The processing of non-native word prosodic cues

A cross-linguistic study

Published by
LOT
Kloveniersburgwal 48
1012 CX Amsterdam
The Netherlands

phone: +31 20 525 2461
e-mail: lot@uva.nl
http://www.lotschool.nl

ISBN: 978-94-6093-366-0
DOI: https://dx.medra.org/10.48273/LOT0556
NUR: 616

The processing of non-native word prosodic cues

A cross-linguistic study

De verwerking van niet-native prosodische woord-cues:
Een cross-linguïstisch onderzoek
(met een samenvatting in het Nederlands)

Proefschrift

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

donderdag 7 januari 2021 des ochtends te 10.30 uur

door

Shuangshuang Hu

geboren op 22 oktober 1988
te Shanghai, China

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

Copromotor: Dr. A. Chen

This dissertation was partly accomplished with financial support from the China Scholarship Council (CSC).

CONTENTS

Acknowledgements

This dissertation would not have been completed without the great support and help of a lot of people. I hope my humble words can express my sincere and deep gratitude.

First and foremost, I would like to express great gratitude for my promoters René Kager and Ao Chen. Thank you both for your meticulous supervision and great support throughout my PhD research, for always being available (although you were very busy) giving me valuable feedback when I had new idea or new results and when I needed advice, for pinpointing the core questions in our meetings, for inspiring me and for teaching me how to become an independent researcher. I would like to thank René for the constant encouragement and comprehensive instruction ever since my Masters. I still remember enthusiastically talking about cross-linguistic word prosody with you on a bus after phonology class at nightfall in the winter of 2013. That's where the research interest that birthed my PhD dissertation began. I benefited a lot from your critical thinking, immense knowledge and insight into phonology and prosody and enthusiasm for research. Thanks also to Ao for the kind support not only with respect to research but also generally in life. Thanks for discussing the dissertation with me even when your baby boy was crying for milk. Thanks for your insightful advice, especially regarding the details of the experimental design and analysis. I was always inspired by your critical and careful thinking from the perspective of psycholinguistics. Thanks for sharing your valuable experience with me regarding the pursuit of a research career. I am so grateful and lucky to have you two as my promoters. You both have shown me what it means to be a great researcher.

Thanks to the China Scholarship Council and UiL-OTS for financially supporting my PhD research and for giving me opportunities to attend courses and conferences.

I owe many thanks to the UiL-OTS lab, as well. Thanks to Maarten Duijndam, Chris van Run and Jacco van Elst for programming the experiments. Thanks to Iris Mulders for helping me to recruit Dutch participants and for suggestions on practical issues with respect to conducting experiments. Thanks to Suki Yu and Ziyi Bai for helping me with the recording of the stimuli. My deep thanks go to all the participants for making the PhD dissertation happen.

Thanks to Kirsten Schutter, Piet van Tuijl, and Hugo Quené for helping me with statistics. Thanks to Xiaochen Zheng for answering my statistics questions even when you were relaxing and drinking beers in a bar on a Friday night. I owe special thanks to Huub van den Bergh. Thank you for always leaving your office door open, welcoming me with my endless statistics questions, for getting back to my emails so quickly, for showing me the fun and beauty of statistics, and for sharing that UFO anecdote with me. I enjoyed talking with you so much. You have inspired and encouraged me a lot.

Thanks to the UiL-OTS coordinator Maaike Schoorlemmer who is always ready to help and who has really helped me a lot. Also thanks to our secretaries, Yvonne van Adrichem, Mariëlle Hilkens and Silvia de Pascalis, and secretaries at the International Office, for their efficiency and help.

Thanks to my lovely officemates Heidi Klockmann, Jorik van Engeland, Silvia Rădulescu (although most times we met in the lab), Maartje Schulpen, Andrea Santana Covarrubias, Marjolein van Egmond, Joanna Wall, Quy Ngoc Thi Doàn, Myrthe Bergstra, Ni Cheng, Sofiya Ros, Anika van der Klis, Eva Poortman and Mariano González for making the working environment so chill and full of fun. Thanks to Emma Everaert for sitting with me at the face-to-face worktable in the office, for your constant help (e.g., serving as my translator for Dutch/English) and for our joyful talks at Tai-Soen Dim Sum, Sanju Ramen, Kounosuke-Don and Taj Mahal restaurant. Thanks for being my paranymph. We still have an Indonesian restaurant to try.

Thanks to my UiL-OTS colleagues, Martijn van der Klis, Imke Kruitwagen, Nina Sangers, Mengru Han, Luying Hou, Ying Liu, Jingwei Zhang, Fang Li, Stavroula Alexandropoulou, Jan Winkowski, Anja Goldschmidt, Rachida Ganga, Iris Selten, Silvia Terenghi, Luana Sorgini, Alberto Frasson, Giada Palmieri, Marlissa Hommel, Illeana Grama, Dominique Blok, Tijn Schmitz, Sonya Nikiforova, Konstantinos Kogkalidis, Bjorn t'Hart, Jolien Scholten, Lisa Bylinina, Brigitta Keij, Anne-France Pinget and Anna Volkova. I enjoyed our nice talks during lunchtime and tea breaks. Thanks to Xin Li for showing me the photos of your rabbit, your chinchillas and your dog to bring me levity and cuteness. Thanks to Zenghui Liu for your encouragement and for always being so sweet and bright. Thanks to Yipu Wei for being my museum companion and for bringing me a spark of art in the research life. How I missed our trip in Denmark during Christmas. Thanks to Jeroen Breteler for sharing your passion for phonology with me, even when we were hungry on our way to a restaurant in Tokyo. Thanks to Willemijn Heeren for your detailed instruction in Praat and valuable advice regarding stimuli design. Thanks to Tom Lentz and Cesko Voeten for discussing experimental paradigms with me. Thanks to Liquan Liu, Shanshan Fan and Zhen Qin for your inspiring feedback on the experiment results. Thanks to Mengxi Cao for designing the fantasy pictures used as stimuli. I owe many thanks to Yuan Xie, Jianan Liu, Na Hu and Chou Mo for your kindness and generosity, for the cheerful chatting we had together and for always being there, helping me without any hesitation. Thanks to Xiaoli Dong, Bert Le Bruyn, Simon Le Bruyn and Anqi Yang for the care and concern, for the fun and laugh, for sharing with me valuable suggestions and experience with respect to research and for always helping and supporting me in so many aspects in daily life. I treasure the time we spent in Plesmanlaan.

Part of the PhD research was carried out in Japan. I would like to thank a group of people for their support and help during my stay in Japan. My special thanks go to Haruo Kubozono for hosting me at the National Institute for Japanese Language and Linguistics in Tokyo, Japan, for introducing me to the NINJAL family, for all the nice talks during lunchtime and at welcome dinner and for your comprehensive and insightful instruction regarding Japanese phonology. The PhD dissertation would not have been able to be completed without your great help. Thanks to Seunghun Lee for welcoming me to conduct experiments at International Christian University. Thanks to the NINJAL colleagues Chisato Ohno, Shu Hirata, Misato Ido and Ai Mizoguchi for helping me not only with the recruitment of Japanese participants but also generally in life in Tokyo. Thanks to Audrey Lai for assisting me with the experiments conducted on campus. Thanks to Manami Hirayama, Yukiko Sugiyama, Mayuki Matsui, Hyunkyung Hwang and Yuki Hirose for the kind suggestions and discussion regarding the experiments and participant recruitment. Thanks to Guofeng Sun, Xiling Liu and Chen Dai at the Inter University Seminar House and my friend Yu Suzuki for your help and support.

Also thanks to Sharon Peperkamp for welcoming me to conduct experiments at Laboratoire de Sciences Cognitives et Psycholinguistique in Paris and for your kind suggestions regarding participant recruitment, although due to practical issues (e.g., the tight schedule), I didn’t manage to collect data of French participants. I look forward to completing the data collection of French group in the future.

Publishing the dissertation was not “easy-peasy”. I would like to thank Sara Court for proofreading the manuscript so efficiently, to WordWizzard Caicaizi for helping me greatly with the dissertation format and to Janacy van Duijn Genet and Jan van Es for being so responsible when publishing the dissertation. Thanks to Di Wu for designing the creative cover page with me.

During the five years of doing PhD, I would like to especially thank my buddies Arnold Kochari, Zhilei Gao and Rachel Gargiulo. Arnold and Rachel, we seemed to immediately become friends on the first day of our Masters program. I cherish our time studying together and having fun together. Arnold, my paranymph, we even got our PhD offers on the same day and started our programs on the same day. Thank you for walking together with me on the road of research, for always encouraging me throughout all the stages of my PhD, and for caring about whether I was happy. Rachel, my chef, thank you for sharing with me the shining moments in your life, for showing me the garden you planted and the art you painted, for helping me with my English and for always making me laugh. Zhilei, my Tietie, thank you for coming to talk to me at the PhD welcome party. That’s how our friendship began. Thank you for celebrating our birthdays together with me, for always saying “take it easy, don’t worry, my Tie” to me, for being there supporting me while I was writing my dissertation and generally in life. You don’t know how grateful I am. Thank you, my buddies.

PhD is more than research. I am grateful to have dear friends around me. I owe special thanks to my dearest friends Wanjing Lu, Yingchao Wu, Milktea Fairy IF-Zhu, Shijia Cao, Shuai Wāng, Shuai Wáng, Pei Liu, Xuelian Wan, Yuelu Wu, Linsha Xu, Yao Chen, Zongxi Li, Shulu Wan, Spitzfan, Di Bao, Yiwei Chen and Xiaoyun Zhu for your great support in so many things, for comforting me and cheering me up, for showing me so much fun outside academia and for sharing the joys and worries of life with me.

My thanks also go to the Chinese friends I met in the Netherlands: Yingtong Liu, Yan Xia, Yuchen Li, Yulong Zhao, Siyang Kong, Yu Zhang, Yiwei Sun & Jun Xiao, Jiaxin Li, Jane Zhang, Ting Zou, Yang Yang, Menghui Shi, Jie Fu, Jie Hu, Xuan Wei, Jia Li, Yanjiao Zhu, Haoyan Ge, Shuwen Chen, Yue Song, Yachan Liang, Yuan Tian and Lei Sun, for inviting me for lunch/dinner/“family feast” and for all your support and encouragement. Thanks to Maya Kuswaya, my housemate, for your help and for sharing with me the delicious Bunda-made Indonesian cuisine.

Last but not the least, I owe the deepest gratitude to my parents who have been supporting me this entire time, and for allowing me to be who I am and do what I like. You are always and will forever be my shelter. I love you more than I can say.

感谢我的家人与亲朋好友的理解、爱与支持。感谢我的父母。

To my parents

至我的父母

Chapter 1 Introduction

1.1 Dissertation goals

Languages differ in their use of prosodic cues to differentiate word meanings. For instance, Mandarin Chinese (henceforth Mandarin), a canonical tone language, exploits pitch variations on each syllable to mark lexical contrasts. Word meaning changes as pitch patterns change. Pitch patterns, based on acoustic perceptual features, consist of two dimensions: pitch level and pitch contour. Four pitch patterns, namely, high-level tone (Tone 1), rising tone (Tone 2), dipping tone (Tone 3) and falling tone (Tone 4) can be carried by a syllable in Mandarin (Chao, 1948; Ho, 1976; Howie, 1976; Lin, 1988). For instance, the meaning of the syllable /ma/ changes in terms of pitch patterns Tone 1, Tone 2, Tone 3 and Tone 4 as “mother”, “numb”, “horse”, and “to blame”, respectively.

Different from Mandarin, Dutch uses word stress to differentiate words. One syllable of any multi-syllabic word is typically marked with higher stress than the other syllable(s) (Kager, 1989; Booij, 1995; Trommelen & Zonneveld, 1999; Rietveld et al., 2004). The meanings of words with identical segments can then be signaled by the position of stress. For instance, *VOORnaam*[1] means “first name” when the stress is located on the first syllable, while *voorNAAM* means “distinguished” when the stress is on the second syllable.

While Mandarin and Dutch exploit pitch variations and word stress position, respectively, in word prosody, Tokyo Japanese (henceforth Japanese) uses the position of pitch accent to contrast lexical meanings. The meaning of a word in Japanese is determined by the presence or absence of an abrupt pitch fall (termed “pitch accent”) and, if present, by the position of pitch accent in any multi-moraic word (Poser 1984; Haraguchi, 1999; Kubozono, 2008; Kawahara, 2015). For instance, the disyllabic (bimoraic) word “hashi” /haʃi/ signals three lexical meanings depending on the presence and position of the pitch accent: “chopsticks” when the pitch accent falls on the first syllable, “bridge” when the pitch accent falls on the final syllable and “edge” when there is no pitch accent.

Prima facie, the three languages discussed above employ different word-prosodic cues (WPCs) in their lexicons. Mandarin uses lexical tones consisting of pitch contour contrasts. Dutch employs the position of word stress, while Japanese uses the position of lexical pitch accent. Nonetheless, the WPCs they use do share commonalities. For instance, Mandarin and Japanese use lexical pitch: pitch variations in Mandarin and pitch accent in Japanese. Japanese and Dutch use the

[1] A capitalized syllable represents that the syllable is stressed.

abstract feature of positional marking: position of pitch accent in Japanese and position of stress in Dutch. Therefore, in terms of the use of WPC, the three languages can be arranged into two intersecting sets, one for lexical use of pitch, and the other for lexical use of position, as arranged in the Venn diagram in Figure 1.

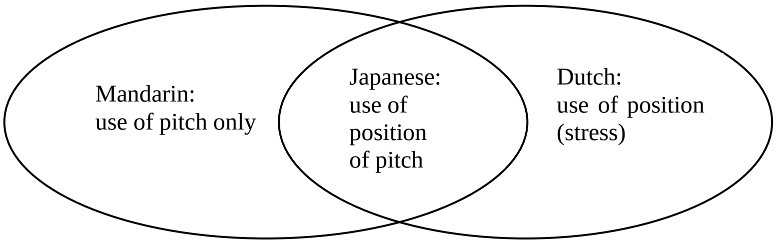


Figure 1.1. *Employment of WPCs in Mandarin, Dutch, and Japanese*

The use of lexical pitch and positional marking in the various languages plays a role in native WPC perception (Hallé et al., 2004; Xu et al., 2006; Liu et al., 2017; Ayusawa et al., 1995; Sugiyama, 2008; Van Heuven, 1985 among others). How do listeners from different language backgrounds perceive non-native pitch-based and position-based cues? How does native word prosody influence perception of non-native WPCs? An extensive body of studies has investigated perceptions of non-native pitch contrasts (lexical tones) in isolated syllables cross-linguistically (Lee et al., 1996; So & Best, 2010; Schaefer & Darcy, 2014; Burnham et al., 2015, among others). These studies reveal language-specific perceptual patterns across listeners from different language backgrounds. However, comparatively less is known about the role of native prosody in the perception of non-native positional marking from a cross-linguistic perspective. Studies on perception of non-native positional marking are very few and have focused on perception of word stress or lexical pitch accent position. For instance, Dupoux et al. (1997) found naive French listeners, who do not use any WPCs in their native language (Rossi, 1980; Vaissière, 1991; Di Cristo, 1998), were able to discriminate contrastive stress position in Spanish. Hu (2015) observed that naive Dutch listeners who use stress lexically were good at discriminating Japanese pitch accent position. It is not fully understood how native word prosody contributes to perceiving non-native positional marking from a cross-linguistic perspective. Specifically, no study to date has elucidated the role native word prosody plays in the perception of non-native pitch contrasts with positional marking, i.e., non-native tone contrasts in position and non-native pitch contrasts occurring in different positions in a word.

Furthermore, previous studies investigating the processing of non-native pitch contrasts and non-native positional marking have mainly centered at the *acoustic* level. However, prosodic processing does not only involve sensory-auditory processing (i.e., processing at the acoustic level) but also involves higher linguistic levels such as the *phonological* level, a more abstract level than the acoustic level where acoustic traces that listeners might rely on are inaccessible (Dupoux et al., 2001, 2008) and the *lexical* level where phonological knowledge is encoded to build lexical representations in word learning (Wong & Perrachione, 2007; Zatorre & Gandour, 2008; Wang & Cooper, 2013; Braun et al., 2014). Compared with cross-linguistic studies on acoustic perception, there is a paucity of studies investigating the perception of non-native WPCs at the two higher linguistic processing levels. For instance, naive French listeners were found able to discriminate contrastive Spanish stress acoustically but failed to process it phonologically (Dupoux et al., 1997, 2001, 2008). Braun et al. (2014) reported that Japanese listeners who were good at perceiving non-native tones acoustically (So & Best, 2010; Schaefer & Darcy, 2014) had difficulty establishing lexical representations of non-native tones mapped with meaning. Crucially, these findings suggest that the capability of perceiving non-native tones or positional pitch marking acoustically does not entail a successful processing at the phonological level or guarantee success in encoding these properties in lexical representations. So far, no study has taken a cross-linguistic perspective to address how listeners from different language backgrounds perceive non-native pitch contrasts with consideration of position at the two linguistic levels or the extent to which native word prosody modulates phonological processing and encoding of non-native WPCs when establishing lexical representations in word learning.

Taken together, the overarching goal of this dissertation is to better understand the role of native word prosody in the processing of non-native WPCs (i.e., non-native pitch contrasts and position), not only at the *auditory-acoustic* level but also at the *phonological* level and *lexical* level (i.e., in word learning). Mandarin, Japanese, and Dutch listeners are selected as participants in the study since the three languages both differ and yet share in certain common properties with respect to their use of WPCs. Hence, the dissertation attempts to address *language-specificities* but also *language-commonalities* relevant to the perception of non-native WPCs at three processing levels within one single study. This helps us to develop a fuller understanding of the dynamic function of WPCs used in human languages across low to high linguistic processing levels and of the ways in which phonological knowledge is used in sound-to-meaning mapping. To this end, three empirical studies will be conducted with respect to the three processing levels in Chapter 2 (acoustic level), Chapter 3 (phonological level), and Chapter 4 (word learning). The current introductory chapter will provide a literature review, elaborate methodologies applied in the study, put forward research questions, and outline the subsequent dissertation chapters.

1.2 Perception of non-native WPCs at the acoustic level

In this section, I will review previous studies on the perception of non-native WPCs (pitch contrasts and position) involved in *identification* and *discrimination* tasks. Identification and discrimination tasks are two general types of tasks used to examine the perception of phonetic contrasts (e.g., segmentals and suprasegmentals)

(Strange & Shafer, 2008). In an identification task, listeners are presented with recorded stimuli one at a time and are required to identify or label each stimulus as an instance of a discrete phonetic category. This task tends to encourage an acoustic mode of listening (Strange & Shafer, 2008). In a discrimination task, two (AX) or more (ABX) stimuli are presented, and listeners are asked to judge the relationship between the stimuli, e.g., whether the stimuli are the same or different. In an AX discrimination task, for example, listeners should determine whether or not target stimulus X matches item A. In an ABX (or AXB) task, listeners should judge whether a target stimulus X matches item A or item B. Compared with ABX tasks, the AX task imposes less *memory load* on listeners in that listeners retain less of an auditory trace of the previous stimulus in short-term memory storage and can thus make a same-or-different judgment based on an immediate acoustic comparison of the two stimuli (Logan & Pruitt, 1995; Strange & Shafer, 2008). In ABX tasks, listeners have to retain more acoustic information of the stimuli, which may, to some extent, tap into the listener's phonological representations (Dupoux et al., 1997). However, ABX tasks are still regarded as phonetically oriented in that the memory load on listeners is relatively small and listeners can still rely on detailed acoustic information in the stimuli (Strange & Shafer, 2008). This is unlike tasks designed for phonological processing such as the sequence recall task, which has a relatively high memory load and much phonetic variability, rendering acoustic traces inaccessible (Dupoux et al., 1997; Dupoux et al., 2001; Strange & Shafer, 2008) (to be discussed in detail in §1.3). Therefore, in the current section ABX discrimination tasks are regarded as tasks that deal with perception at the acoustic level, together with the AX task and identification tasks.

Section 1.2.1 will review studies that have investigated perception of non-native pitch contrasts (pitch contour and pitch level) at the acoustic level. Section 1.2.2 will discuss studies on the perception of non-native positional contrasts at the acoustic level. Each section will focus on investigations into Mandarin, Japanese, and Dutch listeners and compare the three language groups with each other.

1.2.1 Perception of non-native pitch contrasts at the acoustic level

1.2.1.1 Perception of non-native pitch contrasts at the acoustic level by Mandarin listeners

Empirical studies have revealed that linguistic experience with lexical tones in one's native language plays an important role in a listener's acoustic identification and discrimination of non-native tones (Lee et al, 1996; Burnham et al., 1996; Gottfried & Suiter 1997; Burnham & Francis, 1997; Wang et al., 1999; Wayland & Guion, 2004; Wu & Lin, 2008; Schaefer & Darcy, 2014; Burnham et al., 2015; Reid et al., 2015). The current section will review studies that investigated Mandarin listeners' perception of non-native pitch contrasts (pitch level/pitch contour) and will compare Mandarin listeners with listeners from different language backgrounds.

It has been shown that having lexical pitch in their native language can facilitate Mandarin listeners when perceiving non-native pitch contrasts. For instance, Schaefer & Darcy (2014) investigated Mandarin and Japanese listeners, whose native languages have lexical pitch, in comparison with English and Seoul Korean listeners, whose native languages do not use lexical pitch, in discriminating Thai tonal contrasts (pitch contour contrast (rising vs. falling), pitch level contrast (low vs. high, low vs. mid and high vs. mid), and mixed contrast (pitch contour vs. pitch level)) in an AXB discrimination task. They found that Mandarin listeners showed a significant perceptual advantage over listeners who do not use lexical pitch (i.e., English and Seoul Korean listeners). Likewise, Burnham et al. (2015) compared Cantonese, Mandarin, Swedish, and English listeners discriminating Thai tones in an AX task. They reported that Mandarin listeners, together with Cantonese and Swedish listeners who both also use lexical pitch, showed an overall better performance than English listeners without the use of lexical pitch. Their findings show that prior experience with native lexical tones benefits Mandarin listeners in non-native tone perception. Furthermore, neurophysiological studies have observed that Mandarin listeners perceiving Thai tones manifest a more sensitive brain stem mechanism for pitch representation compared with English listeners, reflected by more pitch strength and better pitch tracking accuracy (Krishnan & Gandour, 2009; Krishnan et al., 2009; Krishnan et al., 2010; Krishnan et al., 2005).

On the other hand, various other studies have shown that the employment of lexical pitch in one’s native language does not necessarily benefit perception of non-native pitch contrasts (Qin & Mok, 2013; Lee et al., 1996; Francis et al., 2008; Li & Shuai, 2011; Reid et al., 2015). For instance, Qin & Mok (2013) examined Mandarin, English, and French listeners perceiving Cantonese tonal contrasts in an AX discrimination task and found that Mandarin listeners had difficulty discriminating Cantonese level tones (mid level tone vs. low level tone). Francis et al. (2008) investigated Mandarin listeners and English listeners’ identification of six Cantonese tones (high level tone, high rising tone, mid level tone, low falling tone, low rising tone and low level tone) by an identification task where listeners heard a tone and had to choose the corresponding visualized line pattern representing the tone (e.g., a high flat line pattern represented high level tone). They found that Mandarin listeners showed confusion in identifying the two rising tones (low rising tone and high rising tone) and two level tones (mid level tone and low level tone). In another study, Reid et al. (2015) conducted a category assimilation task where listeners had to map the non-native Thai tone they heard to a native tonal category. They reported that Mandarin listeners perceived Thai high level tone and mid level tone both as the native high level tone. The perceptual difficulties observed in Mandarin listeners might be due to a lack of one-to-one mapping between the non-native tone and the native tone category (Qin & Mok, 2013; Francis et al., 2008; Reid et al., 2015; Burnham et al., 2015). According to the Perceptual Assimilation Model (PAM) proposed by Best (1995) (See also So & Best, 2014; Best & Tyler 2007; So & Best, 2010), if the non-native phonemic contrast can be mapped onto two distinctive categories in the native language, it will yield good discrimination. If the non-native phonemic contrasts are mapped onto one single category in the native language, it will bring about poor discrimination. For instance, the difficulty discriminating between the Cantonese two level tones by Mandarin listeners in Qin & Mok (2013) could be explained as resulting from a mapping of the two Cantonese level tones onto a single category, the high level tone, in Mandarin.

Furthermore, it was observed that Mandarin listeners assigned different weights to acoustic features of pitch in non-native tone perception, compared with listeners from other language backgrounds (Gandour & Harshman, 1978; Gandour, 1983; Francis et al., 2008; Li & Shuai, 2011). For instance, Gandour (1983) examined tone language listeners, Cantonese and Mandarin listeners, and non-tone language listeners, English listeners, perceiving tones (synthesized artificial tones) that differed in pitch contour and pitch level. Listeners had to rate the similarity between the two tones they heard on a 1 (least similar) to 9 (most similar) scale. Via multi-dimensional scaling (MDS) analysis, he found that Mandarin listeners were more sensitive to pitch contour changes than to pitch level changes while English listeners attended more to pitch level than to pitch contour changes. Francis et al. (2008) conducted an identification task where listeners had to identify the tone they heard by choosing a visualized pitch pattern and rate the degree of similarity of the two tones on a 1-10 scale. They observed similar findings that Mandarin listeners attended more to pitch contour contrasts than to pitch level contrasts in non-native Cantonese tone perception and they further proposed that such perception is determined by the relative weight assigned to specific tone features, which in turn is determined by the demands of pitch contrasts in the native language. According to Francis et al. (2008), Mandarin uses lexical tones that are contrastive in pitch contour instead of pitch level, which is why Mandarin listeners give more weight to pitch contour contrasts than to pitch level contrasts during non-native tone perception.

To sum up, the use of lexical pitch in the native language influences Mandarin listeners' perception of non-native pitch contrasts, either as facilitation or as interference. By and large, Mandarin listeners were found to be sensitive to non-native pitch contrasts (e.g., Thai tones, Cantonese tones) and showed an overall advantage over listeners who do not employ lexical pitch in their native language (such as English listeners and Korean listeners) (Wayland & Guion, 2004; Schaefer & Darcy, 2014; Krishnan, Swaminathan et al., 2009; Krishnan et al. 2010; Gandour, 2010). However, Mandarin listeners' non-native tone perception is also influenced by the relationship between the specific tone set in their native language compared to that in the non-native language. Their perception was found to be impeded when there was no one-to-one mapping between the non-native tones and the native tone categories (Francis et al., 2008; Qin & Mok, 2014; So & Best, 2010; Li &Shuai, 2011; Reid et al., 2015). Moreover, the use of contrastive lexical pitch contours in the native language determines their sensitivity to non-native pitch contour changes, more so than to pitch level changes (Gandour, 1983; Francis et al., 2008; Abramson, 1979; Xu et al. 2006).

1.2.1.2 Perception of non-native pitch contrasts at the acoustic level by Japanese listeners

The current section will review studies on the perception of non-native pitch contrasts by Japanese listeners and compare them with listeners from other language backgrounds.

Japanese uses an abrupt pitch fall, termed “pitch accent”, to signal lexical contrasts in a prosodic word (Poser, 1984; Kubozono, 2008; Kitahara, 2001; Kawahara, 2015). More specifically, pitch accent is used in two ways to differentiate word meanings: first, whether or not it is present, and second, if present, where it is located (Haraguchi, 1999; Poser, 1984; Vance, 1987; Kubozono, 2008; Kawahara, 2015). For instance, the disyllabic (here, also bimoraic) word “hashi” /haʃi/ signals three meanings: “chopsticks” (accent on the first syllable), “bridge” (accent on the final syllable) and “edge” (no accent). The accentual patterns are mapped onto surface tonal patterns, that is, in the presence of the accent (marked with an apostrophe (’)). Generally speaking, for words with light syllables (i.e., one syllable contains one mora (C)V)[2] , if the accent falls on the first mora, the accented mora receives a high pitch (H) and the following mora receives a low pitch (L). If the accent falls on the second or later mora, the first mora receives a low pitch and the moras from the second until the accented one all receive a high pitch. When the word is unaccented, the first mora has a low pitch and the remaining moras have a high pitch (Haraguchi, 1999; Kubozono, 2008; Uwano, 1999). Japanese lexically uses only two levels of tonal heights H and L[3] (Kawahara, 2015). Continuing with /haʃi/ (“hashi”) as an example, according to its accentual patterns it has three tonal patterns: H’L (accent on the first syllable), LH’ (accent on the final syllable) and LH (unaccented). Note that minimal pairs due to a contrast in the position and/or absence/presence of pitch accent in Japanese are limited overall, accounting for approximately 20% of all minimal pairs in the language according to Pierrehumbert and Beckman (1988) (see also Wu et al., 2012), or 14% of all minimal pairs according to Shibata & Shibata (1990).

There has been debate about whether languages such as Japanese should be characterized in the typology of word level prosody as a restricted type of tone language or as forming a distinct category of pitch accent languages (McCawley,

---

[2] Note that the rules for tonal patterns discussed here do not include heavy syllables (i.e., one syllable contains two moras). In Tokyo Japanese, except initially-accented, the first two moras (i.e., heavy syllables) can both receive high pitch (Haraguchi, 1977, 1991; Vance, 1987, 2008.)

[3] McCawley (1968) used mid tone (M) to represent downstepped H, a lowered H tone following another H tone. However, it is safe to say that at the lexical level Japanese makes use of only two level tones H and L tone (Kawahara, 2015).

1977, 1978; Beckman, 1986; Ladd, 1996; Hyman, 2006, 2009). The term "*pitch accent*" used in the current dissertation should not be taken to mean that Japanese is regarded as a pitch accent language, but rather as "a subclass of tone languages" (Yip, 2002: 2). Following Hyman (2006, 2009), the term "tone languages" refers to languages in which "pitch enters into the lexical realisation of at least some morphemes" (Hyman, 2006:229), including canonical tone languages such as Mandarin and Cantonese as well as tone languages with a restricted tonal system such as Tokyo Japanese. Note that despite the employment of pitch for lexical contrasts in Japanese, pitch is used in a different manner than in canonical tone languages. In canonical tone languages such as Mandarin, almost[4] each syllable bears pitch variations (tones) that can contrast lexical meanings (Yip, 2002), while in Japanese it is the relative pitch levels between two successive syllables[5] (the HL pitch fall) that is important for differentiating lexical items (Kitahara, 2001).

Studies[6] on Japanese listeners' perception of non-native pitch contrasts yield mixed results. Some studies have reported that due to the employment of lexical pitch accent, Japanese listeners were perceptually sensitive to non-native tones and showed perceptual advantage over listeners who do not employ lexical pitch in their native language (So, 2010, 2012; So & Best, 2010, 2012; Schaefer & Darcy, 2014; among others). For example, Schaefer & Darcy (2014), using an AXB discrimination task, studied perception of Thai tones contrasting in pitch level and pitch contour by Mandarin listeners, Japanese listeners, English listeners, and Seoul Korean listeners. Their findings showed that Japanese listeners had a significantly better performance than the non-tone language listeners (English listeners and Seoul Korean listeners). Interestingly, Japanese listeners were found to be only marginally[7] less accurate than canonical tone language listeners, Mandarin listeners, in discriminating non-native Thai tones, which suggests the pivotal role lexical pitch plays in non-native tone perception (Lee et al. 1996; Wayland & Guion 2004; Burnham et al., 2015). Their findings are in line with So & Best (2010) who found that Japanese listeners performed as well as canonical tone language listeners, Cantonese listeners, and outperformed English listeners in identifying Mandarin tones.

However, it is not always the case that prior experience using lexical pitch accent in the native language facilitates Japanese listeners' perception of non-native pitch contrasts. For instance, Wong (2019) conducted an AXB discrimination task to investigate perception of Cantonese tones by Japanese listeners. She found that

[4] There are a small number of syllables in Mandarin that don't bear tones, the neutral tones, which will be briefly discussed in §1.2.2.1.

[5] "Syllable" in this dissertation refers to a monomoraic (C)V syllable.

[6] How Japanese listeners perceive native pitch accentual patterns will be discussed in §1.2.2.2 regarding the perception of position since Japanese accentual patterns are signaled by the presence and position of pitch accent.

