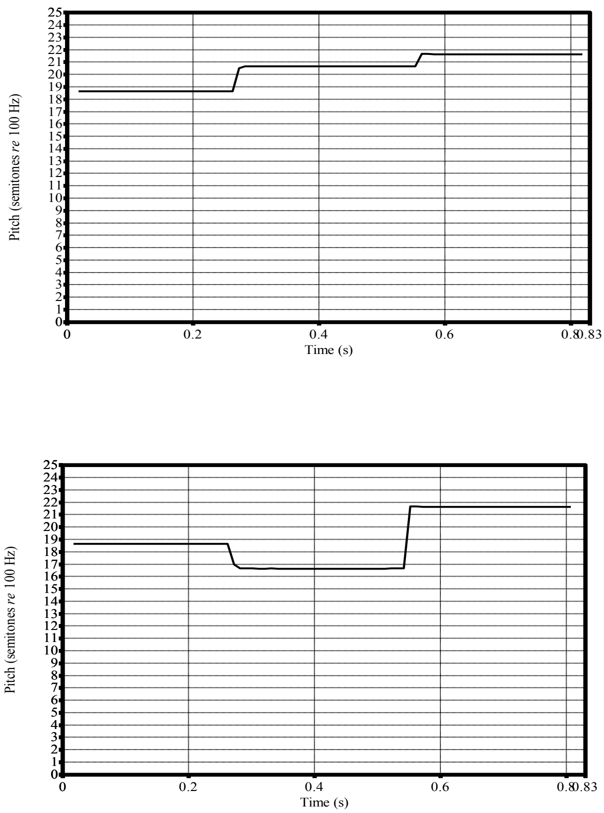


Figure 6.4 Pitch contours of melody 1 (upper panel) and melody 2 (lower panel) in semitones.

Procedure

The identical procedure as in Experiment 1A was used for the current experiment, with the exception being that the auditory stimuli were changed to melody 1 and melody 2.

Results and discussion

The raw looking times for the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times fitted a normal distribution: $D_{LGLT\ old}$ (38) = 0.106, p>0.1, $D_{LGLT\ novel}$ (30) = 0.095, p>0.1. The statistics hereafter are based on the log transformed looking time (LGLT). There was one outlier whose LGLT of the old trial was two standard deviations higher than the average, and this one single participant was excluded from further analysis. The following results were based on the remaining 37 participants.

A 1-tailed paired t test was carried out between the LGLT of the “old” and “novel” trials. There was a significant difference between the LGLT of the old trial and the LGLT of the novel trial: $T_{old\text{-}novel}(36) = -1.93$, $p<0.05$. Figure 6.5 provides the average LGLT of the novel and old stimuli. As can be seen in the figure, the mean LGLT of the novel trial was significantly longer than the mean LGLT of the old trial. It serves as evidence that in the test phase, the listening time of the infants recovered when the new trial was presented to them. Hence, the infants were able to discriminate the two melodies that differed in pitch direction.

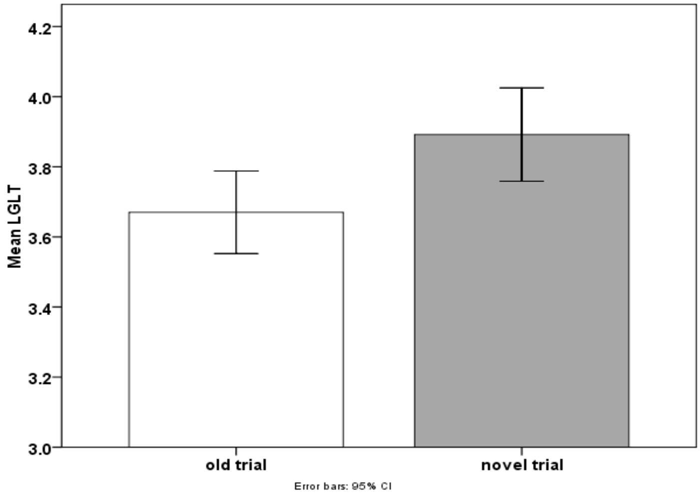


Figure 6.5 Mean LGLT of the old trial and novel trial by 4-month-old Dutch infants in Experiment 3A.

If we compare the performance of the infants on the three experiments mentioned so far, it seems that they discriminated the musical melodies more easily than the Mandarin lexical tones. If we compare the acoustical properties between the musical melodies and the two lexical tones, it can be seen that the middle note of the melodies differed for 5 semitones, while offsets of T2 and T3 also differed for about 5 semitones, and the maximum difference between T2 and T3 was about 6 semitones. Hence, the acoustical saliency of the stimuli in experiment 3A and those in experiment 1A and 2A did not greatly differ. Therefore, it does not seem likely that the failure in discriminating Mandarin T2 and T3 was due to acoustic subtlety. Moreover, the melodies were structured in a way that in terms of pitch dynamics, they were similar to Mandarin T2 and T3, meaning that melody 1 had a rising contour while melody 2 displayed a dipping contour. Therefore, it seems very unlikely that 4-month-old Dutch infants are unable to process the rising pitch direction and the dipping pitch direction efficiently. So far, it seems that the discrimination between a rising pitch contour and a dipping pitch contour is facilitated if the two contours are realized in music rather than in speech.

6.5 EXPERIMENT 4A Discrimination of musical absolute pitch by 4-month-old Dutch infants

The infants succeeded in discriminating musical melodies that differ for 5 semitones, and the magnitude of pitch differences between the musical melodies in Experiment 3A were similar to that observed between the T2 and T3. In order to obtain a clearer view on young infants' ability to detect subtle pitch differences, in the current experiment infants were presented with two musical melodies that had exactly the same pitch contour, but differed in absolute pitch height. As adult listeners tend to process relative pitch patterns in music rather than memorizing the exact absolute pitch (Trainor & Trehub 2006, Schellenberg & Trehub. 2003), it is assumed that the absolute pitch difference is more difficult to perceive. Therefore, the two melodies used in the current experiment are presumably acoustically less salient than the melodies used in Experiment 3A.

Participants

35 healthy, 4-month-old (4:02-4:29) native Dutch infants participated in the experiment. None of the infants had been diagnosed with hearing deficiencies, and

no ear infections were reported for the tested infants. Among the 35 participants, 17 were female and 18 were male.

Stimuli

Another pair of 16th notes, #D4 and #F4, were generated in Nyquist with the same default setting. Next, #D4, F4, and #F4 were concatenated to obtain another three-note melody—#D4F4#F4 (melody 3). In this case, the pitch contours of both melodies displayed rising directions. Importantly, the pitch intervals between each note were the same for both melodies—2 semitones between the first and the second note, and 1 semitone between the second note and the third note. In other words, the two melodies had identical relative pitch shapes but differed in one semitone in terms of absolute pitch height. Both the melodies had a duration of 830 ms. The F0 contours of the melodies are given in Figure 6.6. The visual stimuli were the same as in Experiment 1A.

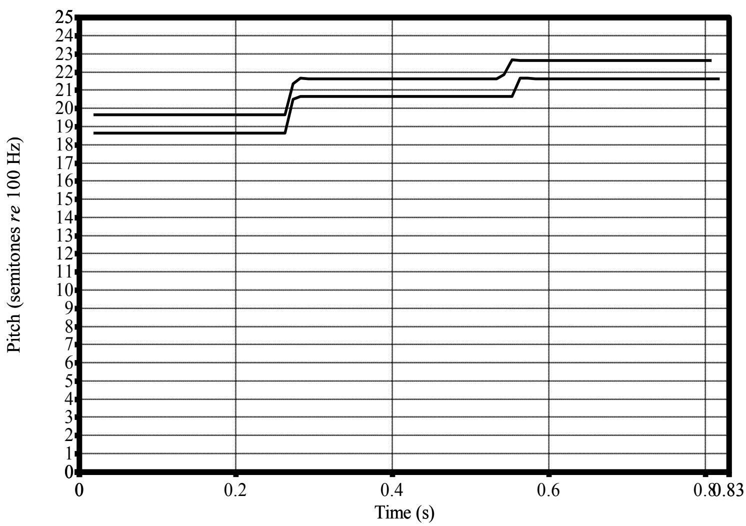


Figure 6.6 Pitch contour of melody 1 (lower line) and melody 3 (upper line).

Procedure

Precisely the same procedure as in Experiment 1A was used for the current experiment, with the exception being that the auditory stimuli were changed to melody 1 and melody 3.

Results and discussion

The raw looking times for the "old" and "novel" trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times fitted a normal distribution pattern: $D_{LGLT\ old}$ (35) = 0.072, p>0.1, $D_{LGLT\ novel}$ (35) = 0.120, p>0.1. The statistics reported hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was carried out between the LGLT of the "old" and "novel" trials. The test revealed a significant difference: $T_{old\text{-}novel}(23) = -1.77$, $p<0.05$. Figure 6.7 illustrates the average the LGLT of novel and old stimuli. From the figure, one can see that the mean LGLT of the novel trial was significantly longer than the mean LGLT of the old trial. It serves as evidence that the listening time of the infants increased when they heard a new melody in the test phase, and hence the infants were able to discriminate the two melodies that differed in absolute pitch.

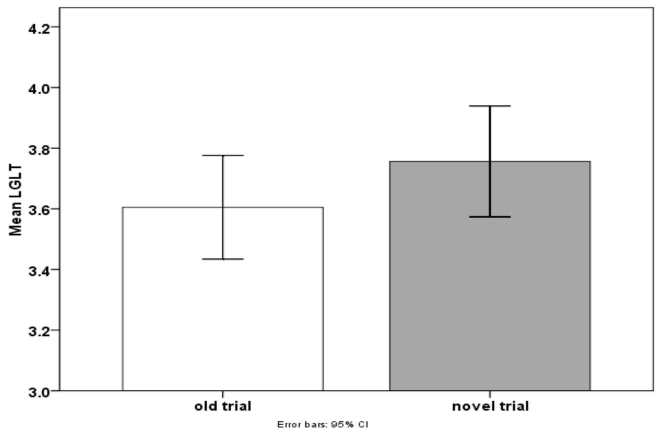


Figure 6.7 The mean LGLT of the old trial and novel trial by 4-month-old Dutch infants in Experiment 4A.

In the current experiment, we generated the stimuli so that the two melodies differed one semitone in absolute pitch, while their pitch direction was controlled to be exactly same. The acoustical difference between melody 1 and melody 3 was markedly smaller than that between the T2 and T3 in Experiment 1 and 2, and it was also much smaller than that between melody 1 and 2. Yet, the infants succeeded in discriminating the acoustical difference. Therefore, infants as young as 4 months possess quite remarkable abilities in processing pitch variations; they are able to detect a subtle pitch difference as small as one semitone. If we look at the infants' performance in Experiment 3A and Experiment 4A, we can see that neither the pitch direction, nor the acoustical saliency between T2 and T3, were enough to account for the failure observed in the Mandarin T2-T3 discrimination tasks. Hence, it seems that the failure in pitch perception is specific to lexical tones.

6.6 CROSS-EXPERIMENT ANALYSIS OF 4-MONTH-OLD DUTCH INFANTS

Among the infants that participated in the four experiments, 29 completed all four tasks. All 29 infants were tested on two separate days, with at least three days in between. On each test day, they participated in two tasks, one lexical tone task and one music task. The order of the music task and the lexical tone task was counterbalanced among the participants (14 of the infants participated in the music task first, and 15 participated in the lexical tone task first). For each infant, the order of the lexical tone task and the music task was the same on the two test days, meaning that if one infant first participated in a lexical tone task and then the music task on the first test day, then on the second test day, he or she was also first tested with the other lexical tone task and then the other music task. On test days, there was a short 2-3 minute break between the two tasks.

In each individual task, these 29 infants displayed the same discrimination pattern as the overall results we have seen in Experiment 1A, 2A, 3A, and 4A. A significant discrimination effect was observed for the musical relative pitch task (Experiment 3A) as well as for the musical absolute pitch task (Experiment 4A): $T_{Exp3A\ old\text{-}novel}\ (28) = -1.98$, $p<0.05$, $T_{Exp\ 4A\ old\text{-}novel} = -2.35$, $p<0.05$. However, they failed to discriminate between Mandarin T2 and T3, both in Experiment 1A and Experiment 2A: $T_{Exp1A\ old\text{-}novel}\ (28) = 0.45$, $p>0.1$, $T_{Exp2A\ old\text{-}novel}\ (28) = -0.1$, $p>0.1$.

In order to gain a better understanding of the effect of the different tasks, a repeated measures ANOVA was conducted. Tasks (Experiment 1A, 2A, 3A, 4A) and trial types (old trial, novel trial) were the within subject factors, while the LGLTs were dependent variables. The analysis indicated a marginal significant effect of trial types: $F_{trial\ type}\ (1, 28) = 3.84$, $p=0.06$. Tasks failed to reach significance level: $F_{tasks}\ (3, 26) = 1.44$, $p>0.1$. There was also a marginal significant interaction between trial type and tasks: $F_{interaction}\ (3, 26) = 2.49$, $p = 0.082$. Figure 6.8 plots the LGLT for the old and novel trials of the four tasks separately.

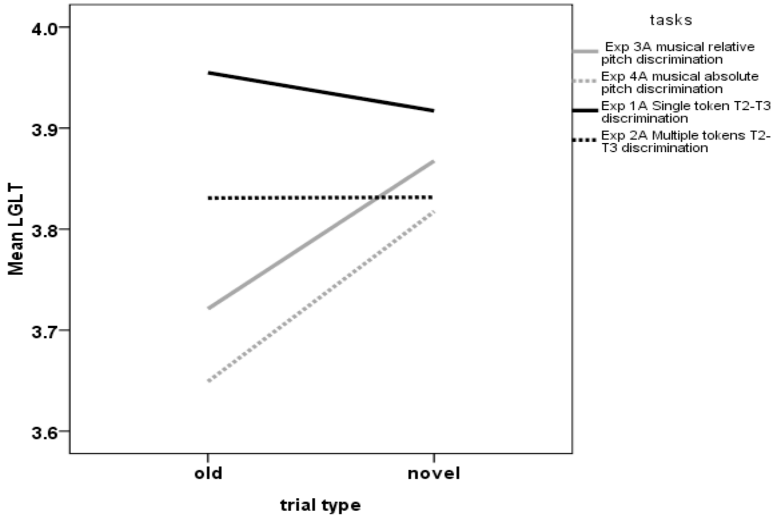


Figure 6.8 The mean LGLT for the old trial and novel trials in experiments 1A, 2A, 3A, 4A for 4-month-old Dutch infants.

As is evident in the above figure, the marginal significant effect of trial types is largely present in Experiments 3A and 4A, in which the infants listened significantly longer to the novel trial than the old trial. Moreover, the two grey lines, which represent the listening time in the two music experiments, could be distinguished from the two black lines, which represents the listening time in the two Mandarin tone experiments. Only in the music experiments (the two grey lines) did the infants demonstrate a clear recovery of listening time when presented with the novel trial, which indices successful discrimination. Based on Figure 6.8, it is very clear that the 4-month-old Dutch infants discriminated the music melodies more easily than the lexical tones.

It is important to bear in mind that in Experiment 3 I have generated the two melodies according to the pitch direction of Mandarin T2 and T3, hence the failure in Mandarin T2-T3 discrimination cannot be due to difficulties in processing a rising contour versus a dipping contour. Moreover, in terms of acoustical salience, the two melodies in Experiment 4 only differed by one semitone, which was much smaller than the acoustical difference between Mandarin T2 and T3. Thus, young infants were able to track very subtle pitch differences, and it is very unlikely that they are not able to hear the pitch difference *per se* between T2 and T3.

The consistent failure of the 4-month-old infants gives clear evidence that the T2-T3 contrast is not salient in perception. Comparing the findings in Liu & Kager (2012), in Chapter 2 and Chapter 3, we can safely draw the conclusion that compared to the T1-T4 contrast, the T2-T3 contrast is less salient in perception. Hence, it is very likely that the sandhi rule targets at the less perceptible tones, as the alternations between these tones may well be unnoticeable. However, the reason as to why the 4-month-old infants consistently fail to discriminate between T2 and T3 remains a mystery. Unfortunately, at this point I cannot give a satisfactory explanation. It could be that the continuity of the pitch contours of the lexical tones makes the tones difficult to discriminate. It is also possible that the segmental information distracted the infants from paying attention to the pitch information. To test these hypotheses, future research should test the infants with sung syllables of the same pitch contour of the melodies that I generated in Experiments 3A and 4A. Moreover, it is important to test infants with the pitch contours of Mandarin T2 and T3 while the segmental information is eliminated.