[7] The difference in perceptual performance between Japanese and Mandarin listeners was marginal ($p = 0.087$) (Schaefer & Darcy, 2014: 502).

Japanese listeners had difficulty discriminating low rising (T5) and high rising (T2) tones in Cantonese, presumably due to a lack of distinctive contour categories in Japanese word prosody. So & Best (2010) examined Japanese listeners' perception of Mandarin tones. They applied a four-choice identification task where listeners should identify the correct tone out of four (all represented in pitch patterns, e.g., high level tone T1 was represented as a flat line pattern) and reported that Japanese listeners displayed confusion when identifying specific tones, that is, they misidentified the Mandarin falling tone T4 as a rising tone T2. The findings contradicted So (2010) in that Japanese listeners were able to assimilate Mandarin T2 and T4 onto their native accentual categories LH and HL. The authors argue that Japanese listeners' confusion when identifying Mandarin T4 and T2 could be due to a preference for rising pattern in Japanese disyllabic (bimoraic) words. Disyllabic accented words in Japanese can have either a LH' or H'L tonal pattern which resemble rising and falling tones, respectively, in Mandarin (So & Best, 2010). However, when the same word is unaccented, Japanese speakers showed a tendency to produce the word with a rising pattern on the last syllable (Fujisaki et al. 1996; Cutler & Otake 1999; Nagano-Madsen, 2003). The preference for the rising pattern in the native language could lead to misidentifying T4 as T2 in Japanese listeners.

Moreover, it was found that native word prosody determines Japanese listeners' perceptual weight assigned to non-native pitch contrasts. Guion & Pederson (2007) examined Japanese listeners' and English listeners' perception of nine synthesized tones contrasting in pitch level (low vs. mid vs. high level tones) and/or contour (rising vs. falling tones). The listeners were required to rate the similarity of a tonal pair on a 1-9 scale (from "no difference" to "extreme difference"). They found that Japanese listeners attended more to non-native pitch level contrasts than contour contrasts, and were especially sensitive to high level tones compared to all other tones. Their sensitivity, in particular to high level tones may be due to the fact that Japanese uses a high pitch (H) as the primary cue to mark an accented syllable in word prosody (Guion & Pederson, 2007; Fujisaki et al., 1996).

Taken together, previous studies have underscored the role of lexical pitch in Japanese listeners' auditory perception of non-native tones. They were found to benefit from the employment of lexical pitch accent: they displayed better performance than non-tone language listeners and they were not worse than canonical tone language listeners (such as Mandarin listeners and Cantonese listeners) at identifying or discriminating non-native tones (Schaefer & Darcy 2014; So & Best, 2010, 2012). On the other hand, a lack of one-to-one mapping between non-native tones and native accentual categories might hinder their perception (So & Best, 2010; Wong, 2019). Moreover, native word prosody shapes the perceptual weighting Japanese listeners assign in perception of non-native pitch contrasts in monosyllables. The contrastive H tone and L tone in word prosody and the employment of a high pitch marked in the accented mora may influence them to be more sensitive to non-native pitch level contrasts than non-native pitch contour contrasts (Fujisaki et al. 1996; Guion & Pederson 2007).

1.2.1.3 Perception of non-native pitch contrasts at the acoustic level by Dutch listeners

Dutch has lexically contrastive stress, the position of which is constrained by phonological properties such as syllable structure (e.g., syllable weight: heavy vs. light) (Kager, 1989; Trommelen & Zonneveld, 1986, 1999; Booij, 1995; Hulst & Kooij, 1992). Dutch uses lexical stress to differentiate word meanings; segmentally identical words can have different meanings due to the position of stress. For instance, *VOORnaam* means “forename” and *voorNAAM* means “distinguished”. Note that the number of minimal pairs that differ only in the position of stress is limited, as few as 14 pairs according to Cutler (1986). In Dutch, the syllable in a polysyllabic word that receives prominent stress is instantiated by acoustic correlates such as longer duration, larger spectral tilt, greater intensity and higher pitch (Van Heuven & Sluijter, 1996; Sluijter, 1995; Sluijter & Van Heuven, 1996; Quené, 1992; Gussenhoven, 2004).

Although Dutch does not utilize pitch at the word level, Dutch listeners are found able to identify or discriminate non-native tonal contrasts acoustically and are perceptually sensitive to subtle pitch differences in non-native tone perception (Leather, 1987, 1990; Chen et al., 2016; Liu et al., 2017). For instance, Leather (1987) conducted an identification task to investigate Dutch and English listeners, compared with Mandarin natives, perceiving Mandarin tones changing from the Mandarin high level tone T1 to the rising tone T2 on a nine-step synthetic continuum. Dutch listeners were found able to identify the tonal steps within the same tone but they did not show a tendency to categorize the tonal steps. Similarly, Liu et al. (2017) created an eight-step continuum from the Mandarin high level tone T1 to the falling tone T4, and applied two identification tasks (forced choice and open choice) and two discrimination tasks (AX and ABX). They found that in identification tasks Dutch listeners were worse than Mandarin listeners and they tended to focus on detailed acoustic properties rather than categorizing tones. However, in discrimination tasks, Dutch listeners showed slightly better performance than Mandarin listeners. These findings suggest that Dutch listeners perceive non-native tones in a psycho-acoustic fashion and a lack of tonal categories in the native language may to some extent facilitate Dutch listeners’ detection of subtle tonal differences (Liu et al., 2017). A sensitivity to pitch differences in non-native tone perception by Dutch listeners can be due to the influence of the use of lexical stress and the intonation system in the native language (Liu et al., 2017). Pitch is one of the acoustic correlates of lexical stress in Dutch, that is, the stressed syllable in a word is realized by a higher pitch together with longer duration and larger amplitude (Gussenhoven, 2004); it is a reliable cue to differentiate stressed vs. unstressed syllable(s) in a word when the word is realized with a nuclear pitch accent (for example, under narrow focus) (Sluijter & Van Heuven, 1996a, 1996b). The employment of stress may improve Dutch listeners' sensitivity to pitch differences. Moreover, Dutch has a rich inventory of nuclear pitch accents in intonation (Gussenhoven, 2004), which may further enhance Dutch listeners' detection of pitch differences in non-native tone perception.

In order to investigate whether native word prosody and/or intonation influence non-native tone perception, Braun & Johnson (2011) compared Dutch and Mandarin listeners' perception of different pitch patterns. They first assigned a pitch rise or a pitch fall resembling Mandarin T2 and T4, respectively (produced by a German speaker) on the first syllable in disyllabic trochaic nonwords and later assigned a pitch rise or a pitch fall on the second syllable. They conducted speeded ABX tasks and found that the Mandarin listeners were sensitive to perceive the pitch patterns on both syllables. The rise and fall contours may signal lexical tones (i.e., rising tone and falling tone) to Mandarin listeners. In contrast, Dutch listeners were observed to be more attentive to the pitch patterns on the final syllable than the initial syllable, which could be due to the fact that final pitch movements may signal post-lexical information to Dutch listeners such as question and statement contour (Haan et al., 1997; Van Heuven & Haan, 2002; Van Heuven & Kirsner, 2004; Van Heuven, 2017).

To conclude, previous studies have revealed that Dutch listeners can discriminate subtle pitch differences in non-native tone perception, although they do not use lexical tones to distinguish words. On the one hand, the lack of tonal categories at word prosody may bring difficulty to Dutch listeners when identifying tones. However, they are not "deaf" to tones. Indeed, they were found to pay more attention to the detailed acoustic properties such as onset and offset of the tones or subtle fundamental pitch differences, which suggests a psycho-acoustic perceptual pattern (Leather 1987, 1990; Wu & Lin, 2008; Liu et al., 2017). Their capability of discriminating tonal contrasts (e.g., Mandarin tones) can be due to the use of lexical stress, of which pitch is an acoustic correlate, and a diversity of nuclear pitch accents in the intonation. Indeed, their perception of non-native tones is influenced by pitch patterns in the native intonation. They were found better at perceiving non-native pitch patterns that can signal linguistic (i.e., post-lexical) meaning in Dutch, that is, they tended to associate rising and falling pitch contour to intonation patterns signaling question and statement (Braun & Johnson, 2011).

1.2.1.4 Summary

The studies reviewed above have underscored the important role native word prosody plays in non-native tone perception at the *acoustic level*. More specifically, tone language listeners, here Mandarin listeners and Japanese listeners, seem to share commonalities in perceiving non-native tones acoustically. On the one hand, they are both found to benefit from the use of lexical pitch (i.e., lexical tones and lexical pitch accent) and to show perceptual advantage over listeners who do not use lexical pitch (i.e., English, Seoul Korean listeners) in the identification and discrimination of non-native pitch contrasts. On the other hand, both groups are observed to be hindered in their perception of specific non-native tones when there is no clear one-to-one mapping between the native tonal category and the non-native tones (Francis et al., 2008; So, 2006; So & Best, 2010; Qin & Mok, 2013; Reid et al., 2015; Wong, 2019). However, despite their linguistic experience with lexical tones in both of the languages, they assign different perceptual weighting to non-native pitch contrasts. Mandarin listeners are found more attentive to non-native pitch contour contrasts than non-native pitch level contrasts in monosyllables, while Japanese listeners were the other way around. Such differences are determined by the demands of pitch contrasts (contour/level) in the native language (Xu et al., 2006; Francis et al., 2008; Li & Shuai, 2011; Guion & Pederson, 2007).

Listeners who do not use lexical pitch, such as Dutch listeners, are overall not as good at non-native tone perception as listeners who use lexical pitch. However, they are not “tone deaf” and are found able to discriminate detailed pitch differences in non-native tone perception (Leather et al., 1987; Hallé et al., 2004; Chen et al., 2016; Liu et al., 2017). The ability to distinguish non-native tonal contrasts acoustically in listeners without use of lexical pitch (i.e., Dutch listeners) can be influenced by the native intonation. For instance, as shown in Braun & Johnson (2011), Dutch listeners tended to perceive non-native pitch patterns as resembling intonation contours that signal post-lexical meanings.

This section discussed the differences and commonalities between listeners with or without lexical pitch in their native language when perceiving non-native pitch contrasts, as well as how native word prosody and intonation may also influence listeners’ non-native tone perception at the acoustic level. However, cross-linguistic studies on the perception of non-native pitch contrasts have mainly focused on monosyllables. The next section will discuss how listeners with (i.e., Japanese and Dutch listeners) or without (i.e., Mandarin listeners) use of lexical positional marking perceive non-native positional contrasts at the acoustic level.

1.2.2 Perception of non-native positional marking at the acoustic level

As introduced before, apart from pitch, *positional marking* is a crucial cue employed at the word level in languages such as Japanese and Dutch to signal lexical contrasts. However, compared with extensive studies on cross-linguistic acoustic perception of non-native pitch contrasts (Gandour, 1983; Hallé et al., 2004; Francis et al., 2008; So & Best, 2010; Schaefer & Darcy, 2014; Liu et al., 2017; among others), there are very few genuine studies that have investigated cross-linguistic perception of non-native positional contrasts (Dupoux et al., 1997; Dupoux et al., 2001; Hu, 2015). Studies on the perception of non-native positional contrasts have mainly focused on second language (L2) listeners (for word stress: Altmann &Vogel, 2002; Altmann, 2006; Wang, 2008; Lai, 2008; Kijak, 2009; Michaux et al., 2014; Chrabaszcz et al., 2014; for lexical pitch accent: Ayusawa & Odaka, 1998; Ayusawa, 2003; Goss, 2015; Shport, 2008, 2011; Ishihara et al., 2011; Tsurutani et al., 2010). The current section will review studies that investigate perception of non-native positional marking (word stress and lexical pitch accent). Due to a small number of studies on the perception of non-native positional marking, L2 studies on Mandarin, Japanese, and Dutch listeners will also be reviewed.

1.2.2.1 Perception of non-native positional marking at the acoustic level by Mandarin listeners

In Mandarin, pitch variations (lexical tones) are the primary cue used to distinguish word meanings. Aside from neutral tones[8] which are regarded as unstressed or having weak stress (e.g., *dongxi* T1T1 "east west" vs. *dongxi* T1-neutral tone "something"), there is, in general, no lexical stress or positional marking in Mandarin, (Chao, 1979; Wang, 1997; Yip, 2002; Lu & Wang, 2005; Duanmu, 2007; Cao, 1992; Bao & Lin, 2014). There is a lack of studies investigating the perception of non-native positional contrasts by native Mandarin listeners. Most studies on the perception of non-native positional marking by Mandarin listeners have focused on L2 listeners, that is, whether Mandarin L2 learners can acquire and thus perceive positional contrasts such as lexical stress (English stress: Altmann & Vogel, 2002; Wang, 2008; Lai, 2008; Altmann, 2006; Yu & Andruski, 2010; Chrabaszcz et al., 2014; Polish stress: Kijak, 2009) or lexical pitch accent (Japanese pitch accent: Ayusawa & Nishinuma, 1997; Ayusawa et al., 1995, 1997a, 1997b, 1999; Ayusawa & Odaka, 1998; Ayusawa, 2003; Goss, 2005; Ishihara et al., 2011).

More specifically, studies on the perception of word stress have mainly centered on 1) what perceptual cues Mandarin L2 learners rely on to perceive stress (e.g., English stress) (Wang, 2008; Lai, 2008; Chrabaszcz et al., 2014) considering that word stress in English has several acoustic correlates such as vowel quality, duration, pitch, and intensity (Bolinger, 1961; Fry, 1958; Lehiste, 1970; Beckman, 1986) and 2) whether Mandarin L2 learners can identify stress position (Yu & Andruski, 2010; Altmann & Vogel, 2002; Altman, 2006; Kijak, 2009). For instance, Wang (2008) investigated perception of contrastive English stress position by

[8] There are a small number of syllables in Mandarin that do not have an underlying tonal pattern, i.e., neutral tones. Neutral tones are often found in function words such as particles or suffixes. On the one hand, they are regarded as contextual tones (Yip, 2002). The realization of neutral tones is highly influenced by the tonal context: the pitch contour of a neutral tone varies depending on the preceding lexical tone (Yip, 2002; Wang, 2004; Duanmu, 2007). On the other hand, neutral tones are regarded as unstressed or having weak stress (Wang, 1997; Lu & Wang, 2005; Duanmu, 2007; Cao, 1992; Bao & Lin, 2014). Compared to their tonal counterparts, neutral tones have a shorter duration and weaker amplitude. They do not occur word initially but often occur in the final unstressed (or weak stressed) syllable of disyllabic words (Li & Thompson, 1977; Zhu & Dodd, 2000; Yip, 2002; Duanmu, 2007). In sum, neutral tones have properties of both lexical stress and lexical tones.

Mandarin advanced L2 learners of English and native English listeners. She strictly manipulated pitch, duration, and intensity with vowel quality held constant in minimal pairs of disyllabic nonwords in English and applied a judgment task where listeners had to judge whether the stress was placed on the first or second syllable after they heard the word. She found that Mandarin advanced L2 learners tended to ignore duration and intensity but solely relied on pitch differences between stressed and unstressed syllables while English native listeners resorted to the three acoustic cues in combination when perceiving contrastive stress position. The findings suggest that Mandarin L2 learners of English may transfer a reliance on pitch to perceive stress position (Wang, 2008). Chrabaszcz et al. (2014) compared Mandarin advanced L2 learners, Russian advanced L2 learners of English, and English natives perceiving disyllabic minimal pairs of nonwords that differed in vowel quality, pitch, duration and intensity. The study used a two-choice forced task where listeners were required to identify whether the stress was on the first or second syllable. The authors found that vowel quality and pitch were the strongest cues for Mandarin L2 advanced learners perceiving English stress. The abovementioned studies regarding perceptual weighting cues of contrastive stress position showed that pitch is a primary acoustic cue that Mandarin advanced L2 learners rely on when perceiving contrastive English stress, despite the different stimuli used in the studies and different definitions of “advanced learners”.

Studies on whether Mandarin L2 learners can identify the overall position of stress have shown mixed results. Altmann (2006) examined advanced L2 learners of English from different language backgrounds including Mandarin, Japanese, Korean, Turkish, Spanish, and Arabic, in comparison with English native listeners. She used disyllabic, tri-syllabic and quadro-syllabic nonwords with word stress always on the first syllable and asked the learners to identify which syllable was stressed after they heard each word. She reported that Mandarin advanced L2 learners achieved as good a performance as English native listeners did and proposed that the absence of stress employment in the native language may not interfere with the perception of L2 stress position, which as a result would yield good perception (see also: Altmann & Vogel, 2002). However, Kijak (2009) reported that a lack of positional cue in the native lexicon hindered Mandarin listeners in identifying Polish word stress. She examined L2 learners of Polish from different proficiency levels and from different language backgrounds, including Mandarin, via a multiple choice identification task where listeners were asked to choose which syllable is stressed in tri-syllabic and quadro-syllabic nonwords in Polish. She found that Mandarin listeners at all proficiency levels showed poor performance, which was attributed to a lack of positional marking used lexically and an inability to rely on pitch differences to identify Polish stress position. Polish word stress is instantiated by spectral tilt and pitch, rather than vowel quality and duration. However, pitch information in Polish is used to a lesser degree than in English and thus Mandarin listeners could not rely on pitch to the same extent as they did when perceiving English stress. The findings in Kijak (2009) therefore provide evidence of the importance of pitch as an acoustic cue that Mandarin listeners use to perceive stress position and suggest that tone language listeners may tend to perceive L2 English stress as tonal patterns (Juffs, 1990; Archibald, 1995; Nguyen et al., 2005, 2008).

It is shown that Mandarin learners of Japanese, another language that employs lexical positional marking, were found sensitive to Japanese pitch accent position. For instance, Ayusawa et al. (1995) investigated advanced L2 learners of Japanese from different language backgrounds including Mandarin learners by applying a listening test (Nishinuma, 1994) where listeners had to circle which syllable (written in hiragana and katakana) was accented in two-, three-, four-, and five-mora real words and phrases after they heard the token. They found that the Mandarin advanced L2 learners achieved good performance (around 85% accuracy) when identifying the correct position of pitch accent. Goss (2015) compared Mandarin and Seoul Korean advanced L2 learners of Japanese perceiving the position of Japanese pitch accent via both a judgment task and a categorization task. Listeners were first tested in the judgment task where they had to judge the three-mora real words (nouns) followed by a postposition that was spoken with the correct or incorrect accent pattern. For the correctly-accented tokens, listeners were then required to categorize the token they heard into one of the four accentual patterns (presented as visualized pitch patterns in line shape) in a four-alternative forced choice task. He found that Mandarin advanced L2 learners outperformed Seoul Korean advanced L2 learners overall in making the correct judgment of the accented words and categorizing accentual patterns. He ascribed such a perceptual advantage to the phonemic use of pitch variations in Mandarin.

To summarize, studies on the perception of non-native positional contrasts by Mandarin listeners have mainly focused on determining whether Mandarin listeners can identify a non-native positional cue (i.e., word stress and lexical pitch accent) as L2 learners and which perceptual cues they rely on when perceiving positions (i.e., word stress). The findings show that Mandarin advanced L2 learners of Japanese can successfully identify Japanese pitch accent positions, which can be accounted for as an augmented sensitivity due to their native language (Ayusawa, 2003; Goss, 2015). Mandarin advanced L2 learners of English were found to primarily rely on pitch when perceiving minimal pairs contrastive in stress position (i.e., English) whereas native listeners rely on multiple acoustic cues (Wang, 2008; Lai, 2008; Chrabaszcz et al., 2014; also see Sluijter & Van Heuven, 1996). They were found able to identify stress position in English (Altmann, 2006) but not in Polish where pitch is employed to mark stress to a lesser degree than in English (Kijak, 2009), suggesting the importance of pitch for Mandarin L2 learners identifying stress position. The studies reviewed in the current section underline the important role pitch plays in Mandarin L2 learners' acoustic perception. Mandarin L2 (mainly advanced level) learners of a language that uses positional marking may use pitch information to perceive L2 position acoustically (see Nguyen & Ingram, 2005;

Nguyen et al., 2008, for discussion of tone language listeners' usage of pitch cues for stress position).

1.2.2.2 Perception of non-native positional marking at the acoustic level by Japanese listeners

Japanese uses the presence/absence and position of pitch accent for lexical contrasts. Japanese listeners are found to identify their native pitch accent positions acoustically (Ayusawa et al. 1998; Ayusawa, 2003; Goss, 2015; Hu, 2015; Byun, 2018). For instance, Ayusawa et al. (1998) used a listening test to examine Japanese listeners to perceive the accentual patterns in three- and four-mora real words and five-mora phrases. Listeners were required to mark where the pitch accent falls in the syllable (written in hiragana and katakana) in the word they heard. Japanese natives overall were found to achieve a nearly ceiling performance when marking the correct position of pitch accent. Note that the perception of a specific accentual minimal pair in Japanese by the natives yielded mixed results (Vance, 1995; Maniwa, 2002; Yoneyama, 2002; Sugiyama, 2006, 2008; Shport, 2008, 2011; Hu, 2015), which is not the main point of discussion in the study.

Since Japanese listeners have been found to show an overall sensitivity to perceive native pitch accent position (Ayusawa et al. 1998; Ayusawa, 2003; Vance, 1995; Yoneyama, 2002; Hu, 2015; Goss, 2015), further studies investigated whether they may transfer the native positional sensitivity to perceive a non-native positional cue. These studies mainly focus on perception of English lexical stress in L2, especially perceptual cue weighting by Japanese L2 learners (Beckman, 1986; Mochizuki-Sudo & Kiritani, 1991; Tokuma, 2003; Ishikawa & Nomura, 2008; Sugahara, 2011, 2016). For instance, Beckman (1986) investigated the effects of duration, pitch, and amplitude on the perception of English word stress in minimal pairs such as *DIgest* vs. *diGEST* (/ˈdaɪdʒest / vs. /daɪˈdʒest/) by English native listeners and Japanese L2 learners of English in different proficiency levels. The listeners were required to use a scale of 1 to 5 to respond after hearing a synthesized word, 1 indicating the word they heard was clearly stressed on the first syllable and 5 indicating the word was clearly stressed on the second syllable. She found that Japanese L2 learners attended more to pitch than other acoustic correlates of stress and their reliance on F0 to perceive contrastive stress position seemed to not be correlated to the degree of their exposure to English, while American English native listeners relied on a combination of pitch, duration, and amplitude to a greater degree. Sugahara (2011, 2016) investigated how Japanese advanced L2 learners of English would perform when pitch is not available to discriminate English stress positions. She used minimal pairs in disyllabic real words by imposing a flat pitch contour on each syllable while keeping the vowel quality, duration and intensity intact, e.g., *IMport* vs. *imPORT* (/ˈɪmpɔːt/ vs. /ɪmˈpɔːt/). Japanese listeners were found to have a strong perceptual bias towards word-final stress (iambic stress) whereas English natives preferred more word-initial stress (trochaic stress) when judging which syllable was stressed. The author ascribed Japanese listeners' preference for final stress to the influence of a native "antepenultimate rule" that words with 5-6 or more morae in Japanese are predominantly accented on the antepenultimate mora (Sato, 1993; Kubozono, 1998, 2006; Shibata, 1994). The final syllable in disyllabic words could be perceived as super-heavy morae in loanwords where it contained the antepenultimate mora. For instance, *IMport*/*imPORT* could be perceived as i.m.po.o.t (V.C.CV.V.C) where the antepenultimate mora "po" in the final syllable nucleus vowel was accented, according to the native accentuation rule.

In sum, Japanese listeners are shown to be sensitive overall to native pitch accent position, despite mixed findings regarding final accented vs. unaccented minimal pairs[9] (Ayusawa, 1998; Vance, 1995; Yoneyama, 2002; Sugiyama, 2006, 2008; Hu, 2015). Their sensitivity to the position of lexical pitch accent seemed beneficial to perception of non-native positional cues such as English lexical stress in L2 studies. Moreover, it has been shown that pitch information is a dominant cue for Japanese L2 learners of English when perceiving contrastive stress positions (Beckman, 1986). However, when pitch is not available, native accentuation in Japanese word prosody played a role in the perception of L2 stress position. For instance, Japanese L2 learners of English were found to be influenced by the "antepenultimate rule" in Japanese word prosody in that they showed a bias for final stress, likely perceiving the final syllable in an English disyllabic word as super-heavy morae in native loanwords that contain the accented antepenultimate mora (Sugahara, 2011, 2016).

1.2.2.3 Perception of non-native positional marking at the acoustic level by Dutch listeners

Dutch uses lexical stress to contrast word meanings. Studies on perception of positional cues have mainly focused on whether Dutch listeners use stress information in native word recognition and on whether the use of native stress assists Dutch L2 learners of another stress language (e.g., English) in L2 word recognition, to be discussed in §2.4.

There is a scarcity of studies that investigate the perception of non-native positional cues by naive Dutch listeners. Among the very few studies, Hu (2015) applied an ABX discrimination task to examine whether naive Dutch listeners could perceive the contrastive position of Japanese pitch accent. She found that Dutch listeners displayed an good performance overall, similar to Japanese natives discriminating contrastive pitch accent positions in two-, three-, four-, and five-mora

[9] Note that the minimal pair, final-accented word vs. unaccented word, have the same tonal pattern when they appear in isolation (Vance 1995; Warner 1997; Sugiyama, 2006). However, a contrast between final-accented and unaccented words emerges when followed by a grammatical particle (such as the nominative *ga*): final-accented words exhibit a pitch fall on the final syllable while unaccented words do not.

(syllable) nonwords, suggesting that Dutch listeners may be able to transfer a sensitivity to positional marking in their native language to discriminating Japanese positional pitch accent contrasts.

Apart from Hu (2015), there seem to be no studies on native Dutch listeners' perception of non-native positional contrasts. It is not fully understood whether listeners who use positional marking in the lexicon, i.e., Dutch listeners, can generalize their sensitivity to native positional marking to non-native positional marking of pitch properties (i.e., tones).

1.2.2.4 Summary

This section reviewed studies on the perception of non-native positional marking. Most studies regarding position perception have focused on L2 learners, especially on determining the perceptual cues learners rely on when perceiving word stress and whether learners can acquire and perceive positional prosodic contrasts (i.e., word stress and lexical pitch accent). Only very few studies have investigated cross-linguistic perception of non-native positional marking by naive listeners (see e.g., Hu, 2015). It is not fully understood how native word prosody, i.e., with the use of positional marking or not, influences perception of non-native positional marking at the acoustic level.

1.2.3 Summary of perception of non-native WPCs at the acoustic level

A large body of studies has investigated cross-linguistic perception of non-native pitch contrasts at the acoustic level to unveil the influence of native word prosody, stemming from intonation in some cases, whether facilitation or interference, or attendance to different pitch contrasts (pitch contour or pitch level) on perception of non-native pitch contrasts (Gandour, 1983; Wayland & Guion, 2004; Francis et al., 2008; So & Best, 2010; Schaefer &Darcy, 2014; among others). However, comparatively less attention has been devoted to perception of another WPC, *position*, which is employed contrastively in languages such as Dutch (for lexical stress) and Japanese (for lexical pitch accent). Studies on perception of non-native positional cues have mainly focused on whether L2 learners can acquire and perceive contrastive position of lexical stress or pitch accent and what perceptual cues they rely on (Wang, 2008; Lai, 2008; Beckman, 1986; Mochizuki-Sudo & Kiritani, 1991; Tokuma, 2003). There are very few true studies on perception of non-native positional contrasts. For instance, Dupoux et al., (1997, 2001) found that French listeners, who do not use any WPCs in their native language, were able to detect contrastive stress positions in Spanish. Hu (2015) found that Dutch listeners who use lexical stress position were overall good at discriminating Japanese pitch accent positions. However, no study to date has investigated the role of native word prosody in acoustic perception of non-native pitch contrasts (tones) regarding position, that is, non-native tone contrasts marked by position and non-native pitch contrasts occurring in different positions. More specifically, it remains unclear whether the use of positional marking in the native language can facilitate acoustic perception of non-native positional tones and non-native pitch contrasts in different positions and, furthermore, how groups of listeners with different language backgrounds differ or share in their perception of non-native pitch contrasts with positions acoustically. These are the gaps the current study aims to bridge.

1.3 Perception of non-native WPCs at the phonological level

The *phonological level* refers to a more abstract processing level where an "acoustic level of representation is not accessible" and which "highlights the phonological level of representation" (Dupoux et al., 2008: 690). It can be examined via a relatively demanding task, namely the *sequence recall task* where "memory and perceptual resources are concerned" (Dupoux et al., 2001: 1607). More specifically, Dupoux et al. (1997) found that French listeners who do not use lexical stress or any other WPCs (Rossi, 1980; Di Cristo, 1988; Vaissière, 1991) had no problem discriminating contrastive stress positions in Spanish in an AX discrimination task. They attributed the French listeners' perceptual success to the nature of the AX task, which may allow listeners to detect fine-grained acoustic details (Dupoux et al., 2001). The AX task, as mentioned before, encourages an acoustic mode of listening. Among perception tasks, it has the least *memory load* on listeners in that listeners spend the least effort to retain auditory information in a short-term memory store and thus listeners can make same-or-different judgments based on an immediate acoustic comparison of the two stimuli (Logan & Pruitt, 1995; Strange & Shafer, 2008; Dupoux et al., 1997).

However, French listeners, different from their success in discriminating Spanish stress in Dupoux et al. (1997), were later observed to show "stress deafness" in the sequence recall task in Dupoux et al. (2001). In the sequence recall task, listeners were required to learn two disyllabic nonwords A and B (produced by different speakers) either contrasting in stress position or segmentally (as a baseline), and then listen to the two words in various sequences (such as A-B or B-A in two-word sequences, A-B-B in three-word sequences). They had to recall the sequence they heard, a task which taps into their memory load of stored representations of stress. By incorporating a high memory load and phonetic variability (tokens spoken by different speakers), the sequence recall task aims to eliminate listeners' employment of acoustic strategies (e.g., focusing on onset or offset of the syllable or a specific voice) and highlight the processing at a "more abstract processing level", the phonological level (Dupoux et al., 2001). The researchers found that French listeners had no problem processing segmental contrasts (as a baseline) but were impaired when perceiving stress contrasts via the sequence recall task. Considering the findings in Dupoux et al. (1997) that French listeners were sensitive to stress positions acoustically, the observed "stress deafness" in Dupoux et al. (2001) likely stems from processing stress at an abstract phonological level instead of a psycho-acoustic level (Dupoux et al., 2001, 2008; Schwab & Llisterri, 2011).

The current section will review studies that apply the sequence recall task to investigate perception of non-native pitch contrasts and positional marking at the phonological level. Section 2.3.1 will deal with perception of non-native pitch contrasts at the phonological level. Section 2.3.2 will discuss existing literature on the perception of non-native positional marking at the phonological level.

1.3.1 Perception of non-native pitch contrasts at the phonological level

Compared with studies on the perception of non-native pitch contrasts at the acoustic level, scant attention has been paid to the processing of non-native pitch contrasts at the phonological level. In particular, there seem to be no cross-linguistic studies to date that have investigated phonological processing of non-native pitch contrasts by naive listeners on a par with Dupoux et al. (2001, 2008) on French naive listeners' phonological perception of Spanish stress contrasts. Studies on the processing of non-native pitch contrasts at the phonological level have mainly focused on L2 acquisition. For instance, Zou (2017) applied a sequence recall task (Dupoux et al., 2008) to examine the developmental trajectory of Mandarin tones in Dutch L2 beginning and advanced learners of Mandarin. She contrasted six tonal pairs (all the possible pairs consisting of Mandarin T1, T2, T3 and T4, such as T1 vs. T2, T1 vs. T3) in disyllabic nonwords (the target tonal pairs on syllable initial position and a fixed neutral tone on syllable final position). She found that compared with Dutch L2 beginning learners of Mandarin who had difficulty (yet still performed above chance level), Dutch L2 advanced learners of Mandarin showed an overall significantly better performance and approximated the performance of Mandarin natives. The findings demonstrated that Dutch L2 learners were able to form Mandarin tonal categories and thus process Mandarin tonal contrasts in a phonological mode after a large amount of input (Zou, 2017; Strange, 2011).