The data collected regarding the 29 infants was also analyzed to discern whether any correlation existed between the discrimination of musical melody and the discrimination of lexical tones in 4-month-old infants. In the visual fixation paradigm, each individual infant's responses were measured in a categorical way, meaning that they either succeeded or failed in discriminating between the two sounds. Considering the rationale behind the experimental paradigm, concerning the successful cases, the infants' listening times to the novel trial should be longer than their listening time to the old trial, while the reversed pattern indicates the failed cases. Therefore, for each task, the infants who obtained longer listening times for the novel trial than for the old trial were denoted as "success", while those who maintained shorter or equal listening times for the novel trial were denoted as "failure". In Experiment 3A and 4A, 20 out of the 29 infants listened longer to the novel trial. In Experiment 1A, 13 out of 29 infants listened longer to the novel trial. In Experiment 2A, 14 of 29 infants listened longer to the novel trial than to the old trial. For all the possible combinations of one music and one lexical tone task, a Chi-square test was carried out with the number of "successes" and "failures" of each task. None of the tests demonstrated a significant effect: between music relative pitch discrimination (Experiment 3) and single token T2-T3 discrimination (Experiment 1), $\chi^2(1) = 0.001$, $p>0.1$, $\phi = 0.05$, $p>0.1$; between relative pitch music discrimination (Experiment 3) and multiple token habituation T2-T3 discrimination (Experiment 2), $\chi^2(1) = 1.16$, $p>0.1$, $\phi = 0.20$, $p>0.1$; between absolute music pitch discrimination (Experiment 4) and single token T2-T3 discrimination (Experiment 1), $\chi^2(1) = 0.61$, $p>0.1$, $\phi = -0.15$, $p>0.1$; and between absolute music pitch discrimination (Experiment 4) and multiple tokens T2-T3 discrimination (Experiment 2), $\chi^2(1) = 0.08$, $p>0.1$, $\phi = 0.05$, $p>0.1$. There was no significant correlation between the two music tasks, χ^2 (1) = 1.10, $p>0.1$, $\phi = 0.19$, and neither was there significant correlation between the two lexical tone tasks, χ^2 (1) = 0.04, $p>0.1$, $\phi = -0.04$, $p>0.1$. For each combination of two tasks, neither is there significant correlation:

Regarding the non-significant correlation (as demonstrated by phi coefficient) between the discrimination of musical melodies and the discrimination of lexical tones among 4-month-old Dutch infants, several issues merit discussion. First, it must be acknowledged that the method used in the aforementioned experiments was quite different from the discrimination tasks used for testing adults in Chapter 3. Only indirect measures, such as listening time, could be used as evidence to infer infants' perception of sound stimuli. Moreover, each infant could only be tested for the discrimination of one or two sound pairs. In infant studies, within one age group, it is practically impossible to test the discrimination of the same tonal contrast realized on multiple tone-carrying segments. As a result, it is not possible to derive an "accuracy rate" in discrimination as it is usually done with adult participants. Therefore, it could be that the currently available paradigms for testing infants limited the sample size, and hence failed to reveal the correlation between the perception of music and lexical tones. Second, the behavior of infants is not as predictable as adults. It could be that one single infant was very attentive to one task while he or she was distracted during another task. Though I have tried to eliminate individual differences as much as possible by testing the same infants on different tasks, it is still possible that the performance of a child in different experiments did not faithfully reflect their perception. Lastly, in the current study, only 29 infants finished all four tasks. A larger sample size may compensate for the variations in infants' behavior mentioned above. The Chi-square test analyzes the correlation between categorical variables and a large sample size would give more robust results.

To summarize, the findings concerning the 4-month-old Dutch infants were three-fold. First, the 4-month-old Dutch infants failed to discriminate Mandarin T2 and T3, regardless of whether they were habituated on multiple tokens or single token of one tonal category. Second, the infants succeeded in discriminating both relative musical pitch differences and absolute musical pitch differences. Third, no evidence of a correlation between music and lexical tone perception has been found.

Seeing the experiments discussed in Chapter 4, and the findings of the aforementioned experiments, it is still hard to infer the developmental pattern of Dutch infants between 4 and 6 months. What is also unclear is whether habituating 6-month-old infants on one single speaker's voice would help them in discriminating the Mandarin T2-T3 contrast. If the aforementioned adjustment of the habituation stimuli improves the performance of the 6-month-old Dutch infants, we may have the chance to observe the categorization and discrimination of the Mandarin T2-T3 contrast at the earliest age when infants start displaying some sensitivity to these tones. One step further, if the same discrimination asymmetry could be found as among the 9 and 11 month-old Dutch infants in Chapter 4, we will get very persuasive evidence about the innate bias in perceiving these two tones.

Additionally, so far no previous studies have tried to reveal how the perception of relative musical pitch patterns and the perception of absolute musical pitch patterns may have changed between 4 and 6 months. In Chapter 3, I found a significant positive correlation between the discrimination of lexical tones and the discrimination of musical melodies among native Dutch adult listeners. However, up to now, for the 4-month-old infants, no evident correlation between the music perception and lexical tone perception has been found. Considering that the acquisition process of infants is the process that they gradually approach the same perception as native adult listeners, it is expected that at some point the Dutch infants or children show some correlation between lexical tone and music perception. It might be the case that, even with the current visual fixation paradigm, the older infants show at least some tendency towards a significant correlation between the discrimination of lexical tones and the discrimination of musical melodies.

In order to shed light on the unsolved issues, in the next section, I proceed with testing the 6-month-old Dutch infants using the same music and lexical tone tasks in an attempt to answer the following four questions: would presenting the infants with one single voice help them in discriminating Mandarin T2 and T3? Would they show confusion asymmetry between T2 and T3? Will the 6-month-old infants perceive the musical melodies differently compared to the 4-month-old infants? Concerning the 6-month-old infants, is there any correlation between the discrimination of lexical tones and the discrimination of musical melodies?

6.7 EXPERIMENT 1B Discrimination of Mandarin T2-T3 under the single-token habituation scenario by 6-month-old Dutch infants

Participants

33 healthy 6-month-old (6:01-7:02) Dutch infants participated in the experiment. None of the infants had been diagnosed with hearing deficiencies, and no ear infections were reported for the tested infants. Among the 33 infants, 16 were male and 17 were female.

Stimuli

The same stimuli used in Experiment 1A were used in the current experiment.

Procedure

The same procedures used in Experiment 1A were used in the current experiment.

Results and analysis

The raw looking times for the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking time of the novel trial and old trial fitted a normal distribution: $D_{LGLT\ novel}$ (33) = 0.1, p>0.1, $D_{LGLT\ old}$ (33) = 0.09, p>0.1. The statistics reported hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was conducted between the LGLT of the “old” and “novel” trials. The test did not reveal a significant difference between the trials:

$T_{old\text{-}novel}$ (32) = -1.27 p>0.1. In order to ascertain whether the infants performed differently when habituated on specific tones, I divided the infants into two sub-groups according to the habituation tone. Within each sub-group, a paired t test was conducted again with the LGLT for the old trial and novel trial. Again, neither subgroup demonstrated a significant discrimination effect: T_{habT2} (16) = -0.342, p>0.1, T_{habT3} (15) = -1.37, p>0.05. Figure 6.9 depicts the overall mean LGLT of the novel trial and the average LGLT of the old trial, alongside the mean LGLT of novel trial and old trial separated by habituation tone. As can be derived from the figure, compared to the looking time when presented with the old trials, the infants did not look longer at the visual stimuli in the novel trial, which suggests that they were not able to discriminate between Mandarin T2 and T3. The same result remained regardless of on which tone they were habituated. Nevertheless, there was a very slight tendency for the 6-month-old Dutch infants to discriminate between Mandarin T2 and T3 if they were habituated on T3.

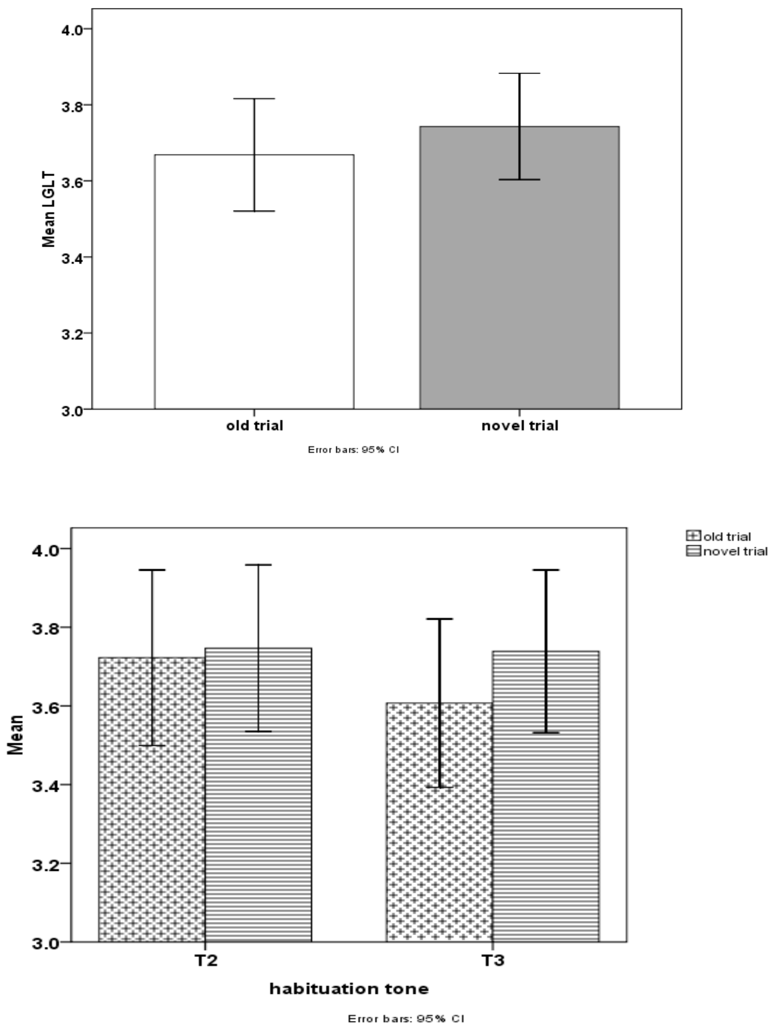


Figure 6.9 Under the single-token-habituation scenario, the overall mean LGLT for the old and for the novel trial (upper panel), and mean LGLT for the old and novel trials separated by habituation tone (lower panel) for 6-month-old Dutch infants.

6.8 EXPERIMENT 2B: Discrimination of Mandarin T2-T3 in the multiple-token habituation scenario by 6-month-old Dutch infants

Participants

32 healthy, 6-month-old (6:01-6:26) Dutch infants participated in the experiment. None of the infants had been diagnosed with hearing deficiencies, and no ear infections were reported for the tested infants. Among the 32 infants, 16 were male and 16 were female.

Stimuli

The same stimuli used in Experiment 2A were used in the current experiment.

Procedure

The same procedures used in Experiment 2A were used in the current experiment.

Results and analysis

The raw looking times for the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times conformed to a normal distribution pattern: $D_{LGLT\ old}\ (32) = 0.091$, $p>0.1$, $D_{LGLT\ novel}\ (32) = 0.102$, $p>0.1$. The log transformed looking time of the novel trial also fitted a normal distribution in the Shapiro-Wilk test: $W_{LGLT\ novel}\ (32) = 0.975$, $p>0.05$. The statistics reported hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was carried out between the LGLT of the “old” and “novel” trials. The test revealed a significant difference: $T_{old\text{-}novel}\ (31) = -1.96$ $p<0.05$. In order to ascertain infants’ performance when habituated on specific tones, I divided the infants into two sub-groups according to the habituation tone. Within each sub-group, I again ran a paired t test with the LGLT for the old trial and novel trials. Only those infants habituated on T3 demonstrated a significant discrimination effect: $T_{habT2}\ (15) = -0.92$, $p>0.1$, $T_{habT3}\ (15) = -1.88$, $p<0.05$. Figure 6.10 gives the overall mean LGLT for the novel trial and the average LGLT for the old trial, along with the mean LGLT for novel and old trials separated by habituation tone. As is evident in the figure, compared to the looking time when presented with the old trials, the infants looked at the visual stimuli longer when they were presented with new stimuli, which suggests that they discriminated between Mandarin T2 and T3. Yet, the overall significant results came principally from those infants who were habituated on T3, while those who were habituated on T2 did not show a discrimination effect.

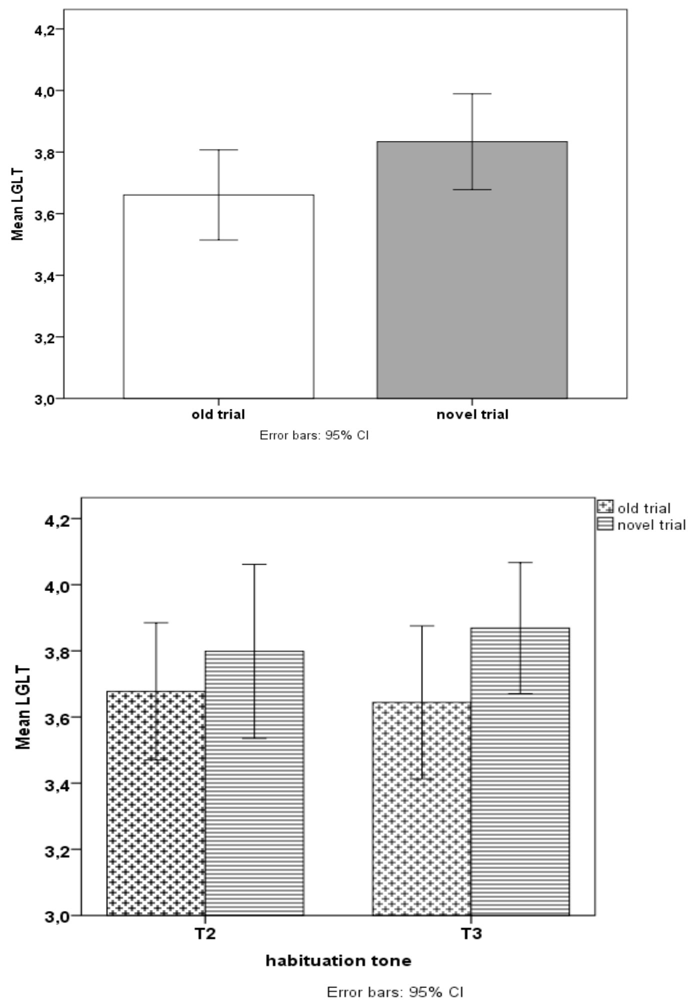


Figure 6.10 Under the multiple-token-habituation scenario, the overall mean LGLT for the old and for the novel trial (upper panel), and the mean LGLT for the old and novel trials separated by habituation tone (lower panel) for 6-month-old Dutch infants.

The most important finding of the 6-month-old infants is that, in the multiple-token scenario, they succeeded in discriminating Mandarin T2 and T3 only if they were habituated on T3. In Chapter 4, we saw that when habituated with two tokens of two speakers, 9 and 11-month-old Dutch infants also succeeded in discriminating T2 and T3 only if they were habituated on T3. Among infants of different ages, it was consistently easier to discriminate a T2 from a pre-given T3 than vice versa. This served as strong evidence that there is a perceptual bias that favors the formation of a T3 category. Thus far, 6 months is the youngest age that Dutch infants displayed at least some sensitivity to the Mandarin contrast T2 versus T3, and from this age on, they consistently demonstrated a discrimination asymmetry between T2 and T3. The Dutch infants in the current study had no exposure to Mandarin whatsoever, and the young infants had barely established any language-specific phonological knowledge, hence their discrimination pattern reflected the natural ease in speech processing. The discrimination patterns found among Dutch infants corroborates the hypothesis that T3 forms a stricter perceptual space compared to T2. It is very likely that the distinctiveness of T3 facilitates the detection of a tone that falls outside its category. This discrimination asymmetry may have triggered T3 sandhi to occur in the asymmetrical way as it does presently.

For the 6-month-old infants, as opposed to the 4-month-old infants in Experiment 2A, it seems that presenting them with multiple tokens of one single speaker assisted them in establishing a tonal category, and hence facilitated discrimination. If they would have discriminated between T2 and T3 in the single-token discrimination scenario, then they needed to be able to perceive the pure acoustic differences between the pitch contour of T2 and T3. The 6-month-old infants' failure in Experiment 1B suggested that the acoustics of T2 and T3 were not salient enough to encourage successful discrimination. On the other hand, presenting the tokens in a more linguistic manner—to habituate them with multiple productions of the same tone—helped the infants in establishing the tonal representations successfully. In Chapter 4, we learned that when 6-month-old Dutch infants were habituated on two tokens of two speakers, the infants failed to discriminate between Mandarin T2 and T3, regardless of on which tone they were habituated. Taking into consideration that 6-month-old infants succeeded in discriminating T2 and T3 in the multiple-token scenario, it seems that speaker normalization was still a formidable task for 6-month-old infants. In general, accurate perception of the acoustics of T2 and T3 was still quite difficult for the Dutch 6-month-old infants. Meanwhile, presenting them with stimuli in a linguistic manner assisted them in extracting the acoustic cues that define a tone, which in turn facilitated discrimination. Comparing these findings to the findings of Chapter 4, it seems that at 6 months of age, the infants are unable to compensate for speaker variation.