Apart from Zou (2017), in which listeners received L2 training for months and even years, there is a lack of studies on perception of non-native pitch contrasts at the phonological level by naive listeners from a cross-linguistic perspective. Previous studies have shown that listeners with different use of WPCs in the native language, in the current study Mandarin listeners, Japanese listeners, and Dutch listeners, show acoustic sensitivity to non-native pitch contrasts on monosyllables (Lee et al., 1996; Wayland & Guion, 2004; Francis et al., 2008; So & Best, 2010; Schaefer &Darcy, 2014; Burnham et al., 2015; Liu et al., 2017; among others). More specifically, differences in pitch at the acoustic level may be available to listeners, even to listeners from non-tone languages, in non-native tone perception (Hallé et al., 2004; Liu et al., 2017), but this does not entail that these listeners may be able to encode non-native pitch contrasts phonologically. Hence, it remains unclear whether the acoustic sensitivity found in non-native tone perception can be retained in phonological processing and to what extent the native language comes into play in processing of non-native pitch contrasts at the phonological level.

1.3.2 Perception of non-native positional marking at the phonological level

There are only very few studies on cross-linguistic perception of non-native position marking at the phonological level (Dupoux et al., 2001; Peperkamp & Dupoux, 2002; Perperkamp et al., 2010; Hu, 2015). For instance, as mentioned above, Dupoux et al. (2001) investigated perception of contrastive stress in Spanish by French naive listeners whose native language does not use lexical stress or any other WPCs (Rossi, 1980; Di Cristo, 1988; Vaissière, 1991) via the sequence recall task that taps into the phonological representation of stress. They manipulated the stressed syllable with a longer duration, higher pitch, and louder intensity compared with the unstressed syllable in disyllabic minimal pairs (segmentally identical, with no vowel reduction) in Spanish. They found that French listeners failed to perceive contrastive stress position at the phonological level, yet successfully discriminated Spanish contrastive stress position acoustically (Dupoux et al., 1997, 2001). Hu (2015) employed a sequence recall task (Dupoux et al., 2001) to examine the discrimination of positional contrasts of Japanese pitch accent by Japanese listeners and naive Dutch listeners. She used three minimal pairs of disyllabic nonwords that differed in the position and presence of a pitch accent: initial accented (H'L) vs. final accented (LH'), initial accented (H'L) vs. unaccented (LH) and final accented (LH') vs. unaccented (LH). Dutch listeners were reported to be able to phonologically discriminate positional contrasts of pitch accent in disyllabic nonwords as well as Japanese listeners did, except for minimal pairs differing (final accented vs. unaccented). Japanese uses the presence versus absence of lexical pitch accent to differentiate word meanings (e.g., final accented *kaki* means "fence" vs. unaccented *kaki* means "permission"), while Dutch always uses the position of stress on a certain syllable for lexical contrasts. This lack of employment of presence vs. absence of word prosodic cues in Dutch may account for Dutch listeners' difficulty in perceiving unaccented vs. final accented minimal pairs in Japanese (Hu, 2015). The findings in Hu (2015) showed that Dutch listeners might benefit from lexical positional marking of word stress in their native language when processing Japanese pitch accent positions phonologically. However, the presence vs. absence of positional marking does not exist in Dutch, which may bring difficulty to Dutch listeners in phonological processing of Japanese pitch accent positions.

Apart from the abovementioned studies, the majority of studies regarding perception of non-native positional contrasts at the phonological level have focused on whether listeners can encode the positional marking in L2 acquisition (Lin et al., 2014; Utsugi et al., 2010; Qin et al., 2017). For instance, Lin et al. (2014) investigated whether Mandarin and Seoul Korean advanced L2 learners of English were able to encode contrastive English stress phonologically using a sequence recall task adapted from Dupoux et al. (2001, 2008). They manipulated minimal pairs of disyllabic nonwords that differed in stress position, with longer duration and higher pitch in the stressed syllable than the unstressed syllable (no vowel quality differences). They found that Mandarin advanced L2 listeners outperformed Seoul Korean advanced L2 learners and the former even showed as good performance as English native listeners when processing stress in minimal pairs phonologically. The authors suggest that the success of Mandarin advanced L2 listeners might be due to a positive transfer from the native language in which they use suprasegmental information for lexical contrasts to perceive another non-native suprasegmental cue, while Seoul Korean does not use lexically contrastive prosodic cues (Sohn, 1999; Jun, 2005). Lin et al. (2014) argued that these listeners could also benefit from the presence of neutral tones in Mandarin, which have stress-like properties (Wang, 1997; Lu & Wang, 2005; Duanmu, 2007; Cao, 1992; Bao & Lin, 2014), although the amount of syllables carrying neutral tones is very limited. Qin et al. (2017) investigated the perceptual cues Mandarin and Taiwanese Mandarin advanced L2 learners of English rely on when perceiving English stress phonologically, compared to English natives. The authors manipulated minimal pairs of disyllabic nonwords under four conditions in which stressed and unstressed syllables were signaled only by pitch (stressed syllable with a higher pitch), only by duration (stressed syllable with a longer duration), by both pitch and duration (stressed syllable with longer duration and higher pitch) and by incongruent correlates (stressed syllable with higher pitch is shorter). They applied the sequence recall task (Dupoux et al., 2008) and found that Mandarin advanced L2 learners attended to pitch more than English natives did when perceiving contrastive stress phonologically, suggesting an influence of pitch sensitivity from the native word prosody in phonological processing of stress.

1.3.3 Summary

This section reviewed the literature on perception at the phonological level of non-native pitch contrasts and non-native positional marking, respectively. Compared with extensive studies on perception of non-native WPCs at the acoustic level (Gandour, 1983; Wayland & Guion, 2004; Francis et al., 2008; So & Best, 2010; Schaefer & Darcy, 2014; Qin & Mok, 2013; Wu et al., 2008; among others), little attention has been given to how non-native WPCs, both non-native pitch contrasts and positional contrasts, are processed at the phonological level in tasks in which acoustic traces are inaccessible to listeners. The “stress deafness” observed in Dupoux et al. (1997, 2001, 2008) of French listeners who do not use lexical stress or any other WPC (Rossi, 1980; Di Cristo, 1988; Vaissière, 1991) is notable. In these studies, French listeners were able to discriminate stress contrasts in Spanish stress acoustically but failed in a more cognitive demanding task that taps into phonological representations of stress. These findings suggest that the capability to discriminate acoustic differences via a non-native prosodic cue does not entail success encoding this information phonologically. So far there seem to exist only a few studies that have looked into phonological perception of a non-native WPC by naive listeners (word stress in Dupoux et al., 2001; Peperkamp & Dupoux, 2002; Perperkamp et al., 2010; lexical pitch accent in Hu, 2015). Studies on the processing of non-native WPCs at the phonological level have instead focused mainly on L2 acquisition where listeners have already been exposed to the WPCs and have been influenced by listeners' "internal and external variables" (Braun et al., 2014: 330) such as the interference of native language, age, study motivation, and teaching strategies (Orie, 2006; Moyer, 1999; Obler, 1989) during their acquisition. Nevertheless, investigating naive listeners may provide a direct view into how and to what extent native word prosody affects the processing of non-native WPCs at the phonological level at the very early stage of acquisition without the influence of other factors.

1.4 Processing of non-native WPCs at the lexical level

Processing of non-native WPCs at the *lexical level* can refer to the encoding of novel WPCs when establishing lexical contrasts (i.e., sound-to-meaning mapping in word leaning) or the employment of non-native WPCs (mostly in L2 acquisition) when accessing lexical activation (i.e., in spoken word recognition). In the current dissertation, processing of non-native WPCs at the *lexical level* refers to the former, that is, how non-native WPCs are encoded when integrating sound to meaning during word learning. Section 1.4.1 will deal with the literature on how listeners from different backgrounds encode non-native pitch contrasts (tones) in word learning, by mainly focusing on Mandarin, Dutch, and Japanese listeners. Compared with studies on the perception of non-native pitch contrasts at the lexical level, there is a lack of studies that examine the encoding of non-native positional contrasts in word learning. Instead, most studies on the processing of non-native positional contrasts at the lexical level have focused on word recognition. Thus, Section 1.4.2 will discuss previous studies that investigated non-native positional contrasts in word recognition, focusing on Mandarin, Dutch, and Japanese listeners.

1.4.1 Processing of non-native pitch contrasts at the lexical level

Previous studies on the processing of non-native pitch contrasts lexically have investigated how native language, more specifically, native word prosody, plays an important role in building non-native tones paired with meanings in word learning. It has been shown that prior experience with lexical tones in the native language can benefit listeners when associating novel tones with meanings in word learning. For example, Poltrock et al. (2018) investigated Mandarin and French listeners learning minimal pairs of disyllabic pseudo-words with Cantonese native listeners as controls. The minimal pairs were contrastive in either a consonant, a vowel or a Cantonese tone on the first syllable while the second syllable was kept constant, carrying the Cantonese high level tone T1. Listeners had to learn to associate the word with the object (a fictional object presented in an animation) during the training phase and then had to look at the correct object after they heard the word in the test phase. It was found that all three groups were able to learn the sound-to-meaning associations (they all performed above chance level when learning all three types of contrasts) and that Cantonese native listeners outperformed the two non-native groups. Mandarin listeners showed an overall better performance than French listeners when learning tonal contrasts, but not consonant or vowel contrasts, which suggests that a general property of phonological representations of tones in the native language can enhance non-native tone-to-meaning mapping. Cooper & Wang (2013) compared naive Thai listeners, naive English listeners, and English listeners with previous Cantonese tone training to learn five monosyllabic real words (word-object associations) differing in five Cantonese tones (high-level, high-rising, low-falling, low-rising, and low-level tone). Before learning, they were first presented with a tone identification task where they had to associate a visualized pitch pattern diagram (with rising, falling, and flat arrows) with the tone they heard. It was found that the tone-trained English listeners showed better tone identification performance than naive Thai listeners and naive English listeners. They were then guided to learn Cantonese tones (carried by monosyllables) associated with pictures during a period of four days, after which it was found that the Thai listeners and the tone-trained English listeners performed equally well, better than non-tone trained English listeners, in a picture matching task where participants had to match a word they heard with a corresponding picture. Thai listeners were especially good at learning Cantonese high-rising, high-falling and high-level tones, which could be due to a similarity with tonal counterparts in Thai. This suggests that tone language listeners may rely on specific lexical pitch representations in the native language when building non-native tone to meaning mappings during word learning (Cooper & Wang, 2013).

On the other hand, it was found that a lack of lexical pitch in the native language may impede the encoding of non-native tones when building lexical representations. For instance, Ramachers et al. (2017) applied a name-labelling paradigm (Quam & Swingley 2010; Singh et al., 2014) to investigate whether Dutch listeners could attend to pitch differences when learning novel words in Limburgian, a dialect with a restricted tonal system consisting of a binary tone contrast between accent 1 HL and accent 2 LHL[10] (Gussenhoven, 2000a, 2000b). In Ramachers et al. (2017), non-Limburgian Dutch listeners learned monosyllabic (bimoraic CVVC) novel words either carrying accent 1 or accent 2 with paired objects. They were first presented with the target word-object association and a distracter (an unlabelled object) during training. There were two conditions in the test, correct pronunciation (CP) and mispronunciation (MP) conditions. In both conditions, listeners were presented with the target and the distracter. In the MP condition, listeners heard the

[10] Any primary stressed bimoraic syllable in Limburgian carries either accent 1 HL or accent 2 LHL (Gussenhoven, 2000b).

mispronounced counterpart (e.g., accent 2) of the target word (e.g., accent 1). It was supposed that if listeners noticed the tone changes, they would look at the distracter (Markman, 1990; Markman & Wachtel, 1988; Quam & Swingley, 2010; Kalashinikova et al., 2015). They found that the non-Limburgian Dutch listeners showed as good performance as the natives did in the CP condition. However, in the MP condition, they still preferred to look at the target word-object instead of looking at the distracter when the word was mispronounced with another accent, indicating that they neglected the pitch changes during word learning of Limburgian. Their insensitivity to non-native pitch contrasts during word learning can be due to an influence of the native prosody at the word level. Pitch is one of the acoustic correlates of word stress in Dutch (Van Heuven & Sluijter, 1996; Sluijter & Van Heuven, 1996) and can be relied on to distinguish stressed vs. unstressed syllable(s) in a word when the word is accented in narrow focus (Van Heuven & Sluijter, 1996; Sluijter & Van Heuven, 1996; Van Heuven & Haan, 2002). A lack of contrastive pitch at the word-level in Dutch therefore impeded Dutch listeners from attending to non-native pitch contrasts (Limburgian tones) in word identification (Ramachers et al., 2017).

The abovementioned studies have shed some light on the influence of native word prosody, whether facilitation or interference, on encoding non-native pitch information when establishing lexical representations. Apart from the influence of native word prosody, it has also been shown that prosodic features at the *utterance level* can shape the sound-to-meaning mapping in word learning. Braun et al. (2014) examined Japanese, German, and French listeners with Mandarin controls learning non-words differing in Mandarin tones with paired objects. They used two disyllabic non-words in which the tone on the first syllable was fixed (rising tone, T2) and the second syllable carried different tones, high-level, rising, dipping and falling tones (T1 to T4), respectively. They found that German listeners, following the Mandarin controls, outperformed French and Japanese listeners in judging whether the picture matched the word in “tonal-mismatch condition” where the word did not match the object. Interestingly, Japanese listeners who have a restricted tone system did not differ from French listeners who do not use prosodic cues lexically, both showing no sensitivity to the tonal contrasts in lexical encoding. According to Braun et al. (2014), the differences among the three groups of listeners in word learning can be due to the different sizes of pitch contrasts at the utterance level. German has a larger inventory of pitch accent types than Japanese and French do in intonation. French and Japanese listeners relied little on pitch variation to signal post-lexical contrasts (Abe, 1998; Post, 2000; Turco et al. 2012; Asano & Braun, 2011), which can account for their insensitivity to encode non-native tonal contrasts in the mental lexicon. Braun et al. (2014) argues that it is the prosodic features at the *utterance level* in the native language rather than the *word level* prosody that can be beneficial for building non-native tone to meaning mapping at the lexical level.

In addition to the above-mentioned studies on the influence of native word prosody and/or intonation in processing non-native tones lexically, there are studies that have focused on what factors can contribute to enhancing or predicting word learning by non-native listeners (Wong & Perrachione, 2007; Chandrasekaran et al., 2010; Showalter & Hayes-Harb, 2013; Bowles et al., 2016). For instance, Wong & Perrachione (2007) investigated whether musical experience and phonological awareness can affect English listeners learning Mandarin tones. English listeners were first tested for phonological awareness in a pitch pattern identification task where they had to associate the visualized pitch pattern (represented by level, rising, and falling arrows) with the three corresponding Mandarin tones (high level, rising, and falling tone) they heard. They were then trained to learn six sets of English-like monosyllabic pseudo-words superimposed with three Mandarin tones (high level T1, rising T2 and falling T4) paired with real object pictures. It was observed that the English listeners were able to identify the picture out of three with the corresponding word they heard. Despite a large individual difference, those who achieved higher accuracy in identification of tone-to-meaning associations were found to have higher pitch pattern identification ability and prior musical experience. The findings in Wong & Perrachione (2007) showed that metalinguistic awareness and musical experience could assist non-tone language listeners (i.e., English listeners) to integrate novel tones with meanings in word learning and such facilitation was also observed in neurophysiology studies (Wong et al., 2007; Wang et al., 2003; Kaan et al., 2008; Shen & Froud, 2019).

In sum, behavioural and psycholinguistic studies have shown that native prosody at the word level and/or the utterance level, as well as factors such as prior musical experience and metalinguistic awareness, influence the encoding of non-native pitch contrasts in building lexical representations in word learning.

1.4.2 Processing of non-native positional marking at the lexical level

Compared with studies on lexical processing of non-native pitch contrasts, few studies have investigated how non-native positional marking is encoded during? in sound to meaning mapping. Studies on the processing of position at the lexical level have been mainly conducted in spoken word recognition, which include investigation into whether listeners who use positional marking in their native language, e.g., Japanese and Dutch listeners, can exploit the positional cue in native word recognition (for Japanese: Otake & Cutler, 1999; Sekiguchi & Nakajima, 1999; Shibata & Hurtig, 2008; Goss, 2015; for Dutch: Van Heuven, 1985; Van Heuven, 1988; Jongenburger & van Heuven, 1995a; Jongenburger & van Heuven, 1995b; Van Leyden & Van Heuven, 1996; Jongenburger, 1996; Van Kuijk, 1996; Quené & Koster, 1998; Cutler & Koster, 2000; Koster & Cutler, 1997; Cutler & Van Donselaar, 2001; Van Donselaar et al., 2005) and furthermore, how the native prosody (with or without positional marking) affects perception of the non-native positional cue (e.g., lexical stress) in L2 word recognition (Archibald, 1997; Cooper et al., 2002; Zhang & Francis, 2012; Connell et al., 2018).

To start with, studies on whether the positional marking of word prosody can be employed in lexical activation and selection in the native language will be discussed. It is shown that English natives, whose native language has word stress, do not use stress information (suprasegmental, e.g., *FORbear* vs. *forBEAR*, /ˈfɔːbɛə/ vs. /fɔːˈbɛə/) to constrain lexical activation and competition in word recognition, which could be because English predominantly uses segmental cues (vowel reduction) to cue word stress (Cutler, 1986; Slowiaczek, 1990, 1991). How about Japanese and Dutch listeners who also employ positional marking at the word level? Several studies have been devoted to answering this question. For instance, Otake & Cutler (1999) investigated whether pitch accent information in Japanese was employed in word recognition by means of three different experiments. They first applied a two-choice classification experiment where Japanese native listeners should choose the word from two words contrasting in accentual pattern H vs. L, based on the extracted syllable they heard (e.g., whether *ka* is from “kage” (H’L) or “kagi” (LH’)). They found that Japanese listeners succeeded in making correct judgments of the word based on the pitch accent information of the syllable. In the second experiment, a gating experiment in which listeners heard one fragment at a time (from the initial consonant, initial consonant and vowel, to the end of the word), Japanese native listeners were found able to accurately identify the word of the same pitch accentual structure based on only the initial mora (e.g., *na* from “nagasa” (H’LL) or “nagashi” (LHH’)). In the third experiment, a lexical decision experiment where listeners should decide as quickly as possible whether a word was preceded by the same word or the word contrastive in accentual pattern was a real word or not. They observed that Japanese listeners made speeded response to the word (e.g., “ame” H’L) presented after the same prime (e.g., “ame” H’L) but not the prime contrasting in accentual pattern (e.g., “ame” LH’), suggesting that Japanese listeners employed pitch accentual information to constrain lexical activation.

Studies on whether Dutch listeners use stress information in native word recognition have displayed mixed results. For instance, Jongenburger & Van Heuven (1995b) used Dutch minimal pairs differing in stress positions (carried in a semantically neutral sentence) and applied a gating experiment where listeners had to write down and say out loud what they thought the word was after hearing the fragment(s) in each gate. The first gate contained the preceding context and the initial consonant(s) and the first vowel onset of the target word. Each subsequent gate contained more diphones until the whole word was complete. For instance, the second gate consisted of the initial consonant(s), a vowel, and the onset of the following consonant(s). They found that Dutch listeners did not differentiate based on stressed vs. unstressed initial syllables in minimal pairs but they could make correct stress judgments when the initial syllable and part of the following vowel were presented. Their findings suggested that stress position did not have a strong effect on constraining lexical activation (see also: Jongenburger & Van Heuven, 1996; Jongenburger & Van Heuven, 1995a). Cutler & Van Donselaar (2001) provided different observations via a series of experiments. They conducted a two-choice classification task where Dutch listeners should circle the word (in minimal pairs) from which the syllable was extracted. They were found to be able to correctly judge the word based on the syllable they heard (e.g., whether /ka/ is from *CAnon* "norm" or *kaNON* "gun"). They further applied a repetition priming paradigm where listeners were required to decide whether the disyllabic word they heard was a real word or not. They found that presenting a disyllabic word contrasting in stress position did not speed up Dutch listeners' response to their decision, which means, for instance, the stressed syllable *VOOR* (in *VOORnaam* "first name") only activated the lexical representation of "*VOORnaam"* while the unstressed syllable *voor* (in *voorNAAM* "respectable") only activated the lexical representation of *voorNAAM*. In a follow-up word-spotting experiment where Dutch listeners were required to decide whether there was a real word embedded in a nonsense letter string, it was found that they constrained lexical activation by an early use of stress in recognizing spoken words (e.g., *zee* "sea" embedded in a string received greater activation from *zee* in *muZEE* than the same segmental *zee* in mismatching stress *MUzee*). These combined findings showed that Dutch listeners can effectively exploit stress position to constrain lexical activation in word recognition. Moreover, studies on perception of mis-stressed words have shown that such words impeded Dutch listeners when recognizing words, suggesting an important role stress position plays in lexical activation in the native language (Van Heuven, 1985; Van Leyden & Van Heuven, 1996; Koster & Cutler, 1997; Cutler & Koster, 2000).

The studies reviewed above show that both Japanese and Dutch listeners use positional marking in native word recognition. Will the employment of native positional marking be transferred to perception of a non-native positional cue in L2 word recognition? Cooper et al. (2002) investigated whether Dutch L2 learners at a proficient level of English were able to use native stress sensitivity to perceive English word stress in word recognition, with English natives as controls. They conducted a series of experiments. In a two-alternative forced-choice task, they used truncated fragments from real words in English and listeners had to circle (on an answer sheet) the correct word based on the fragment they heard, e.g., whether *mus* is from *MUsic* or *muSEum.* They found that Dutch proficient L2 listeners of English outperformed English natives in correctly assigning a monosyllabic fragment to one of the two words that differed only in stress position. Their findings suggested that the importance of stress information in native word recognition can assist Dutch listeners in differentiating stressed vs. unstressed initial syllable in L2 word activation.

A recent study conducted by Connell et al. (2018) addressed how listeners whose native language does not use a lexical positional cue recognized English words in L2. They examined Mandarin and Seoul Korean advanced L2 learners of English in comparison with English native listeners in how they used segmental (i.e., vowel reduction) and suprasegmental information (i.e., without vowel reduction) of stress in word learning. They constructed two experimental conditions: a vowel reduction condition where the initial syllable of the target word and the competitor differed segmentally (full vowel vs. reduced vowel), e.g., stressed /pæɹ/ in *PArrot* vs. unstressed /pəɹ/ in *paRADE*, and no-vowel-reduction condition where the initial syllable of the target and competitor differed suprasegmentally (stressed vs. unstressed), e.g., target stressed /kɑɹ/ in *CARpet* vs. unstressed /kɑɹ/ in *carTOON*. Via an eye-tracking experiment, the listeners were presented with four printed words (the target, the competitor, and two distracters) on the screen and had to click on the corresponding word they heard. The results showed that English native listeners used both suprasegmental and segmental information together, as well as suprasegmental information alone to recognize words. Mandarin advanced L2 learners of English employed stress in lexical access only when a stress difference was realized by suprasegmental cues, while Seoul Korean advanced L2 learners of English did so when stress was signaled by suprasegmental and segmental cues together. Their findings reveal the important role native word prosody plays in L2 word recognition. The use of lexical tones in Mandarin as suprasegmental information to differentiate words in the native language may enhance their perception of another suprasegmental cue (stress) in L2 word recognition. In contrast, Seoul Korean does not use suprasegmental for lexical contrast (Jun, 2005; Sohn, 1999) and thus the Seoul Korean L2 learners had to rely both on segmental (vowel reduction) and suprasegmental (stressed vs. unstressed) to constrain word activation in English.

To sum up, listeners with lexical positional marking in their native language, i.e., Dutch and Japanese listeners, can exploit positional marking in lexical activation in their native languages. The native positional marking can also be beneficial for them to recognize words signaled by a non-native positional cue (i.e., English) in an L2 (Cooper et al., 2002). Listeners who do not use position but rather tones lexically in the native language, i.e., Mandarin listeners, were shown to be able to employ position information (stressed vs. unstressed) when there was no vowel reduction in L2 word recognition in English, suggesting that Mandarin listeners may be facilitated by the use of a suprasegmental cue for lexical contrasts in the native language (Connell et al., 2008). These studies have mainly centered on L2 learners of a positional cue (e.g., English stress) in word recognition. The target L2 learners in these studies have more or less achieved an advanced level which can be regarded as an "end-state" of acquisition. They may be ultimately affected by internal and external factors such as learning motivation, age, and learning and teaching strategies (Orie, 2006; Moyer, 1999) during acquisition. It is important to address how and to what extent native word prosody can facilitate naive listeners who are not driven by the L2 learning process to encode a non-native positional cue (i.e., pitch) in building lexical representations, i.e., mapping the positional cue with meaning in word learning.

1.4.3 Summary

The current section reviewed studies with respect to the processing of non-native tones and non-native positional cues at the lexical level. Studies on lexical processing of non-native tones revealed that the ability to perceive acoustic differences of non-native tones does not entail success encoding this information in lexical representations for word learning (Braun et al., 2014). In fact, the capability of encoding non-native tones is influenced by native prosody, both at the word level and the intonation level, possibly in combination with other factors such as prior musical experience and metalinguistic awareness (Wong & Perrachione, 2007; Chandrasekaran et al., 2010; Braun et al., 2014). Since there is a lack of prior investigation into the encoding of non-native positional cues in lexical representations, the section reviewed studies on the processing of non-native positional cues in L2 studies, mostly focusing on L2 advanced learners in word recognition. It is shown that native word prosody plays a crucial role in perceiving non-native positional cues (i.e., English stress) in L2 word recognition. For instance, the use of lexical positional marking in the native language could benefit Dutch L2 advanced learners of English to use English stress cues for constraining lexical activation in English. However, the L2 advanced learners are at the "end-state" of acquisition. They may have been influenced by not only the native language but also other variables such as learning period, learning motivation, age, and teaching strategies (Orie, 2006; Moyer, 1999). In contrast, studying naive listeners who have never been exposed to the non-native WPCs may allow us to have a better understanding of how and to what extent one's native language (i.e., native word prosody) plays a role in encoding non-native WPCs when establishing sound-to-meaning associations.

1.5 Research gaps, research aim and methods

Based on the reviewed literature, four main research gaps are identified with respect to the role of native prosody in perception of non-native WPCs (pitch contrasts and positional marking). First, no study to date has investigated the *acoustic* perception of non-native WPCs (i.e., non-native pitch contrasts depending on positions, which is to say, non-native pitch contrasts occurring in different positions in word and non-native tones contrastive in position) from a cross-linguistic perspective. It is not fully understood whether the use of positional marking in the native language can contribute to the perception of non-native pitch contrasts depending on position. Second, it remains unclear how listeners from different language backgrounds perceive non-native WPCs when fine-grained acoustic details are inaccessible, that is, at the *phonological* level. Third, no studies have addressed how and to what extent native word prosody influences listeners at the *lexical* level, that is, encoding non-native WPCs in establishing lexical representations. Last but not the least, it is unknown how the perception at the three linguistic processing levels relate to each other, that is, to what extent acoustic sensitivity can facilitate phonological processing or sound-to-meaning mapping.

To bridge the gaps, the aim of the dissertation is to develop a comprehensive understanding of the role of native prosody in the perception of non-native WPCs from auditory-acoustic perception, phonological processing, and sound-to-meaning mapping by listeners from three different language backgrounds within a single study. To be more specific, the dissertation attempts to address *language-specificities* and *commonalties* in non-native WPC perception at the three levels as well as how perception at each level relates to the other levels. In order to achieve these goals, three experiments will be conducted and described in three empirical chapters, Chapter 2, 3 and 4, as outlined below.

In Experiment 1, an AX discrimination task will be applied to examine *acoustic* perception of non-native WPCs. AX tasks encourage an acoustic mode of listening. Among perception tasks, it has the least memory load and stimulus uncertainty because listeners spend the least amount of effort to retain auditory information in a short-term memory store and thus listeners can make same-or-different judgments based on immediate acoustic comparison of the two stimuli (Strange & Shafer, 2008; Logan & Pruitt, 1995; Dupoux et al., 1997). In the AX discrimination task, Mandarin, Dutch, and Japanese participants are required to judge whether the two stimuli, A and X, they have heard are the same or not.

In Experiment 2, in order to investigate *phonological* processing of non-native WPCs, the same participants of Experiment 1 will be tested in a sequence recall task adapted from Dupoux et al. (2001). The sequence recall task incorporates a high memory load and phonetic variability so as to eliminate acoustic traces listeners may rely on and highlight the phonological level (Dupoux et al., 2001). Participants in the sequence recall task will learn stimulus A and B. They have to recall the sequence (e.g., A-B in a two-word sequence, A-B-A in a three-word sequence and A-B-A-B in a four-word sequence) after hearing the sequence.

In Experiment 3, the same participants of Experiments 1 and 2 will be examined in an associative word learning task. They will first learn sound-object associations and then be tested in a picture selection task where they are required to choose the corresponding picture (representing the object) when hearing the stimulus (sound).

1.6 Dissertation outline and research questions

Chapter 2 sets out to investigate the general issue of to what extent one’s native language influences perception of non-native WPCs, pitch contrasts with positional marking at the acoustic level from a cross-linguistic perspective. Non-native pitch contrasts with positional marking can be disentangled in two aspects: non-native pitch contrast occurring in different positions and non-native tones contrastive in position. Chapter 2 addresses the following questions:

RQ 1a: How will Mandarin, Japanese and Dutch listeners differ in perceiving non-native pitch contrasts occurring in different positions? More specifically, will the native WPCs facilitate the listeners' perception of non-native pitch contrasts that occur in different positions?

RQ 1b: What perceptual patterns will Mandarin, Japanese, and Dutch listeners display when perceiving non-native pitch contrasts occurring in different positions? More specifically, will listeners show language-specific patterns or will they share similarities?

RQ 1c: When perceiving the non-native pitch contrasts in different positions acoustically, to what extent will position interact with non-native pitch contrast in each language group? Will listeners show specific preferences for any positions when perceiving non-native pitch contrasts?

RQ 1d: How will Mandarin, Japanese, and Dutch listeners differ when perceiving non-native tones that occur in contrastive positions? More specifically, will Japanese and Dutch listeners share commonalities, that is, transfer a native positional sensitivity to perceive non-native positional tones? Will Mandarin listeners who use tones but not positional marking rely on the tonal sensitivity to perceive non-native positional tones?

The ability to perceive a non-native WPC acoustically does not entail success in perceiving it at the *phonological* level where acoustic traces are not available. From this motivation, Chapter 3, following Chapter 2, asks the following questions:

RQ 2a: If listeners are found acoustically sensitive to perceive non-native WPCs in Chapter 2, will the listeners remain sensitive to the non-native WPCs (i.e., pitch contrasts with positions) at the *phonological* level? More specifically, to what extent will native word prosody (or utterance prosody) affect listeners *phonologically* processing non-native WPCs?

RQ 2b: What specific perceptual patterns will the three groups of listeners display when processing non-native WPCs phonologically? More specifically, how differently will they perform? Will listeners who share commonality in their native use of WPCs share similar perceptual patterns at the phonological level?

RQ 2c: Will position play a role in the perception of non-native WPCs at the phonological level? To what extent will position influence each language group in their encoding of non-native WPCs at the phonological level? Will listeners show preference for specific positions in their phonological processing of non-native pitch contrasts?

In acoustic perception of non-native WPCs, listeners are able to rely on acoustic details to achieve perceptual success. However, the capability of perceiving non-native WPCs acoustically does not guarantee success establishing *lexical representations* of sound (non-native WPCs) to meaning. From this starting point, following Chapters 2 and 3, Chapter 4 investigates to what extent native word prosody influences listeners in *sound-to-meaning mapping* from a cross-linguistic perspective by addressing:

RQ 3a: Will native word prosody (and/or native intonation system) facilitate the encoding of non-native WPCs when building lexical representations in word learning? How will the three groups differ in word learning?

RQ 3b: What language-specific patterns will Mandarin, Japanese, and Dutch listeners show in their encoding of non-native WPCs in word learning? Will they share commonalities in sound-to-meaning mapping?

RQ 3c: How will position influence, i.e., interfere or enhance the processing of non-native WPCs in word learning for each language group? To what extent will position interact with the processing of non-native pitch contrasts in word learning?

Chapter 5 will conclude the major findings from the three empirical chapters and offer a general discussion with the findings from the acoustic level, the phonological level, and in word learning. It aims to provide insights into theoretical perceptual mechanisms from a cross-linguistic perspective and from different processing levels. Limitations and future work will also be discussed in Chapter 5.

Chapter 2
Perception of non-native WPCs at the acoustic level

2.1 Introduction

As introduced in Chapter 1, Chapter 2 sets out to investigate how native language influences the perception of non-native word prosodic cues (WPCs), namely, pitch contrasts and positional marking, at *the acoustic level* by Mandarin, Dutch and Japanese listeners.

To recapitulate, a large body of previous studies has demonstrated that native word prosody plays a pivotal role in discriminating and/or identifying non-native pitch contrasts (i.e., pitch contour and/or pitch level contrasts). It is documented that prior experience of using pitch for lexical contrasts in the native language facilitates non-native tone perception. For instance, Mandarin listeners were found to be able to discriminate and/or identify non-native tones (e.g. Cantonese tones, Thai tones) and showed an advantage over non-tone listeners (e.g., English and Seoul Korean listeners) (Lee et al., 1996; Gottfried & Suiter, 1997; Wayland & Guion, 2004; Schaefer & Darcy, 2014). Neurophysiology studies further observed that Mandarin listeners, compared with non-tone language listeners (e.g., English listeners), displayed a more sensitive brain stem mechanism for pitch representation, reflected by more pitch strength and better pitch tracking accuracy when perceiving non-native Thai tones (Krishnan & Gandour, 2009; Krishnan et al., 2010; Krishnan et al., 2005). Similarly, the employment of lexical pitch accent in Japanese helps Japanese listeners to be sensitive to non-native tones (e.g., Thai tones and Mandarin tones) and manifest better perceptual performance than non-tone listeners (e.g., English listeners, French listeners and Seoul Korean listeners) (Guion & Pederson, 2007; So & Best, 2010, 2012, 2014; Schaefer & Darcy, 2014).

On the other hand, the use of lexical pitch in the native language is not always beneficial in non-native tone perception. For instance, Mandarin listeners were found to encounter difficulties in discriminating non-native level tones (e.g. Thai level tones and Cantonese level tones), which is likely due to the lack of pitch level contrasts in Mandarin (Francis et al., 2008; Qin & Mok, 2013; Wu et al., 2014). Japanese listeners were reported to show confusion in discriminating or identifying non-native tonal contrasts when the non-native tonal pairs share phonetic similarities (e.g., low rising tone T5 vs. low level tone T6 in Cantonese) and when one-to-one mapping between the native accentual patterns and the non-native tones is lacking (So & Best, 2010; Tsukada et al., 2016; Wong, 2019) (see §1.2.1.2 for detailed discussion).

Moreover, studies have shown that it is not only word prosody but also sometimes intonation in the native language that influences non-native tone perception. For example, Braun & Johnson (2011) examined the perception of Mandarin tones by Dutch listeners. They found that Dutch listeners were more sensitive to the Mandarin rising and falling tones on the final syllable than on the initial syllable, which could be due to the fact that final pitch movements can signal post-lexical information to Dutch listeners, such as question and statement contour (Haan et al., 1997; Van Heuven & Haan, 2002; Van Heuven & Kirsner, 2004; Van Heuven, 2017).

Mounting literature has captured cross-linguistic perception (i.e., discrimination and/or identification) of non-native pitch contrasts at the acoustic level and has unveiled the influence of native word prosody, sometimes with intonation prosody, (whether facilitation or interference). However, it seems that a scarcity of attention is paid to *position*, another pivotal cue which is employed contrastively at the word level in languages such as Japanese (the position of lexical pitch accent) and Dutch (the position of lexical stress). How do listeners with the use of different WPCs differ in perceiving non-native positional contrasts? Will listeners who use positional marking at the word level (i.e., Japanese and Dutch listeners) utilize positional sensitivity in non-native position perception? Will listeners who do not use positional marking in their native language but use pitch variations lexically (i.e., Mandarin listeners) be impeded when perceiving non-native positional contrasts?

Studies on perception of non-native position have mainly been centered on whether L2 learners can acquire and perceive contrastive position of lexical stress or pitch accent and what perceptual cues they rely on (mostly in perceiving lexical stress) (Beckman, 1986; Mochizuki-Sudo & Kiritani, 1991; Altmann &Vogel, 2002; Tokuma, 2003; Wang, 2005; Altmann, 2006; Lai, 2008; Kijak, 2009; Michaux et al., 2014; Chrabaszcz et al., 2014). There are very few cross-linguistic perception studies with naive listeners on perception of non-native positional contrasts. For instance, Dupoux et al., (1997, 2001) found that French listeners who do not use any WPCs (Rossi, 1980; Di Cristo, 1988; Vaissière, 1991) were able to discriminate contrastive stress positions in Spanish. Hu (2015) found that Dutch listeners, who use lexical stress, were overall good at discriminating Japanese pitch accent positions. The very few studies on perception of non-native positional contrasts have not accounted for a fuller understanding of how native word prosody influences perceiving non-native positional marking from a cross-linguistic perspective. Furthermore, no study to date has investigated the role of native word prosody in acoustic perception of non-native pitch contrasts (tones) with positions, that is, non-native tones contrastive in position and non-native pitch contrasts occurring in different positions. More specifically, it remains unclear whether the use of positional marking in the native language can facilitate acoustic perception of non-native positional tones and non-native pitch contrasts in different positions and how listeners differ in perceiving non-native pitch contrasts with positions acoustically. These are the gaps the current study aims to bridge.

Taken together, the present chapter taps into the general issue of the extent to which native language influences cross-linguistic perception of non-native WPCs, pitch contrasts and positional marking at the acoustic level by addressing four specific questions:

(1) How will Mandarin, Japanese and Dutch listeners differ in perceiving non-native pitch contrasts occurring in different positions? More specifically, will the native WPCs facilitate the listeners' perception of non-native pitch contrasts that occur in different positions?

(2) What perceptual patterns will Mandarin, Japanese and Dutch listeners display when perceiving non-native pitch contrasts occurring in different positions? More specifically, will they show language-specific patterns or will they share similarities?

(3) When perceiving the non-native pitch contrasts in different positions, to what extent will position interact with non-native pitch contrast in each language group? Will listeners show specific preferences for any positions when perceiving non-native pitch contrasts?

(4) How will Mandarin, Japanese, and Dutch listeners differ when perceiving non-native tones that occur in contrastive positions? More specifically, will Japanese and Dutch listeners share commonalities, that is, transfer a native positional sensitivity to perceive non-native positional tones? Will Mandarin listeners who use tones but not positional marking rely on the tonal sensitivity to perceive non-native positional tones?

In order to investigate these questions, three experiments (three AX discrimination tasks) will be conducted. Experiments 1 investigates perception of non-native pitch contrasts in isolation (in monosyllables), which serves as the baseline for Experiment 2. Experiment 2 inquires perception of non-native pitch contrasts occurring in different position (word-initial, -medial and -final in trisyllables), which aims to answer research question (1), (2) and (3). Experiment 3 addresses research question (4) that investigates perception of non-native tones in contrastive positions (word-initial vs. -medial position, word-initial vs. -final position and word-initial vs. -final position).The hypotheses and predictions are as followed:

In Experiment 1 (non-native pitch contrasts in monosyllables), it is predicted there will be an overall hierarchical order of performance: Mandarin listeners >> Japanese listeners >> Dutch listeners. This is based on the Functional Pitch Hypothesis proposed by Schaefer & Darcy (2014) (derived from the Feature Hypothesis (McAllister et al., 2002)) that the degree to which pitch functions to differentiate lexical items in the native language shapes non-native tone perception. More specifically, the higher the functionality of pitch used for lexical contrasts, the better performance the listeners will display in non-native tone perception. Such functionality of pitch is determined by several properties: whether or not a language uses lexical pitch, the prosodic domain where pitch variations are realized, whether the lexical pitch is used alone (and not in combination with other acoustic parameters to contrast words), and by the number of minimal pairs differentiated in the native language (Schaefer & Darcy, 2014). Accordingly, Mandarin has the highest functionality of pitch among the three languages in that pitch variations occur on individual syllables and they are used alone to signal lexical contrasts for most words (Chao, 1948; Ho, 1976; Howie, 1976; Lin, 1988; Duanmu, 2007). Japanese has intermediate functionality of pitch because, though pitch is used for lexical contrasts, pitch is used in a different manner from a canonical tone language. An abrupt pitch fall (pitch accent) is used to signal lexical contrasts in a prosodic word (Poser, 1984; Vance, 1987; Kubozono, 2008; Kawahara, 2015). The accented mora receives a high pitch and the following mora receives low pitch and thus it is the relative pitch between successive moras that is important for differentiating lexical items. Moreover, compared with Mandarin, Japanese has fewer minimal pairs differentiated by lexical pitch accent, 14% of minimal pairs according to Shibata & Shibata (1990), 20% according to Pierrehumbert & Beckman (1988). Dutch has low pitch functionality in that Dutch uses word stress for lexical contrasts. Pitch is not the sole acoustic correlate of lexical stress in Dutch, but is one of many. That is, the stressed syllable in a word is realized by a higher pitch (Kager 1989; Sluijter, 1995; Sluijter & van Heuven, 1996); it is a reliable cue to differentiate stressed vs. unstressed syllable(s) in a word when the word is marked with an accent in narrow focus; however, pitch is absent as a cue to lexical stress in accented words (Sluijter & Van Heuven, 1996a, 1996b). The functionality of pitch from high to low across the three languages would predict a hierarchical perceptual performance of Mandarin listeners >> Japanese listeners, Dutch listeners in perception of non-native pitch contrasts.

The predictions for the performance pattern of each language group are based on an approach in terms of selective perceptual weighting given to specific pitch features in long-term memory (Lee et al., 1996; Xu et al., 2006; Francis et al., 2008). Francis et al. (2008) extended the "multi-store memory model" (Xu et al., 2006) together with the perceptual weighting of features of tone (Lee et al., 1996) and proposed that listeners store long-term memory representations of pitch-based phonological categories (tone and/or intonation). The stored pitch representations in the native language determine listeners' language-specific weighting of perceptual features of pitch in cross-linguistic tone perception. Accordingly, Mandarin has contrastive contour tones but not contrastive level tones at the word prosody. Mandarin listeners may have long term representations of contrastive contour tones but not contrastive level tones in storage, which as a result predicts that they will be more sensitive to non-native pitch contour contrasts than pitch level contrasts. Dutch listeners will be more attentive to non-native pitch level contrasts than pitch contour contrasts because Dutch has a rich inventory of nuclear tones in its intonation system, which consists of H and L tones. Japanese listeners will be more sensitive to non-native pitch level contrasts than to non-native pitch contour contrasts because Japanese uses H and L tone in lexical words.

In Experiment 2 (non-native pitch contrasts occurring in trisyllables), it yields two divergent predictions for research question (1) (the overall perceptual performance across language groups). 1a) If lexical pitch alone is the dominant factor in influencing the perception of non-native pitch contrasts occurring in different positions, based on the Functional Pitch Hypothesis (Schaefer & Darcy, 2014), it is predicted that, although Mandarin listeners do not use positional cues for lexical contrasts, they will still outperform Japanese and Dutch listeners in that Mandarin has higher functionality of pitch than Dutch or Japanese. 1b) Alternatively, according to the Feature Hypothesis (McAllister et al., 2002), if a phonetic or phonological dimension is not used for lexical contrast in the native language, it will be difficult for naive listeners to perceive that non-native feature. So, if both lexical pitch and position influence the perception of non-native pitch contrasts occurring in different positions, it is predicted that Japanese listeners (who use position of pitch accent lexically) will overall outperform Dutch listeners (who use position of lexical stress) and Mandarin listeners (who use lexical tones) in perceiving non-native pitch contrasts that occur in different positions.

To make predictions for research question (2) (perceptual patterns in each language group), by grouping the positions where the non-native pitch contrasts occur, based on Francis et al.'s (2008) proposal in terms of selective perceptual weighting of pitch features in long-term storage, Mandarin listeners will be more sensitive to non-native pitch contour contrasts than pitch level contrasts, while Dutch listeners will attend more to non-native pitch level contrasts than to pitch contour contrasts. Japanese listeners will be more sensitive to non-native pitch level contrasts (see Experiment 1 above for details).

The predictions for research question (3) (interaction of position with the perception of non-native pitch contrasts in trisyllables) are made in light of the influence of the native word prosody. Mandarin uses lexical tones but not positional marking at the word level. Since contrastive positional cues are absent in the word prosody, Mandarin listeners may ignore the position and only use tonal sensitivity to perceive non-native pitch contrasts in different positions. Therefore it is expected that Mandarin listeners will not show preference for any position.

Dutch does not use lexical tones but lexical stress, of which the position is predictable to some degree. The primary stress in Dutch is, within a three-syllable window, generally placed at the right-hand word edge (Kager, 1989; Trommelen & Zonneveld, 1986, 1999)[11]. Due to the right-edge location of stress in Dutch, Dutch listeners are predicted to be sensitive near the end of the word when perceiving non-

[11] The stress assignment in Dutch mentioned here is a generalized pattern. The placement of stress in Dutch is more complex, based on phonological properties of words such as syllable structure and syllable weight (see Trommelen & Zonneveld, 1986, 1999; Kager, 1989 for details).

native pitch contrasts. However, taking into account that the prefinal syllable, not the final one, is the canonical position for lexical stress (Kager, 1989), Dutch listeners are predicted to prefer the word-medial position to the word-initial and/or -final position in perceiving non-native contrasts in trisyllables.

Japanese uses the presence/absence and the location of pitch accent at the word level. *n*-mora words theoretically have *n+1* possible accentual patterns (McCawley, 1968; Haraguchi, 1999; Uwano, 1999). Studies on statistical analyses of the proportion of Japanese accentual patterns have shown that all accentual patterns exist in 3-4 mora words (Sato, 1993; Suzuki, 1995; Kitahara, 2001). Due to the influence of the native accentuation, Japanese listeners are expected to show no preference for positions in perceiving non-native pitch contrasts in different positions (initial-medial-final) since unaccented form is dominant in 3-4 mora words in Japanese.

In Experiment 3 (perception of non-native positional tones in trisyllables) for research question (4), it can be predicted: 4a) based on the Feature Hypothesis (McAllister et al., 2002), Japanese listeners who use position for lexical pitch accent will show better performance than Dutch listeners (who use position for lexical stress) and Mandarin listeners (who use lexical tones). 4b) if perception of non-native positional tones is by nature the perception of position, based on the Feature Hypothesis (McAllister et al., 2002), Japanese and Dutch listeners will outperform Mandarin listeners in perceiving non-native tones contrastive in positions. 4c) if perception of non-native positional tones is by nature the perception of tone, based on Functional Pitch Hypothesis (Schaefer & Darcy, 2014), Mandarin listeners will display perceptual advantage over Dutch and Japanese listeners.

Note that the Functional Pitch Hypothesis (Schaefer & Darcy, 2014), Feature Hypothesis (McAllister et al., 2002) and Francis et al., (2008) do not strictly tease apart processing levels but generally aim to shed light on the understanding of perception of non-native sounds (i.e., cross-linguistic tone perception) and its acquisition, which entails that these hypotheses can be applied to predict perceptual performance at the acoustic level.

2.2 Method

Identification and discrimination tasks are regarded as two general types of tasks used to examine perception of phonetic contrasts (e.g., segmentals and suprasegmentals) (Strange & Shafer, 2008). In an identification task, listeners are presented with recorded stimuli one at a time and are required to identify or label the stimulus as an instance of a discrete phonetic category (Strange & Shafer, 2008). In a discrimination task, two (AX) or more stimuli (ABX) are presented and listeners should judge the relationship between the stimuli. More specifically, in an AX discrimination task, listeners should determine whether the target stimulus X matches the item A or not. In an ABX (or variation AXB) task, listeners should judge whether the target stimulus X matches the item A or the item B, which may, to some extent, tap into the listener's phonological representations (Dupoux et al., 1997; Strange & Shafer, 2008). Comparatively, AX task imposes less *memory load* on listeners in that listeners retain less auditory trace of previous stimulus in their short-term memory stores and thus listeners can make same-or-different judgments based on the immediate acoustic comparison of the two stimuli, which encourages an acoustic mode of listening to a larger extent (Logan & Pruitt, 1995; Strange & Shafer, 2008). Therefore, AX discrimination task will be applied in the current study to investigate perception of non-native WPCs at the acoustic level. In the AX paradigm, the order of the two items to be discriminated can be AA, XA and AX.

2.2.1 Participants

Thirty native speakers of Dutch (eight male and twenty-two female) and thirty-two native speakers of Mandarin Chinese (thirteen male and nineteen female) were recruited from Utrecht University. Thirty-two native speakers of Tokyo Japanese (ten male and twenty-two female) were recruited from International Christian University in Tokyo. The average age of Dutch, Mandarin and Japanese listeners was 22 years old (SD = 4.1), 25 years old (SD = 3.7) and 25 years old (SD = 5.6), respectively. All the Mandarin participants speak native northern Chinese dialects, with no prior experience of Cantonese or any other tonal languages including southern Chinese dialects. None of the Dutch or Japanese listeners had been exposed systematically to any tonal languages including tonal dialects (e.g., Limburgian for Dutch listeners and Osaka dialect for Japanese listeners). None of the Mandarin participants spoke Dutch or have learned Dutch. The only foreign language all the three groups of participants have been exposed to is English. None of the participants had any systematic musical training and none reported any hearing impairment.

2.2.2 Stimuli

Cantonese tones were selected as the stimuli material in that Cantonese tones are contrastive in pitch level and pitch contour, which is non-native to Dutch, Japanese and Mandarin listeners. There are six contrastive tones in Cantonese (Fok Chan, 1974; Bauer & Benedict, 1997), three level tones, namely, high level tone (T1) , mid level tone (T3) and low level tone (T6), and three contour tones, namely, high rising tone (T2), low falling tone (T4) and low rising tone (T5), as displayed in Figure 2.1. I selected three level tones high vs. mid vs. low level tone (T1, T3 and T6) as pitch level contrasts and two contour tones low rising and low falling tone (T4 and T5) as pitch contour contrasts.

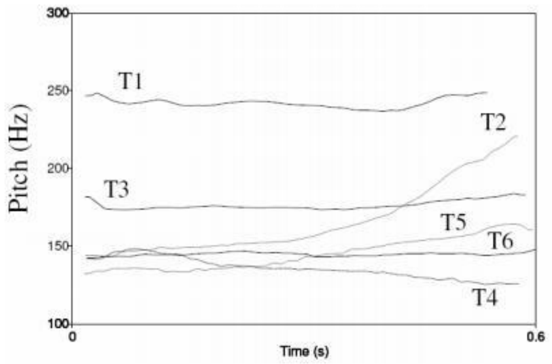


Figure 2.1. *Cantonese tones* (reproduced from Qin & Mok (2013) with permission).

2.2.2.1 Experiment 1: Perception of non-native pitch contrasts in monosyllables

The stimuli in Experiment 1 consisted of the monosyllable /ba/ carrying three level tones and two contour tones. Six repetitions of /ba/ in each tone were produced and recorded in isolation in Praat (Boersma & Weenink, 2008) by a 26-year-old native Cantonese female speaker who is a well-trained phonetician. Three tokens of the best quality of each tone were selected. All the tokens were normalized in Praat with an intensity of 70 dB and duration of 400 ms. The five tones result in 10 contrastive tone pairs as shown in Table 2.1. The overall design of Experiment 1 was: target pairs (10) x the order of X (2: AX, XA) with fillers (10 AA pairs), which contained 30 trials in total.

Table 2.1. *Non-native pitch contrasts in monosyllables (Experiment 1).*

Pitch contrasts	**Tonal pattern**
Pitch level contrast	T1-T3 (high vs. mid level tone)
	T1-T6 (high vs. low level tone)
	T3-T6 (mid vs. low level tone)
Pitch level	T1-T4 (high level vs. low falling tone)
vs.	T1-T5 (high level vs. low rising tone)
pitch contour contrast	T3-T4 (mid level vs. low falling tone)
	T3-T5 (mid level vs. low rising tone)
	T6-T4 (low level vs. low falling tone)
	T6-T5 (low level vs. low rising tone)
Pitch contour contrast	T4-T5 (low falling vs. low rising tone)

2.2.2.2 Experiment 2: Perception of non-native pitch contrasts in trisyllables

The stimuli in Experiment 2 consisted of trisyllabic nonwords /bababa/. All tokens of /bababa/ were concatenated sequences consisting of three monosyllables /ba/ the same as in Experiment 1. Six tokens of the monosyllable /ba/ in each tone were recorded by the same speaker in Experiment 1. Three tokens of the best quality of /ba/ were selected and normalized in Praat with an intensity of 70 dB and duration of 300ms. In order to maximally preserve the word-like character of the /bababa/ sequence, it was presented to five Cantonese native speakers, four Mandarin native speakers, two Japanese native speakers and four Dutch native speakers in four conditions: no silence inserted among each syllable in /bababa/, 15 ms, 25 ms, 35ms and 50ms silence inserted among each syllable in /bababa/, respectively. They were required to score to what extent /bababa/ sounded like a natural word on a 5-point scale (where 1 means ‘not word-like’ and 5 ‘extremely word-like’). /bababa/ with 25 ms silence inserted among each syllable was scored highest [12] .None of the participants regarded or noticed the 25 ms as a salient pause, which might influence their perception. Thus 25 ms silence was selected, yielding a total duration of all trisyllabic tokens of (300+25+300+25+300=) 950 ms. Figure 2.2 shows an example of the stimuli /bababa/.

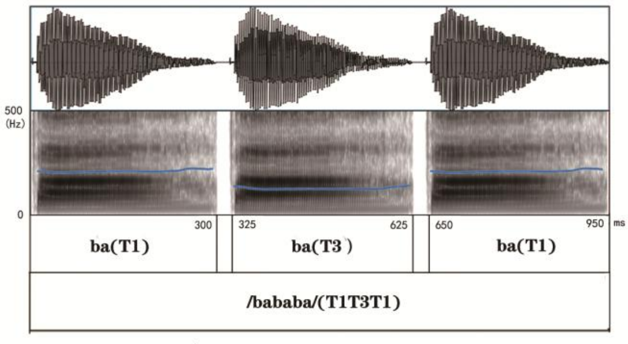


Figure 2.2 Stimulus /bababa/ (T1T3T1).

To investigate perception of non-native pitch contrasts depending on positions, the 10 contrastive tonal pairs in Experiment 2 occurred in word-initial, -medial or -final position. Three level tones were used as contextual tones in the other two positions. The whole design is displayed in Table 2.2. For instance, a target pair could be: T1-T1-T5 (high level tone-high level tone-low rising tone) vs. T1-T1-T6 (high level tone-high level tone-low level tone), where T1 (high level tone) was the contextual tone and T5-T6 (low rising vs. low level) tone was the contrastive tone pair on word-final position.

---

[12] See Appendix A for details.

The overall design in Experiment 2 was: contextual tone (3: T1 high level tone, T3 mid level tone or T6 low level tone) x tone pair (the same 10 contrasts in Part a) x position (3: word-initial, -medial, and -final) x the order of X (2: AX or XA), resulting in 180 trials in total. 65 fillers were used. The fillers consisted of AA pairs of trisyllabic nonwords /bababa/ in which each syllable carried the same tone such as T1-T1-T1 vs. T1-T1-T1, or different tones such as T1-T4-T4 vs. T1-T4-T4 or T4-T5-T6 vs. T4-T5-T6.

Table 2.2. *Non-native pitch contrasts in trisyllables (Experiment 2).*

Pitch contrasts	**Tonal pattern**	**Examples**	**Position**
Pitch level contrast	T1-T3	T1T1T1-T3T1T1	Initial
		T1T1T1-T1T3T1	Medial
		T1T1T1-T1T1T3	Final
	T1-T6	T1T1T1-T6T1T1	Initial
		T1T1T1-T1T6T1	Medial
		T1T1T1-T1T1T6	Final
	T3-T6	T3T1T1-T6T1T1	Initial
		T1T3T1-T1T6T1	Medial
		T1T1T3-T1T1T6	Final
Pitch level vs. pitch contour contrast	T1-T4	T1T1T1-T4T1T1	Initial
		T1T1T1-T1T4T1	Medial
		T1T1T1-T1T1T4	Final
	T1-T5	T1T1T1-T5T1T1	Initial
		T1T1T1-T1T5T1	Medial
		T1T1T1-T1T1T5	Final
	T3-T4	T3T1T1-T4T1T1	Initial
		T1T3T1-T1T4T1	Medial
		T1T1T3-T1T1T4	Final
	T3-T5	T3T1T1-T5T1T1	Initial
		T1T3T1-T1T5T1	Medial
		T1T1T3-T1T1T5	Final
	T4-T6	T4T1T1-T6T1T1	Initial
		T1T4T1-T1T6T1	Medial
		T1T1T4-T1T1T6	Final
	T5-T6	T5T1T1-T6T1T1	Initial
		T1T5T1-T1T6T1	Medial
		T1T1T5-T1T1T6	Final
Pitch contour contrast	T4-T5	T4T1T1-T5T1T1	Initial
		T1T4T1-T1T5T1	Medial
		T1T1T4-T1T1T5	Final

2.2.2.3 Experiment 3: Perception of non-native positional tones in trisyllables

Experiment 3 examined the perception of non-native positional tones, where the five tones (T1, T3, T4, T5 and T6) were kept constant while the position was contrastive. Level tones were used as contextual tones. Table 2.3 shows the whole design. For instance, contrastive positional pair could be /bababa/ in T1-T1-T5 vs. T1-T5-T1 (high level tone - high level tone - low rising tone vs. high level tone - low rising tone - high level tone), in which the positional contrast of T5 is in word-final versus word-medial position. The overall design is: contextual tone (3: T1 high level tone, T3 mid level tone or T6 low level tone) x target tone (5: T1, T3, T4, T5 and T6) x positional contrast (3: word-initial vs. -medial position, word-initial vs. -final position and word-medial vs. -final position) x the order of X (2: AX or XA), leading to 90 pairs in total. 30 fillers were used, consisting of AA pairs of trisyllabic nonwords /bababa/ in which each syllable carries the same tone such as T4-T4-T4 vs. T4-T4-T4, or different tones such as T1-T5-T3 vs. T1-T5-T3 or T3-T3-T4 vs. T3-T3-T4.

Table 2.3. *Positional contrasts of non-native tones in trisyllables (in Experiment 3).*

Tone	**Examples**	**Positional contrast**
T1	T1T3T3-T3T1T3	Initial vs. medial
(high level tone)	T1T3T3-T3T3T1	Initial vs. final
	T3T1T3-T3T3T1	Medial vs. final
T3	T3T1T1-T1T3T1	Initial vs. medial
(mid level tone)	T3T1T1-T1T1T3	Initial vs. final
	T1T3T1-T1T1T3	Medial vs. final
T4	T4T3T3-T3T4T3	Initial vs. medial
(low falling tone)	T4T3T3-T3T3T4	Initial vs. final
	T3T4T3-T3T3T4	Medial vs. final
T5	T5T1T1-T1T5T1	Initial vs. medial
(low rising tone)	T5T1T1-T1T1T5	Initial vs. final
	T1T5T1-T1T1T5	Medial vs. final
T6	T6T3T3-T3T6T3	Initial vs. medial
(low level tone)	T6T3T3-T3T6T3	Initial vs. final
	T6T3T3-T3T6T3	Medial vs. final

The ISI (inter-stimulus-interval) in each trial was 800 ms in Experiment 1 (perception of non-native pitch contrasts in monosyllables) and 1200 ms in Experiment 2 (perception of non-native pitch contrasts in trisyllables) and Experiment 3 (perception of non-native positional tones in trisyllables). All the trials in each experiment were randomized.

2.2.3 Procedure

The three experiments were programmed and conducted in Zep software (Veenker, 2012). All the participants participated in the three experiments in one go. They all participated in Experiment 1 first. After finishing Experiment 1, half of them participated in Experiment 2 next and half of them participated in Experiment 3 next. They were tested individually in a sound-attenuated room, equipped with a computer, headphones and a two-button button box. Dutch and Mandarin listeners were tested in the UiL-OTS phonetics lab at Utrecht University. Japanese listeners were tested in a quiet room in the Research and Education Centre at the International Christian University in Tokyo. All the participants were instructed to listen to a pair of tokens that only differed in "melody" and to judge whether the tokens were same or different by pressing the button "Same" or "Different" on the button box as quickly and accurately as possible.

Each experiment consisted of a practice phase (6 trials in each part) and a test phase (30, 245, 120 trials in each part, respectively). Participants were instructed to take a 3-5 minutes' break after finishing each experiment. In the test phase of Part b in Experiment 1, when the trials proceeded to 120, there was a three-minute break so that the participants could take a rest. After the break, they could proceed by themselves. The trial proceeded immediately once the response was made. If the participant failed to respond within 1500 ms, the trial would proceed automatically. Completing all three experiments took approximately 45 minutes.

2.3 Results and discussion

2.3.1 Experiment 1: Perception of non-native pitch contrasts in monosyllables

2.3.1.1 Perceptual performance across language groups

A correct response made by a participant was marked as "1" while an incorrect response was marked as "0". Any missing response was regarded as an incorrect response. To analyze whether the three language groups differ in perceiving non-native tonal contrasts, a generalized linear mixed model (GLMM) was computed in SPSS 25. Language (3 levels) and Contrast (10 levels) were taken as fixed factors. Intercepts for participants and items were added as random effects[13]. The F-tests show that Contrast (F (9, 1,850) = 5.652, $p < 0.001$) was significant and Language (F (2, 1,850) = 0.008, $p = 0.992$) was not significant, as reported in Table 2.4. An interaction of Language and Contrast was significant (F (18, 1,850) = 3.225, $p < 0.001$), suggesting that the performance of language groups differed on the pitch contrasts. Figure 2.3 displays estimated proportion of accuracy of each pitch contrast of each language group.

[13] Random slope was not added in the analysis in the three empirical chapters in the dissertation because the models with random slope failed to converge.

Table 2.4. *Parameters for fixed effects and random effects in Experiment 1.*

Fixed effects	F	df	p
Language group	0.019	2	0.992
Tonal contrast	5.652	9	<0.001
Language group * Tonal contrast	3.225	18	<0.001

Random effects	Est.	Std. Error	z	p
1 | Participant	1.900	0.424	4.481	<0.001
1 | Item	1.157	0.267	4.324	<0.001

Firstly, to compare the perceptual differences across language groups, a post hoc pair-wise analysis with Bonferroni adjustment was conducted. It showed that the three language groups did not differ in perceiving pitch contrasts located in high-mid register. They differed significantly in perceiving pitch contrasts located in mid to low register zone (T4-T5, T4-T6 and T5-T6), as can be seen in Figure 2.3. More specifically, Mandarin listeners showed significantly better performance than Japanese listeners (F (9, 1,850) = 14.689, $p = 0.013$) and Dutch listeners (F (9, 1,850) = 31.802, $p < 0.001$) in perceiving non-native T4-T5. Mandarin listeners did not differ from Japanese listeners (F (9, 1,850) = 7.984, $p = 0.068$) in perceiving T4-T6 while they both significantly outperformed Dutch listeners (F (9, 1,850) = 62.173, $p < 0.001$). When perceiving T5-T6, Mandarin listeners significantly outperformed Japanese listeners (F (9, 1,850) = 207.132, $p < 0.001$) and Dutch listeners (F (9, 1,850) = 212.365, $p < 0.001$), while Japanese listeners significantly outperformed Dutch listeners (F (9, 1,850) = 13.712, $p = 0.017$).