If we look at the Dutch infants' development between 4 and 6 months, it is evident that 4-month-old Dutch infants were unable to discriminate Mandarin T2 and T3, no matter in which scenario they participated. However, by 6 months of age, the infants succeeded in discriminating T2 and T3 if they were habituated on multiple tokens of the same tone. Hence, tracking the pure acoustics of T2 and T3, as in Experiment 1A and 1B, seems to remain difficult for Dutch infants. Yet by 6 months of age, Dutch infants' ability to extract a tonal category based on various tokens improved, and with the encouragement of within-category variations, they became more aware of the linguistic functions of the lexical tones. Hence, it can be speculated that the linguistic processing of speech sounds improves between 4 and 6 months.

6.9 EXPERIMENT 3B Discrimination of musical relative pitch by 6-month-old Dutch infants

Participants

32 healthy, 6-month-old (6:01-7:02) Dutch infants participated in the experiment. None of the infants had been diagnosed with hearing deficiencies, and none of the tested infants were reported to have ear infections. Among the 32 infants, 16 were male and 16 were female.

Stimuli

The same stimuli used in Experiment 3A were used in the current experiment.

Procedure

The same procedure used in Experiment 3A was used in the current experiment.

Results and discussion

The raw looking times for the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking time of the novel trial and old trial fitted a normal distribution: $D_{LGLT\ novel}$ (32) = 0.11, p>0.1, $D_{LGLT\ old}$ (32) = 0.12, p>0.1. The statistics reported hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was carried out between the LGLT of the “old” and “novel” trials. The test revealed a significant difference: $T_{old\text{-}novel}$ (31) = -2.82, p<0.01. Figure 6.11 plots the mean LGLT of the old and novel trial in Experiment 3B. From the figure, we can see that the infants listened significantly longer in the novel trial than in the old trial, which serves as evidence that they discriminated between melody 1 and melody 2.

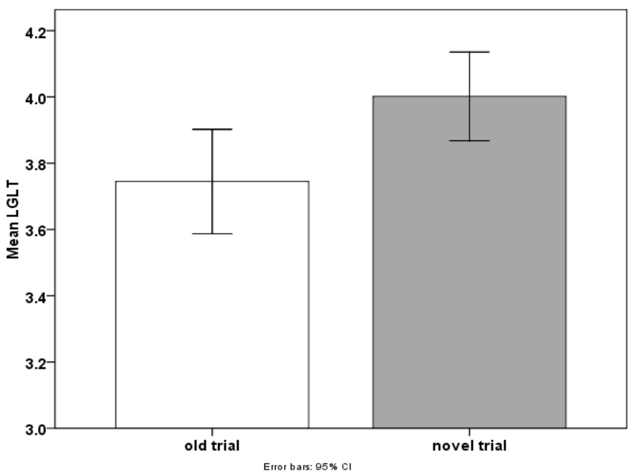


Figure 6.11 The mean LGLT for the old trial and the novel trial for 6-month-old Dutch infants in Experiment 3B (musical relative pitch).

In reference to Experiment 3A, it is evident that the Dutch infants were sensitive to the pitch direction change at both 4 and 6 months. Moreover, 6-month-old infants demonstrated a stronger discrimination effect as compared to the 4-month-old group. Therefore, the relative musical pitch difference seems to be quite salient for infants from very early on, and the pitch direction change realized in music is highly detectable for infants.

6.10 EXPERIMENT 4B Discrimination of absolute musical pitch by 6-month-old Dutch infants

Participants

31 healthy, 6-month-old (6:01-6:26) Dutch infants participated in the experiment. None of the infants had been diagnosed with hearing deficiencies, and no ear infections were reported for any of the tested infants. Among the 31 infants, 15 were male and 16 were female.

Stimuli

The same stimuli used in Experiment 4A were used in the current experiment.

Procedure

The same procedure used in Experiment 4A was used in the current experiment.

Results and analysis

The raw looking times for the "old" and "novel" trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking time of the novel and old trials fitted a normal distribution: $D_{LGLT\ novel}$ (31) = 0.09, p>0.1, $D_{LGLT\ old}$ (31) = 0.09, p>0.1. The statistics reported hereafter are based on the log transformed looking time (LGLT). There was one infant whose LGLT for the old trial was two standard deviations higher than the mean LGLT of the old trial, and this infant was excluded from further analysis.

A 1-tailed paired t test was conducted between the LGLT of the "old" and "novel" trials. The test revealed a significant difference: $T_{old\text{-}novel}$ (29) = -2.38, p<0.01. Figure 6.12 plots the mean LGLT for the old and novel trials in Experiment 4B. From the figure, it is clear that the infants listened significantly longer to the novel trial than to the old trial, which serves as evidence that they discriminated between melody 1 and melody 3.

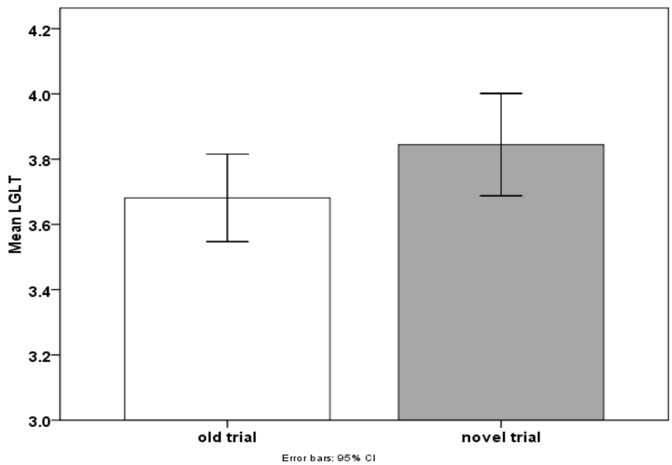


Figure 6.12 The mean LGLT for the old trial and the novel trial by 6-month-old Dutch infants in Experiment 4B (musical absolute pitch).

In the test phase of Experiment 4B, the 6-month-old Dutch infants listened significantly longer to the novel trial than to the old trial, which indicated that they were able to discriminate between the two musical melodies that differed with only one semitone in absolute pitch. Compared to the performance of 4-month-old Dutch infants in Experiment 4A, one can see that the Dutch infants were sensitive to the absolute pitch difference at both 4 and 6 months of age. Moreover, as was the case for the two melodies that differed in relative pitch, 6-month-old infants displayed a stronger discrimination effect than the 4-month-old infants.

6.11 CROSS-EXPERIMENT COMPARISON OF 6-MONTH-OLD DUTCH INFANTS

Among the infants that participated in the four experiments, 30 completed all four tasks. All 30 infants were tested on two separate days, with at least three days in between. On each test day, they participated in two tasks, one lexical tone task and one music task. The order of the music task and the lexical tone task was controlled among the participants (14 of the infants participated in the music task first, and 16 participated in the lexical tone task first). For each infant, the order of the lexical tone task and the music task was the same on the two test days, meaning that if one infant first participated in a lexical tone task and then in the music task on the first test day, then on the second test day, he or she was also first tested with the other lexical tone task and then the other music task. On test days, there was a short 2-3 minute break between the two tasks.

For these 30 infants, a paired t test (1 tailed) comparing the LGLTs of the old and novel trials revealed the same pattern found in Experiments 1B, 3B, 4B. However, only a marginal significant difference between the LGLT of the old trial and the LGLT of the novel trial could be observed for Experiment 2B: $T_{\text{old-novel Exp 1B}}$ (29) = -0.55, p>0.1; $T_{\text{old-novel Exp 3B}}$ (29) = -3.16, p<0.01; $T_{\text{old-novel Exp 4B}}$ (29) = -1.71, p<0.05; $T_{\text{old-novel Exp 2B}}$ (29) = -1.59, p = 0.062. The reason that the data from Experiment 2B failed to show significance is that two infants who were habituated on T3 failed to finish one of the other three tasks. However, if we only look at the infants who were habituated on T3 in Experiment 2B, a significant discrimination effect does not change: $T_{\text{old-novel}}$ (13) = -2.24, p<0.05.

Towards the end of better understanding the effect of different tasks, a repeated measures ANOVA was conducted with the data of these 30 infants. Tasks (Experiment 1B, 2B, 3B, 4B) and trial types (old trial, novel trial) were the within subject factors, and the LGLTs were dependent variables. The analysis indicated a significant effect of trial types, $F_{\text{trial type}}$ (1, 29) = 21.4, p<0.001. Tasks failed to reach significance level: F_{tasks} (3, 27) = 1.85, p>0.1. There was no significant interaction between trial type and tasks, $F_{\text{interaction}}$ (3, 27) = 0.97, p>0.1. Figure 6.13 plots the LGLT for the old and novel trials of the four tasks separately.

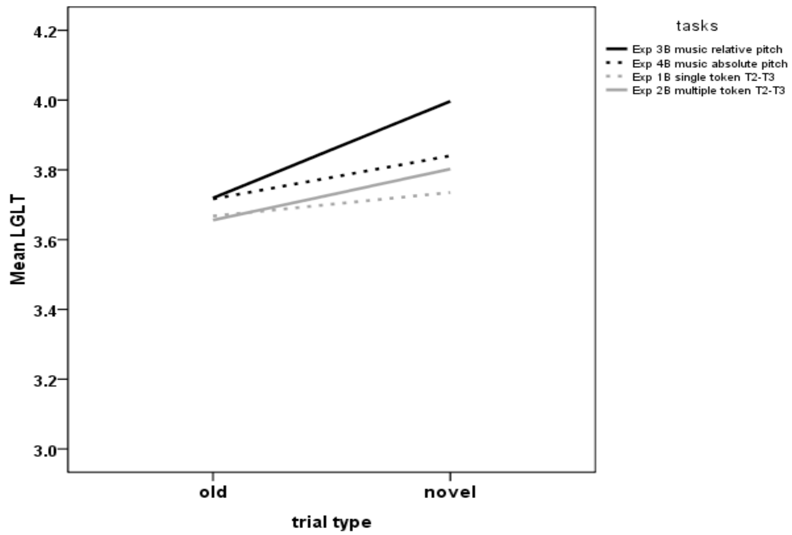


Figure 6.13 Mean LGLT for the old trial and the novel trial for the four tasks by 6-month-old Dutch infants.

From the figure, it can be seen that in general the infants' listening times were longer in the novel trials than in the old trials in all tasks. Only in Experiment 2A, the single-token scenario for Mandarin T2-T3 discrimination, their listening time to the novel trial was too short to achieve significance level. Hence, the tasks failed to be a significant factor, and there was no significant interaction between tasks and trial types.

The data of the 6-month-old infants was analyzed, just as it was for the 4-month-old infants, to discern whether any correlation existed between the discrimination of musical melody and the discrimination of lexical tones. The discrimination was denoted as a "success" if one infant listened longer to the novel trial, and a "failure" if reverse was the case. Among the 30 infants who finished all four tasks in Experiment 1B there were 16 "success" cases, and in Experiment 2B there were 19 "success" cases. In Experiment 3B there were 18 "success" cases, while in Experiment 4B, there were 10 "success" cases. For all the possible combinations of one music and one lexical tone task, a Chi-square test was carried out using the number of "success" and "failure" of each task. None of the tests demonstrated a significant effect: between relative musical pitch discrimination

(Experiment 3) and single token T2-T3 discrimination (Experiment 1), χ^2 (1) = 0.001, $p>0.1$, ϕ = -0.22, $p>0.1$; between relative musical pitch discrimination (Experiment 3) and multiple-token habituation T2-T3 discrimination (Experiment 2), χ^2 (1) = 1.16, $p>0.1$, ϕ = -0.06, $p>0.1$; between absolute music pitch discrimination (Experiment 4) and single token T2-T3 discrimination (Experiment 1), χ^2 (1) = 0.61, $p>0.1$, ϕ = -0.05, $p>0.1$; between absolute musical pitch discrimination (Experiment 4) and multiple tokens T2-T3 discrimination (Experiment 2), χ^2 (1) = 0.08, $p>0.1$, ϕ = 0.1, $p>0.1$. There was no significant correlation between the two lexical tone tasks: χ^2 (1) = 0.1, $p>0.1$, ϕ = -0.02, $p>0.1$. However, the Chi-square test did reveal a significant effect between the two music experiments: χ^2 (1) = 10.00, $p<0.01$, ϕ = 0.58, $p<0.01$.

Based on these statistics, it can be seen that in general, the 6-month-old infants tended to look longer at the novel trial in both the music and lexical tone tasks. For the music tasks, they detected the relative music pitch difference fairly easily. Yet, though there was also a significant discrimination effect regarding the absolute music pitch discrimination task, only 10 of the 30 infants obtained longer listening times in the novel trial. The listening time difference of these 10 infants between the old and novel trials was so large that the overall effect was significant.

Regarding the lexical tone tasks, the 6-month-old infants generally discriminated between Mandarin T2 and T3 when habituated on multiple tokens of one tonal category. Yet, the overall significance was due to those infants who were habituated on T3, while those habituated on T2 failed to show any discrimination effect. In Chapter 4, we saw that 6-month-old Dutch infants failed to discriminate between Mandarin T2 and T3 when they were presented with voices of two speakers in the habituation phase. At 9 and 11 months of age, the infants succeeded in discriminating in the multiple-speaker scenario, but, again, only if they were habituated on T3. By now, we have seen the discrimination data of 4, 6, 9, 11, and 14- month-old Dutch infants regarding Mandarin T2 and T3, with 6 months being the earliest age that the infants displayed any discrimination effect.

Dutch infants' discrimination patterns of the Mandarin T2-T3 contrast can be discussed from two perspectives. First, the discrimination failure of the 4-month-old infants goes against the hypothesis that tonal perceptual reorganization occurs later than 6 months. The 4-month-old infants failed to discriminate in instances where they heard pure acoustics (Experiment 1A), as well as when the tones were presented in a more linguistic manner (Experiment 2A). Moreover, the 4-month-old infants did not demonstrate any discrimination effect regardless of whether they were habituated on T2 or T3. Only at 6 months of age, the Dutch infants started showing some sensitivity to the Mandarin T2-T3 contrast. Hence, infants are not endowed to discriminate all the contrasts, and probably only those non-native contrasts that are salient enough are perceptible for infants. Second, regarding the normalization difficulty, it seems that young infants (4 and 6 months) are not able to cope with multiple speaker productions of the same phonological category. Although I did not test the 4-month-old infants with the multiple-speaker habituation scenario as I did with older infants in Chapter 4, as they failed in both the single-speaker-single-token scenario and single-speaker-multiple-token scenario, there is no reason to suspect that 4-month-old infants would succeed in a more complicated scenario. The ability to normalize different speakers' voices seems to improve between 6 and 9 months. Third, to present the 6-month-old infants with multiple tokens of one single speaker assisted them in establishing tonal categories. The 6-month-old infants were tested in all three scenarios, which differed in terms of the amount of variation. In the single-speaker-single-token scenario, they were only provided with pure acoustics of the tones; in the single-speaker-multiple-token scenario, they were habituated on 12 different tokens of one single tone, which included within-category variations; in the multiple-speaker-multiple-token scenario, they were presented with voices of two speakers. The 6-month-old infants only succeeded in the second scenario. The single-speaker-multiple-token scenario was the case most similar to real language acquisition. In real life, the caregiver(s) may well produce the same word multiple times for the infant, while the young language learners have quite a limited diversity of speakers. Hence, it seems that variations within the same speaker are quite beneficial for infants in establishing phonological categories.

Also, it seems that the infants get more aware of the linguistic function of phonological categories between 4 and 6 months of age.

If we combine the results in the current chapter with those from Chapter 4, the most crucial finding is that infants without any input of Mandarin discriminated Mandarin T2 and T3 asymmetrically right from the earliest age that they displayed some sensitivity to this tonal contrast. As mentioned above, so far there has been no evidence of the perceptual reorganization of lexical tones occurring earlier than 6 months (Mattock & Burnham 2006, 2008, Gao et al. 2011, Shi et al. 2012, Liu & Kager 2012), hence, non-tone language infants younger than or equal to 6 months are supposed to be open to tonal contrasts, and the influence from the native language is marginal. In the current dissertation, the earliest age at which Dutch infants began showing sensitivity to Mandarin T2 and T3 was 6 months. From this earliest age on, Dutch infants discriminated the T2-T3 contrast more easily if they were habituated on T3 than on T2, and this discrimination asymmetry persisted until at least 11-12 months. The performance of the 6-month-old infants serves as strong evidence for the innate biases in perceiving T2 and T3 without influence from the native language. It is a consistent finding that infants establish a T3 category more easily, which facilitates the recognition of a token that falls outside the T3 category.

The ease of setting up a T3 category also suggests that T3 is naturally more specified and distinctive in perception. The specificity of T3 may form the basis of the discrimination asymmetry that we have observed among native and non-native adults, shown in Chapter 2 and Chapter 3. In other words, the listeners benefited in discriminating T2 and T3 if a T3 preceded a T2, presumably since detecting a tone that deviates from a distinctive representation is easier than detecting a tone that may deviate from a more general representation. For future research, cross-linguistic evidence is needed to examine whether the asymmetry in the discrimination of distinctive versus less distinctive sounds holds more generally. Moreover, it may also be interesting to test whether the same kind of discrimination asymmetry can be found in segment discrimination and in other perceptual domains, such as vision. Possible further avenues for this topic may be the consistently observed vowel discrimination asymmetry (Polka & Bohn 2011) and consonant asymmetry (Fikkert 2006, Altvater-Mackensen, & Fikkert 2010).