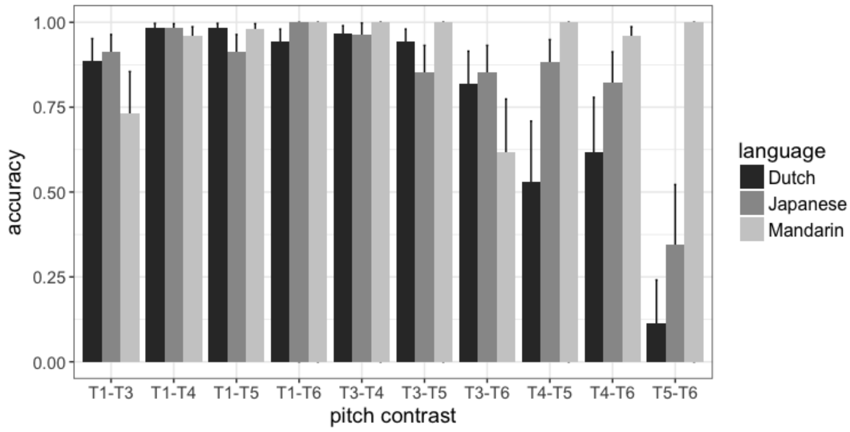


Figure 2.3. *Performance of each contrast by each group in monosyllables (Experiment 1). Error bar*[14]*: 95%CI*

Secondly, a post hoc pair-wise analysis of pitch contrasts (with Bonferroni adjustment) for each language group was conducted to investigate language-specific perceptual patterns. As shown in Figure 2.3, Dutch listeners performed significantly worse in perceiving T5-T6 than all the other pitch contrasts, with around 13% accuracy (F (9, 1,850) = 41.920, $p < 0.001$). Apart from the most troublesome T5-T6, they also displayed worse performance in perceiving T4-T5 (F (9, 1,850) = 38.105, $p < 0.001$) and T4-T6 (F (9, 1,850) = 34.376, $p < 0.001$) than the other contrasts.

Japanese listeners had significantly more trouble in perceiving T5-T6, with only 36.9% accuracy, well below chance level (50%), compared with their good performance in perceiving other pitch contrasts all reaching more than 80% accuracy (F (9, 1,850) = 6.710, $p < 0.001$).

Different from Dutch and Japanese listeners, Mandarin listeners showed significant vulnerability in perceiving T3-T6, below 75% accuracy, than the other contrasts (F (9, 1,850) = 5.981, $p = 0.031$) which all achieved above 90% accuracy, except for T1-T3 (73.5% accuracy) (F (9, 1,850) = 2.418, $p = 0.895$).

2.3.1.2 Summary

To summarize, the three language groups differed in perceiving non-native pitch contrasts in monosyllables. All three groups achieved equally good performance in perceiving non-native pitch contrasts paired with T1 and T3 in high-mid register, which could be due to the relatively high acoustic salience of T1 and T3 (Qin & Mok, 2013). However, they differed in perceiving non-native pitch contrasts located

[14] The error bar reported in the current dissertation refers to 95% CI.

in mid-low register (T4-T5, T4-T6 and T5-T6). More specifically, Mandarin listeners showed better performance than Dutch and Japanese listeners in perceiving T4-T5 and T5-T6. Mandarin and Japanese listeners performed equally well in discriminating T4-T6, both displaying advantage over Dutch listeners. The advantage of Mandarin listeners over Japanese and Dutch listeners in discriminating non-native pitch contrasts in mid-low register could be due to the maximal functionality of pitch in the native language (Schaefer & Darcy, 2014).

Moreover, language-specific perceptual patterns were observed as well. Dutch listeners had relatively worse performance in perceiving pitch contrasts located in mid-low register (T4-T5, T4-T6 and T5-T6) compared with their good performance in perceiving other contrasts. They were still able to perceive T4-T5 and T4-T6, both above chance level (50%). However, they seemed especially unable to perceive T5-T6, with merely 13.7% accuracy. T5-T6 brought the most trouble to Japanese listeners as well. Japanese listeners achieved good performance in discriminating all the other contrasts except for the problematic T5-T6. Unlike Dutch and Japanese listeners, Mandarin listeners were relatively worse in discriminating non-native pitch level contrasts T1-T3 (high level vs. mid level tone) and T3-T6 (mid level vs. low level tone) compared with their good performance in discriminating other contrasts. The differences in perceptual patterns between language groups can be attributed to influence of the native word prosody, which will be further discussed in detail in §2.4.

2.3.2 Experiment 2: Perception of non-native pitch contrasts in trisyllables

2.3.2.1 Perceptual performance across language groups

Experiment 2 investigated the specific perceptual patterns of each language group in perceiving non-native pitch contrasts (the same 10 non-native pitch contrasts in Experiment 1) occurring in different positions (word-initial, -medial and –final position). A GLMM was computed in SPSS 25 in which Language (3 levels), Contrast (10 levels) and Position (3 levels) were computed as fixed factors. Intercepts for participants and items were added as random effects. Table 2.5 reports that Language (F (2, 8,424) = 12.412, $p < 0.001$), Contrast (F (9, 8,424) = 15. 292, $p < 0.001$) and Position (F (2, 8,424) = 7.673, $p < 0.001$) were significant. Furthermore, the interaction between Language and Contrast (F (18, 8,424) = 10.045, $p<0.001$) and between Language and Position (F (4, 8,424) = 3.349, $p = 0.010$) were significant as well, indicating that the differences between language groups depended on the contrast and position, respectively. No interaction between Contrast and Position was found (F (18, 8,424) = 0.487, $p = 0.965$), nor was a three way interaction between Contrast, Position and Language (F (54, 8,370) = 0.542, $p = 0.998$).

Table 2.5. *Parameters for fixed effects and random effects in Experiment 2.*

Fixed effects	F	df	p
Language group	12.412	2	<0.001
Tonal contrast	15. 292	9	<0.001
Position	7.673	2	<0.001
Language group * Tonal contrast	10.045	18	<0.001
Language group * Position	3.349	4	0.010
Tonal contrast * Position	0.487	18	0.965

Random effects	Est.	Std. Error	z	p
1 | Participant	0.824	0.154	5.348	<0.001
1 | Item	0.922	0.184	5.020	<0.001

The observed interaction effects between Language and Contrast, as well as between Language and Position, allowed us to further investigate, 1) perceptual differences across the three groups for non-native pitch contrasts and positions and, 2), perceptual patterns within each group in perception of non-native pitch contrasts and positions. First, a Bonferroni-corrected post hoc pair-wise analysis was carried out to compare differences across groups in perception of non-native pitch contrasts and positions. Figure 2.4 provides a detailed scenario of the performance of the three groups in perceiving non-native pitch contrasts. It can be seen that, although all the groups performed equally well in discriminating contrasts in high-mid register (T4-T5, T4-T6 and T5-T6) with nearly ceiling performance, they differed from each other in other pitch contrasts. Mandarin listeners outperformed Dutch listeners in discriminating all three mid-low tone contrasts (T4-T5: F (2, 8,424) = 9.247, $p < 0.001$), T4-T6: (F (2, 8,424) = 4.05, $p = 0.020$) and T5-T6: (F (2, 8,424) = 33.757, $p < 0.001$). Mandarin listeners showed significantly better performance than Japanese listeners in discriminating contrast T4-T5 (F (2, 8,424) = 9.247, $p < 0.001$) and T5-T6 (F (2, 8,424) = 33.757, p < 0.001). Japanese listeners significantly outperformed Dutch listeners in discriminating contrast T4-T6 (F (2, 8,424) = 4.050, $p = 0.020$).

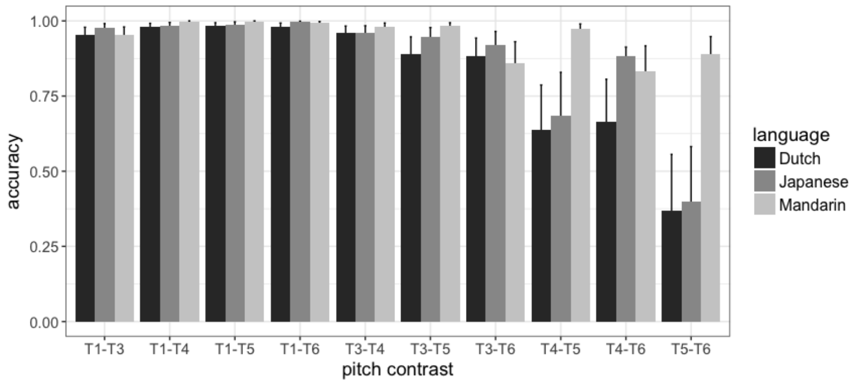


Figure 2.4 *Performance of perception of non-native pitch contrasts of each group in trisyllables (Experiment 2).*

Figure 2.5 displays the performance of the three language groups in each position. When the non-native pitch contrasts occurred in word-initial position, Mandarin listeners outperformed Dutch (*F* (2, 8,424) = 6.799, *p = 0.002*) and Japanese listeners (*F* (2, 8,424) = 4.601, *p = 0.042*). Similarly, when perceiving non-native pitch contrasts in word-medial position, Mandarin listeners outperformed Dutch (*F* (2, 8,424) = 6.437, *p=0.005*) and Japanese listeners (*F* (2, 8,424) = 6.314, *p = 0.009*). In word-final position, both Mandarin listeners (*F* (2, 8,424) =6.179, *p* = 0.007) and Japanese listeners (*F* (2, 8,424) = 5.581, *p = 0.013*) showed better performance than Dutch listeners in word final position, while the former two groups did not differ (*F* (2, 8,424) = 2.917, *p = 0.105*).

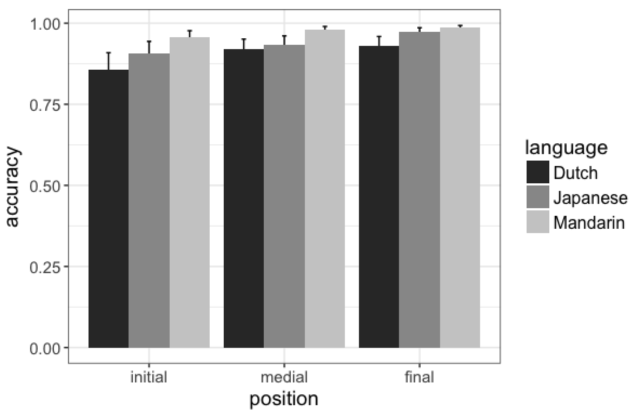


Figure 2.5. *Performance of perception in each position by each group in trisyllables (Experiment 2)*

Secondly, a Bonferroni-corrected post hoc pair-wise analysis with contrast (10 levels) and position (3 levels) was conducted in GLMM for each language group to investigate language-specific perceptual patterns. For Dutch listeners, Contrast (F (9, 8,424) = 17.761, $p < 0.001$) and Position (F (2, 8424) = 5.383, $p = 0.005$) were both significant. Dutch listeners reached a great performance in discriminating contrasts with T1, with accuracy above 95%. T5-T6, compared with the other contrasts (F (9, 8,424) =6.858, $p < 0.001$ to all the other contrasts), was significantly difficult for Dutch listeners, with accuracy below chance level (37% accuracy). Besides T5-T6, T4-T6 and T4-T5 (F (9, 2,688) = 8.692, all $p < 0.001$) were troublesome for Dutch ears as well; however, the discrimination of these two contrasts were still above chance level. Regarding position, Dutch listeners performed significantly better on word-final (93.1% accuracy) than word-initial position (85.6% accuracy) (F (2, 8,424) = 3.016, $p = 0.043$). Moreover, no significant interaction between Contrast and Position was observed (F (18, 8,424) = 0.332, $p = 0.996$).

For the Japanese group, both Contrast (F (9, 8,424) = 12.510, $p<0.001$) and Position (F (2, 8,424) = 4.935, $p = 0.007$) was significant. Contrast did not interact with Position significantly (F (18, 8,424) = 0.467, $p = 0.972$). It was found that Japanese listeners' performance of perceiving T5-T6 was significantly worse than that of the other pitch contrasts (F (9, 8,424) = 5.998, $p < 0.001$ to all the other contrasts). T4-T5 was worse than the others (F (9, 8,424) = 5.173, $p < 0.05$ to all the other contrasts) except for T5-T6. With regard to the performance on positions, they performed significantly better on word-final (97.5% accuracy) (F (2, 8,424) = 6.266, $p = 0.005$) than word-medial (93.3% accuracy) and word-initial (90.8% accuracy) (F (2, 8,424) = 5.937, $p = 0.024$).

For Mandarin listeners, Contrast was not significant (F (9, 8,424) = 1.511, $p = 0.138$). They performed equally well in perceiving all the pitch contrasts (F (9, 8,424) = 2.510, $p>0.05$ in all pair-wise comparisons). With respect to perceiving pitch contrasts on different positions (Position: F (9, 8,424) = 3.850, $p = 0.041$), they performed significantly better when pitch contrasts occurred on word- final (98.7% accuracy) than word-initial position (95.8% accuracy) (F (2, 8,424) = 3.360, $p = 0.030$). No interaction was found between Contrast and Position (F (18, 8,424) = 0.138, $p = 0.999$), suggesting that position does not interfere with Mandarin listeners' performance in differentiating contrasts.

2.3.2.2 Summary

Taken together, when perceiving non-native pitch contrasts in trisyllables, the three language groups differed on pitch contrasts and on positions. With regard to non-native pitch contrasts, the three groups performed equally well when perceiving pitch contrasts located in high-mid register. What made a difference was their performance of pitch contrasts in mid-low register. Mandarin listeners showed advantage over Dutch and Japanese listeners in perceiving contrasts located in mid-low register (T4-T5, T4-T6 and T5-T6), which was in line with the findings in Experiment 1 (perception of non-native pitch contrasts in monosyllables).

Regarding the performance for each position, Mandarin listeners overall showed a better performance than Japanese and Dutch listeners. Specifically, when the non-native pitch contrasts occurred in word-initial and -medial positions, Mandarin listeners outperformed Dutch listeners and Japanese listeners. Mandarin listeners and Japanese listeners outperformed Dutch listeners in perceiving non-native pitch contrasts in word-final position. The better performance of Mandarin listeners in perceiving non-native contrasts in each position could be a benefit from lexical pitch being used to the maximal degree in Mandarin (Schaefer & Darcy, 2014).

Furthermore, position did not interact with the perception of non-native pitch contrasts in trisyllables. All three groups showed better performance when the non-native pitch contrasts occurred word-finally than word-initially, which could be due to the recency effect such that acoustic information on the final syllable is easier to recall than that on non-final syllables (Demany & Semal, 2008).

Language-specific perceptual patterns regarding non-native pitch contrasts were observed as well. Dutch listeners showed relatively worse performance in discriminating non-native pitch contrasts located in mid-low register (T4-T5, T4-T6 and T5-T6), compared with their good perception of other contrasts. They seemed unable to discriminate T5-T6, with only near 14% accuracy. Japanese listeners had good perception in all the pitch contrasts except for T5-T6. The perceptual patterns of Dutch and Japanese listeners were consistent with their performance in perceiving the non-native pitch contrasts in isolation. Unlike Dutch and Japanese listeners, Mandarin listeners performed equally well in perceiving all the non-native pitch contrasts, which was inconsistent with their perception of the pitch contrasts in isolation. They had comparatively difficulty in perceiving non-native pitch level contrast T1-T3 and T3-T6 in isolation. However, their equally good performance of non-native pitch level and pitch contour contrasts in trisyllables seemed suggest that Mandarin listeners might need contextual tones to discriminate non-native pitch level contrasts.

2.3.3 Experiment 3: Perception of non-native positional tones

2.3.3.1 Performance across language groups

In order to investigate whether the three groups differ in perceiving positional contrasts, a GLMM analysis was conducted, with fixed factors Language (3 levels), Contrast (5 levels: T1, T3, T4, T5 and T6), Positional Contrast (3 levels: initial vs. medial, initial vs. final and medial vs. final) and Contextual Tone (3 levels: T1, T3 and T6). Intercepts for participants and items were added as random effects. Table 2.6 shows that Language (F (2, 3,370) = 14.052, $p<0.001$), Tone (F (4, 3,370) = 6.828, $p < 0.001$) and Contextual Tone (F (2, 3,370) = 16.833, $p < 0.001$) were significant, while Positional Contrast had no significant main effect (F (2, 3,375) = 0.032, $p = 0.968$). An interaction between Tone and Contextual Tone (F (5, 3,370) = 3.922, $p = 0.002$) was found to be significant. There was no significant interaction between Language and Tone (F (8, 3,367) =0.573, $p = 0.801$) or Language and Contextual tone (F (4, 3,371) = 1.312, $p = 0.263$). No three-way interaction was found among Language, Contextual tone and Tone (F (22, 3,348) =0.455, $p = 0.986$).

Table 2.6. *Parameters for fixed effects and random effects in Experiment 3.*

Fixed effects		F	df	p
Language		14.052	2	<0.001
Tone		6.828	4	<0.001
Positional contrast		0.032	2	0.968
Contextual tone		16.833	2	<0.001
Tone & Contextual tone		3.922	5	0.002
Language * Tone		0.573	8	0.801
Language * Contextual tone		1.312	4	0.263
Language * Contextual tone * Tone		0.455	22	0.986
Random effects	Est.	Std. Error	z	p
1 | Participant	1.193	0.328	3.637	<0.001
1 | Item	0.026	0.076	0.347	0.729

A post hoc pair-wise analysis (with Bonferroni adjustment) of Language showed that although all the three groups reached over 95% accuracy, a graded order of overall performance was still observed: Mandarin listeners (99.6% accuracy) outperformed Japanese listeners (98.9% accuracy) (F (2, 3,370) = 7.354, $p = 0.002$) and Dutch listeners (96.6%) (F (2, 3,370) = 4.650, $p = 0.030$). Moreover, Japanese listeners outperformed Dutch listeners (F (2, 3,370) =5.192, $p = 0.018$).

An interaction between Contextual tone and Tone suggests that the perception of non-native positional tones in general depended on the contextual tone. By a Bonferroni-corrected post-hoc pair-wise analysis, when the contextual tone was T1 and T3, the performance of the perception of all the tones achieved accuracy above 95%. When the contextual tone was T6, the overall performance of perceiving T1, T3 and T4 in different positions was still above 95% accuracy. T5 was the exception in that the overall performance of perceiving T5 when the contextual tone was T6 dropped to 79.3% accuracy, significantly worse than when the contextual tone was T1 (F (2, 3,370) = 13. 185, $p < 0.001$) and T3 (F (2, 3,370) = 15. 767, $p < 0.001$).

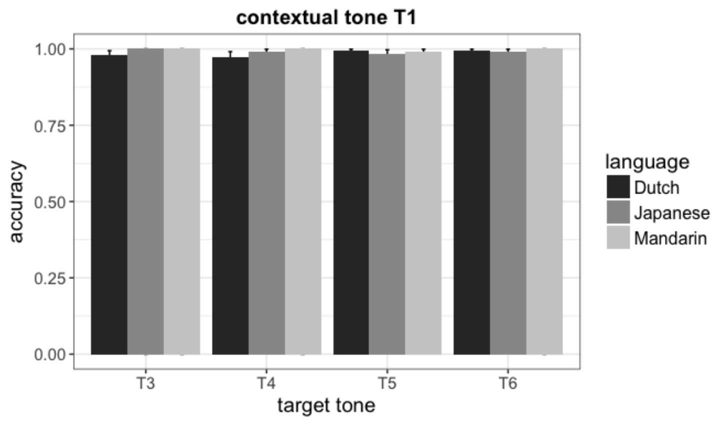


Figure 2.6. *Performance of each group in perception of non-native tones with contextual tone T1 (Experiment 3).*

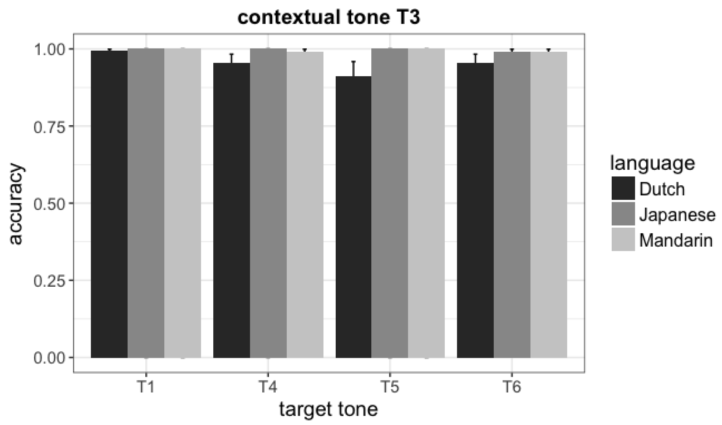


Figure 2.7. *Performance of each group in perception of non-native tones with contextual tone T3 (Experiment 3)*

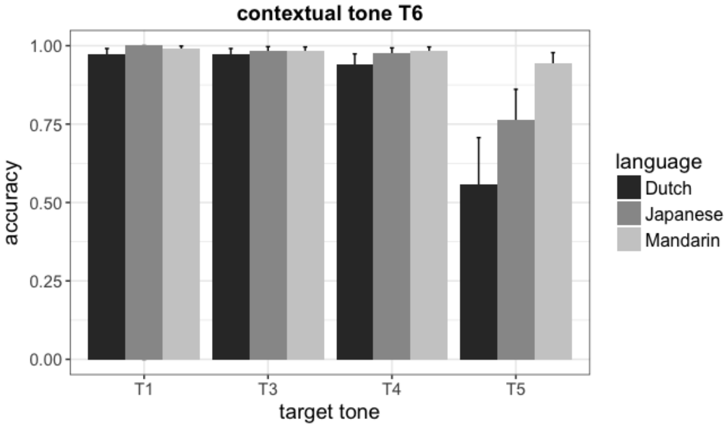


Figure 2.8. *Performance of each group in perception of non-native tones with contextual tone T6 (Experiment 3).*

Figure 2.6, 2.7 and 2.8 show the performance of each language group in discriminating non-native tones with contextual tone T1, T3 and T6, respectively. Although no significant interaction was observed among Language, Tone and Contextual Tone, driven by the hypothesis, GLMM analysis (with Tone and Contextual tone as fixed factors, and participant and item as random effects) was conducted for each language group for exploratory purposes. For Dutch listeners, Tone (F (4, 1,068) = 3.166, $p = 0.013$), Contextual Tone (F (2, 1,068) = 9.357, $p < 0.001$) and an interaction between Tone and Contextual Tone (F (5, 1,068) = 3.208, $p = 0.07$) were significant. When the contextual tone was T6, their performance of discriminating T5 in contrastive positions, was significantly worse (55.9% correct) than their performance when the contextual tone was T1 (96.2% correct) (F (2, 1,068) = 9.795, $p < 0.001$) and T6 (97.2% correct) (F (2, 1,068) = 7.393, $p < 0.001$).

For Japanese listeners, Tone (F (4, 1,176) = 5.329, $p < 0.001$) and Contextual Tone (F (2, 1,176) = 8.534, $p < 0.001$) and an interaction between Tone and Contextual Tone (F (5, 1,176) = 3.062, $p = 0.03$) were found significant. They were found relatively worse in perception of T5 in contrastive positions when the contextual tone was T6 (with 76.2% accuracy), compared with when the contextual tone was T1 (100% accuracy) (F (2, 1,176) = 6.063, $p = 0.003$) and T3 (98.5% accuracy) (F (2, 1,176) = 6.180, $p = 0.002$).

For Mandarin listeners, none of the fixed factors were found significant (Tone: F (4, 1,176) = 0.151, $p = 0.96$; Contextual Tone: F (2, 1,176) = 0.776, $p = 0.46$; interaction between Tone and Contextual Tone: (F (5, 1,176) = 0.058, $p = 0.95$). They performed equally well on discriminating different tones that were contrastive in position, regardless of contextual tone.

2.3.3.2 Summary

To conclude, all three language groups showed good performance in discriminating non-native tones when they occurred in contrastive positions. Despite their overall good performance, a gradient fashion was still observed: Mandarin listeners >> Japanese listeners >> Dutch listeners. The relatively poor performance of Japanese and Dutch listeners was presumably caused by T5 interacting with contextual tone T6. More specifically, for Japanese and Dutch listeners, when the contextual tone was T6, perceiving T5 contrastively in different positions was significantly more difficult than discriminating it when the contextual tone was either T1 or T3. Note that T5 and T6 share similarities in their onset but differ towards the ending point; these tones were shown to be crowded psycho-acoustically for non-canonical tone listeners in previous studies (Qin & Mok, 2013; Wong, 2019; Francis et al., 2008). Presumably due to a lack of lexical tones in their native language, Dutch and Japanese listeners had difficulty in discriminating T5 and T6 across the board, whereas the presence of native lexical tones may benefit Mandarin listeners in discriminating T5 and T6 regardless of the position. Therefore, all three language groups were found to have overall good performance in discriminating non-native positional tones. The confusion of discriminating T5 in contrastive position with contextual tone T6, observed in Dutch and Japanese listeners, was not due to their inability to discriminate positional tones but to their general perceptual difficulty in discriminating T5 from the contextual tone T6. Note that considering the interaction of among contextual tone, tone and language was not significant, the current results with regard to the confusion of T5-T6 in Dutch and Japanese listeners shoud be taken with caution.

2.4 General discussion

The present chapter investigated perception of non-native WPCs via three experiments, namely, perception of non-native pitch contrasts in monosyllables (in isolation) (Experiment 1), perception of non-native pitch contrasts in trisyllables (in different positions) (Experiment 2) and perception of non-native positional tones in trisyllables (Experiment 3). It attempted to tap into the role native word prosody plays in processing non-native WPCs at the acoustic level by addressing 1) whether the three language groups overall make differences in perception of non-native WPCs, 2) whether the three groups display language-specific patterns, and 3) whether the position influences the perception of non-native pitch contrasts in trisyllables. The current section will first discuss the perception of non-native pitch contrasts in isolation (in monosyllables) and in different positions (in trisyllables) and then discuss the perception of non-native positional tones, from the perspective of comparisons both across groups and within groups.

2.4.1 Perception of non-native pitch contrasts in monosyllables

To start with, a hierarchical order of overall performance was predicted: Mandarin listeners >> Japanese listeners, Dutch listeners in perception of non-native pitch contrasts in monosyllables (Experiment 1). However, the findings in Experiment 1 were different from the prediction. It was found that the perceptual differences among the three groups depended on non-native pitch contrasts. The three language groups performed equally well in perception of non-native pitch contrasts located in high-mid register (tonal pairs with T1 and T3) but they differed in discriminating non-native pitch contrasts in mid-low register (T4-T5, T4-T6, T5-T6). Their equally good performance in discriminating non-native high-mid pitch contrasts could be due to the intrinsic acoustic salience of T1 and T3 in contrast with other tones. T1, the high level tone, is located in the high register area, well-separated from other tones in the acoustic space, which makes it distinctive from other Cantonese tones (Qin & Mok, 2013; Qin, 2014; Mok & Wong, 2010; Wong, 2019). T3, the mid level tone, located in the high-mid register area, is relatively distinct from low tones in the acoustic perceptual space (Francis et al., 2008; Qin & Mok, 2013).

Where the three language groups differed in their overall performance was the non-native pitch contrasts in mid-low register (T4-T5, T4-T6 and T5-T6). It was observed that Mandarin listeners showed an advantage over Dutch listeners and Japanese listeners in discriminating the mid-low tone pairs. The low falling tone T4, low rising tone T5 and low level tone T6 are crowded into a small acoustic space. They are acoustically similar in that they have acoustically close starting points, and they differ slightly in their ending points (Qin &Mok, 2013). The acoustic similarity could result in more difficulty for Japanese and Dutch listeners than Mandarin listeners because Japanese and Dutch have more limited functionality of pitch compared to Mandarin, according to the Functional Pitch Hypothesis (Schaefer & Darcy, 2014) as mentioned before. In contrast, the high functionality of pitch in Mandarin may benefit Mandarin listeners in discriminating non-native pitch contrasts when they are not acoustically salient. The differences observed across the three language groups may also suggest that when non-native tonal contrasts are not acoustically salient, Mandarin listeners may approach a discrimination task more phonologically than acoustically in that they may perceive the non-native tones based on their similarity to their native tonal categories and hence they may not need to focus on memorizing detailed acoustic information when making the same-or-different decision.

It is notable that when discriminating T4-T6, Japanese listeners were as good as Mandarin listeners, and they both outperformed Dutch listeners. This could be due to the sensitivity to a pitch fall HL in Japanese listeners. Japanese uses pitch accent (a pitch fall from H to L tone in two successive moras) to differentiate lexical items (Poser, 1984; Kubozono, 2008). Previous studies have shown that Japanese listeners are sensitive to the location of a pitch accent and discriminate contrastive accentual patterns in Japanese (Vance, 1995; Yoneyama, 2002). Sensitivity to the

HL pitch fall in their native language may aid their discrimination of a non-native falling tone from a level tone in the mid-low register.

The current findings that Mandarin listeners outperformed Japanese and Dutch listeners in perception of non-native pitch contrasts, not generally, but limited to the mid-low register, seem to be in discrepancy with previous cross-linguistic studies that have reported an overall advantage for tone language listeners (i.e., Mandarin) over non-tone/non-canonical tone language listeners (Qin & Mok, 2013; Schaefer & Darcy, 2014). For instance, Qin & Mok (2013) applied an AX discrimination task to investigate cross-linguistic perception of Cantonese tonal pairs by tone (Mandarin) and non-tone (English and French) language listeners. They examined the perception of all the possible tonal pairs consisting of six Cantonese tones and reported an overall perceptual advantage of Mandarin listeners over English and French listeners. However, in their study, they seemed to neglect an observed interaction between the language group and tonal pairs, which suggested the perceptual differences across the groups depended on tonal contrasts. In Schaefer & Darcy (2014), they conducted AXB identification task to examine perception of Thai tones by listeners of Mandarin, Japanese, English and Seoul Korean listeners which differ from high to low in the functionality of pitch. The listeners examined in their study were similar to those in prosodic typology in the current study. Schaefer & Darcy (2014) found a clear overall hierarchical order of Mandarin >> Japanese >> English >> Seoul Korean listeners. However, the current study was not able to observe an overall perceptual hierarchy. This could be due to the use of different tasks. The current study used an AX discrimination task where listeners have to judge whether the two items are the same or different, which encourages an acoustic mode of processing (Logan & Pruitt, 1995; Strange & Shafer, 2008). Schaefer & Darcy (2014) applied an AXB discrimination task which requires listeners to determine whether item X belongs to A or B. Listeners have to retain more acoustic information of the stimuli, which may, to some extent, tap into phonological representations (Dupoux et al., 1997; Strange & Shafer, 2008). Compared with an AX discrimination task, an AXB discrimination task may encourage listeners to approach the task more phonologically than acoustically and thus lead to a more robust perceptual hierarchy based on the high to low functionality of lexical pitch in the native language.

Apart from the differences among groups, language-specific patterns were also observed in perception of non-native pitch contrasts in monosyllables. Different from the prediction that Dutch listeners will attend more to non-native pitch level contrasts than pitch contour contrasts, it was found that Dutch listeners' perceptual weighting was influenced by where the contrasts were located in acoustic space. They were less sensitive to the pitch contrasts in the mid-low register (T4-T5, T4-T6 and T5-T6), compared with their good discrimination of the other contrasts. As mentioned before, T4, T5 and T6 share similar starting points and only differ slightly at the ending point. Presumably due to a lack of contrastive tones in the native word prosody, when the non-native pitch contrasts share similarity in acoustic space, the acoustic information may not be distinctive enough for them to rely on.

Note that although Dutch listeners were less accurate when discriminating non-native pitch contrasts in mid-low register compared with their good performance in high-low register, they were still able to discriminate T4-T5 and T4-T6 (accuracy above chance level at 50%). The investigations that Dutch listeners, by and large, were sensitive to non-native pitch contrasts are in line with previous studies (Leather, 1987, 1990; Liu et al., 2017; Chen et al., 2015). The ability to perceive pitch differences in non-native tones can presumably be due to the influence of the rich inventory of nuclear tones in Dutch intonation and/or the use of lexical stress at the word prosody level. Dutch has 24 types of nuclear tones in intonation (Gussenhoven, 2004) and such diversity of pitch movements in intonation may benefit Dutch listeners in perceiving non-native pitch patterns (Braun & Johnson, 2011). In addition, Dutch uses lexical stress which is instantiated by acoustic correlates such as pitch. The employment of lexical stress with pitch as one of the acoustic cues may to some extent attune Dutch listeners' tone perception acoustically. However, T5-T6 was the exceptional contrast that they seemed unable to discriminate, with their perceptual performance much below chance level (less than 15% accuracy). The perceptual asymmetry of T5-T6 could be due to psychoacoustic similarity to non-tone language listeners. Qin & Mok (2013) used a Multidimensional Scaling (MDS) task to investigate how tone language and non-tone language listeners rate the dissimilarity of six Cantonese tones (T1, T2, T3, T4, T5 and T6). They found that non-tone language (English and French) listeners rated T5 and T6 as closer at the ending point than T4-T6 and T4-T5 in perceptual space while Mandarin listeners rated the three tonal contrasts to be quite dissimilar.