Comparing the infants' performance in the discrimination of relative musical pitch and absolute musical pitch, it seems that at 6 months of age, relative pitch difference is more salient for infants than absolute pitch difference. Most adults are able to recognize a familiar song when it is played at varying pitch heights, and the perception and enjoyment of music mainly resides in tracking the relative pitch (Schellenberg 2003, Trainor & Trehub 2006). My findings suggest that at 6 months of age, infants already show some of the tendencies present in the music processing of adults. They are able to differentiate relative musical pitches, and the discrimination of absolute musical pitch is less robust than the discrimination of relative pitch. Moreover, the results also suggest that in early infancy, pitch direction forms a salient cue in perception.

The fact that there was only a significant correlation between relative music pitch discrimination and absolute music pitch discrimination suggests that at 6 months of age, infants are consistent in perceiving musical melody. The correlation also unearths the possibility that 6-month-old infants may already have distinctive domains of music and of speech. Nevertheless, more points need to be enumerated and expanded upon in order to fully understand the domain-general processing of pitch.

First, it must be noted that "success" and "failure" in discrimination are defined by the looking times of the infants, which is a very indirect measure. Moreover, the habituation-dishabituation paradigm differs from adult online discrimination tasks such as AX or AXB in which a discrimination accuracy rate is calculated. So it could be that the current paradigm used to test the auditory discrimination is not sensitive enough to detect other correlations which may exist, such as those between musical task and lexical tone task, or between lexical tone tasks. If we are to draw reliable conclusions about the correlation (or lack of correlation) between the perception of music prosody and lexical tones, it is necessary to collect the infant discrimination data in a more direct manner, which in turn calls for the development of more direct testing paradigms that are suitable for infants.

Second, the music stimuli may need to be diversified in order to fully trigger musical processing. Though the stimuli were generated in a way that assures the distinctive characteristics of music and speech, I was quite conservative in creating the musical stimuli. The three-note melodies did not include any rhythmic patterns, and the melodies could not be categorized as musical phrases. I used three-note melodies because 4-month-old infants are very limited in cognitive capacity. If I had made the stimuli more complex (such as four or five notes), and if infants had then failed to demonstrate a discrimination effect, it would be difficult to know whether this failure occurred because they were unable to track the pitch patterns, or because they failed to adequately process the complicated musical structure. For a more complete picture of pitch processing across different domains, future research should create more diversified musical stimuli, such as adding rhythmic or durational information, and then observe how the infants process these stimuli that fully embody musical characteristics.

Third, more information is needed to better understand the processing of pitch direction. In my relative music pitch discrimination task, the two melodies differed in pitch direction, meaning the DEF had a rising contour while the DCF had a dipping contour. Though infants succeeded in discriminating these two melodies, it is unknown how they would perform if two melodies shared the same direction, but differed in interval. For example, DEF and DEG both have rising contours, but DEG ends up one semitone higher than DEF. In this case, the two melodies also differ in relative pitch, but share the same pitch direction. An infant's ability to cope with this kind of relative pitch difference needs to be examined.

By now, three questions still remain unanswered. First, native Dutch infants consistently displayed a high sensitivity to Mandarin T1 and T4 from 6 months until 14-15 months (Liu & Kager 2010, 2012), but they only succeeded in discriminating Mandarin T2 and T3 under specific conditions. Hence, the issue here is finding out what makes the discrimination of Mandarin T2 and T3 so difficult. It may be that the continuous contour of Mandarin T2 and T3 are naturally difficult to perceive. It may also be that to perceive speech pitch correctly calls for larger acoustic differences than those present in music. In other words, one semitone may be enough for successful discrimination of musical melodies, but a larger difference is needed for correctly discriminating speech pitch. These hypotheses need further examination to better understand the role of acoustical salience in early auditory perception. Second, though it is highly likely that the complex contour of T3 makes it naturally distinctive in perception, what factors exactly contribute to the distinctiveness of T3 is not clear. Whether the complex contour of T3 which at the same time includes low, rising and dipping makes it easily distinguishable from other contours, or whether there are other factors that contribute to the specificity of T3 needs further research. Specifically, whether complexity naturally indicates distinctiveness may need cross-linguistic evidence. Third, there is not a clear answer to the question of whether the perception of lexical tones and perception of music are correlated from early infancy onwards. In my experiments, there is no evidence demonstrating this correlation among Dutch infants. As mentioned earlier, if we accept that the language acquisition is the process that infants gradually attain the same perception as adults, then, as Dutch adult listeners do display the music-lexical tone correlation, at certain point, Dutch infants or children are expected to show such a correlation too. A substantial amount of work still needs to be done to better understand the perception of lexical tones and non-linguistic pitch, and whether there are general auditory constraints that leave an impact on pitch processing in both domains. As mentioned earlier, both the paradigm and the stimuli need further diversification to fully represent the music processing and the lexical tone processing in early infancy.

6.12 LONGITUDINAL DEVELOPMENT OF DUTCH INFANTS

Among the infants that participated in the aforementioned Experiments, 24 completed all Experiments—1A, 2A, 3A, 4A— at 4 months, and— 1B, 2B, 3B, 4B— at 6 months. Using the same participants largely reduced individual variability, and their data were taken out for the analysis of longitudinal development of Dutch infants. Among the 24 infants, 12 were female and 12 were male. For the lexical tone tasks, 11 were habituated on T2 and 13 were habituated on T3. At each age, the infants finished the four tasks on two separate days with at least 3 days in between. On each test day, a music task was always coupled with a lexical tone task. The order of the music and lexical tone tasks was always the same on each test day. 12 of the infants first participated in a music task, and the other 12 first participated in a lexical tone task. For each infant, both the tasks and the order of the tasks were exactly the same at 4 and at 6 months.

As the number of the infants was quite small, the age effect was analyzed for each task individually. For each task, a repeated measures ANOVA was conducted, where the LGLT was the dependent variable, age (4 months, 6 months) and trial types (old trial, novel trial) were the within subject variables. In Experiment 1A and 2A, the single-token scenario of Mandarin lexical tone discrimination, there was a significant effect of age, $F_{age}(1, 23) = 7.74$, $p<0.01$. However, there was neither a significant effect of trial types, nor interaction between age and trial types: $F_{trial\ types}(1, 23) = 0.14$, $p>0.1$, $F_{interaction}(1, 23) = 0.62$, $p>0.1$. Figure 6.14 plots the mean LGLT for the old and novel trials of the single-speaker-single-token discrimination scenario, for Experiment 1A (4 months) and Experiment 1B (6 months). As is evident in the figure, the infants listened to both trials much longer at 4 months than at 6 months, while at each age, their looking time for the old and the novel trials did not greatly differ.

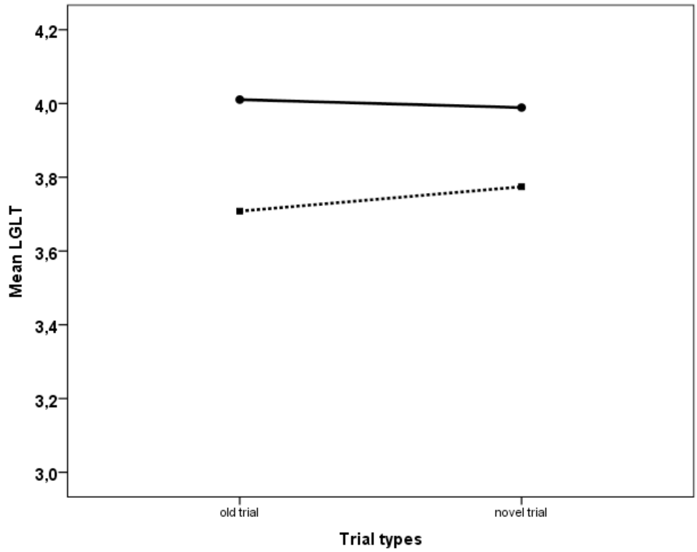


Figure 6.14 The mean LGLT for the old and novel trials of the musical relative pitch experiment. Solid line represents the performance of 4-month-old Dutch infants and dashed line represents the performance of 6-month-old Dutch infants.

For Experiment 2A and 2B, the single-speaker-multiple-token scenario of Mandarin T2 and T3 discrimination, both age and trial types failed to be a significant factor: $F_{age}(1, 23) = 2.51$ $p>0.1$, $F_{trial\ types}$ $(1, 23) = 0.65$, $p>0.1$. There was no significant interaction between the two factors, $F_{interaction}$ $(1, 23) = 0.71$, $p>0.1$. Figure 6.15 plots the mean LGLT for the old and novel trials of the single-speaker-multiple-token discrimination scenario, for Experiment 2A (4 months) and Experiment 2B (6 months). From the figure, it can be seen that as was the case in Experiment 1A and 2A, compared to 6 months, the infants tended to listen longer to the old trial and novel trial overall at 4 months of age, but the difference did not reach significance level. Again, the response to the old and novel trial remained the same at 4 months and 6 months.

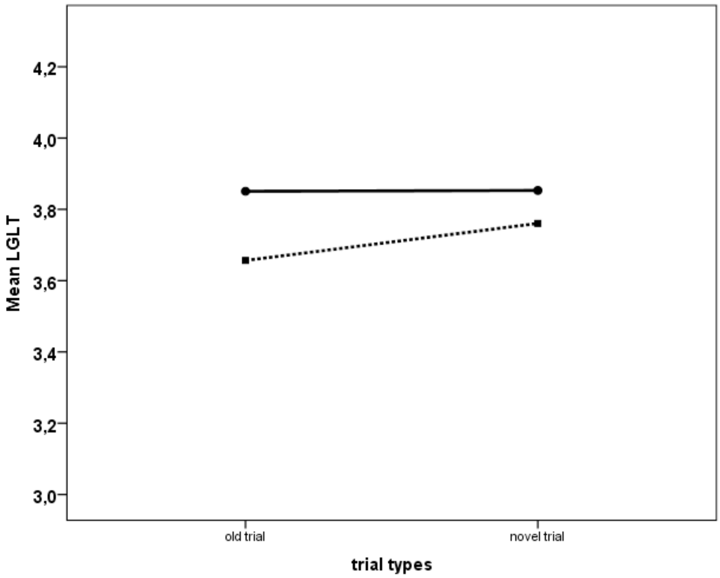


Figure 6.15 The mean LGLT for the old and novel trials of the infants in Experiment 2A (4 months, solid line) and Experiment 2B (6 months, dashed line).

For Experiments 3A and 3B, the discrimination of music relative pitch difference, trial types revealed to be a significant factor, $F_{trial\ types}$ (1, 23) = 17.62, p<0.001. Age failed to demonstrate any significant effect, F_{age} (1, 23) = 1.37, p>0.1, and there was no significant interaction between the two factors, $F_{interaction}$ (1, 23) = 1.62, p>0.1. Figure 6.16 plots the mean LGLT for the old and novel trials for the musical relative pitch discrimination in Experiment 3A (4 months) and Experiment 3B (6 months). From the figure, it is evident that the infants were able to discriminate the two music melodies that differed in relative pitch at both 4 months and 6 months. At 6 months, the discrimination effect is more evident, but the difference between 4 months and 6 months failed to reach significance.

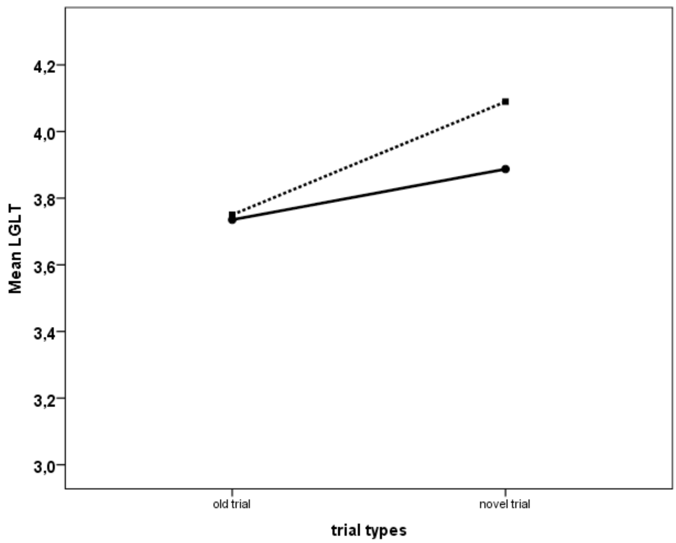


Figure 6.16 The mean LGLT for the old and novel trials of the infants in Experiment 3A (4 months, solid line) and Experiment 3B (6 months, dashed line).

For Experiments 4A and 4B, the discrimination of musical absolute pitch difference, trial types revealed marginal significance, $F_{\text{trial types}}(1, 23) = 3.98$, $p = 0.058$. Age failed to yield any significant effect, $F_{age}(1, 23) = 1.40$, $p>0.1$, and there was no significant interaction between the two factors, $F_{interaction}(1, 23) = 0.23$, $p>0.1$. Figure 6.17 plots the mean LGLT for the old and novel trials of the discrimination of musical absolute pitch difference in Experiment 4A (4 months) and Experiment 4B (6 months). As is evident from the figure, at 4 months and 6 months Dutch infants displayed a discrimination effect between the two musical melodies that differed one semitone in absolute height, yet, the discrimination effect was not as evident as in Experiment 3A and 3B.

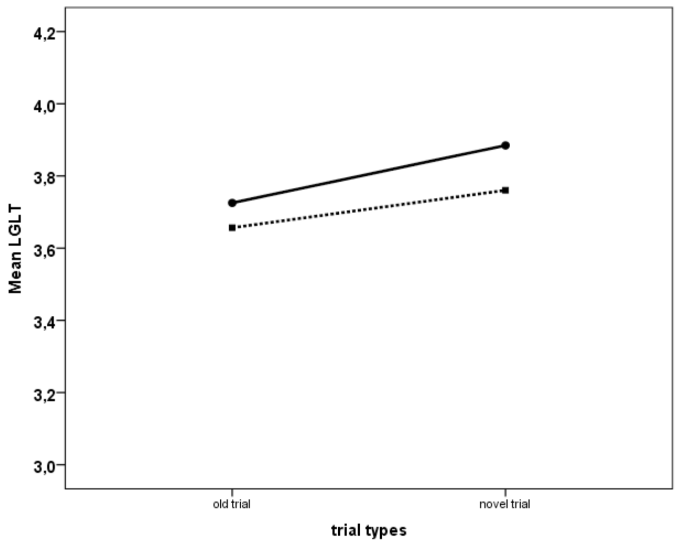


Figure 6.17 The mean LGLT for the old and novel trials of the infants in Experiment 4A (4 months, solid line) and Experiment 4B (6 months, dashed line).

Analyzing Figure 6.17, we see that the pitch perception development between 4 and 6 months is quite marginal. Overall, the 4-month-old infants tended to listen to the stimuli longer than the 6-month-old infants. The infants across the two age groups consistently perceived the musical pitches more easily than the speech stimuli. Specifically, a discrimination effect could be found between the two musical melodies that differed in absolute pitch.

The results of the longitudinal analysis can be interpreted from the following two perspectives. First, it seems that from very early on, infants perceive music pitch and speech separately, and processing lexical tones is more difficult than processing music pitch. This separation implies a speech-specific processing of lexical tone. Infants seem to be highly sensitive to the pitch variations realized in music. For future research, it is worth testing infants with other kinds of linguistic pitch variation in speech, such as intonation, and to see whether infants still perceive musical pitch more easily. Having a more complete picture of early pitch perception in multiple tasks across different domains would help us to pinpoint where the limitations of pitch processing initially lie. Second, my findings are consistent with previous research in that I did not find evidence for perceptual reorganization of lexical tones occurring earlier than 6 months (Mattock & Burnham 2006, 2008, Shi et al. 2010, Gao et al. 2011, Liu & Kager 2010, 2012). It seems that there is no radical change between 4 and 6 months in terms of pitch processing, both in the domain of music and in the domain of speech. Hence, the tonal perceptual reorganization probably occurs after the age of 6 months. It also seems that the auditory development between 4 months and 6 months is marginal. Yet, more evidence is needed from the processing of segmental and intonational information to draw a firm conclusion. Future research should test at what age the infants begin to show language-specific perception of linguistic pitch.