Japanese listeners were predicted to be more sensitive to non-native pitch level contrasts than contour contrasts. Different from the prediction, they performed equally well in perceiving non-native pitch contrasts in monosyllables except for one problematic contrast T5-T6 with the performance below chance level. Their overall good discrimination of non-native pitch contrasts can be beneficial from the use of lexical pitch accent in the native language, which was also observed in previous studies such as in So & Best (2010) (Japanese listeners' perceiving Mandarin tones) and in Schaefer & Darcy (2014) (Japanese listeners' perceiving Thai tones). However, the difficulty of discriminating T5-T6 seems to suggest that the restricted tonal system in the native language may not be sufficient for Japanese listeners to rely on it to discriminate the two tones T5 and T6 when they are similar in acoustic perceptual space. Another account of the overall good performance of Japanese listeners could be due to facilitation from the various pitch patterns in Japanese words. In Japanese one mora is usually around 100-150 ms in normal speech (Hoequist, 1983). The monosyllables used in the study had duration of 400 ms, and thus Japanese listeners might have perceived the monosyllables as words with 3-mora instead of one mora. 3-mora words in Japanese display diverse pitch patterns, i.e., rising, falling, level high and rising+falling contour (Kubozono, 2008), which may help Japanese listeners to perceive non-native contour and level tones. However, recall that the T5-T6 (low rising vs. low level tone) contrast was problematic to Japanese listeners. If the diverse pitch patterns in 3-mora Japanese words could have benefited Japanese listeners to be sensitive to pitch contour and pitch level contrasts in non-native tone perception, they should have also shown good performance in discriminating T5-T6. This seems to suggest that the assumption that Japanese listeners may perceive the non-native tones carried by monosyllables as native pitch patterns carried by three moras may not be convincing enough.

Unlike Japanese and Dutch listeners, Mandarin listeners were more sensitive to non-native pitch contour contrast and pitch level vs. pitch contour contrasts than to pitch level contrasts (T1-T3 and T3-T6) in monosyllables, as predicted. The differences of perceptual weighting they assigned to non-native pitch contrasts were consistent with previous studies as well (Gandour 1983; Francis et al., 2008; Qin & Mok, 2013). Mandarin uses contrastive contour tones lexically, which influences them to be more sensitive to contour changes in non-native tone perception (Xu et al., 2006; Francis et al., 2008). The comparative vulnerability in discriminating level contrasts can be due to the fact that Mandarin uses one level tone (high level tone) at word prosody and does not use contrastive level tones to differentiate lexical items.

2.4.2 Perception of non-native pitch contrasts in trisyllables

Experiment 2 investigated the perception of non-native contrasts that occurred in word-initial, -medial and -final position. The observations in perception of non-native pitch contrasts in trisyllables are in line with those in monosyllables. The three language groups did not differ in discriminating non-native pitch contrasts located in high-mid register, but differed in discriminating non-native pitch contrasts located in mid-low register. More specifically, similar to the findings in perception of non-native pitch contrasts in monosyllables, Mandarin listeners showed better performance than Japanese and Dutch listeners in discriminating non-native low tones (T4-T5 and T5-T6). Japanese listeners performed as well as Mandarin listeners in discriminating T4-T6, while both outperformed Dutch listeners.

The findings were different from the predictions in 1(a) and 1(b). Firstly, 1(a), based on the Functional Pitch Hypothesis (Schaefer & Darcy, 2014), predicted that the high functionality of pitch in Mandarin would facilitate Mandarin listeners to have an overall better performance than Japanese and Dutch (relatively low functionality of pitch) listeners in non-native tone perception. However, it was found that the perceptual advantage of Mandarin listeners over Japanese and Dutch listeners was only observed when the non-native tones are not acoustically salient (for detailed discussion, see 2.4.1). Secondly, prediction 1(b), based on the use of lexical pitch and positional marking in the native language (Feature Hypothesis), predicted that Japanese listeners would outperform Mandarin and Dutch listeners overall, which was rejected by the findings. Instead, the findings seem to support that the relative linguistic experience with tones (i.e., the use of lexical pitch), rather than the use of positional marking, in the native language was a dominant factor that influenced perception of non-native pitch contrasts that occurred in different positions.

Language-specific patterns in the perception of non-native pitch contrasts in trisyllables were observed. The findings were, by and large, consistent with those in perception of non-native pitch contrasts in monosyllables. Different from the prediction that Dutch listeners would attend more to non-native pitch level contrasts than to contour contrasts, Dutch listeners were influenced by where the contrasts were located in acoustic space. They displayed better performance in discriminating non-native pitch contrasts in high-mid than in mid-low acoustic space (T4-T5, T4-T6 and T5-T6). Moreover, they seemed especially unable to discriminate T5-T6 with their performance below chance level (only around 14% accuracy). Japanese listeners showed good performance in discriminating all the non-native pitch contrasts except for T5-T6 which was below chance level. The different perceptual patterns of the two language groups can be explained by the influence of the native prosody and the acoustic features of the non-native tones (see §2.4.1 for detailed discussion). For Japanese listeners, another account of their overall good performance could be attributed to the benefit from the various pitch patterns in Japanese words. In Japanese one mora is usually around 100-150 ms in natural speech (Hoequist, 1983). The trisyllables used in the study had duration of 950 ms, which might have been perceived as words with longer moras (e.g., six moras) instead of three moras by Japanese listeners. Multi-mora words in Japanese manifest various pitch patterns, i.e., rising, falling, level high and rising+falling contour (Kubozono, 2008), which may in general help Japanese listeners to perceive non-native contour and level tones. However, Japanese listeners had difficulty in discriminating T5-T6 contrast in trisyllables (performance below chance level). If the various pitch patterns in Japanese words with six or longer moras could have faciliated Japanese listeners to be sensitive to pitch contour and pitch level contrasts in non-native tone perception, they should have also had good performance in discriminating T5-T6 in trisyllables. This seems to suggest that the assumption that Japanese listeners may perceive the non-native tones carried by trisyllables as native pitch patterns carried by six or longer moras may not be convincing enough.

Unlike Dutch and Japanese listeners, Mandarin listeners showed good performance in discriminating all the non-native pitch contrasts, which was different from the prediction that they would be more sensitive to non-native pitch contour contrasts than to level contrasts. Recall that Mandarin listeners were more sensitive to non-native pitch contour contrasts than pitch level contrasts in monosyllables. However, interestingly, when the non-native pitch level contrasts occurred in trisyllables, Mandarin listeners regained the sensitivity to both non-native pitch contour and level contrasts. Presumably due to a lack of contrastive level tones in Mandarin, Mandarin listeners may need contextual level tones to build tonal references when perceiving the non-native pitch level contrasts. The level tonal references in trisyllables thus may have assisted them in performing just as well as they did in discriminating non-native pitch contour contrasts.

Furthermore, position did not interact with the perception of non-native pitch contrasts. None of the three groups showed preference for a specific position when perceiving a specific non-native pitch contrast. Instead, all three groups showed overall better performance when the non-native pitch contrasts occurred word-finally than word-initially and/word-medially, which was in conflict with the predictions that Mandarin and Japanese would show no positional preferences while Dutch listeners would prefer word-medial to word-initial and word-final position. The observed preference for word-final position over word-initial and/or medial position in all three groups could be due to the recency effect – that listeners tend to recall acoustic information in final position better than in initial position since acoustic information is retained in memory longer in the offset than in the onset (Demany & Semal, 2008). Among the performances in each position across the three groups, it was found that Mandarin listeners showed better performance than Japanese listeners and Dutch listeners word-initially and word-medially. Mandarin listeners and Japanese listeners outperformed Dutch listeners in discriminating non-native contrasts on word-final position, while the former two groups performed equally well. The findings were mainly consistent with the prediction in 2a that Mandarin listeners will overall outperformed Japanese and Dutch listeners in all the three positions. The advantage that Mandarin listeners displayed over Dutch (in all the three positions) and Japanese listeners (word-initially and word-medially) suggests that the higher functionality of pitch facilitates perception of non-native tonal contrasts when the contrasts occurred in different positions. It is worthy remarking that when the contrasts occurred word-finally, Japanese listeners performed as well as Mandarin listeners and they both outperformed Dutch listeners. It is documented that listeners tend to recall acoustic information better on the offset than the onset of a sound string due to the memory trace decay (Demany & Semal, 2008) and therefore the non-native pitch contrasts on the word-final position may leave more acoustic information for listeners than those on the word-initial and –medial position do. In the word-final position where more acoustic traces are retained, Japanese listeners were able to discriminate the non-native pitch contrasts as well as canonical tone language (Mandarin) listeners, and both showed advantage over non-tone language (Dutch) listeners. This suggests that the lexical pitch used in the native language may still augment the perception of non-native tonal contrasts when less attention is required (i.e., contrasts on word-final position compared with those on word-initial and word-medial position).

2.4.3 Perception of non-native positional tones in trisyllables

All three groups achieved overall good perception, each group with an overall accuracy over 90% in perceiving non-native tones in contrastive positions. However, a hierarchical perceptual order was still observed: Mandarin listeners >> Japanese listeners >> Dutch listeners, which supported the prediction in 4c but not 4a and 4b. To begin with, based on the use of both lexical pitch and positional marking in the native language, 4a predicted that Japanese listeners would have an overall advantage over Dutch and Mandarin listeners, which was rejected by the finding. Based on the use of positional marking for lexical contrast in the native language, 4b predicted that both Japanese and Dutch listeners would outperform Mandarin listeners, which contradicted the current finding as well. Based on the use of lexical pitch and its relative degree of the use in the native language, 4c predicted that Mandarin listeners would outperform Japanese and Dutch listeners overall, which was upheld by the finding. On the one hand, although Mandarin does not use positional cue lexically, it has higher functionality of pitch than Japanese and Dutch do, according to the Functional Pitch Hypothesis (Schaefer & Darcy, 2014) derived from the Feature Hypothesis (McAllister et al., 2002). The higher functionality of pitch successfully predicted a perceptual advantage over Japanese and Dutch listeners. On the other hand, Japanese and Dutch share commonality in the use of positional cues for lexical contrasts, differing, however, in the degree of pitch functionality. Japanese has higher functionality than Dutch, which could lead to the better performance of Japanese listeners compared to Dutch listeners.

The observations in Experiment 3 seem to suggest that the use of lexical pitch and/or to what degree lexical pitch is used in the native language, rather than the positional marking, is a prominent factor that matters in perception of non-native positional tones. One potentially confounding factor in the study is that all the listeners who participated in the study are college students who have been exposed to English, a language that has word stress. The exposure to a stress language (English) might be helpful for Mandarin listeners who do not have positional marking in the native language in perceiving a non-native positional contrast. However, L2 studies on perception of contrastive stress position by Mandarin listeners have shown that Mandarin L2 learners primarily relied on pitch to perceive English stress, which may relieve this concern to some extent. For instance, Wang (2005) manipulated duration, intensity and pitch, with vowel quality kept constant, in minimal pairs of disyllabic nonwords in English. She found that pitch was the decisive cue for Mandarin listeners to correctly judge whether the stress was placed on the first or second syllable, which implied that Mandarin L2 learners of English transfer their reliance on pitch to perceive stress position (Wang, 2005). Also, L2 studies on perception of lexical pitch accent positions show that Mandarin learners of Japanese are able to detect the position of Japanese pitch accent (Ayusawa et al., 1998). These findings in L2 studies that investigate Mandarin learners perceive a positional marking cue (i.e., stress and pitch accent) may shed some light on a transfer of sensitivity to pitch to perceive non-native positional tones.

In addition to the overall differences among groups, based on analyses of exploratory purposes, a specific tone seemed to be problematic that brought perceptual difficulty to Japanese and Dutch listeners, which is T5 when the contextual tone was low level tone T6. Japanese listeners and Dutch listeners might have trouble in perceiving T5 in contrastive positions when the contextual tone was T6. This can be due to the fact that T5 and T6 share similar onset and are crowded into a small acoustic space. In perceiving non-native pitch contrasts in monosyllables and trisyllables, T5-T6 was least accurate in Japanese listeners and Dutch listeners. This problem in discriminating this T5-T6 contrast influenced the perception of T5 when the contextual tone was T6, which suggests the difficulty originates from confusion between the tones rather than position alone. Note that considering the insignificant interaction among contextual tone, tone and language group, the current results regarding the problematic T5-T6 observed in Dutch and Japanese listeners should be taken with caution.

2.5 Conclusion

The present study served to investigate to what extent native language influences acoustic perception of non-native WPCs, pitch contrasts, and position, for Mandarin, Japanese and Dutch listeners in three AX discrimination tasks. It addressed perception of non-native Cantonese pitch contrasts in monosyllables and in trisyllables, and perception of non-native tones in contrastive positions. It observed that the three groups performed equally well in perceiving non-native pitch contrasts that were acoustically salient, while differing in their perception of non-salient, non-native pitch contrasts (contrasts, to some extent, that share similarity in acoustic space). Mandarin listeners showed an advantage over Japanese and Dutch listeners in discriminating non-salient, non-native pitch contrasts both in monosyllables and in trisyllables. The findings suggest that both the acoustic salience and the use of lexical pitch, and/or to what degree the lexical pitch is used in the native language, can influence the perception of non-native pitch contrasts. More specifically, higher functionality of pitch in the native language can be an aid in non-native tone perception when the non-native pitch contrasts are not acoustically salient.

In the perception of non-native positional contrasts, a perceptual gradient order was observed: Mandarin listeners (who do not use positional cue but use pitch to a high degree (Schaefer & Darcy, 2014)) >> Japanese listeners (who use the position of pitch accent and use pitch to an intermediate degree) >> Dutch listeners (who use position of word stress and use pitch to a low degree). The findings seem to indicate that the use of lexical pitch, and/or to what degree lexical pitch is used in the native language, rather than positional marking, is a robust factor that matters in perception of non-native positional tones. Moreover, perceiving T5 contrastive in positions when the contextual tone was T6 seemed to be problematic to Dutch and Japanese listeners but not to Mandarin listeners, based on analysis of exploratory purposes.

Language-specific patterns were also observed in perception of non-native contrasts in the present study. Dutch listeners were influenced by whether the non-native pitch contrasts were acoustically salient (T4-T5, T4-T6 and T5-T6) or not, both in monosyllables and trisyllables. Japanese listeners did not show perceptual preference except that they were relatively vulnerable in one specific non-salient non-native contrast (T5-T6). Mandarin listeners were more attentive to non-native pitch contour contrasts than pitch level contrasts in monosyllables, which was in line with earlier studies (Gandour, 1983; Francis et al., 2008; Qin & Mok, 2013). However, an interesting finding in the current study was that, when discriminating non-native pitch contrasts in different positions, Mandarin listeners attended equally well to both non-native pitch contour and pitch level contrasts in trisyllables, unlike their perceptual weighting in perception of non-native pitch contrasts in monosyllables. This suggests that Mandarin listeners may need context tones to establish a tonal reference to discriminate non-native level contrasts.

To summarize, the current chapter investigated the perception of non-native WPCs at the acoustic level. It observed perceptual differences across language groups and language-specific patterns as well. It revealed that native prosody, especially the use of lexical pitch, together with the acoustic salience of tones influenced perception of non-native WPCs. Notably, even though Dutch and Japanese listeners showed worse performance than Mandarin listeners in discriminating non-salient non-native pitch contrasts, they were still able to discriminate them (above chance level) (except for one specific tonal pair T5-T6), which entails that they may be able to rely on acoustic information. The following chapter, Chapter 3, will serve to investigate the perception of non-native WPCs at the phonological level when acoustic strategies are inaccessible (Dupoux et al., 2001, 2008).Chapter 4 will examine the processing of non-native WPCs at the lexical level when the non-native WPCs are mapped with meaning in word learning.

Chapter 3
Processing of non-native WPCs at the phonological level

3.1 Introduction

The previous chapter investigated cross-linguistic perception of non-native WPCs at the acoustic level and found that Mandarin, Japanese, and Dutch listeners were all able to discriminate non-native pitch contrasts in different positions, achieving ceiling performance when discriminating non-native positional tones (non-native tones that occurred at contrastive positions). Chapter 3 focuses on cross-linguistic perception of non-native WPCs, more specifically non-native pitch contrasts in different positions, at the phonological level.

Recall that the “phonological level” refers to a relatively abstract processing level at which phonological representation is highlighted. This level can be investigated via a short-term memory task, the sequence recall task (Dupoux et al., 2001, 2008). In a sequence recall task, listeners are required to learn two items (a minimal pair contrastive in one phonological dimension) and recall sequences of items they have heard, which taps into a high memory load. The items are produced by different speakers so as to prevent listeners from relying on a specific voice. Hence, the sequence recall task aims to eliminate listeners’ reliance on fine-grained acoustic details and taps into phonological representation. Applying the sequence recall task, Dupoux et al. (2001) found that naive French listeners who do not use lexical stress or any other WPCs (Rossi, 1980; Di Cristo, 1988; Vaissière, 1991) showed “stress deafness” since they were unable to perceive Spanish contrastive stress, while they had no perceptual problem in an AX stress discrimination task (Dupoux et al., 1997). The anecdotal notion of “stress deafness” in Dupoux et al. (2001, 2008) suggests that the capability of perceiving a word’s prosodic cue acoustically does not guarantee success processing the word’s prosodic cue at an abstract phonological level. This motivates us to ask: will listeners who managed to acoustically discriminate non-native WPCs (pitch contrasts in different positions) maintain this capability at the phonological level? How will native language come into play in the processing of non-native WPCs at the phonological level?

Previous studies have shown that listeners across various native languages show acoustic sensitivity to non-native pitch contrasts on monosyllables (the previous chapter; Lee et al., 1996; Wayland & Guion, 2004; Hallé et al., 2004; Francis et al., 2008; So & Best, 2010; Schaefer &Darcy, 2014; Burnham et al., 2015; Liu et al., 2017 among others). More specifically, pitch differences at the acoustic level may be generally available to listeners in non-native tone perception, not only for tone language listeners, but also for non-tone language listeners. For instance, Mandarin listeners who use lexical tones were found adept at discriminating and/or identifying non-native tones (the previous chapter; Lee et al., 1996; Wayland & Guion, 2004; Schaefer &Darcy, 2014; Burnham et al., 2015, among others). French listeners who do not use any WPC were able to discriminate tonal differences in Mandarin (Hallé et al., 2004). Dutch listeners who do not use lexical tones but do use lexical stress were reported to perform well when discriminating and identifying tonal differences in Mandarin (Liu et al., 2017). Chapter 2 further reported that tone language (Mandarin and Japanese) and non-tone language (Dutch) listeners were overall able to discriminate non-native pitch contrasts in different positions. However, a capability of perceiving non-native pitch contrasts in different positions acoustically does not necessarily entail successful encoding at the phonological level. It remains unclear whether acoustic sensitivity to non-native pitch contrasts in different positions can be retained in a task with high memory load (in particular in the sequence recall task, which taps into phonological representation). This is an empirical gap the current study aims to fill.

Compared to the bulk of studies on cross-linguistic perception of non-native pitch contrasts at the acoustic level (Lee et al, 1996; Burnham et al., 1996; Gottfried & Suiter 1997; Wang et al., 1999; Wayland & Guion, 2004; Wu & Lin, 2008; So & Best, 2010, 2014; Schaefer & Darcy, 2014; Burnham et al., 2015; Reid et al., 2015, among others), scarce attention has been paid to the processing of non-native pitch contrasts at the phonological level. No cross-linguistic studies seem to have investigated the processing of non-native pitch contrasts at the phonological level on a par with Dupoux et al. (2001) on the perception of Spanish stress by naive French listeners. One recent study on the perception of non-native pitch contrasts at the phonological level focused on L2 acquisition. Zou (2017) applied the sequence recall task to examine whether Dutch L2 learners of Mandarin could acquire and encode Mandarin tones phonologically. However, an investigation into how and to what extent naive listeners are influenced by the native language in the perception of non-native pitch contrasts at the phonological level is still lacking.

Furthermore, the current study also investigates the influence of positional marking as another WPC on non-native tone perception at the phonological level. Only very few studies have focused on the perception of non-native position marking at the phonological level by naive listeners (Dupoux et al., 2001; Peperkamp & Dupoux, 2002; Hu, 2015). For instance, Hu (2015) conducted a sequence recall task (Dupoux et al., 2001) to examine the perception of positional contrasts of Japanese pitch accent by Japanese natives and naive Dutch listeners. She found that Dutch listeners were overall able to perceive accentual patterns contrastive in pitch accent position in disyllabic nonwords in the sequence recall task, suggesting a positive transfer from the native lexical positional marking in processing non-native positional cues phonologically. This leads us to ask, will positional sensitivity benefit naive listeners with lexical use of positional marking in their native languages (i.e., Japanese and Dutch listeners) in the perception of non-native pitch contrasts in different positions at the phonological level? Will listeners without the use of positional cues in their native language (i.e., Mandarin listeners) encounter difficulty perceiving non-native pitch contrasts in different positions at the phonological level?

Prior studies investigating whether listeners whose native language does not use positional marking can encode positional contrasts (i.e., lexical stress or pitch accent) at the phonological level are mainly centered on L2 acquisition (Lin et al., 2014; Utsugi et al., 2010; Qin et al., 2017). For instance, Lin et al. (2014) investigated whether Mandarin and Seoul Korean advanced L2 learners of English were able to phonologically perceive minimal pairs of English disyllabic nonwords contrastive in stress position (no vowel quality differences) using a sequence recall task adapted from Dupoux et al. (2001, 2008). They found that Mandarin advanced L2 listeners outperformed Seoul Korean advanced L2 learners and the former even showed as good performance as English native listeners in processing stress in minimal pairs phonologically. They suggested that the success of Mandarin advanced L2 listeners might be due to a positive transfer from the native language in which they use suprasegmental information for lexical contrasts to perceive another non-native suprasegmental cue.[15] Qin et al. (2017) investigated the perceptual cues Mandarin and Taiwanese Mandarin advanced L2 learners of English rely on to perceive English stress (without vowel deduction) phonologically, compared to English natives. They applied a sequence recall task (Dupoux et al., 2008) and reported that Mandarin advanced L2 learners attended to pitch more than English natives did when perceiving contrastive stress phonologically, indicating an influence of pitch sensitivity from native word prosody. These L2 studies, despite their limited number, suggest the pivotal role lexical pitch plays for Mandarin L2 learners of English to perceive positional contrasts at the phonological level. More specifically, Mandarin L2 learners rely on pitch to process positional cues (e.g., English stress) phonologically. It is unknown whether naive listeners whose native language makes use of lexical pitch but not positional marking, i.e., Mandarin listeners, benefit from pitch sensitivity when perceiving non-native pitch contrasts occurring in different positions, or whether they will be hindered by position.

In all, the present study considers non-native pitch contrasts occurring in different positions, focusing at the phonological level from a cross-linguistic perspective. It attempts to disentangle the role native language plays in the phonological processing of non-native WPCs by addressing the following questions:

[15] Lin et al., (2014) also argues that the better performance of Mandarin listeners could be attributable to neutral tones that have stress properties in Mandarin (Wang, 1997; Lu & Wang, 2005; Duanmu, 2007; Cao, 1992; Bao & Lin, 2014). However, the number of syllables carrying neutral tones in Mandarin is very few, so neutral tones are not discussed further in this chapter. See §1.2.2.1 for more details regarding the properties of Mandarin neutral tones.

(1) How will Mandarin, Japanese,[16] and Dutch listeners differ in their perception of non-native WPCs (non-native pitch contrasts occurring in different positions) at the phonological level? More specifically, to what extent will a listener's native WPCs benefit the perception of non-native pitch contrasts that occur in different positions at the phonological level?

(2) What specific perceptual patterns will the three groups of listeners display with respect to non-native pitch contrasts in different positions at the phonological level? More specifically, how differently will they perform when perceiving non-native pitch contrasts in different positions phonologically? Will listeners who share commonality in their native use of WPCs share similar perceptual patterns at the phonological level?

(3) To what extent will position influence each language group in their encoding of non-native pitch contrasts in different positions at the phonological level? Will listeners show preference for specific positions in their phonological processing of non-native pitch contrasts?

The predictions are made based on the same hypotheses used in Chapter 2 to investigate the perception of non-native WPCs at the acoustic level. The Functional Pitch Hypothesis (Schaefer & Darcy, 2014), the Feature Hypothesis (McAllister et al., 2002), and the approach based on perceptual weighting to pitch features in long-term memory (Francis et al., 2008; Lee et al., 1996; Xu et al., 2006) do not refer to a specific processing level or strictly tease apart differences among acoustic perception, perception at the more abstract level (the phonological level), or processing at the lexical level (in word learning). These hypotheses are proposed to account for non-native speech perception and provide "a baseline for L2 tone acquisition" (e.g., Schaefer & Darcy, 2014: 514), which entails that their predictions may carry over to processing involving phonological information or lexical information. Thus, corresponding to the research questions, the predictions for the perception of non-native pitch contrasts occurring in different positions (trisyllables) at the phonological level are by and large identical to those at the acoustic level (in Chapter 2) as follows:

(1a) If both lexical pitch and positional marking can influence the perception of non-native pitch contrasts occurring in different positions, based on the Feature Hypothesis (McAllister et al., 2002) it is predicted that Japanese listeners (who use of position of pitch accent lexically) will outperform Dutch listeners (who use of position of lexical stress) and Mandarin listeners (who use of lexical tones) in perceiving non-native pitch contrasts at the phonological level (grouping positions).

(1b) Alternatively, if lexical pitch is more dominant in influencing the perception of non-native pitch contrasts occurring in different positions, according to the Functional Pitch Hypothesis (Schaefer & Darcy, 2014), it is predicted that a

[16] "Japanese" in this dissertation refers to Tokyo Japanese.

hierarchical performance order will be observed: Mandarin >> Japanese >>/= Dutch listeners.

(2) By grouping the positions where the non-native pitch contrasts occur, it is predicted that each language group will show different perceptual patterns with regard to non-native pitch level and contour contrasts. This is based on the proposal that suggests the long-term stored pitch representations (tones and/or intonation) in the native language determine the relative weight listeners given to specific perceptual features of tones in cross-linguistic tone perception (Francis et al., 2008; Lee et al., 1996; Xu et al., 2006). Mandarin listeners will be more sensitive to non-native pitch contour contrast than to pitch level contrast at the phonological level, since Mandarin has contrastive contour tones but not contrastive level tones at the word prosody.

Dutch does not have lexical tones but enjoys a rich inventory of nuclear tones that consist of H and L tones in the intonation system (Gussenhoven, 2005). Dutch listeners may have stored representations of nuclear tones consisting of H and L tones in long-term memory and thus they are expected to be more sensitive to non-native pitch level contrasts than pitch contour contrasts at the phonological level.

Japanese words have accentual patterns (determined by the presence/absence and location of lexical pitch accent) that consist of H and L tones. Japanese listeners may have stored in long-term memory representations of H and L tones in the accentual patterns. Therefore, they are predicted to be more attentive to non-native pitch level contrasts than to non-native pitch contour contrasts at the phonological level.

(3) Positional cues are employed in word prosody to distinguish word meanings in languages such as Dutch and Japanese. I assume that the feature of position (representation of positional feature) in the native word prosody also plays a role to some extent in the perception of non-native pitch contrasts, not only at the lower linguistic processing level (the acoustic level) but also at the phonological level (as well as the lexical level, to be discussed in Chapter 4). Hence, predictions are made in terms of the influence of position in the native phonology, similar to the predictions for acoustic discrimination. Due to a lack of positional marking in the native language, Mandarin listeners may only use tonal sensitivity to perceive non-native pitch contrasts in different positions instead of relying on a specific position. Thus, it is expected that Mandarin listeners will not show preference for any position.

Dutch uses word stress for lexical contrast. Word stress (primary stress) in Dutch is contained within a three-syllable window generally placed at the right-hand word edge (Kager, 1989; Trommelen & Zonneveld, 1986, 1999).[17] However, it should be noted that the dominant position for word stress is the prefinal, not the

---

[17] The stress assignment in Dutch mentioned here is a generalized pattern. The placement of stress in Dutch is more complex, based on phonological properties of words such as syllable structure and syllable weight (see Trommelen & Zonneveld, 1986, 1999; Kager, 1989 for details).

final position (Kager, 1989). Hence, Dutch listeners are predicted to prefer word-medial position to word-initial and/or word-final position when perceiving non-native pitch contrasts in different positions at the phonological level.

Japanese uses the presence/absence and the location of pitch accent at the word level. *n*-mora words theoretically have *n+1* possible accentual patterns (McCawley, 1968; Akinaga, 1985; Haraguchi, 1999; Uwano, 1999). Studies on statistical analyses of the proportion of Japanese accentual patterns have shown that all accentual patterns exist in 3- to 4-mora words (Sato, 1993; Suzuki, 1995; Kitahara, 2001). Due to the influence of native accentuation, Japanese listeners are predicted to show no preference for any position in perceiving non-native pitch contrasts since unaccented form is dominant in 3- to 4-mora words in Japanese.

3.2 Method

To answer the research questions above, a sequence recall task (Dupoux et al., 2001) will be used. As previously discussed, the sequence recall task provides a robust paradigm for investigating the processing of novel suprasegmental contrasts at the more abstract, phonological level. In the sequence recall task, listeners are required to learn a minimal pair of nonwords A and B (produced by different speakers) that are phonemically contrastive (here, non-native pitch contrasts occurring in different positions) and segmentally contrastive (as baseline), and listen to the two words in various sequences (such as A-B or B-A in a two-word sequence, A-B-B in a three-word sequence). Participants must then recall the sequences they have heard. The difficulty of processing non-native WPCs at the phonological level will be manifested by a comparison between performance perceiving segmental contrasts and that perceiving non-native pitch contrasts occurring in different positions.

3.2.1 Subject

Participants in the current study were the same as in Chapter 2. Two Mandarin participants and two Japanese participants from the study discussed in Chapter 2 withdrew from the present study. Thus, thirty native speakers of Dutch (eight male and twenty-four female), thirty native speakers of Mandarin Chinese (thirteen male and seventeen female), and thirty native speakers of Tokyo Japanese participated in the current study. The average age of the Dutch, Mandarin, and Japanese listeners was 22 years old (SD = 4.1), 25 years old (SD = 3.8), and 25 years old (SD = 5.2), respectively. All of the Mandarin participants speak native northern Chinese dialects, with no prior experience with Cantonese or any other additional tone languages (including southern Chinese dialects). None of the Dutch or Japanese participants had been systematically exposed to any other tonal languages or other tonal dialects. None of the participants had undergone systematic musical training or reported hearing impairment. Dutch and Mandarin participants participated in the present study nearly eight months after the acoustic experiment in Chapter 2. Japanese participants participated in the study nearly one month and a half after the previous acoustic experiment.

3.2.2 Materials and design

In Chapter 2 we used 10 Cantonese tonal contrasts T1-T3, T1-T4, T1-T5, T1-T6, T3-T4, T3-T5, T3-T6, T4-T5, T4-T6, and T5-T6 (see Figure 3.1 for Cantonese tones) to investigate acoustic perception of non-native tonal contrasts in isolation (in monosyllables) and in different positions (in trisyllables).

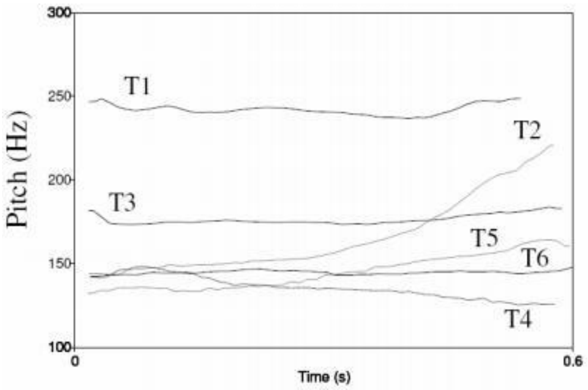


Figure 3.1. *Cantonese tones* (Qin & Mok, 2013)

Chapter 2 reported that all the tonal contrasts paired with T1, contrast T3-T4, and contrast T3-T5 were all "optimally" perceived at the acoustic level, with an overall accuracy reaching ceiling effect. The overall performance when acoustically perceiving contrasts T4-T5, T4-T6, and T3-T6 was "suboptimal" with around 80% accuracy. Contrast T5-T6 was overall not favoured in that overall performance was slightly above chance level (50%) and particularly difficult for Dutch and Japanese listeners, both displaying performances around 15% accuracy. Since the sequence-recall task is more demanding than an acoustic AX discrimination task, the "non-optimal" contrast T5-T6 was not included further in the stimuli. In order to select the most suitable contrasts for the sequence recall task, a pilot study using the "optimal" contrasts (contrasts with T1, T3-T4, and T3-T5) and "suboptimal" contrasts (T4-T5, T4-T6, and T3-T6) was conducted. Five Mandarin listeners, five Dutch listeners and two Japanese listeners from Utrecht University participated in the pilot study, the procedure of which was the same as that discussed in §3.2.3 in the current study.

The vowel set of the non-words in the pilot consisted of /a/, /i/, and /u/. The consonant set contained three pairs of stops, labial /b/-/p/, velar /g/[18]-/k/, and alveolar /d/-/t/ (/p/-/pʰ/, /k/-/kʰ/, /t/-/tʰ/). Monosyllables of /ba/, /ta/, /ga/, /ka/, /da/, /pa/, /gu/, /ku/, /bu/, /bi/, and /pi/ were selected since these monosyllables are phonetically and phonotactically legitimate in all three languages. The selected syllables each carrying T1, T3, T4, T5, and T6 were produced six times each by three female and three male native speakers of Cantonese. For each speaker, three items of the best quality for each tonal syllable were selected and manipulated to 300 ms. All the tokens were created in a similar fashion as those in Chapter 2, normalized with the same duration (300 ms) and intensity (70dB). The monosyllables were concatenated to the trisyllabic non-words, with 25 ms of silence between syllables in accordance with Chapter 2, making each trisyllabic non-word 950ms. None of the trisyllabic-nonwords exist in any of the three languages. T3 was used as a companion tone, since T1 as the companion tone was more acoustically salient than the other tones as shown in Chapter 2 and in previous studies (Qin & Mok, 2013; Wong, 2019), while T6 as the companion caused confusion when the target tone was T5, as reported in Chapter 2. To eliminate the possible influence of position where the contrasts occur, the tonal and segmental contrasts were fixed at the initial position in the pilot study. The tokens in the pilot were a subset of those used in the current study. Two-, three- and four-word sequence lengths were used in the pilot.

Table 3.1. *Stimuli in the pilot study.*

	Contrast	Token	Tonal pattern
	segmental	tapibu-gapibu	T3T3T3-T3T3T3
optimally perceived at the acoustic level	T1 – T3	budapi	T1T3T3-T3T3T3
	T1 – T4	gutapi	T1T3T3-T4T3T3
	T1 – T5	gapibu	T1T3T3-T5T3T3
	T1 – T6	tapibu	T1T3T3-T6T3T3
	T3 – T4	pibuga	T3T3T3-T4T3T3
	T3 – T5	kupiga	T3T3T3-T3T3T3
sub-optimally perceived at the acoustic level	T3 – T6	tabupi	T3T3T3-T5T3T3
	T4 – T5	bigupa	T4T3T3-T5T3T3
	T4 – T6	pibaku	T4T3T3-T6T3T3

The overall performance of participants perceiving contrasts with T1, T3-T4, and T3-T5 reached ceiling effect in the sequence-recall task in the pilot study, with all three groups showing comparable good performance around 90% accuracy. The

[18] Although Dutch has no phonemic contrast /g/ versus /k/, it does have loan words borrowing from English, French and German that contain the sound /g/.

overall performance perceiving contrasts T3-T6, T4-T5, and T4-T6 was around 70% to 80%, and the performance of the three language groups varied from 40% to 90% accuracy (see Appendix C for detailed data in the pilot study). The results in the pilot study suggest that contrasts with T1, T3-T4, and T3-T5 were not only optimally perceived at the acoustic level to the listeners but also seemed comparably easy in the more demanding task, the sequence-recall task. Comparatively, contrasts T3-T6 (pitch level contrast), T4-T5 (pitch contour contrast), and T4-T6 (pitch contour vs. pitch level contrast) were more robust for teasing apart the groups with respect to their perception of non-native pitch contrasts at the phonological level. Therefore, these three contrasts were selected for use in the present study.

Table 3.2 shows the stimuli used in the current study. All of the non-native pitch contrasts occurred in word-initial, word-medial, and word-final positions, respectively. Segmental contrasts were used as a baseline.

Table 3.2. *Stimuli in the sequence-recall task in the current study.*

Contrast	Pair	Token (carrying tonal pattern)	Position
Segmental	1	tapibu-gapibu (T3T3T3-T3T3T3)	initial
	2	gutapi-gukapi (T3T3T3-T3T3T3)	middle
	3	pibuda-pibuga (T3T3T3-T3T3T3)	final
T4 vs. T5	4	budapi (T4T3T3-T5T3T3)	initial
	5	pabigu (T3T4T3-T3T5T3)	middle
	6	kupiga (T3T3T4-T3T3T5)	final
T4 vs. T6	7	tabupi (T4T3T3-T6T3T3)	initial
	8	gubapi (T3T4T3-T3T6T3)	middle
	9	bidagu (T3T3T4-T3T3T6)	final
T3 vs. T6	10	pibaku (T3T3T3-T6T3T3)	initial
	11	kapubi (T3T6T3-T3T3T3)	middle
	12	bigupa (T3T3T3-T3T3T6)	final

The experiment was adapted from the sequence-recall task from Dupoux et al. (2001). Two-word, three-word, and four-word sequences were used. There are four possible combinations of two-word sequences, namely AB, BB, AA, and BA. Eight three-word combinations of sequences are possible, AAA, BBB, ABA, ABB, ABA, BAB, BBA, BAA, while there were 16 possible combinations of four-word sequences, among which eight sequences with two and three transitions from A to B or B to A were selected. These were AABA, ABAA, ABBA, BAAB, BABB, BBAB, ABAB, BABA, as in Dupoux et al. (2001). The inter-stimulus-interval was set to 80 ms, as in Dupoux et al. (2001). In each sequence, the non-words were produced by different speakers. The order of the voices was counterbalanced over the sequences. There were eight trials of each sequence length, as follows:

Two-word length: AA, BB, AB, BA, AA, BB, AB, BA

Three-word length: AAA, AAB, ABA, ABB, BBB, BBA, BAB, BBB

Four-word length: AABA, ABAA, ABBA, BAAB, BABB, BBAB, ABAB, BABA

The overall design of the experiment is: pair (12) x trials (24), yielding 288 trials in total.

3.2.3 Procedure

The experiment contained 12 pairs of non-words in total, each pair taking approximately 10 minutes. Each participant was required to come to the lab twice, with a three-to-seven-day interval separating the visits. The Dutch and Chinese participants completed the experiments in the phonetics lab at UiL-OTS; Japanese participants participated in a sound-attenuated room in the Education and Research Centre at International Christian University in Tokyo. All of the participants were instructed that they were going to learn six pairs of new words in a foreign language each time. The learning order of the 12 pairs was counterbalanced across participants. However, all of the participants learned one of the three segmental contrasts first so that they could get an idea of the experiment.

For each pair, there were 5 phases. The participants were instructed that they were going to learn new words A and B that differed only in "melody" or in one segment. Note that the same instructions were presented to the participants both in English and their native language (See Appendix F for details). The two words were associated with buttons A and B, respectively. In Phase 1, participants first listened to 12 tokens of word A followed by 12 tokens of word B. After learning words A and B, they proceeded to Phase 2 where they could press button A or B to listen to the words repeatedly as many times as they wished to make sure they had memorized the two words. In Phase 3, they did a practice trial in which they heard a word and judged whether the word was word A or B by pressing the corresponding button. There were 20 trials in the practice, in which they were required to reach a criterion of 75% accuracy. If they failed in the practice, they would go back to Phase 2 to learn words A and B again until they could reach the passing criterion for Phase 3. They then proceeded to Phase 4 where they listened to words A and B in two-word and three-word sequences. They were required to recall the sequences by pressing the corresponding buttons. For instance, if they heard A-B, they should press button A firstly and B secondly in an A-B order. There were four trials in Phase 4: two trials of two-word sequences and two trials of three-word sequences. After finishing Phase 4, they could redo Phase 4 for more practice if they wished or proceed directly to Phase 5, the test phase. Phase 5 contained 3 blocks. Blocks 1, 2, and 3 were words A and B in two-word, three-word, and four-word sequences, respectively. A feedback message “OK” would appear on the screen, simultaneously presented with a sound message “okay” once they responded to each trial. The “okay” was recorded by a female native speaker of English to eliminate the possible use of echoic memory by the participants (Morton et al., 1981; Morton et al., 1991; Dupoux et al., 2001). Each block had 8 trials, yielding 24 trials in total in each pair.

3.3 Results & Discussion

3.3.1 Practice

Table 3.3 shows the number of practice times in Phase 3 that were needed to reach the criterion in order to proceed to the test phase by each language group. All three groups needed only one or two times for practice when learning segmental contrasts (Pair 1, 2, and 3). When learning non-native pitch contrasts (Pair 4 to Pair 12), Mandarin listeners needed one to three times to pass the practice phase. Japanese listeners needed one to four times. The number of practice times needed to reach the criterion by Dutch listeners varied from one time to six times.

Table 3.3. *Number of practice times needed to proceed to the test phase in each language group.*

Number of participants in each language group		Number of practice times needed to proceed to the test phase					
		one	two	three	four	five	six
Pair 1	Dutch	24	6				
tapibu-gapibu	Japanese	24	6				
	Mandarin	24	5	1			
Pair 2	Dutch	20	10				
gutapi-gukapi	Japanese	20	10				
	Mandarin	21	9				
Pair 3	Dutch	22	8				
pibuda-pibuga	Japanese	23	7				
	Mandarin	23	7				
Pair 4	Dutch	6	8	9	5	1	1
budapi	Japanese	14	11	4			
(T4T3T3-T5T3T3)	Mandarin	21	9				
Pair 5	Dutch	6	6	9	7	1	1
pabigu	Japanese	12	14	4			

(T3T4T3-T3T5T3)	Mandarin	22	8				
Pair 6	Dutch	5	7	11	7		
kupiga	Japanese	13	14	3			
(T3T3T4-T3T3T5)	Mandarin	24	6				
Pair 7	Dutch	6	5	8	9	1	1
tabupi	Japanese	9	19	2			
(T4T3T3-T6T3T3)	Mandarin	15	14	1			
Pair 8	Dutch	7	6	8	8	1	
gubapi	Japanese	13	13	4			
(T3T4T3-T3T6T3)	Mandarin	14	16				
Pair 9	Dutch	7	10	8	4	1	
bidagu	Japanese	12	15	3			
(T3T3T4-T3T3T6)	Mandarin	18	12				
Pair 10	Dutch	6	6	11	6	1	
pibaku	Japanese	12	13	4			
(T3T3T3-T6T3T3)	Mandarin	14	15	1			
Pair 11	Dutch	7	8	11	4		
kapubi	Japanese	12	14	4			
(T3T6T3-T3T3T3)	Mandarin	16	13	1			
Pair 12	Dutch	8	9	11	2		
bigupa	Japanese	11	14	4	1		
(T3T3T3-T3T3T6)	Mandarin	13	16	1			

Figure 3.2 further reports the differences in practice times that were necessary for each language group to proceed to the test phase, based on a cross-classified multilevel model with language and pair as fixed factors and subject as random effect in SPSS 25. The F-tests showed that Language (F (2, 990) = 17.370, $p < 0.001$) and Pair (F (11, 990) = 39.557, $p < 0.001$) were significant, as shown in Table 3.4. An interaction between Language and Pair (F (22, 990) = 13.045, $p < 0.001$) was found significant, suggesting that the times needed by the three language groups differed significantly depending on specific pairs. Post-hoc analysis with

Bonferroni-corrected showed that the three groups did not differ significantly in the practice times they needed to meet the criterion to learn segmental contrasts (Pair 1, 2 and 3) ($p > 0.05$). However, when learning non-native pitch contrasts (Pair 4 to 10), Dutch listeners needed significantly more practice than Mandarin and Japanese listeners ($p < 0.05$), while the latter two groups did not differ significantly ($p > 0.05$) (for detailed data please see Appendix B).

Table 3.4. *Parameters for fixed factors and random effects in practice phase.*

Fixed effects	F	df	p	
Language group	17.370	2	<0.001	
Pair	39.557	11	<0.001	
Language group * Pair	13.045	22	<0.001	
Random effects	Est.	Std. Error	z	p
1 | Participant	0.207	0.009	22.248	<0.001
Residual	0.922	0.515	6.327	<0.001

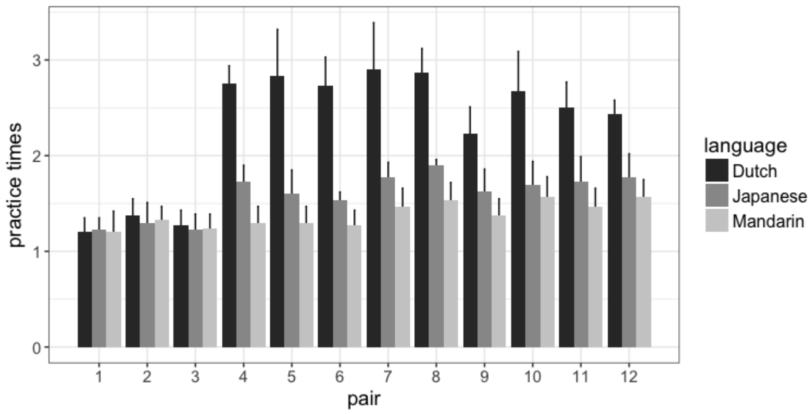


Figure 3.2. *Practice times of each language group needed to proceed to the test phase. Error bar[19]: 95% CI.*

3.3.2 Overall performance among language groups

3.3.2.1 Results

A response is correct if the entire sequence is recalled correctly. Correct responses were marked as “1” while incorrect responses were marked as “0”. Any missing response was regarded as an incorrect response. To analyze whether the three language groups differ in perception of the non-native contrasts, a generalized linear

[19] All error bars reported in the present chapter refer to 95% CI.

mixed model (GLMM) was computed in SPSS 25. Contrast (4 levels: one segmental contrast and three tonal contrasts), Position (3 levels: contrasts occurring word-initially, -medially, and -finally), Sequence Length (3 levels: 2-word, 3-word, and 4-word length sequence) and Language (3 levels) were taken as fixed factors, and intercepts for participant and item were taken as random effects into the model. The F-tests showed that Contrast (F (3, 25,884) = 57.841, $p < 0.001$), Position (F (2, 25,884) = 48.996, $p < 0.001$), Sequence Length (F (2, 25,884) = 20.436, p < 0.001), and Language (F (2, 25,884) = 37.528, $p < 0.001$) were significant, as shown in Table 3.5. Two-way interactions between Language and Contrast (F (6, 25,884) = 10.871, $p<0.001$), Language and Position (F (4, 25,884) = 60.124, $p < 0.001$), and Contrast and Position (F (6, 25,884) = 7.385, $p < 0.001$) were found significant. Moreover, a three-way interaction was found significant among Language, Contrast, and Position (F (12, 25,884) = 18.396, $p < 0.001$). Sequence Length did not interact with Language (F (4, 25,884) = 1.747, $p = 0.137$), Contrast (F (4, 25,884) = 0.725, $p = 0.\ 630$), or Position (F (4, 25,884) = 0.507, $p = 0.731$), indicating that the differences among language groups, contrasts, and positions were comparable in each sequence length.

Table 3.5. *Parameters of fixed effects and random effects in sequence-recall task.*

Fixed effects	F	df	p
Language group	37.528	2	<0.001
Contrast	57.841	3	<0.001
Position	48.996	2	<0.001
Sequence length	20.436	2	<0.001
Language group * Contrast	10.871	6	<0.001
Language group * Position	60.124	4	<0.001
Language group * Sequence length	1.747	4	0.137
Contrast * Position	8.344	6	<0.001

Random effects	Est.	Std. Error	z	p
1 | Participant	1.090	0.174	6.266	<0.001
1 | Item	0.233	0.028	8.209	<0.001

Figure 3.3 shows the overall performance of each language group in each sequence length. The chance level for the 2-word, 3-word, and 4-word sequences are 25% ($1/2^2$), 12.5% ($1/2^3$), and 6.25% ($1/2^4$), respectively (Dupoux et al., 2001). All three groups performed better than chance level and showed similar perceptual patterns in each sequence length. Their performance decreased with the increase of sequence length. Furthermore, in each sequence length, Mandarin listeners significantly outperformed Japanese (in two-word sequences: F (2, 25,896) = 20.858, $p = 0.003$; in three-word sequences: F (2, 25,896) = 25.7607, $p = 0.015$; in four-

word sequences: *F* (2, 25,896) = 30.709, *p = 0.010*) and Dutch listeners (two-word sequences: *F* (2, 25,896) = 20.858, *p* < 0.001; three-word sequences: *F* (2, 25,896) = 25.7607, p < *0.001*; four-word sequences: *F* (2, 25,896) = 30.709, *p* < *0.001*), while Japanese listeners significantly outperformed Dutch listeners (two-word sequences: *F* (2, 25,896) = 20.858, *p* < 0.001; three-word sequences: *F* (2, 25,896) = 25.7607, *p* < *0.001*; four-word sequences: *F* (2, 25,896) = 30.709, *p* < *0.001*).

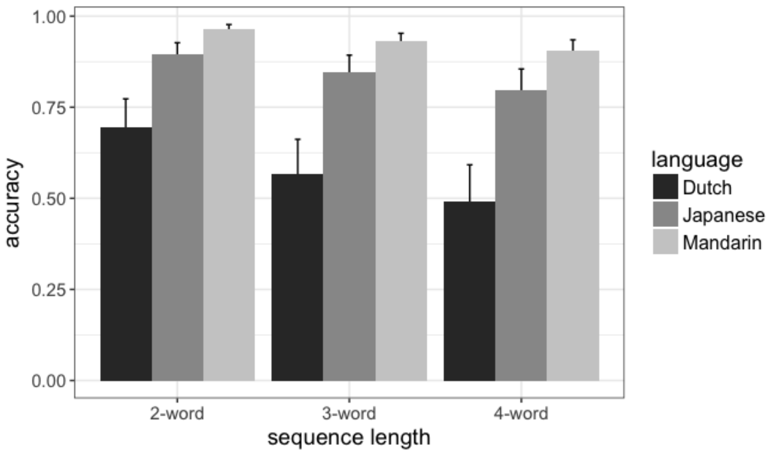


Figure 3.3 *Performance of each language group in each sequence length.*

Since Sequence length had no interaction with Contrast, Position or Language group, it was grouped into Contrast and Position with respect to the performance of each language group. Figure 3.4 displays the performance of each language group in perceiving segmental contrast (control condition) and pitch contrasts (tonal condition) with sequences of different lengths collapsed. In perceiving segmental contrasts, all three groups achieved good performance, with accuracy more than 85% (Dutch listeners: 88.76%; Mandarin listeners: 95.01%; Japanese listeners: 92.38%). The three groups had no difficulty encoding segmental contrasts in the sequence recall task and did not differ from each other significantly (*F* (2, 25,884) = 8.775, all *p* > *0.05*). However, they showed differences in the tonal condition. In perceiving non-native pitch contour contrast (T4 vs. T5), overall both Mandarin and Japanese listeners significantly outperformed Dutch listeners (both *F* (2, 25,884) = 47.708, *p* < *0.001*), while Mandarin listeners had significantly better performance than Japanese listeners (*F* (2, 25,884) = 47.708, *p = 0.004*), displaying a hierarchy: Mandarin listeners >> Japanese listeners >> Dutch listeners. A similar hierarchy was also found in perceiving non-native pitch level contrast (T3 vs. T6): both Mandarin and Japanese showed significantly better performance than Dutch listeners (both *F* (2, 25,884) = 24.197, *p* < *0.001*). Meanwhile, Mandarin listeners significantly outperformed Japanese listeners (*F* (2, 25,884) = 24.197, *p = 0.001*). With respect to their perception of pitch contour vs. pitch level contrast (T4 vs. T6), Mandarin and Japanese listeners did not differ from each other (*F* (2, 25,884) = 39.565, *p = 0.055*) and they both significantly outperformed Dutch listeners (both F (2, 25,884) = 39.565, $p < 0.001$).

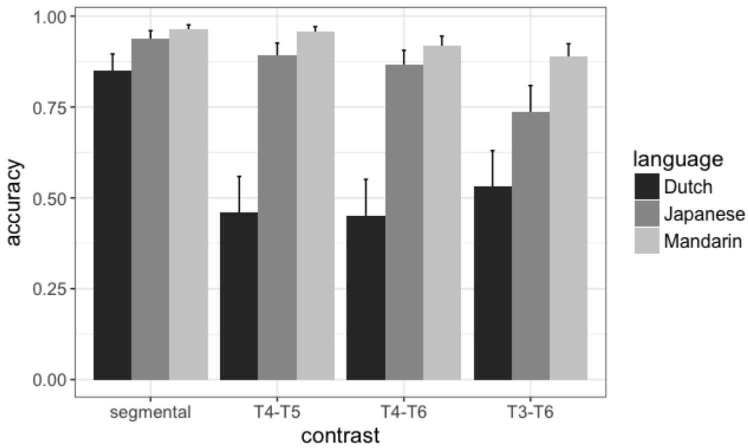


Figure 3.4 *Performance of each language group on each contrast.*

3.3.2.2 Summary

To summarize, all three language groups showed comparable performance patterns in each sequence length both in the control condition (segmental contrast) and tonal condition (non-native pitch contrasts in different positions). Their accuracy decreased when the sequence length increased from two words, to three words, to four words. In each sequence length, a gradient perceptual hierarchy was observed: Mandarin listeners >> Japanese listeners >> Dutch listeners. Moreover, sequence length did not influence their perception with respect to contrast (segmental and tonal) or position (word-initial, -medial, and -final).

Secondly, the three language groups achieved comparably good performance in the control condition. They all had no problem perceiving segmental contrasts at the phonological level. They all performed better when the segmental contrast occurred word-finally than word-initially.

However, the three groups showed differences with respect to tonal condition. Overall, Mandarin listeners outperformed Japanese listeners, while Japanese listeners outperformed Dutch listeners when perceiving the non-native pitch contrasts. More specifically, when perceiving non-native pitch contour contrast (T4 vs. T5) and pitch level contrast (T3 vs. T6), the three groups displayed a hierarchical order: Mandarin listeners >> Japanese listeners >> Dutch listeners. When perceiving pitch contour vs. pitch level contrast (T4-T6), Mandarin and Japanese listeners performed comparably well and both outperformed Dutch listeners.

The advantage of Mandarin and Japanese listeners over Dutch listeners was also observed with regard to the positions where the non-native pitch contrasts occurred at the phonological level. When the non-native pitch contrasts occurred word-initially, a gradient perceptual order was found: Mandarin listeners >> Japanese listeners >> Dutch listeners. When the non-native pitch contrasts occurred word-medially and word-finally, both Mandarin and Japanese listeners outperformed Dutch listeners, while the former two groups did not differ from each other.

3.3.3 Performance of each language group

To investigate the perceptual patterns of each language group, GLMM analysis was conducted for each language group. In each language group, Contrast (4 levels: one segmental contrast and three pitch contrasts) and Position (3 levels: word-initial, -medial, and -final position) were taken as fixed factors. Intercepts for participant and item were added as random effects. Post hoc comparisons were analyzed with Bonferroni adjustment in each language group.

3.3.3.1 Results of Dutch listeners' performance

Figure 3.5 displays the performance of Dutch listeners with respect to segmental contrast (control condition) and pitch contrast (tonal condition) in different positions. For Dutch listeners, both the fixed factors, Contrast (F (3, 8.628) = 89.129, $p < 0.001$) and Position (F (2, 8,628) = 23.413, $p < 0.001$), were significant. Contrast and Position had a significant interaction (F (6, 8,628) = 7.385, $p < 0.001$), indicating that perceptual differences in performance depended on position.

As can be seen in Figure 3.5, Dutch listeners had a significantly overall better performance perceiving segmental contrast than all the non-native pitch contrasts (F (3, 8,628) = 62.466, $p < 0.001$), indicating that they had more difficulties encoding non-native pitch contrasts than segmental contrasts at the phonological level. In the tonal condition, Dutch listeners showed significantly better performance overall when perceiving pitch level contrast T3-T6 (55.3% accuracy) than pitch contour T4-T5 contrast (46.2% accuracy) (F (2, 6,471) = 3.268, $p = 0.039$) and pitch level vs. pitch contour contrast T4-T6 (45.4% accuracy) (F (2, 6,471) = 3.131, $p = 0.031$).

Moreover, Dutch listeners performed differently with regard to position. Specifically, when perceiving pitch contour vs. pitch level contrast T4 vs. T6, they achieved significantly better performance at the word-final position, reaching over 70% accuracy than at word-initial (around 35% accuracy) and word-medial positions (around 25% accuracy) (F (2, 8,628) = 23.546, both $p < 0.001$). However, when perceiving pitch contour contrast T4-T5 and pitch level contrast T3-T6, there was no significant difference at any of the three positions (T4-T5: F (2, 8,628) =

1.694, $p > 0.05$ in all comparisons; T3-T6: F (2, 8,628) = 2.939, $p > 0.05$ in all comparisons).

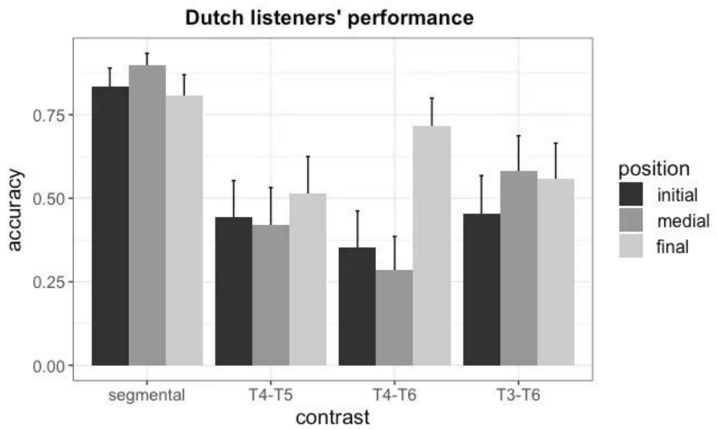


Figure 3.5. *Dutch listeners' performance for each contrast and position.*

3.3.3.2 Results of Mandarin listeners' performance

Figure 3.6 shows the performance of Mandarin listeners perceiving segmental contrast and non-native pitch contrasts in different positions. They achieved more than 75% accuracy in each contrast at each position. For Mandarin listeners, Contrast (F (3, 8,628) = 17.794, $p<0.001$) but not Position (F (2, 8,628) = 0.400, $p = 0.670$) was found significant. Post hoc pair-wise comparison showed that Mandarin listeners performed better in encoding pitch contour contrast T4-T5 than in pitch contour vs. level T4- T6 (F (3, 8,628) = 8.860, $p = 0.006$) and pitch level contrast T3-T6 (F (3, 8,628) = 10.987, $p < 0.001$), while their performance perceiving T4-T6 was significantly better than perceiving T3-T6 (F (3, 8,628) = 7.683, $p = 0.035$) at the phonological level. Notably, Mandarin listeners perceived non-native pitch contour contrast T4-T5 as well as they did perceiving segmental contrast (F (3, 8,628) = 0.475, $p = 0.379$), with an accuracy of approximately 96%.

Regarding the influence of position on pitch contrast, a significant interaction between Contrast and Position was found (F (6, 8,628) = 11.766, $p < 0.001$), suggesting that there were differences among the performance for contrasts depending on certain positions. Post hoc pair-wise comparison further showed that Mandarin listeners better perceived pitch contour vs. pitch level contrast T4-T6 word-finally than word-initially (F (2, 8,628) = 7.249, $p = 0.047$) and word-medially (F (2, 8,628) = 9.319, $p = 0.001$). When perceiving pitch level contrast T3-T6, they showed significantly better performance when the contrast occurred word-medially than word-initially (F (2, 8,628) = 8.968, $p = 0.006$) or word-finally (F (2, 8,628) = 10.042, $p < 0.001$), as shown in Figure 3.6. However, when perceiving pitch contour contrast T4-T5, they achieved comparably good performance at each position (F (2, 8,628) = 0.408, p = *0.665*).

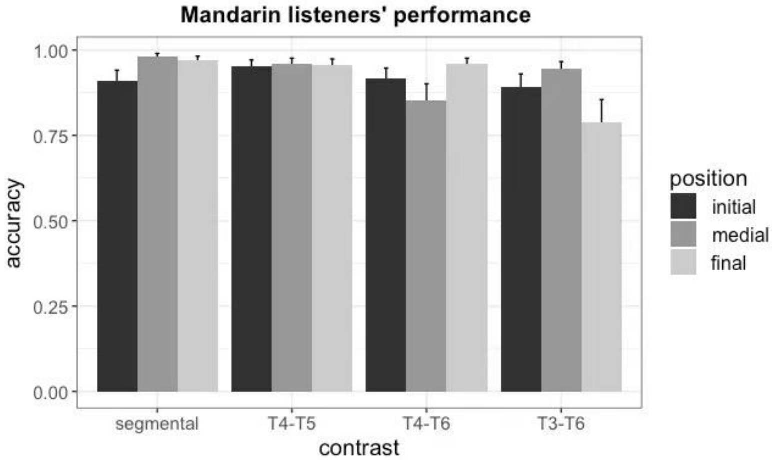


Figure 3.6. *Mandarin listeners' performance for each contrast at each position.*

3.3.3.3 Results of Japanese listeners' performance

For Japanese listeners, both Contrast (F (3, 8.63) = 52.224, p < *0.001*) and Position (F (2, 8,628) = 133.202, p = *0.003*) were found significant. Post hoc pair-wise comparison reported that the overall performance perceiving pitch contour T4-T5 (90.1% accuracy) and pitch contour vs. pitch level contrast T4-T6 (87.3% accuracy) was comparable (F (2, 6,471) = 15.053, p = *0.063*), both significantly better than that of pitch level contrast T3-T6 (75.1% accuracy) (F (2, 6,471) = 16.306, both p < *0.001*).

Contrast interacted with Position significantly (F (6, 8,628) = 26.432, p < *0.001*), implying that the performance for non-native pitch contrasts by Japanese listeners differed depending on position. As shown in Figure 3.7, when perceiving pitch contour contrast T4-T5, Japanese listeners achieved around 97% accuracy at word-medial position, significantly better than word-initial (58.4% accuracy, F (2, 8,628) = 19.062, p < *0.001*) and word-final position (F (2, 8,628) = 19.062, p = *0.001*). When perceiving pitch contour vs. level contrast T4-T6, they showed comparably good performance word-medially and word-finally (F (2, 8,628) = 0.013, p=*0.943*), both significantly better than word-initially (F (2, 8,628) = 15.600, both p < *0.001*). Their perception of pitch level contrast T3-T6 was significantly better in word-medial than word-initial and word-final positions (F (2, 8,628) = 20.580, both p < *0.001*).