6.13 EXPERIMENT 1C Discrimination of T2-T3 under the sinlge-token habituation scenario by 6-month-old Mandarin infants

So far, we have seen that at 6 months of age, Dutch infants are able to perceive differences in both relative and absolute musical pitch, but fail to discriminate between Mandarin T2 and T3 if they were presented with the pure acoustics of the two tones (the single-speaker-single-token scenario). If we want to understand how language experience may shape the early perception of pitch, it is necessary to test infants whose language background differs from Dutch in terms of prosodic property. More specifically, Mandarin is a tone language that calls for accurate processing of pitch patterns realized as lexical tones. Previous studies have argued that an indicator of musical gift, namely Absolute Pitch (AP), which is the ability to recognize a note without any referent notes, tends to occur with higher probability among tone language speakers than among non-tone language speakers (Deutsch et al. 2004, 2006). Hence, investigating whether Mandarin learning infants are more accurate than their non-tone language learning peers in perceiving pitch, both lexical tones and musical pitches, would help us to better understand how a tone language shapes pitch perception early in life. Moreover, to compare the performance of Mandarin and Dutch infants before the tonal perception reorganization begins, would give more information about how infants are endowed in pitch perception.

Due to the limited access to native Mandarin infants living in China, in the current dissertation, only 6-month-old Mandarin infants participated in the experiments. The infants only participated in one session of experiments taking place on one day, which included a musical task and a lexical tone task. As the Dutch infants demonstrated a clear discrimination effect between the musical relative pitch, it seems that absolute musical pitch discrimination was more demanding than the musical relative pitch discrimination for Dutch infants. Therefore, if Mandarin infants had any benefit in pitch perception, we would expect them to outperform their Dutch counter-parts in the more demanding task, namely the musical absolute pitch discrimination task. Hence, the experiments with Mandarin infants were targeted at musical absolute pitch discrimination as a way of finding out whether native tonal language infants have an advantage in perceiving pitch. Regarding the lexical tone task, the single-speaker-single-token scenario aims to test the ability to discriminate between pure acoustics, which is more comparable with the musical tasks than the single-speaker-multiple-token scenario. Moreover, considering that Dutch infants also showed difficulties in perceiving the acoustics of Mandarin T2 and T3, how Mandarin infants perform in this task would be informative to find out whether experience with a tone language enhances lexical tone perception. Based on these reasons, Mandarin infants were tested for their musical absolute pitch discrimination and their discrimination of the T2-T3 contrast under the single-speaker-single-token scenario.

Participants

12 healthy, 6-7 month-old (5:23-7:15) Mandarin infants participated in this task. Among the infants, 6 were male and 6 were female. All the infants were raised in Beijing, and the language spoken at the infants' home was Mandarin. Another two infants were tested, but excluded from analysis due to crying.

Stimuli

The same auditory and visual stimuli used in Experiments 1A and 1B were used in the current experiment.

Experiment settings and procedure

All the infants were tested at the infant lab of the Chinese Academy of Social Sciences' Institute of Linguistics in Beijing, China. In the lab, there was a testing room where the infants and the caregivers participated in the experiment, and there was a separate control room for the experimenter. The testing room was sound-proof. There were black curtains covering the four walls of the testing room, and a sofa was placed against one of the walls, where the caregiver sat and held the infant on their laps during the experiment. In front of the infant, at a distance of about 1.5 meters, there was a 42” screen that displayed the visual stimuli. A speaker was put behind the screen in order to play the auditory stimuli. Beneath the screen, there was a hidden camera that recorded the behavior of both the infant and the parent during the experiment. The live video was transferred to a computer in the control room. Outside the testing cabin, there was a separate room for the experimenter. During the experiment, the experimenter could not see what was going on in the test cabin, and the experimenter was not able to hear the auditory stimuli. By looking at the transferred video, the experimenter recorded the eye movement of the infant on a different control computer from that which displayed the video.

There was a pre-test and a post-test before and after the experiment. Once the caregiver and the infant sat down on the sofa and reported they were ready for the experiment, the experimenter came out of the test cabin and started the experiment. The experimenter then pressed a key on the keyboard of the control computer, beginning the pre-test. After the pretest, once the infant looked at the screen, the experimenter started the first test trial by pressing the same key again, and the control computer started counting the looking time of the infant. Once the infant looked away, she released the key, the auditory stimuli stopped, and the control computer stopped counting the looking time. The looking and non-looking of the infant were always recorded by pressing and releasing the same key. If the infant looked away and then looked back to the screen within two seconds, the same trial proceeded. If the infant looked away for more than two seconds, the trial stopped, and the attention getter was automatically activated. Once the infant looked at the attention getter, the experimenter pressed the same key again to initiate the next trial. In the habituation phase, once the total looking time of the last three trials was equal or less than the total looking time of the first three trials, the habituation phase finished, and the test phase automatically began. The post-test followed the test phase. There could be maximally 12 habituation trials.

In the test phase, the infants were presented with one “old” trial and one “novel” trial. The “old” trial included the same auditory stimuli as they had heard in the habituation phase, and the “novel” trial was composed of a previously unheard sound. The order of old and novel trials, as well as the habituation stimuli, were counter-balanced across the infants. If the infants are able to discriminate between the two sounds, then their looking times to the visual stimuli in the novel trial should be significantly longer than their looking times to the visual stimuli in the old trial, due to the recovery of interest upon hearing something new.

Results and discussion

The raw looking times of the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times fitted a normal distribution: $D_{LGLT\ old}$ (12) = 0.23, $p>0.05$; $D_{LGLT\ novel}$ (12) = 0.14, $p>0.05$. The statistics reported hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was conducted with the LGLT of the old trial and the LGLT of the novel trial, and there was no significant difference between the LGLT of these two trials: T (11) = 1.34, $p>0.1$. Due to the small sample size, the participants were not divided into subgroups according to the habituation tone. Figure 6.18 depicts the LGLT of the old and novel trials for the 6-month-old Mandarin infants.

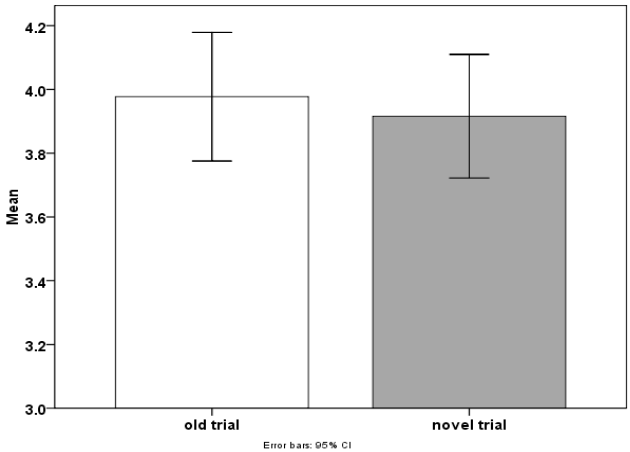


Figure 6.18 The mean LGLT of the old trial and of the novel trial in the single-speaker-single-token T2-T3 discrimination task for 6-month-old native Mandarin infants.

As can be seen from the figure, there was no significant increase in looking time when the infants were presented with novel trial, which suggests the lacking of a discrimination effect. It has to be acknowledged that the number of participants was very small, much smaller than that of the Dutch infants, which may render the statistics less powerful. Yet, only 6 out of the 12 infants listened longer to the novel trial, and there was no tendency for successful discrimination. Hence, it is justifiable to conclude that 6-month-old Mandarin infants do not show discrimination effect between the acoustics of T2 and T3.

6.14 EXPERIMENT 4C discrimination of musical absolute pitch by 6-month-old Mandarin infants

Participants

The same 12 6-month-old infants that have participated in the previous experiment participated in the present experiment. Among them, one boy was tested, but he failed to finish the experiment due to parental interference. The following analysis will be based on the data of the remaining 11 infants.

Stimuli

The same stimuli used in Experiment 4A and Experiment 4B were used in the current experiment.

Procedure

The infants first finished Experiment 1C, and right after this experiment they participated in Experiment 4C. The same procedure used in Experiment 1C was used.

Results and discussion

The raw looking times of the “old” and “novel” trials were logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking times fitted a normal distribution: $D_{LGLT\ old}$ (11) = 0.15, p>0.1; $D_{LGLT\ novel}$ (11) = 0.2, p>0.1. The statistics reported hereafter are based on the log transformed looking time (LGLT).

A 1-tailed paired t test was carried out with the LGLT of the old trial and the LGLT of the novel trial, and there was no significant difference between the LGLT of the old trial and the LGLT of the novel trial: T (10) = -0.25, p>0.1. Figure 6.19 depicts the LGLT of the old trial and the LGLT of the novel trial of the 6-month-old Mandarin infants.

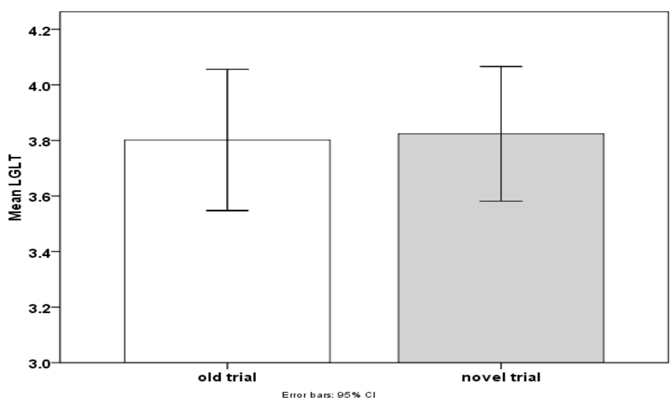


Figure 6.19 The mean LGLT of the old trial and the novel trial of the absolute music pitch discrimination task by 6-month-old native Mandarin infants.

As the figure illustrates, the looking time of the infants did not recover when presented with the novel trial, suggesting that they did not hear the difference between the two melodies that differed by one semitone in absolute pitch. Though the number of infants was quite limited in the experiment, there was not even tendency towards successful discrimination. Thus, it is very unlikely that enlarging the number of the participants would change the results.

Considering the small number of participants, I have to be cautious in drawing firm conclusions from the null results. Yet, so far, it seems that at 6 months of age, native Mandarin infants did not display sensitivity to the acoustical difference between Mandarin T2 and T3. Neither did they show a discrimination effect between the two three-note melodies that differed in absolute pitch. Referring to the results of Experiment 1B and 4B, the 6-month-old Dutch infants failed to discriminate between the acoustics of T2 and T3, but they succeeded in discriminating the musical absolute pitch difference. Therefore, it seems that at 6 months of age, Mandarin infants did not outperform their Dutch counterparts, neither in the lexical tone discrimination task nor the music task. In other words, at 6 months of age, the experience with a tone language does not facilitate pitch perception.

These patterns could be discussed from two perspectives. First, the failure of Mandarin infants, again, serves as evidence for the lack of salience of the T2-T3 contrast. The Mandarin infants had been exposed to Mandarin lexical tones for six months, and although the input during this period may not have been substantial, considering the highly informative status of lexical tones, the infants were expected to display some sensitivity to the tones. However, they did not. In other words, even experience with these two tones did not successfully encourage the discrimination between them. The insensitivity of the Mandarin infants may indicate that the acoustical difference between the two tones was too subtle for the infants to discern. Second, it seems unlikely that acquaintance with a tonal language would assure a better perception of pitch in general. As we have seen, Dutch infants discriminated between the two melodies that differed in absolute pitch at both 4 months and 6 months. It has to be acknowledged that the number of Mandarin infants was quite small, but even if I had doubled the number of participants, it would not seem likely that the Mandarin infants would have displayed a significant discrimination effect.

In Chapter 3, we learned that native Mandarin adults did not outperform native Dutch adults in the MBEA, and in this chapter, we see that the infants' performance in the music task was consistent with the adults' performance. In other words, the need to perceive pitch contrastively in language does not necessarily lead to a better perception of pitch realized in music. However, previous research has found that compared to non-tone language listeners, native Mandarin listeners were better in maintaining the same absolute pitch height in enunciating words (Deutsch et al. 2004, 2006). The authors have hypothesized that the acquisitional process of a tone language may have helped the listeners to reach a higher accuracy in pitch manipulation. To disentangle what may have contributed to these different findings, and how the acquisition of a tone language may help the processing of pitch in general, a longer observation timeframe, such as until childhood, may be needed.

6.15 CONCLUSION

In this chapter, I tested native Dutch infants at 4 and 6 months longitudinally on Mandarin T2-T3 discrimination tasks, one musical relative pitch discrimination task, and one musical absolute pitch discrimination task. To tease apart the perception of pure acoustics present in T2 and T3 and the perception of the linguistic function of the T2 and T3, the lexical tone tasks were divided into two scenarios: one single-speaker-single-token discrimination scenario targeting acoustics, and one single-speaker-multiple-token discrimination scenario targeting the linguistic function of the tones. At 4 months of age, the Dutch infants succeeded in discriminating both the relative musical pitch difference and absolute musical pitch difference while they failed on both of the lexical tone tasks. By 6 months of age, the Dutch infants still succeeded in both of the musical tasks while they failed to show a discrimination effect between the T2-T3 contrast under the single-speaker-single-token scenario. However, they did succeed in discriminating between T2 and T3 under the single-speaker-multiple-token scenario. In order to have a clear view of the results, the findings of each experiment are summarized in Table 6.2, in the same way as in Table 6.1.

Expt	N	Task	Tokens	Age	Lang	Results
1A	39	T2-T3	Single token	4m	Dutch	n.s.
2A	37	T2-T3	Multiple tokens	4m	Dutch	n.s.
3A	38	Musical rel		4m	Dutch	sig
4A	35	Musical abs		4m	Dutch	sig
1B	33	T2-T3	Single token	6m	Dutch	n.s
2B	32	T2-T3	Multiple tokens	6m	Dutch	sig, hab on T3
3B	32	Musical rel		6m	Dutch	sig
4B	31	Musical abs		6m	Dutch	sig
1C	12	T2-T3	Single token	6m	Mandarin	n.s.
4C	12	Musical abs		6m	Mandarin	n.s.

Table 6.2 Summary of the results of the experiments reported in the current chapter.

The most important finding of the current chapter is that at 6 months, in the single-speaker-multiple-token scenario, only those infants who were habituated on T3 showed a clear discrimination effect between T2 and T3, while those who were habituated on T2 failed to do so. So far, 6 months is the earliest age at which the infants showed some sensitivity to the T2-T3 contrast, and from this age on, the Dutch infants consistently perceived the T2-T3 contrast in an asymmetrical manner. It seems that with the same amount of variation, establishing a T3 category was easier than establishing a T2 category. As a result, to distinguish a T2 from a T3 category was easier than to distinguish a T3 from a T2 category. This discrimination pattern was consistent among Dutch 6, 9, and 11-month-old infants, suggesting that T3 may be easier to categorize. It is likely that the distinctiveness of T3 in perception excludes a T2 from it easily, which makes T3 a good perceptual referent. The distinctiveness of T3 may be the reason of the discrimination asymmetry that we have observed in Chapter 2 and Chapter 3, namely to discriminate between T2 and T3 is easier in a T3T2 order than in a T2T3 order. This innate discrimination asymmetry may well be the motivation that triggers T3 sandhi; the listeners tend to accept the two tones in a T2T3 sequence as the same, and confuse the T2T3 and T3T3.

In reference to the 6-month-old native Mandarin infants, there was no tendency towards successful discrimination in either the absolute musical pitch discrimination task or the single-speaker-single-token T2-T3 discrimination task. Moreover, the native Mandarin infants did not outperform their Dutch peers in either the lexical tone or music tasks at 6 months of age. Even assisted with the existing presence of these two tones in the input, the infants were still not able to pick up the acoustic difference between the two tones. The failure of native Mandarin infants to discriminate between the Mandarin T2-T3 adds more evidence to the acoustical non-salience of the two tones. Their failure in discriminating musical absolute pitch differences suggests that, at least in early infancy, the acquaintance with lexical tones does not necessarily lead to superior pitch processing.

In order to find out why the T2-T3 contrast is so difficult to perceive, the acoustical attributes of T2 and T3 were examined across the speech and music domain. The two musical melodies that differed in relative pitch shared the same pitch direction with Mandarin T2 and T3, and the acoustical difference between the two melodies used in the absolute pitch discrimination task was much smaller than the acoustical difference between T2 and T3. Hence, the infants do have a substantial capacity in tracking pitch patterns, and neither the pitch direction nor the acoustical difference alone is enough to account for the failure observed in lexical tone discrimination. The pitch processing failure is specific to lexical tones.

By 6 months of age, the Dutch infants displayed a stronger discrimination between the two melodies that differed in relative pitch, while remaining sensitive to the musical absolute pitch difference. Meanwhile, just as at 4 months of age, they failed to discriminate Mandarin T2 and T3 in the single-speaker-single-token scenario. However, they did succeed in discriminating T2 and T3 when they participated in the single-speaker-multiple-token scenario. Under the single-speaker-multiple-token scenario, the multiple tokens used in the habituation phase provided more information about the phonemic function of lexical tones. Therefore, compared to 4 months, at 6 months the infants became more aware of the linguistic usage of the tonal acoustics, and they were more capable of incorporating variations into one phonological category.

Regarding the development in perception, the results presented here showed that for all the four tasks, the difference between the performance at 4 months and at 6 months was marginal. In general, the infants tended to pay longer attention at 4 months. A stronger discrimination effect by the older infants was only observed in the musical relative pitch discrimination task. The highly successful discrimination of 6-month-old Dutch infants when presented with melodies differing in relative pitch also serves as evidence that at least in music, pitch direction difference is quite salient in perception.

Up to this point, I have not observed any correlation between musical perception and lexical tone perception among 4-month-old Dutch infants. However, by 6 months of age, the infants showed a significant correlation between the discrimination of musical relative pitch and the discrimination of musical absolute pitch. The performance of the 6-month-old infants implies that they may process pitch in a domain-specific manner, and it seems that infants have disentangled speech pitch from music. Moreover, the infants consistently displayed a higher sensitivity to musical pitch than to lexical tones.

By now, several questions may need further investigation. First, why is the acoustical difference between T2 and T3 so difficult for infants to perceive? More experiments are needed to discern whether it is the continuous pitch contour, the segmental information, or other factors that hinder successful discrimination. Second, how the perception of lexical tones and the perception of musical pitch (un)correlate in infancy remains unclear. Though the results of my experiments suggest that infants process musical melodies and linguistic tones separately, more efforts need to be made across the separate domains to discover whether the perception of pitch is correlated. Referring to the different music-lexical tone correlation patterns among Dutch and Mandarin listeners in Chapter 3, establishing a clear picture of early musical pitch and lexical tone perception across infants with different language backgrounds is of great importance for us to understand the initial state and developmental course of the human auditory system. As was mentioned earlier, the musical pitch stimuli may need to be diversified in order to encourage comprehensive musical processing. Moreover, the testing paradigm needs to be improved to obtain a more direct and extensive measurement of infants' perception of pitch.

Chapter 7 Conclusion

The current dissertation has explored possible perceptual biases that may shape the Mandarin T3 sandhi rule. T3 sandhi restricts the co-occurrence of T3s, applying in a unidirectional and asymmetrical way, in that it requires the first T3 in a T3T3 sequence to change into T2 obligatorily while the second T3 stays intact. In order to identify a perceptual motivation for this asymmetry, multiple tasks were carried out, including categorical perception experiments with native and non-native adults (Chapter 2) and speeded discrimination experiments with native and non-native adults (Chapter 3), infant T2-T3 discrimination experiments (Chapter 4 and Chapter 6), and infant lexical tone learning experiments (Chapter 5). In addition, by testing native and non-native, adult and infant listeners on their perception of musical pitch, the perception of lexical tones was studied from a domain-general perspective across different populations. By examining these different populations in different tasks, a more general issue was addressed: how biases intrinsic to human perception interact with the language-specific perception of lexical tones, and how such universal biases may shape language-specific phonological grammars of lexical tones.

Chapter 2 demonstrated that adult native Dutch listeners perceived the Mandarin T2-T3 and T1-T4 contrasts in a psycho-acoustical manner whereas native Mandarin listeners perceived these contrasts categorically, in accordance with the phonemic function of lexical tones. Importantly, in Chapter 2 and Chapter 3, it was found that in spite of their different ways of processing the lexical tones, native Dutch listeners and Mandarin listeners showed the same discrimination patterns: they discriminated between T2 and T3 more easily if T3 preceded T2 in the to-be-discriminated pair than in the reverse order T2-T3; they discriminated between bisyllabic sequences T3T2 and T3T3 more accurately than between T2T3 and T3T3; their discrimination of bisyllabic tonal sequences was more successful if the T3T3 sequence occurred first in the to-be-discriminated pair, rather than last. Moreover, all these asymmetries are specific to T2 and T3, whereas the T1-T4 contrast is exempt from any discrimination asymmetries. The listeners also consistently reached higher accuracy in discriminating between T1 and T4 and in bisyllabic sequences involving T1 and T4 than between T2 and T3 or in bisyllabic sequences involving T2 and T3. The fact that the Dutch adult listeners had no prior exposure to Mandarin and hence, no knowledge of the tonal phonology, serves as strong evidence that there are innate perceptual biases restricting the discrimination between T2 and T3 in certain orders. Dutch listeners and Mandarin listeners do not differ in terms of having or lacking the discrimination asymmetry, but only differ in terms of its magnitude: native listeners' knowledge of T3 sandhi amplifies the aforementioned asymmetries in the Mandarin listeners' responses.

In Chapter 4, native Dutch infants participated in a Mandarin T2-T3 discrimination task, in which two tokens of either tone of two native Mandarin speakers served as stimuli in the habituation phase. It was demonstrated that Dutch infants failed at the discrimination at 6 and 14 months of age, while they succeeded at 9 and 11 months of age. Importantly, the 9 and 11-month-old infants succeeded in discriminating between T2 and T3 only if they were habituated on T3.

In Chapter 5, native Dutch infants were trained with naturally produced Mandarin speech, in which either the naturally produced confusing T2T3 or the non-confusing T3T2 were embedded. After the training, they discriminated between the underlying T3T2 and T3T3 more readily than between underlying T2T3 and T3T3. Taking together the findings of adult and infant participants, it seems that T3 sandhi targets a contrast of low perceptibility, namely the T2-T3 contrast, and it merges perceptually similar sequences of tones, namely T2T3 and T3T3. The biases have been phonologized into Mandarin tonal grammar as a T3 sandhi process, rendering the perceptual neutralization of T2T3 and T3T3 into a categorical process. As a result of phonologization, the 'natural' perceptual confusion between T2T3 and T3T3 as well as between T2 and T3 in a T2-T3 pair becomes more evident among native Mandarin listeners.

Considering the fact that pitch variations function in a meaningful way across different domains, in particular speech and music, this dissertation investigated the perception of Mandarin T2-T3 and T3 sandhi from a domain-general perspective as well. Chapter 3 found that Dutch but not Mandarin listeners show a significant positive relationship between the discrimination of Mandarin tones and the discrimination of musical melodies. The Dutch listeners were divided into two groups, namely the 'music high group' and the 'music low group', according to their accuracy in perceiving a change of one single note of a melody. Although the difference in accuracy between the groups of listeners regarding the discrimination of T2 and T3 is evident, these groups do not differ substantially qua accuracy when discriminating between T1 and T4. Crucially, although the 'music high group' reached higher accuracy in the discrimination of T2 and T3 in general, still both groups consistently show the aforementioned asymmetries in the discrimination of Mandarin T2 and T3.

Chapter 6 addressed the topic how infants perceive lexical tones and musical melodies. Regarding musical perception, it was found that infants as young as four-month-olds succeeded in discriminating three-note melodies that differed in pitch direction as well as three-note melodies differing in terms of absolute pitch by one semitone. Yet these infants were not able to discriminate between Mandarin T2 and T3, no matter whether they were encouraged to pay attention to the pure acoustics of the two tones (when only one token was presented in the habituation phase), or were provided with hints about the linguistic function of the tones (when multiple tokens occurred in the habituation phase). By 6 months of age, these infants retained their sensitivity in discriminating the two types of musical melodies; meanwhile, they had increased their sensitivity in discriminating Mandarin T2 from T3 when habituated on multiple tokens of T3. Taken together, these results suggest that although young infants are capable of processing subtle pitch differences, they nevertheless failed in the discrimination of T2-T3. Specifically, the musical melodies that differed in pitch direction were generated in such a way that they shared the pitch direction of T2 and T3, while the musical melodies that differed in absolute pitch displayed a much smaller pitch difference compared to T2 and T3. Hence, the infants' failure in T2-T3 discrimination cannot be accounted for by pitch direction, nor by acoustical salience alone. Taking together the performance of 6-month-old infants in Chapter 4 (habituated on voices of two speakers of either T2 or T3), and in Chapter 6 (habituated on single token of either tone, or multiple tokens of either tone produced by one single speaker), it can be seen that 6-month-old infants display difficulties in normalizing the voices of multiple speakers, but multiple productions by one single speaker helps infants to categorize lexical tones. The fact that 4-month-old infants failed to discriminate between T2 and T3 under the single-speaker-multiple-token scenario while 6-month-old infants succeeded in doing so suggests that infants may become more aware of the linguistic function of the phonological category between 4 and 6 months. Mandarin infants of 6 months of age also performed in the T2-T3 discrimination where only one token of either tone was used as habituation stimuli, as well as the musical absolute pitch discrimination task. No discrimination effect could be observed in both tasks, indicating that at 6 months of age, the Mandarin infants did not outperform their Dutch peers, neither in discrimination of the Mandarin T2-T3 contrast nor in discrimination of musical absolute pitch. These results suggest that immersion in a tonal language environment need not guarantee enhanced pitch processing. Moreover, no correlation was observed between musical melody discrimination and lexical tone discrimination among Dutch infants at either 4 or 6 months. Such a correlation is also absent among Mandarin infants. Yet 6-month-old Dutch infants did show a correlation between the two musical tasks. In sum, the findings of Chapter 6 suggest that pitch variations are processed in a domain-specific way from early infancy on.

In Chapter 4 and Chapter 6, the finding that cannot be over-emphasized is that, with equal amount of variations in the habituation phase (four tokens for 9 and 11-month-old infants in Chapter 4, and 12 tokens for 6-month-old infants in Chapter 6), the Dutch infants consistently categorized T3 more easily compared to T2. One speculation is that the more acoustically specified a sound is, the more distinctive it is, and hence the easier it should be to categorize it. In the face of the discrimination asymmetry among the Dutch infants, the facilitated categorization of T3 implies that T3 is more distinctive in perception compared to T2.

Taking together all these findings, it seems that Mandarin T3 sandhi is motivated perceptually. First, the complex pitch contour of T3, which features three low, dipping and rising components, makes it acoustically specified and hence, perceptually distinctive. Thus, to detect a tone that falls outside the acoustic space of the distinctive tone is relatively easy. As a result, in the discrimination tasks, a change from T3 to T2 was detected fairly accurately, when T3 preceded T2. In comparison, the rising contour of T2 is less acoustically complex and more general than T3, and as a result, a T3 may be erroneously perceived as an instantiation of a T2. Hence, the discrimination of T2 and T3 was less successful when T2 preceded T3, compared to the reverse order. Second, compared to other bisyllabic tonal sequences, T2T3 and T3T3 are highly similar in perception. Based on the aforementioned confusion asymmetry of monosyllabic T2 and T3, it might be that the tones in T2T3 are mis-perceived as being identical. As a result, T2T3 shares the same structure with T3T3 in being composed of identical tones, which might make listeners mis-perceive these two sequences as the same. Yet it has to be acknowledged that what makes T2T2 more distinguishable from T3T3 remains unclear. Moreover, the fact that discrimination between T2T3/T3T2 and T3T3 is more accurate when T3T3 occurs first rather than last is consistent with the findings about the confusion asymmetry between monosyllabic T2 and T3. To detect a change from T3 to T2 in a bisyllabic tonal sequence is fairly easy while a change from T2 to T3 in a bisyllabic sequence is less detectable. Hence, T2T3 tolerates more variations and listeners may perceive T3T3 as an instantiation of T2T3. The fact that perceptually, T2T3 and T3T3 are highly similar and listeners are more willing to accept a T3T3 as a T2T3 than vice versa may trigger T3 sandhi to occur in the way as it does now: T3T3 is sandhied to T2T3, as the alternation is not obviously detectable.

To put the findings into a bigger picture, it seems that the listener plays a crucial role in shaping the phonological grammar of lexical tones, as innate perceptual biases are likely to set up pressure for T3 sandhi to occur. In other words, the language-specific tonal grammar may function to suit human auditory perception; a tonal contrast of low perceptibility is merged in the specific context in which the contrast is least perceptible. So far, my findings fit into an emergentist view in that a consistent confusion asymmetry re-occurs among native and non-native adult listeners, as well as young infants. To put it differently, the phonological grammar including T3 sandhi suits the functional needs of humans, in that merging T2T3 and T3T3 reduces the perceptual burden of discriminating between two elements of low perceptibility (e.g. Steriade 1995, 1999, Hume & Johnson 2001, Huang 2001, Seo 2001). Moreover, it may well be the case that evolutionarily, T3 sandhi originated from a misperception by listeners (Ohala 1981, 2012). These considerations support the view that innate perceptual biases restrain the formation of phonology, both in terms of segments and lexical tones. In other words, both segmental phonology and prosody may be structured in ways that suit the human perceptual system.

At this point, several questions still need further investigation. First, although I speculated that the specificity and distinctive nature of T3 relative to T2 may be the factor that triggers the discrimination asymmetry, more research is needed to establish more generally whether the acoustic complexity of a speech sound is the primary factor determining its distinctiveness, causing perceptual asymmetries with other sounds, or whether other factors come into play to contribute to perceptual asymmetry. Second, cross-linguistic research is needed to understand what exactly constitutes the perceptual bias in shaping tone sandhi and whether perceptual asymmetries identified in the current study can be widely observed across different sandhi patterns. Third, for infants, it is not yet clear whether a correlation holds [that is, at the level of individuals?] between the perception of musical pitch and the perception of lexical tones, or more generally, the perception of speech pitch. As was discussed in Chapter 6, my findings suggest that infants process musical and speech pitch differently, but the question of whether there is a correlation between the two domains still has no clear answer. Both lexical tone stimuli and musical stimuli need to be diversified in order to arrive at a fuller understanding of the perception of pitch patterns by infants. Fourth, there is still a long way to go to understand how language experience may shift the initial pitch perception in infancy. Thorough investigation is needed to answer questions concerning the age at which non-tone-language learning language infants (or children) start displaying a correlation between the perception of lexical tones and musical pitch, and concerning the mechanism behind the change in pitch perception from infancy until adulthood.

References

Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech, 21*, 319-325.

Alexander, J. A., Wong, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Proceedings of Ninth Conference on Speech Communication and Technology*, Lisbon.

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua*, 1898-1909.

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83*(2), 111-130.

Aquilina, J. (1959). *The structure of Maltese*. Valleta, Malta: Progress press.

Aslin, R. N., & Pisoni, D. B. (1980). Effects of early linguistic experience on speech discrimination by infants: A critique of Eilers, Gavin, & Wilson. *Child Development, 51*, 107-112.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia. *Brain, 125*(2), 238-251.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology, 26*(01), 75.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition, 104*(3), 591-630.

Bertoncini, J., Floccia, C., Nazzi, T., & Mehler, J. (1995). Morae and syllables: Rhythmical basis of speech representation in neonates. *Language and Speech, 38*, 311-329.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171–204). Baltimore: York Press.

Best, C., Singh, L., Bouchard, J., Connelly, G., Cook, A., & Faber, A. (1997). Developmental changes in infants' discrimination of non-native vowels that assimilate to two native categories. *Meeting of the Society of Research in Child Development,* Providence.

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

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, 14*(3), 345-360.

Belotel-Greni, A., Greni, M. (1995). Consonants and vowels influence on phonation types in isolated words in Standard Chinese. *Proceedings of XIIIth International Congress on Phonetic Sciences*, 400-403.

Bidelman, G., Gandour, J., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*, 425-434.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do 4-day-old infants categorize multisyllabic utterances? *Developmental Psychology, 29*, 711–721.

Blevins, J. (2006). A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics, 32*(2), 117-166.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*, 37-49.

Bluhme, H., & Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tones. *Studies in Linguistics, 22*, 51–57.

Bohn, O. (2007). Exploring the natural referent vowel hypothesis in infant perception experiments. *Journal of Acoustic Society of America, 121*(5), 3187-3187.

Bohn, O., & Polka, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *Journal of Acoustic Society of America, 110*, 504-515.

Bortfeld, H., Morgan, J., Michnick Golinkoff, R., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science, 16*(4), 298-304.

Bortfeld, H., & Morgan, J. L. (2010). Is early word-form processing stress-full? how natural variability supports recognition. *Cognitive Psychology, 60*(4), 241-266.

Burnham, D., & Brooker, R. (2002). Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers. In *Proceedings of The 7th International Conference on Spoken Language Processing,* pp. 257-260.

Burnham, D., Kirkwood, K., Luksaneeyanawin, S., & Pansottee, S. (1992). Perception of central Thai tones and segments by Thai and Australian adults. In *Proceedings of the Third International Symposium on Language and Linguistics,* pp.546-560

Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., Kasisopa, B., & Luksaneeyanawin, S. Are tones phones? *Journal of Experimental Child Psychology 108* (4), 693-712.

Cao, J. (2002). The relationship between tone and intonation. *Report of Phonetic Research.* Retrieved from ling.cass.cn/yuyin/report/report_2002_3

Chan, M. K. M. (1995). An autosegmental analysis of Danyang tone sandhi: Some historical and theoretical issues. In E. Zee (ed.), *Studies of the Wu dialects* (pp.146-184). Hong Kong: New Asia Academic Publishing.

Chan, M. K. M., & Ren, H. (1989). Wuxi tone sandhi. *Acta Linguistica Hafniensia, 21*(2), 35-64.

Chang, S., Plauche, M., & Ohala, J. (2001). Markedness and consonant confusion asymmetries. In E. Hume, & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 79-101). Academic Press.

Chao, Y. (1930). A system of tone letters. *Le Maitre Phonetique, 45*, 24-27.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Berkeley, CA: Cambridge University Press.

Chen, Y., Shen, R., & Schiller, N. (2011). Representation of allophonic tone sandhi variants. In *Proceedings of Psycholinguistic Representation of Lexical Tones 2012,* Hong Kong.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech -- evidence from F(0) patterns of neutral tone in standard Chinese. *Phonetica, 63*(1), 47-75.

Chen, Y., & Yuan, J. (2007). A corpus study of the 3rd tone sandhi in standard Chinese. In *Proceedings of Interspeech* 2007, 2749-2752.

Cheng, C. (1973). *A synchronic phonology of Mandarin Chinese*. Den Haag: Mouten.

Cheng, R. L. (1968). Tone sandhi in Taiwanese. *Linguistics, 41*, 19-42.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Chuang, C. K., & Hiki, S. (1972). Acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *Journal of Acoustic Society of America, 52*(1A), 146-146.

Ciocca, V., & Liu, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders, 1*(2), 141-147.

Cole, R., Jakimik, J., & Cooper, W. (1978). Perceptibility of phonetic features in fluent speech. *Journal of Acoustic Society of America, 64*, 44-56.

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Neurobiology of Learning and Memory, 92*(2), 225-234.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

De Lacy, P. (2006). *Markedness: Reduction and preservation in phonology*. Rutgers University, New Jersey: Cambridge University Press.

DeCasper, A., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sound. *Infant Behavior and Development, 9*, 133-150.

Delogu, F., Lampis, G., & Belardinelli, M. O. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing, 7*(3), 203-207.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., & Pierce, A. (1997). Perception of synthetic /ba/--/wa/ speech continuum by budgerigars (melopsittacus undulatus). *Journal of the Acoustical Society of America, 102*, 1891-1897.

Deutsch, D. (2002). The puzzle of absolute pitch. *Current Directions in Psychological Science, 11*(6), 200-204.

Deutsch, D., Henthorn, T., & Dolson, M. (1999). Absolute pitch is demonstrated in speakers of tone languages. *Journal of Acoustical Society of America, 106*, 2267.

Dreher, J., & Lee, P. C. (1968). Instrumental investigation of single and paired Mandarin tonemes. *Monumenta Serica, 27*, 343-373

Duanmu, S. (2000). *The phonology of standard Chinese*. Oxford: Oxford University Press.

Duanmu, S. (2006). Chinese (Mandarin): Phonology. In Brown, K. (Ed.) Encyclopedia of Language and Linguistics, 2nd Edition, (pp. 351-355). Oxford: Elsevier Publishing House.

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

Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning, 48*(4), 631-664.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech, 46*(2-3), 245-264.

Fikkert, P. (2006). Developing representations and the emergence of phonology: Evidence from perception and production. *Laboratory Phonology.* Retrieved from http://www.basesproduced.com/699/readings/FikkertinpressLabPhon.pdf.

Fok, Y. Y. C. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: University of Hong Kong.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & D. Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying ‘tone deafness’. *Brain, 127*(4), 801-810.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Attention, Perception, & Psychophysics, 65*(7), 1029-1044.

Fu, Q. -., & Zeng, F. -. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language, and Hearing, 5*, 45-57.

Gandour, J. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41-76). New York: Academic.

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics, 9*, 20-36.

Gandour, J., & Harshman, R. (1978). Cross-language difference in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1-33.

Gandour, J. (1983). Tone perception in far eastern languages. *Journal of Phonetics, 11*(2), 149-175. Retrieved from SCOPUS database.

Gao, J., Shi, R., & Li, A. (2011). Lexical tone perception in non-tone-learning infants. *12th International Congress for the Study of Child Language,* Montreal.

Gerrits, E., & Schouten, M. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*, 363-376.

Granier-Deferre, C., Bassereau, S., Jacquet, A. Y., & Lecanuet, J. P. (1998). Fetal and neonatal cardiac orienting response to music in quiet sleep. *Developmental Psychobiology, 33*, 372.

Greenberg, J. H. (1978). Some generalizations concerning initial and final consonant clusters. In E. A. Moravcsik (Ed.), *Univsersals of human language* (pp. 243-279). Stanford: Stanford University Press.

Grieser, D. L., & Kuhl, P. K. (1989). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese *Developmental Psychology, 24*, 14-20.

Hallé, P. A., Chang, Y., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua, 110*(8), 581-616.

Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain, 126*(8), 1838-1850.

Hollich, G., Jusczyk, P., & Brent, M. (2002). Talker variation and word learning. *Acoustical Society of America Journal, 111*, 2455-2455.

Holt, L. L., & Lotte, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics, 72*(5), 1218-1227.

Hombert, J. M., Ohala, J., J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones.*55*, 37-58.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26*(5), 1570-1582.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.

Howie, J. M. (1974). On the domain of tone in Mandarin. some acoustical evidence. *Phonetica, 30*(3), 129-148. Retrieved from SCOPUS database.

Hsieh, F. (2010). Gesture reorganization in Mandarin tone 3 sandhi. *Journal of Acoustical Society of America, 128*, 2458.

Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*(1), 3-42.

Huang, T. (2001). The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. *Working Papers in Linguistics Ohio State University, 55*, 23-42.

Hume, E. (1996). Coronal consonant, front vowel parallels in Maltese. *Natural Language and Linguistic Theory, 14*, 163-203.

Hume, E., & Johnson, K. (2001). A model of the interplay of speech perception and phonology. In E. Hume, & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 3-26). Academic press.

Hume, E. (1998). Metathesis in phonological theory: The case of Leti. *Lingua, 104*(3–4), 147-186.

Hume, E., & Johnson, K. (2003). The impact of partial phonological contrast on speech Perception1.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356-360.

Jakobson, R. (1941). *Kindersprache, aphasie, und allgemeine lautgezetze* [Child Language, Aphasia and Phonological Universals, Den Haag, Mouten]. Uppsala: Almqvist & Wiksell.

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia, 48*(9), 2630-2639.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*, 159-207.

Jusczyk, P. W., & Thompson, E. J. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. *Perception & Psychophysics, 23*, 105-109.

Kager, R. (1999). *Optimality theory*. Cambridge: Cambridge University Press.

Keating, P.A., Esposito, C. (2006). Linguistic voice quality. Paper presented at the *11th Australasian International Conference on Speech Science and Technology*.

Kiriloff, C. (1969). On the auditory discrimination of tones in Mandarin. *Phonetica, 20*, 63-67.

Kratochvil, P. (1971). An experiment in the perception of Peking dialect tones. In I-L. Hansson (Ed.), *A symposium on Chinese Grammar* (pp. 7-31). Lund Sweden: Curson Press.

Kuang, G., & Wang, H. (2006). Acoustic cues of successive tone 3 sandhi at different prosodic boundaries. In *Proceedings of Phonetic Conference in China,* Beijing.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of Acoustic Society of America, 66*, 1668-1679.

Kuhl, P. K. (1991). "Perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Attention, Perception, & Psychophysics, 50*(2), 93-107.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics, 21*, 125-139.

Kuhl, P. K. (1995). Mechanisms of developmental change in speech and language. In *Proceedings of the XIllth International Congress of Phonetic Sciences,* Stockholm: KTH and Stockholm University. pp. 132-139.

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), 13-21.

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*, 606-608.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*(2–3), 263-285.

Kuhl, P., & Padden, D. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Attention, Perception, & Psychophysics, 1982*, 542-550.

Kuhl, P., & Miller, J. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*(4209), 69-72.

Kulh, P. (1988). Auditory perception and the evolution of speech. *Human Evolution, 3*, 19-43.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics, 11*(4), 373-382.

Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America, 124*, 3235-3248.

Lee-Schoenfeld, V. (2002). *Third tone prominence.* Manuscript, Swarthmore College. Retrieved from http://www.swarthmore.edu/SocSci/Linguistics/verawebpage/Papers/paper7.pdf

Lee-Schoenfeld, V., & Kandybowicz, J. (2008). Third tone patterns in Mandarin Chinese: A new perspective. *UCLA Working Papers in Linguistics,* Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences, 9*(1), 26-33.

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*(02), 185.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1967). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4*, 175-195.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61*(5), 379-388.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language, 48*, 839-862.

Lin, H. (2006). Mandarin neutral tone as a phonologically low tone. *Journal of Chinese Language and Computing, 16*(2), 121-134.

Lin, M., & Wang, W. S. (1985). Shengdiao ganzhi wenti [tone perception]. *Zhongguo Yuyan Xuebao, 2*, 59-69.

Lin, M. (1988). The acoustical properties and perceptual characteristics of Mandarin tones. *Zhongguo Yuwen, 3*, 182-193.

Lin, H., & Repp, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech, 32*(1), 25-44.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403-440). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Lindblom, B., & Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics, 17*, 107-121.

Liu, F., Jiang, C., Thompson, W., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *Plos One, 7*, e30374.

Liu, L. (2010). *Is perceptual reorganization affected by statistical learning? Dutch infants' sensitivity to lexical tone.* Unpublished Utrecht University Master thesis.

Liu, L., & Kager, R. (2011). Is perceptual reorganization affected by statistical learning? Dutch infants' sensitivity to lexical tones. *BUCLD 35,* pp. 404-413.

Liu, L., & Kager, R. (2012). Non-native tone perception from infant to adult: How consistent and flexible is it? *Proceedings of Speech Prosody 2012,* Shanghai, China.

Lu, Z. (1986). Tonal changes: Interplay between tone and tone sandhi. A case study of the Shanghai dialect. *Studies in the Linguistic Sciences, 16*(1), 97-111.

MacWhinney, B. (2001). Emergence from what? *Journal of Child Language, 28*, 726-732.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18*(2), 199-211.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Jounral of Cognitive Neuroscience, 23*, 2701-2715.

Massaro, D. W., Cohen, M. M., & Tseng, C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13*, 267-290.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367-1381.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*(1), 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-B111.

Mehler, J. & Christophe, A. (1995) Maturation and learning of language in the first year of life. In Gazzaniga, Michael S. (Ed), *The Cognitive Neurosciences* (pp. 943-954). Cambridge, MA, US: The MIT Press.

Mehler, J., & Dupoux, E. (1994). *What infants know: The new cognitive science of early development.* Blackwell Publishers.

Michnick Golinkoff, R., & Hirsh-Pasek, K. (2011). Methods for studying language in infants: Back to the future. In E. Hoff (Ed.), *Research methods in child language: A practical guide*, (pp. 60-75). Oxford: Wiley-Blackwell.

Milliken, S. (1989). Why there is no third tone sandhi rule in standard Mandarin. In *proceedings of Tianjin International Conference on Phonetics and Phonology,* Tianjin Normal University.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*, 495-500.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of Acoustical Society of America, 102*(3), 1864-1877.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*, 911-936.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain, 133*(9), 2635-2642.

Narayan, C. R. (2004). A phonological asymmetry and its perceptual consequence: The case of the velar-nasal onset. *Journal of Acoustic Society of America, 116*(4), 2572.

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

Nelson, K., Deborah, G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., et al. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111-116.

Nettl, B. (1983). *The study of ethnomusicology: 29 issues and concepts.* Urbana: University of Illinois Press.

Newman, R. S. (2008). The level of detail in infants' word learning. *Current Directions in Psychological Science, 17*(3), 229-232.

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences, 1169*(1), 490-493.

Nordenhake, M., & Svantesson, J. O. (1983). Duration of standard Chinese word tones in different sentence environments. *Lund University Working Papers in Linguistics, 25*, 105-111.

Norman, J. (1988). *Chinese*. Cambridge, UK: Cambridge University Press.

Ohala, J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178-203). Chicago: Chicago Linguistic Society.

Ohala, J. (1996). Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America, 99*(3), 1718-1725.

Ohala, J. (2012). The listener as a source of sound change, an update. In M. Sole, & D. Recasens (Eds.), *The initiation of sound change, perception, production, and social factors* (pp. 21-35). Philadelphia: John Benjamins.

Ohala, J. J. (1986). Consumer's guide to evidence in phonology. *Phonology, 3*, 3-26.

Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication, 13*(1-2), 155-161.

Patel, A., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns a neuropsychological investigation. *Brain and Language, 61*, 123-144.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition, 59*(3), 310-313.

Pater, J., Stager, C., & Werker, J. (2004). The perceptual acquisition of phonological contrasts. *Language, 80*(3), pp. 384-402. Retrieved from http://www.jstor.org/stable/4489718

Peng, S. (1996). *Phonetic implementation and perception of place coarticulation and tone sandhi.* Unpublished PhD dissertation, Ohio State University, Department of Linguistics,

Peretz, I., Champod, A., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences, 999*(1), 58-75.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*, 688–691.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., & Demeurisse, G. (1994). Functional dissocations following bilateral lesions of auditory cortex. *Brain, 117*, 1283-301.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*(2), 185-191.

Pinker, S. (1997). *How the mind works*. London: Allen Lane.

Pisoni, D. (1971). *On the nature of categorical perception of speech sounds.* Unpublished PhD dissertation, Michigan University at Ann Arbor,

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Attention, Perception, & Psychophysics, 13*(2), 253-260.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition, 98*(1), 1-11.

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. Retrieved from SCOPUS database.

Polka, L., & Bohn, O. (2003). Asymmetries in vowel perception. *Speech Communication, 41*(1), 221-231.

Polka, L., & Bohn, O. (2011). Natural referent vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*(4), 467-478.

Polka, L., Colantonio, C., & Sundara, M. (2001). A crosslanguage comparison of /d/-/ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America, 109*, 2190–2201.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology, 20*(2), 421-435.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar* Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, N.J., and Computer Science Department, University of Colorado, Boulder.

Querleu, D., Lefebvre, C., Titran, M., Renard, X., Morillion, M., & Crepin, G. (1984). Reactivity of newborns less than two hours old to the maternal voice. *Journal De Gynecologie, Obstetrique Et Biologie De La Reproduction, 13*, 125-134.

Repp, B. H. (1984). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. *Language and Speech, 27*(3), 245-254.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*(2), 339-349.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology, 37*(1), 74-85.

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*(3), 484-494.

Saffran, J. R. (2003). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science, 6*(1), 35-43.

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

Saffran, J. R., Reeck, K., Niebuhr, A., & Wilson, D. (2005). Changing the tune: The structure of the input affects infants? Use of absolute and relative pitch. *Developmental Science, 8*(1), 1-7.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science, 14*, 262-266.

Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception*. (pp. 546-566). Oxford: Blackwell.

Seo, M. (2001). The realization of l/n sequences in Korean. *The Interplay of Speech Perception and Phonology*, Ohio State University Working Papers in Linguistics, 55.

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association, 24*, 27-47.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*(2), 145-156.

Shen, R. (1994). Beijinghua shansheng liandu de diaoxing zuehe he jiezou xingshi [The tone and rhythmic configuration of sandhi in Beijing Mandarin]. *Zhongguo Yuwen, 4*, 274-281.

Shi, R., Gao, J., & Li, A. (2011). Perception of lexical tones in infants. Poster presentation in *36th Annual Boston University Conference on Child Language Development,* Boston. *, 36.*

Shi, C. (1986). The prosodic domain of tone sandhi in Chinese. PhD dissertation, University of California, San Diego.

Shih, C. (1997). The representation of the neutral tone in Chinese Putonghua. In Wang J. & Smith, N. (Eds.), *Studies in Chinese phonology*. (pp. 81-123). Berlin: Mouton de Gruyter.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition, 106*(2), 833-870.

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

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17*(8), 675-681.

Speer, S., Shih, C., & Slowiaczek, M. (1989). Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin. *Language and Speech, 32*, 337-354.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381-382.

Stampe, D. (1972). How I spent my summer vacation. Ph.D. dissertation, Chicago: University of Chicago.

Steriade, D. (1999). Phonetics in phonology: The case of laryngeal neutralization. In M.Gordon (Ed.) *UCLA working papers in linguistics 2,* Los Angeles: Department of Linguistics, University of California.

Steriade, D. (1995). Underspecification and markedness. In J. Goldsmith (Ed.), *Handbook of phonological theory.* Cambridge, Massachusetts: Blackwell.

Strange, W. (Ed.), (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-/ in monolingual and bilingual acquisition of English. *Cognition, 100*(2), 369-388.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology, 54*(2), 99-132.

Swoboda, P. J., Kass, J., Morse, P. A., & Leavitt, L. A. (1978). Memory factors in infant vowel discrimination of normal and at-risk infants. *Child Development, 49*, 332-339.

Swoboda, P. J., Morse, P. A., & Leavitt, L. A. (1976). Continuous vowel discrimination in normal and at risk infants. *Child Development, 47*, 459-465.

Tew, S., Fujioka, T., He, C., & Trainor, L. (2009). Neural representation of transposed melody in infants at 6 months of age. *Annals of the New York Academy of Sciences, 1169*(1), 287-290.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language, 56*(1), 16-34.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 1*(4), 46-64.

Trainor, L. J., & Trehub, S. (1992). A comparison of infants' and adults' sensitivity to western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 394-402.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science, 7*(3), 289-296.

Trehub, S. (2000). Human processing predispositions and musical universals. In N. Wallin, B. Merker & S. Brown (Eds.), *The Origins of Music.* (pp. 427-448). MIT Press.

Trehub, S. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences, 930*(1), 1-16.

Trehub, S. (2003). Toward a developmental psychology of music. *Annals of the New York Academy of Sciences, 999*(1), 402-413.

Trehub, S., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development, 55*, 821-830.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*(1), 73-99.

Trubetzkoy, N. S. (1931). Die phonologischen systeme. *Travaux Du Cercle Linguistique De Prague, 14*, 96-116.

Trubetzkoy, N. S. (1939). *Grundzuge der phonologie* (A. M. Baltaxe Trans.). Gottingen: Vandenhoeck & Reprecht.

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

Tsao, F. M. (2008). The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology, 50*(2), 111-124.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., & Koroshetz, W. J. (1997). Neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*, 266-276.

Volkova, A. (2004). *Long-term memory for absolute pitch level of songs in infancy.* Unpublished MA thesis, University of Toronto,

Wang, W. S. (1976). Language change. *Annals of the New York Academy of Sciences, 28*, 61-72.

Wang, W.S., Li, K.P. (1963) Tone 3 in Pekinese. *Journal of Speech and Hearing Research, 10*, 629-636.

Wang, W. S., Li, K. P. & Brotzman, R. L. (1963). *Research on Mandarin phonology*. http://www.dtic.mil/cgi-bin/GetTRDoc?AD=AD0409748

Wang, W. S. (1967). Phonological features of tone. *International Journal of American Linguistics, 33*(2), pp. 93-105. Retrieved from http://www.jstor.org/stable/1263953

Wang, W. S. (2012). Adventures with tones. In *Proceedings of Tonal Aspects of Language 2012,* Nanjing, China.

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic perception of Mandarin tones by Chinese and American listeners. *Brain and Language, 78*(3), 332-348.

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics, 24*, 113-129.

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*, 1289-1309.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development, 52*, 349-355.

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.

Wong, C. M. P., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*, 420-422.

Wong, P., C.M., & Perrachione, T., K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*(4), 565.

Wu, Z. (1982). Tonal variations in Mandarin sentences. *Chinese Linguistics, 6*, 439-450.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics, 5*, 757-797.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61-83.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*, 179-203.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61-83.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95*(4), 2240-2253.

Xu, Y., & Emily Wang, Q. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33*(4), 319-337.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America, 120*(2), 1063-1074. Retrieved from http://link.aip.org/link/?JAS/120/1063/1

Yeh, C. H. (2010). Comparison of phonetic naturalness between rising-falling and falling-rising tonal patterns in Taiwan Mandarin. In *Proceedings of Speech Prosody 2010,* Chicago

Yip, M. (1980). *The tonal phonology of Chinese*. Bloomington, Indiana: the Indiana University Linguistics Club.

Yip, M. (2000). The complex interaction of tone and prominence. Talk Given at *NELS 31,* Georgetown University.

Yip, M. (2001). Segmental unmarkedness versus input preservation in reduplication. In L. Lombardi (Ed.), *Segmental phonology in optimality theory* (pp. 206-230), Cambridge: Cambridge University Press.

Yue-Hashimoto, A. O. (1987). Tone sandhi across Chinese dialects, In Chinese Language Society of Hong Kong (Ed.), *Wang Li memorial volumes, English volume* (pp. 445-474), Joint Publishing Co., Hong Kong.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*(1), 37-46.

Zee, E. (1980). A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Linguistics.*

Zhang, J. (2007). A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*, *16*(4), 259-302.

Zue, W. (1976). Some perceptual experiments on the Mandarin tones. *Journal of Acoustical Society of America, 60*(S1), S45-S45.

Samenvatting

Dit proefschrift onderzoekt de mogelijke perceptuele voorkeuren ten gunste van het voorkomen van het fonologische proces van T3 toon sandhi in het Mandarijn. Er is overvloedig bewijs dat het idee ondersteunt dat luisteraars mogelijk een cruciale rol spelen in het vormen van fonologieën. Ohala (1981, 1986, 1993) bijvoorbeeld linkte de fonologische grammatica expliciet aan perceptie door te pleiten voor de cruciale rol die luisteraars kunnen hebben in het vormen van fonologieën. Er is cross-linguaal een neiging in klinkersystemen om die klinkerruimte in beslag te nemen die ervoor zorgt dat maximale contrasten worden behouden (Liljencrants & Linblom 1972, Lindblom & Engstrand 1989, Lindblom 1990). Ondertussen worden contrasten met lage perceptie veelal of gerepareerd voor een sterker contrast, of gefuseerd. In het eerste geval kunnen meerdere strategieën worden toegepast, zoals epenthesis, dissimilatie of metathesis (Hume & Johnson 2001). Betreffende lexicale tonen is er echter weinig onderzoek gedaan om de interactie tussen tonale grammatica en perceptuele voorkeuren aan te tonen.

In het Mandarijn worden lexicale tonen vooral gerealiseerd door variaties in toonhoogte, en tonen spelen een fonemische rol en onderscheiden lexicale betekenis. De vier lexicale tonen van het Mandarijn zijn hoog niveau (T1), mid-stijgend (T2), laag dalend (T3) en hoog vallend (T4). In het Mandarijn is er een fonologisch proces T3 sandhi waarbij een T3T3 volgorde verplicht verandert in een T2T3. T3 sandhi voorkomt het voorkomen van twee opeenvolgende T3 tonen, door op een asymmetrische manier alleen de eerste T3 te veranderen in een T2 terwijl de tweede T3 intact blijft. Het huidige proefschrift kijkt vooral naar mogelijke perceptuele voorkeuren die de Mandarijnse T3 sandhi regel kunnen vormgeven. Om de perceptuele motivatie te identificeren voor de hiervoor genoemde asymmetrie, zijn meerdere taken uitgevoerd, waaronder categorische perceptie experimenten met native en niet-native volwassenen (hoofdstuk 2) en snelle discriminatie experimenten met native en niet-native volwassenen (hoofdstuk 3), T2-T3 discriminatie experimenten met jonge kinderen (hoofdstuk 4 en hoofdstuk 6) en lexicale toon leerexperimenten met jonge kinderen (hoofdstuk 5). Bovendien worden de volwassen en jonge luisteraars ook getest op hun perceptie van muzikale toonhoogten, waardoor de perceptie van lexicale tonen niet alleen is bestudeerd door gebruik te maken van verschillende populaties, maar ook wordt de perceptie bekeken vanuit een domeinoverstijgend perspectief. Hierdoor is er ook gekeken naar hoe voorkeuren intrinsiek aan de menselijke perceptie in interactie staan met de taalspecifieke perceptie van lexicale tonen, en hoe zulke universele voorkeuren taalspecifieke fonologische grammatica's van lexicale tonen mogelijk vormgeven.

Hoofdstuk 2 demonstreerde dat volwassen native Nederlandse luisteraars Mandarijnse T2-T3 en T1-T4 contrasten op een psycho-akoestische manier waarnemen terwijl de native Mandarijnse luisteraars de contrasten categorisch waarnemen. Hoofdstukken 2 en 3 lieten zien dat ondanks het feit dat native Nederlandse luisteraars en Mandarijnse luisteraars verschillende manieren hebben om lexicale tonen te verwerken, beide groepen luisteraars dezelfde discriminatiepatronen lieten zien; beiden onderscheidden de T2 en T3 makkelijker als de T3 voor een T2 kwam in een te onderscheiden paar dan in de omgekeerde volgorde T2-T3; beiden onderscheidden de tweelettergrepige reeks T3T2 en T3T3 nauwkeuriger dan de reeks T2T3 en T3T3; beiden maakten een nauwkeuriger onderscheid als de T3T3 reeks als eerst kwam in een tweelettergrepige reeks in plaats van als laatst. Bovendien waren al deze asymmetrieën specifiek voor het T2-T3 contrast, omdat er in de T1-T4 contrast geen discriminatie asymmetrie is gevonden. Het feit dat volwassen Nederlandse luisteraars een asymmetrie vertoonden zonder enige eerdere blootstelling aan het Mandarijn en daardoor dus zonder enige kennis van een toonfonologie, dient als sterk bewijs dat er aangeboren perceptuele voorkeuren zijn die het onderscheiden van T2 en T3 in bepaalde volgorden beperkt.

In hoofdstuk 4 namen native Nederlandse jonge kinderen deel aan een Mandarijn T2-T3 discriminatie taak, waarbij een van de twee tonen diende als stimuli in de gewenningsfase. De resultaten lieten zien dat de Nederlandse jonge kinderen faalden in het onderscheid bij de leeftijd van 6 en 14 maanden, terwijl de kinderen van 9 en 11 maanden slaagden. Hierbij viel op dat de kinderen van 9 en 11 maanden alleen slaagden in het onderscheiden van T2 en T3 als ze waren gehabitueerd op T3.

In hoofdstuk 5 zijn native Nederlandse jonge kinderen getraind op natuurlijke

Mandarijnse spraak waarin of het van nature verwarrende T2T3 of het niet verwarrende T3T2 voorkwam. Na de training bleek dat de jonge kinderen de onderliggende T3T2 en T3T3 makkelijker van elkaar konden onderscheiden dan de onderliggende T2T3 en T3T3. Als men de bevindingen van de volwassen en jongere deelnemers samenneemt, lijkt het erop dat T3 sandhi zich richt op een contrast met een lage perceptie, namelijk het T2-T3 contrast, en dat perceptueel op elkaar lijkende tonen worden gefuseerd, namelijk de T2T3 en T3T3. Door deze perceptuele neutralisatie van T2T3 en T3T3 zorgt T3 sandhi voor een categorisch proces.

Gezien het feit dat variaties in toonhoogte op een betekenisvolle manier functioneren tussen verschillende domeinen, voornamelijk in spraak en muziek, heeft deze dissertatie ook gekeken naar de perceptie van Mandarijnse toonhoogten perceptie vanuit een domeinoverstijgend perspectief. Hoofdstuk 3 liet zien dat Nederlandse maar niet Mandarijnse luisteraars een positief significante correlatie toonden tussen de discriminatie van Mandarijnse tonen en de discriminatie van muzikale melodieën. De Nederlandse luisteraars die hoger scoorden op de muzikale taak scoorden echter alleen beter in de discriminatie van T2 en T3, maar niet in de discriminatie van T1 en T4. Zowel de luisteraars die hoog scoorden op de muzikale test als de luisteraars die laag scoorden lieten een discriminatie asymmetrie zien in de Mandarijnse T2 en T3.

Hoofdstuk 6 richtte zich op jonge kinderen en hun perceptie van lexicale tonen en muzikale melodieën. Aangaande muzikale perceptie bleken Nederlandse jonge kinderen al bij 4 maanden in staat om 3 noten melodieën te onderscheiden die verschilden in richting van toonhoogte alsook 3 noten melodieën te onderscheiden die maar één semitoon verschilden in absolute toonhoogte. Toch waren deze jonge kinderen niet in staat om de Mandarijnse T2 en T3 te onderscheiden van elkaar, of ze nu werden aangemoedigd om te letten op de pure akoestiek van de tonen (door het aanbieden van één token in de gewenningsfase) of dat ze hints kregen over de taalkundige functie van de tonen (door het aanbieden van meerdere tokens in de gewenningsfase). Bij de leeftijd van 6 maanden behielden de jonge kinderen hun gevoeligheid voor het discrimineren van de twee typen muzikale melodieën, terwijl hun gevoeligheid voor het onderscheiden van de Mandarijnse T2 en T3 groter werd. Samengevat suggereren deze bevindingen dat, ook al zijn jonge kinderen in staat om subtiele verschillen in toonhoogten waar te nemen, jonge kinderen niet in staat zijn om T2-T3 van elkaar te onderscheiden. Het niet kunnen onderscheiden van de T2-T3 valt dus niet te wijten aan alleen de richting van toonhoogte of akoestische saillantie. Het feit dat jonge kinderen van 6 maanden wel in staat waren om T2 en T3 te onderscheiden bij tokens gesproken door één spreker terwijl kinderen van 4 maanden hier niet toe in staat waren, suggereert dat jonge kinderen tussen 4 en 6 maanden zich meer bewust worden van de taalkundige functie van de fonologische categorie. Mandarijnse kleine kinderen van 6 maanden namen ook deel aan de T2-T3 discriminatie taak en de musikale absolute toonhoogte discriminatie taak. In beiden taken is geen discriminatie effect waargenomen, waaruit blijkt dat Mandarijnse jonge kinderen niet beter presteren dan hun Nederlandse leeftijdsgenoten, noch in de discriminatie van het Mandarijnse T2-T3 contrast, noch in de discriminatie van muzikale absolute toonhoogte. Deze resultaten suggereren dat immersie in een toontaal omgeving geen garantie is voor een makkelijkere verwerking van toonhoogten. Tevens is geen correlatie gevonden tussen muzikale melodie discriminatie en lexicale toon discriminatie, voor zowel de Nederlandse als Mandarijnse jonge kinderen bij 4 en 6 maanden. Samengevat suggereren de bevindingen in hoofdstuk 6 dat variaties in toonhoogte al op vroege leeftijd op een domein-specifieke manier worden verwerkt.

De bevindingen in hoofdstuk 4 en 6 die niet genoeg kunnen worden benadrukt is dat met dezelfde hoeveelheid aan variatie in de gewenningsfase, de Nederlandse jonge kinderen de T3 consequent makkelijker categoriseerden dan de T2. Een mogelijke reden hiervoor is dat hoe akoestisch specifieker een geluid is, hoe disctinctiever het is, en dus gemakkelijker te categoriseren. Omdat in de experimenten de T3 makkelijker werd gecategoriseerd, impliceert dit dat de T3 distinctiever is in perceptie dan de T2.

Al deze bevindingen samen genomen, lijkt het erop dat T3 sandhi in het Mandarijn wordt gemotiveerd door perceptie. Ten eerste is de tooncontour van de T3 complexer en dus akoestisch specifieker en hierdoor makkelijker waarneembaar. Om een toon waar te nemen die anders is dan de distinctieve toon is relatief makkelijk, waardoor een verandering van T2 en T3 nauwkeuriger werd opgemerkt als een T3 vooraf ging aan een T2. Omdat een T2 minder complex is dan een T3, kan een T3 ten onrechte worden waargenomen als een vorm van een T2. Ten tweede zijn de tweelettergrepige toonreeksen T2T3 en T3T3 bijzonder gelijk in perceptie. Door de eerder genoemde verwarring tussen T2 en T3, zou het kunnen dat de tonen in T2T3 worden waargenomen als identiek aan elkaar. Het feit dat de discriminatie van T2T3/T3T2 van T3T3 nauwkeuriger is als de T3T3 eerst voor komt, is overeenkomstig met de bevindingen in de eenlettergrepige taak van T2 en T3. Hierdoor tolereert T2T3 meer variatie en luisteraars kunnen T3T3 waarnemen als een vorm van T2T3. Omdat T2T3 en T3T3 perceptueel erg gelijk zijn en luisteraars eerder geneigd zijn om T3T3 als een T2T3 aan te zien dan andersom kan dit ervoor zorgen dat T3 sandhi plaatsvindt op de manier waarop het nu doet; T3T3 wordt geassimileerd tot T2T3, aangezien de alternatie niet duidelijk waarneembaar is.

Om de bevindingen in een breder perspectief te plaatsen, lijkt het erop dat de luisteraars een cruciale rol spelen in het vormen van de fonologische grammatica van lexicale tonen, aangezien aangeboren perceptuele voorkeuren waarschijnlijk de druk creëren voor het doen voorkomen van T3 sandhi. Met andere woorden, de taalspecifieke toongrammatica functioneert ten behoeve van de menselijke gehoorperceptie; een tooncontrast van lage perceptie wordt gefuseerd in de specifieke context waar het contrast het minst waarneembaar is. Tot dusver passen mijn bevindingen in een emergentistische opvatting in de zin dat een consequente verwarring asymmetrie optreedt onder native en niet-native volwassen luisteraars en onder jonge kinderen. Anders gezegd, de fonologische grammatica waaronder T3 sandhi, past in de functionele behoeften van mensen, omdat het fuseren van T2T3 en T3T3 de perceptuele belasting vermindert bij het onderscheiden van twee elementen met lage waarneembaarheid (e.g. Steriade 1995, 1999, Hume & Johnson 2001, Huang 2001, Seo 2001). Bovendien zou het het geval kunnen zijn dat evolutionair T3 sandhi zijn oorsprong kent in de misperceptie van luisteraars (Ohala 1981, 2012). Deze overwegingen ondersteunen de opvatting dat aangeboren perceptuele voorkeuren de formatie van fonologie inperken, zowel in segmenten als in lexicale tonen. Met andere woorden, zowel segmentale fonologie als prosodie kunnen worden gestructureerd op manieren die het meest geschikt zijn voor het menselijk perceptuele systeem.