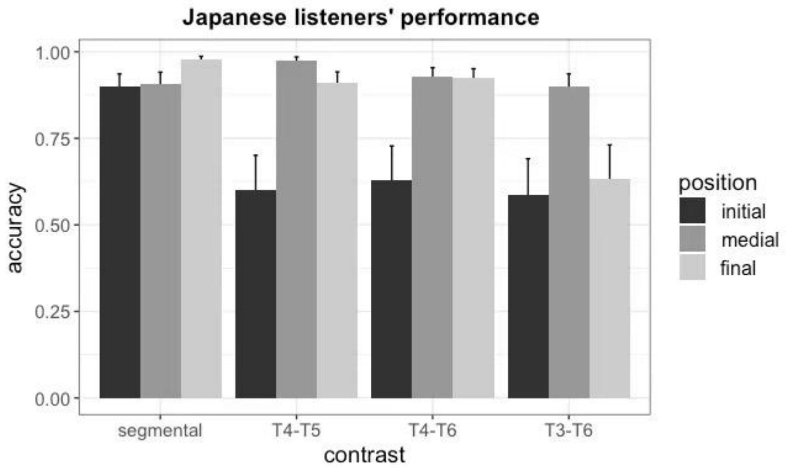


Figure 3.7. *Japanese listeners' performance for each contrast at each position.*

3.3.3.4 Summary

Taken together, language-specific and language-general patterns were observed across the three groups. All three groups showed a preference for word-final position over word-medial and/or word-initial position when perceiving pitch contour vs. pitch level contrast T4-T6.

Apart from the language-general pattern, the three language groups displayed different perceptual patterns when encoding non-native pitch contrasts in different positions at the phonological level. Dutch listeners had more difficulty perceiving non-native pitch contrasts than segmental contrasts (control condition). In general, Dutch listeners favoured perceiving pitch level contrast T3-T6 over the other non-native pitch contrasts. Position had a robust influence on Dutch listeners' performance only when perceiving pitch contour vs. pitch level contrast T4-T6. More specifically, their performances rocketed to around 70% correct when T4-T6 occurred word-finally compared to word-initially (35% accuracy) and word-medially (28% accuracy). However, Dutch listeners had no preference for any position when perceiving pitch contour contrast T4-T5 and pitch level contrast T3-T6.

Mandarin listeners perceived the non-native pitch contour contrast T4-T5 as accurately as the segmental contrast at the phonological level. Pitch level contrast T3 vs. T6 was the most difficult for Mandarin listeners to encode at the phonological level, compared with other non-native pitch contrasts. Moreover, the differences among their performance perceiving the non-native pitch contrasts were dependent on position. When perceiving non-native pitch level contrast T3-T6, they showed significantly better performance word-medially than word-finally and word-initially, while they were significantly better at encoding pitch contour vs. pitch level contrast

T4-T6 word-finally than word-initially and word-medially. However, they didn't show any preference with respect to position when perceiving pitch contour contrast T4-T5.

Japanese listeners performed better perceiving segmental contrast than all the non-native pitch contrasts. Among the non-native pitch contrasts, they encountered most difficulty perceiving non-native pitch level contrast T3-T6, while they had comparable performance when perceiving pitch contour contrast T4-T5 and pitch contour vs. level contrast T4-T6. They displayed a strong preference for word-medial position over word-initial and word-final position for all the pitch contrasts.

3.4 General discussion

The present study investigated the perception of non-native pitch contrasts in different positions at the phonological level by Dutch, Mandarin, and Japanese listeners. It aims to shed some light on to what extent native prosody (word prosody or intonation prosody if it exists) influences the encoding of non-native WPCs, specially, non-native pitch contrasts that occur word-initially, -medially, and -finally at the phonological level, by putting forward 1) how the three language groups differ from each other in perceiving non-native WPCs at the phonological level, 2) whether the three groups show differences or share commonalities when encoding non-native WPCs phonologically, and 3) to what extent position interacts with the perception of non-native pitch contrasts in each language group.

3.4.1 Overall performance across language groups

To begin with, in general, a hierarchical perceptual pattern of perceiving non-native pitch contrasts in different positions at the phonological level was observed: Mandarin listeners >> Japanese listeners >> Dutch listeners. The perceptual advantage of Mandarin listeners over Dutch and/or Japanese listeners was also observed at each position where the non-native pitch contrasts occurred. The findings were in accordance with predictions in (1b), but not (1a). (1a) was based on the Feature Hypothesis (McAllister et al., 2002) which states that if a phonetic or phonological dimension is not used for lexical contrast in the native language it will be difficult for naive listeners to perceive. Accordingly, Japanese listeners (whose native language uses lexical pitch accent and positional marking) would be expected to show advantage over Mandarin listeners (who use only lexical tones) and Dutch listeners (who only use lexical stress), contrary to the current findings. Instead, the predictions in (1b) held. These predictions were based on Functional Pitch Hypothesis stating that the functionality of pitch used for lexical contrasts will determine non-native tone perception (Schaefer & Darcy, 2014). According to this hypothesis, the higher the functionality of pitch in the native language, the better the performance will be in non-native tone perception. Mandarin, Japanese, and Dutch differ in the degree of functionality of lexical pitch in the language ranging from maximal, to intermediate, to low. This shapes the perceptual performance of the three groups in a gradient order. The current findings seem to suggest that the use of lexical pitch and/or to what degree lexical pitch is used in the native language, rather than positional marking, is a prominent factor that matters in the perception of non-native pitch contrasts occurring in different positions.

3.4.2 Language-specific perceptual patterns

Besides the differences of overall performance across groups, each language group showed specific patterns when perceiving non-native WPCs at the phonological level. Dutch listeners displayed more troubles encoding non-native pitch contrasts, compared with their good performance (ceiling effect) encoding segmental contrast (baseline) phonologically. Among the non-native pitch contrasts, they showed a perceptual preference for pitch level contrast T3-T6 over pitch contour T4-T5 and pitch contour vs. pitch level contrast T4-T6. Unlike Dutch listeners, Mandarin listeners were in favour of non-native pitch contour contrast T4-T5. They perceived non-native pitch contour contrast as well as they had segmental contrast at the phonological level, both reaching ceiling performance. Moreover, based on performance accuracy, the non-native pitch contrasts they attended to were found in a hierarchical order: pitch contour contrast T4-T5 >> pitch contour vs. pitch level contrast T4-T6 >> pitch level contrast T3-T6, i.e., the non-native pitch contour contrast was their favourite while the non-native level contrast was the most troublesome. Japanese listeners showed comparable performance perceiving non-native pitch contour contrast T4-T5 and pitch contour vs. pitch level contrast T4-T6. However, comparatively, they had more difficulty perceiving non-native pitch level contrast T3-T6 than perceiving T4-T6 and T4-T5.

The language-specific perceptual patterns were by and large in line with the predictions based on the approach of relative perceptual weighting to features of tones (Lee et al., 1996; Xu et al., 2006; Francis et al., 2008) proposing that stored representations of pitch-based phonological categories in long-term memory influence listeners' perceptual weight given to non-native pitch features. The predictions are supported by the perceptual patterns observed in Dutch and Mandarin listeners but not Japanese listeners. Japanese listeners were predicted to be more sensitive to perceive non-native pitch level contrast than pitch contour contrast at the phonological level as Japanese listeners may have long-term storage of representations of the accentual patterns of prosodic words consisting of H tone and L tones. However, the current findings contradict this prediction. It was found instead that Japanese listeners were more sensitive to perceive the non-native pitch contour contrast T4-T5 and non-native pitch contour vs. level contrast T4-T6 than the non-native pitch level contrast. This could be accounted for by their native sensitivity to detect a falling tone. The non-native pitch contour contrast T4-T5 (falling vs. rising) and pitch contour vs. pitch level T4-T6 both contained a falling tone T4. Recall that Japanese employs a lexical pitch fall to distinguish word meanings. According to a proposal of relative perceptual weighting to features of tones (Lee et al., 1996; Xu et al., 2006; Francis et al., 2008), Japanese listeners may have stored long term memory representation of the lexical pitch fall, which may facilitate Japanese listeners to distinguish a falling pitch contour from a non-native pitch contrast. Indeed, Ustugi et al. (2010) via a sequence recall task (Dupoux et al., 2001) reported that Japanese listeners were good at detecting the lexical pitch fall in their native language at the phonological level. This sensitivity to pitch fall in one's native language may be transferred to perceive a non-native falling tone. Moreover, the findings on Japanese listeners' perceptual pattern seem to suggest that the representations of pitch-based phonological categories stored in Japanese listeners' long-term memory are associated with lexical pitch fall, rather than the basic structure of H and L tones in the surface tonal pattern.

Despite the rejection of the prediction for Japanese listeners, the predictions for Dutch and Mandarin listeners held. Dutch listeners, as predicted, were more sensitive to perceive the non-native pitch level contrasts than contour contrasts at the phonological level. Dutch does not have lexical tones. According to a proposal of relative perceptual weighting to features of tones (Lee et al., 1996; Xu et al., 2006; Francis et al., 2008), the long-term storage of pitch-based phonological categories in Dutch listeners are intonational nuclear tones consisting of H and L tones. This may equip Dutch listeners to show more sensitivity and encode the non-native pitch level contrast compared to the non-native pitch contour contrast and non-native pitch contour vs. level contrast at the phonological level.

As for Mandarin, the language has three contour tones and one level tone. It uses pitch contour contrasts but not pitch level contrasts to signal lexical meanings. Mandarin listeners may have stored pitch representations of tonal categories, i.e., categories of contrastive contour tones, in long-term memory, which may benefit them when encoding non-native pitch contour contrasts at the phonological level. However, due to a lack of contrastive level tones in Mandarin, they were relatively at a disadvantage when perceiving non-native pitch level contrasts, compared with their good performance perceiving non-native pitch contrasts paired with pitch contour (T4-T5, T4-T6). More specifically, a lower sensitivity to the non-native pitch level contrasts could be further attributed to a lack of clear correspondence between the native and non-native tone categories. Francis et al. (2008) further develops this proposal in light of relative perceptual weighting given to tone features (Lee et al., 1996) and the multi-store memory model (Xu et al., 2006) and proposes that when a clear correspondence holds between native and non-native category representations, the non-native categories will be assimilated to the native categories. When there is no clear correspondence, the relative weight of perceptual features will be distributed in accordance with the native tone categorization. Mandarin has three contrastive contour tones and one level tone. Accordingly, the non-native pitch contrast T4-T5 (falling vs. rising tone) may be assimilated to the falling and rising tone category in Mandarin, respectively. The non-native pitch contour vs. pitch level contrast T4-T6 (falling vs. low level tone) may be assimilated to the falling and level tone categories, respectively. However, there is no clear correspondence between the non-native pitch level contrast T3-T6 (mid level vs. low level tone) and any native level tone category in Mandarin. Thus, native lexical tone categories would lead Mandarin listeners to be less accurate in encoding this non-native pitch level contrast, compared with the non-native pitch contour contrast and pitch contour vs. pitch level contrast.

In terms of the correspondence between the native category and the non-native category representations, more specifically, So & Best (2010, 2011, 2012, 2014) extended the Perceptual Assimilation Model (PAM) to PAM-S (suprasegmental) to propose that if the non-native phonemic contrast (here, suprasegmentals) is mapped onto the same category in the native language, it will yield poor perception. If the non-native phonemic contrast is mapped onto two different categories in the native language, it will lead to good perception. Based on PAM-S (So & Best, 2014), Mandarin listeners may perceive the non-native pitch contour contrast T4-T5 and pitch contour vs. pitch level contrast T4-T6 as falling tone vs. rising tone and falling tone vs. level tone, respectively, in the native tonal categories, leading to good perception. In contrast, due to a lack of contrastive level tones in Mandarin, the non-native T3 (mid level tone) and T6 (low level tone) in non-native pitch level contrast could be mapped onto the same tonal category high level tone in Mandarin, which would give rise to perceptual difficulty for Mandarin listeners.

The PAM-S perceptual assimilation framework could account for Mandarin listeners' perceptual pattern. However, PAM-S seems not supported by the perceptual patterns observed in Dutch and Japanese listeners. Dutch does not have lexical tones but has a rich repertoire of nuclear tones in an intonation system. According to PAM-S, the non-native pitch contrasts can be mapped onto Dutch intonation categories, that is, the non-native contrast T4 vs. T5 and T4 vs. T6 can be mapped onto H*L L% (fall) vs. L*H H% (low rise) and L*H H% vs. H* % (level tone) (Gussenhoven, 2004, 2005) in Dutch intonational nuclear tonal categories, respectively, which would result in good perception. The non-native pitch level contrast T3-T6, would be mapped onto the same nuclear tone H* % (level tone), leading to poor perception. The expectations based on PAM-S contradict the current finding that Dutch listeners were more sensitive to non-native pitch level contrast than non-native pitch contour contrast and pitch contour vs. pitch level contrast. The conflict between the predictions of PAM and the current findings seems to suggest that the categorical status of intonational contours is not as distinctive as tonal categories, although there are studies suggesting some intonational pattern may be perceived categorically (Pierrehumbert & Steele, 1989; Grabe et al., 2003).

According to PAM-S, Japanese listeners may assimilate non-native pitch contrasts onto native accentual patterns in word prosody. Recall from §1.2.1.2 that in Japanese the accentual pattern of a word is determined by the presence/absence and the location of the lexical pitch accent (marked with an apostrophe (’). If the accent falls on the first mora[20], the accented mora receives a H tone and the following mora receives a L tone. If the accent falls on the second or later mora, the first mora receives a L tone and the moras from the second until the accented ones all receive a H tone. When the word is unaccented, the first mora has a L tone and the remaining moras receive a H tone (Haraguchi, 1999; Kubozono, 2008). Based on the accentual rules (McCawley, 1968; Haraguchi, 1999; Uwano, 1999), there can be H’L (e.g., a H’L pattern in an initial-accented bimoraic word), LH’ (e.g., LH’ pattern in a final-accented bimoraic word), HH’ (e.g., HH’ pattern in a LHH’, final-accented trimoraic word) and LL pattern (e.g., LL pattern in a HLL, initial-accented trimoraic word). In light of the PAM-S, the non-native pitch contrast T4-T5 (falling vs. rising tone) may be assimilated to the H’L and LH’ accentual categories in Japanese, respectively. The non-native pitch contour versus pitch level contrast T4-T6 (falling vs. low level tone) may be assimilated to LH’ and HH’ category, respectively. The non-native pitch level contrast T3-T6 (mid level vs. low level tone) may be assimilated to accentual categories HH’ and LL, respectively. The one-to-one assimilation between the three pairs of non-native pitch contrasts and the native accentual patterns would yield good perception in Japanese listeners, which was in conflict with the finding that Japanese listeners were more sensitive to encode T4-T5 and T4-T6 than T3-T6. It seems that the mapping between Japanese accentual patterns and the non-native tones may not be as clear as that between Mandarin lexical tones and the non-native tones. This opens up a question about the correspondence relation between the Japanese accentual patterns and non-native lexical tones (here Cantonese tones). For instance, So (2010) investigated the assimilation of Japanese accentual patterns and Mandarin tones by Japanese listeners. She found that Japanese listeners were able to categorize the Mandarin tones they heard onto the native accentual categories, that is, they associated the Mandarin high level tone, rising tone, and falling tone with the native accentual pattern HH’, LH’, and H’L, respectively. However, in So (2010), the LL pattern which is legitimate in Japanese trimoraic words (e.g., *megane* , “glasses” in Japanese carries H’LL pattern) was not investigated. Japanese listeners’ relatively poor performance encoding the non-native pitch level contrast T3-T6 seems to suggest that the status of the accentual patterns HH’ and LL is not as categorical as that of H’L and LH’, which may need further investigation in the future study.

Note that the equivalence relation between the non-native Cantonese tones with Mandarin tones, Japanese accentual categories, and Dutch intonation contours in the current study is in nature speculative. It is necessary to conduct overt studies to investigate the correspondence between the native categories (e.g., tonal, accentual, and intonational categories) and the non-native Cantonese tonal categories as in previous studies on the mapping between Mandarin tones and Japanese accentual

[20] Recall that the rules for tonal patterns in Tokyo Japanese discussed in the dissertation are words with light syllables (see §1.2.1.2 for details).

patterns (So, 2010) and between Mandarin tones and Australian English and French intonation patterns (So & Best, 2014).

3.4.3 Influence of position on perception of non-native pitch contrasts phonologically

For each language group, position was found to cast influence on their perception of non-native pitch contrasts at the phonological level. It was predicted that Dutch listeners would favour the final position while Mandarin listeners and Japanese listeners would show no preference for any position when perceiving non-native pitch contrasts in different positions. However, the current findings were not in line with these predictions.

First, we didn't predict a language-general pattern. The three groups were observed to share perceptual commonality, that is, when perceiving the non-native pitch contour vs. pitch level contrast T4 vs. T6, all the three groups achieved better performance on word-final position than on word-initial position and/or word-medial position. One parsimonious account could be simply attributed to the recency effect that the auditory information on the offset of a word retains more than that on the onset to listeners (Demany & Semal, 2007). However, if this finding is due to the recency effect, a preference for word-final position rather than word-initial and word-middle position should also be observed in the other contrasts, which on the contrary was not found in the current study. Hence, the recency effect (Demany & Semal, 2007) seems to not hold as an explanation for the language-general pattern observed in all the three groups.

Another account could be the influence of intonation contour. It is generally deemed that intonation patterns are associated with non-lexical linguistic (e.g., pragmatic, paralinguistic) meanings. For instance, rising contour at the end of an utterance is generally associated with interrogative meaning, while falling contour signals assertive meaning (Ladd, 1996; Cruttenden, 1981, 1997). The low falling tone T4 on the final position in the tonal pattern T3-T3-T4 may resemble a falling contour at the end of an utterance. Such a terminal fall in an utterance, most commonly in a statement, is generally observed across languages (Bolinger, 1978, 1982). The final falling contour in an utterance can refer to an effect of declination which is a downtrend of pitch declining over the course of an utterance and towards the end ('t Hart & Cohen, 1973; Maeda, 1976; Pierrehumbert, 1979; Cohen et al. 1982; Ladd, 1984; Strik & Boves, 1995), or a final lowering which is an additional lowering of pitch at the end of an utterance (Maeda 1976; Liberman & Pierrehumbert, 1984; Pierrehumbert & Beckman, 1988; Herman, 1996). Both effects of the terminal fall in an utterance are found in Dutch (Gussenhoven & Rietveld, 1988), Japanese (Sugito, 1982; Poser, 1984; Pierrehumbert & Beckman, 1988) and Mandarin (Shih, 1997, 2000; Lai et al., 2014). Listeners may associate the falling tone at the final position with a terminal fall signalling assertive intonation, and thus it might be distinctive enough to be differentiated from the word-final low level tone T6, compared with the T4-T6 contrast at word-initial and word-medial positions. However, such perceptual association between the non-native tone categories and intonation categories is in nature speculative. It would be necessary to conduct a true study on overt mapping between the native intonation categories and non-native tone categories (for instance, as is done for the mapping between Mandarin tonal categories and intonation categories in Australian English and French in So & Best (2014).

Secondly, for each language, position interacted differently with the perception of non-native pitch contrasts. Dutch listeners were found to only be affected by position when perceiving the pitch contour vs. pitch level contrast T4-T6. They were more in favour of perceiving T4-T6 when it occurred word-finally than word-initially and word-medially. In contrast, when perceiving pitch contour contrast T4-T5 and pitch level contrast T3-T6, they were not influenced by the position at all. They were comparatively vulnerable in each position when perceiving T4-T5 and T3-T6 at the phonological level. Their confusion of T4-T5 and T3-T6 could be attributable to a lack of lexical tones in the native language. Dutch does not have lexical tones. Dutch listeners may store long-term memory representations of pitch-based phonological categories, which are nuclear tones in intonation. Due to the lack of lexical tones, Dutch listeners may resort to intonation patterns. However, when the tonal contrast belonged to the same category, i.e., both T4 and T5 are contour tones or both T3-T6 are level tones, Dutch listeners seemed unable to efficiently rely on the native intonation contours. It seemed they could rely on native intonation only when the two tones in the contrast were in different tonal categories, i.e., pitch contour vs. pitch level T4-T6 and when such contrast occurred at the final position (see above for detailed discussion).

For Mandarin listeners, position played different roles in their perception of non-native pitch contrasts. When perceiving the non-native pitch contour contrast T4-T5 phonologically, position did not influence Mandarin listeners. They performed equally well at each position, suggesting that they were sensitive to the non-native pitch contour contrast at the phonological level regardless of the position where it occurred. This can be due to facilitation from the native tonal categories. Mandarin uses contrastive contour tones to signal lexical items. The distinctive lexical tonal categories of falling and rising tones may enhance their perception of non-native rising and falling tones. In other words, Mandarin listeners were so adept at perceiving the non-native pitch contour contrasts that the positions where the contrast occurred did not matter in their perception.

Nonetheless, unlike the good performance at each position when perceiving T4-T5, Mandarin listeners showed significantly better performance at word-medial position than at word-initial and word-final position when perceiving the non-native pitch level contrast T3-T6. Presumably due to a lack of pitch level contrasts in the native word prosody (Mandarin does not have contrastive level tones), Mandarin listeners may require tonal references for them to perceive non-native pitch level contrasts efficiently. To be more specific, when the T3-T6 contrast occurred word-initially (T3-T3-T3 vs. T6-T3-T3) or word-finally (T3-T3-T3 vs. T3-T3-T6), it only had one neighbouring level tone that Mandarin listeners may rely on to perceive the level contrast. In contrast, when the T3-T6 contrast occurred word-medially (T3-T3-T3 vs. T3-T6-T3), it was surrounded by two neighbouring level tones. The two contextual level tones could "highlight" the T3-T6 contrast at the medial position to make it more salient for Mandarin listeners, compared with when T3-T6 occurred in the initial or final position with only one contextual tone. This reliance on contextual tones by Mandarin listeners is in line with previous studies which found that level tones are more context-dependent than contour tones to tone language listeners (Fox & Qi, 1990; Moore & Jongman, 1997; Francis et al., 2003; Wong & Diehl, 2003; Xu et al., 2006).

Japanese listeners showed a perceptual preference for word-medial position over word-initial and word-final positions. We predicted that Japanese listeners would not favour any positions in that tri-mora words are mostly unaccented in Japanese, which was based on the assumption that they perceived trisyllabic non-words as 3-mora words. The current finding suggests that Japanese listeners may perceive the trisyllabic non-words CVCVCV not as 3-mora words, but rather as words with longer moras such as CVV.CV.CVV or CV.CV.CVV (or even longer).[21] One mora in Japanese is usually around 100-150 ms at normal tempo in natural speech (Hoequist, 1983). The trisyllabic non-words used in the study were each 950 ms, much longer than the phonetic duration of Japanese 3-mora words. Hence, they may have perceived the trisyllabic non-words as six or more than six moras. According to their native accentuation (Sato, 1993; Kubozono, 1998, 2006; Suzuki, 1995; Kitahara, 2001), pitch accent is placed dominantly on the antepenultimate mora in 5- to 7-mora words in Japanese, which is the medial position in this case, for instance /ne/ at the medial position (the antepenultimate mora) in /ka.a.ne'e.syon/ "carnation". The preference for the medial position (the antepenultimate mora) of pitch accent in a word consisting of six or more than six moras might account for the Japanese listeners' preference for word-medial position when perceiving non-native pitch contrasts.

To sum up, position interacted with the perception of specific non-native pitch contrasts in the perception of non-native pitch contrasts at the phonological level. Both language-general and language-specific patterns were found. The language-general pattern could be due to the influence of the intonation patterns shared across languages. The language-specific patterns regarding the effect of position in non-

---

[21] Note that analysis was conducted after the experiment took place, by which time I had already returned to the Netherlands. Due to practical issues, I was unable to recall the Japanese participants for a further report on whether they perceived the words in the experiment as words with longer moras.

native pitch contrasts could be attributed to features in the native phonology (word prosody and intonation).

3.5 Conclusion

The current study investigated the cross-linguistic perception of non-native pitch contrasts in different positions at the phonological level by Dutch, Mandarin, and Japanese listeners. It attempted to answer how native word prosody and/or intonation prosody modulate listeners encoding non-native pitch contrasts that occur word-initially, -medially, and -finally at the phonological level.

First, it observed an overall perceptual hierarchy: Mandarin listeners >> Japanese listeners >> Dutch listeners. Second, language-specific patterns were found with different weighting given to non-native pitch contrasts. Dutch listeners attended more to non-native pitch level contrast than pitch contour and pitch level vs. pitch contour contrast. Japanese listeners were more sensitive to perceive non-native pitch contour and pitch level vs. pitch contour contrast. Mandarin listeners displayed a gradient order favouring non-native pitch contour contrast >> non-native pitch level vs. pitch contour contrast >> non-native pitch level contrast. The observed language-specific patterns were accounted for in terms of the selective perceptual weighting to features of tones (Lee et al., 1996) and long-term memory representations of pitch-based phonological categories (tones and/or intonation) (Xu et al., 2006) and assimilation between the native categories and non-native categories (So & Best, 2010, 2011, 2014; Francis et al., 2008). Thirdly, the current study further examined the role position played in perceiving non-native pitch contrasts in different positions. On the one hand, interestingly, a language-general pattern was found that all three language groups preferred word-final position over word-initial and/or word-final position when perceiving non-native pitch level vs. pitch contour (falling tone), which could be due to the influence of intonation contours generally across languages. On the other hand, language-specific patterns were observed. Position played different roles in each language group in the perception of non-native pitch contrasts at the phonological level, which can be accounted for by the influence of the native phonology (word prosody and intonation).

The current study also discussed its observations in light of different perceptual frameworks (e.g., pitch-based feature and assimilation-based) and attempted to extend theoretical perceptual mechanisms, mostly discussed at the acoustic and phonological levels. Note that the comparisons between the findings in Chapter 2 (acoustic level) and in the current chapter (phonological level) will be discussed, together with Chapter 4 (word learning) in Chapter 5 for a general discussion.

The following chapter, Chapter 4, will continue to study how non-native WPCs, more specifically non-native pitch contrasts in different positions, are phonologically encoded at the lexical level, i.e., in word learning, from a cross-linguistic perspective.

Chapter 4
Processing of non-native WPCs in word-learning

4.1 Introduction

The processing of word prosodic cues does not only involve sensory-auditory processing (i.e., processing at the acoustic level) but also involves higher linguistic levels. Among these are the abstract phonological level, where acoustic traces that listeners might rely on are inaccessible (Dupoux et al., 2001, 2008), and the lexical level, where phonological knowledge is encoded to build lexical representation in word learning (Wong & Perrachione, 2007; Zatorre & Gandour, 2008; Wang & Cooper, 2013; Braun et al., 2014). Chapters 2 and 3 addressed the role of native word prosody in the cross-linguistic perception of non-native WPCs at the acoustic and phonological levels, respectively. The current chapter further taps into how and to what extent one's native language, i.e., native word prosody and/or intonation prosody if there is any, plays a role in the processing of non-native WPCs in integrating sound-to-meaning associations in word learning.

As mentioned earlier, the bulk of previous studies have addressed how native language influences naive listeners in acoustic discrimination or identification of non-native WPCs, mostly with respect to non-native pitch contrasts (Lee et al, 1996; Burnham et al., 1996; Gottfried & Suiter 1997; Burnham & Francis, 1997; Wu & Lin, 2008; Schaefer & Darcy, 2014; Burnham et al., 2015; Reid et al., 2015; among others). Comparatively less is known about how naive listeners encode non-native pitch contrasts in establishing lexical representations. Earlier studies on the lexical encoding of non-native pitch contrasts have mainly focused on whether naive listeners can use non-native lexical tones to learn words after receiving short-term training (Wong & Perrachione, 2007; Chandrasekaran et al., 2010; Perrachione et al., 2011; Cooper & Wang, 2012, 2013), whether L2 learners of a tone language can acquire non-native lexical tones to recognize words (Zou, 2017), and what factors (e.g., musical experience, phonological awareness) can contribute to improved learning of non-native tones paired with meanings (Wong et al., 2007; Wang et al., 2003; Kaan et al., 2008; Showalter & Hayes-Harb, 2013; Bowles et al., 2016; Shen & Froud, 2019). For instance, Wong & Perrachione (2007) found that naive English listeners were able to learn to identify nonwords contrastive in Mandarin tones in a picture selection task after receiving training (three to four sessions per week). Zou (2017) reported that Dutch L2 advanced learners of Mandarin showed an improvement in identifying Mandarin words compared with beginning learners. The above mentioned studies suggest that training or natural L2 acquisition can enhance non-tone language listeners' use of non-native lexical tones when building lexical representations. However, they centred on non-naive listeners who had received either short-term training or long-term exposure to the non-native pitch contrasts. The process of such training involves both internal and external variables (e.g., motivation and learning strategy) that may influence listeners (Orie, 2006; Moyer, 1999; Obler, 1989). How, then, do naive listeners at the very earliest stage of acquisition, who have not been influenced by other factors, encode non-native pitch contrasts when building sound-to-meaning associations in word learning? To what extent does one's native language play a role in word learning?

There seems to be a lack of investigation into how one's native language may modulate naive listeners' encoding of non-native tones in building sound-to-meaning mappings. Only very few studies exist; for instance, a recent study conducted by Ramachers et al. (2017) investigated whether naive Dutch listeners could attend to pitch differences when learning novel words in Limburgian, a dialect with a restricted tonal system consisting of a binary tonal contrast of accent 1 HL and accent 2 LHL[22] (Gussenhoven, 2000a, 2000b). The authors applied a name-labeling paradigm with eye-tracking (Quam & Swingley 2010; Singh et al., 2014) in which listeners had to choose the corresponding object (presented in animation) when hearing the target word. Their findings showed that Dutch listeners could overall recognize target word-object associations when hearing correctly pronounced words. However, when the word was mispronounced in another accent, the participants tended to look at the target word-object instead of looking at the distracter, indicating that they neglected the pitch changes during word learning in Limburgian. Their insensitivity to non-native pitch contrasts in word learning can be accounted for by a lack of lexical tones in Dutch (Ramachers et al., 2017). The findings in Ramachers et al. (2017) suggest that a lack of lexical tones in the native language may impede naive listeners from encoding non-native tones in sound-to-meaning mapping.

In contrast, two recent studies have shown that prior experience with lexical tones can benefit tone language listeners when mapping novel tones to meanings in word learning. For example, Poltrock et al. (2018) conducted an eye-tracking study to investigate Mandarin and French listeners learning minimal pairs of disyllabic pseudo-words in comparison with Cantonese native listeners as controls. Listeners had to learn the word (non-native tone)-object (fantasy object) associations and look at the correct object after each word they heard. They observed that Mandarin listeners outperformed French listeners with respect to learning tonal contrasts, which suggests that the phonological representations of tones in one's native language may facilitate the establishment of non-native tone-to-meaning mapping. Cooper & Wang (2013) compared naive Thai listeners, naive English listeners, and English listeners with previous Cantonese tone training to learn five monosyllabic

[22] Any primary stressed bimoraic syllable in Limburgian carries either accent 1 HL or accent 2 LHL (Gussenhoven, 2000b).

real words (word-object associations) differing in five Cantonese tones (high-level, high-rising, low-falling, low-rising and low-level). They found that after learning, Thai listeners and the tone-trained English listeners performed equally well, better than non-tone trained English listeners, in a picture matching task where they had to choose the picture corresponding with each word they heard. Thai listeners were especially good at identifying words carrying Cantonese high-rising, high falling, and high-level tones, which could be due to similar tonal counterparts in Thai. This suggests that tone language listeners may have relied on specific lexical pitch representations in the native language to build non-native tone-to-meaning mapping in word learning (Cooper & Wang, 2013).

The abovementioned studies have shed some light on the influence of native word prosody, whether facilitation or interference, on the encoding of non-native pitch contrasts when naive listeners are establishing lexical representations. In addition to the influence of native word prosody, it has also been shown that prosodic features at the *utterance level* come into play in building sound-to-meaning associations in word learning. Braun et al. (2014) examined Japanese, German, and French listeners as well as Mandarin controls learning non-words that differed in Mandarin tones with paired objects. They used two disyllabic non-words with four tones (Mandarin high-level, rising, dipping, and falling tones) and rising tone fixed on the second syllable. They found that German listeners outperformed French and Japanese listeners in judging whether the picture matched the word in a "tonal-mismatch condition" where the word did not match the object. Interestingly, Japanese listeners whose native language has a restricted tone system did not differ from French listeners who do not use any prosodic cues lexically (Rossi, 1980; Di Cristo, 1988; Vaissière, 1991), both showing no sensitivity to the tonal contrasts in lexical encoding. According to Braun et al. (2014), the differences between German listeners and French and Japanese listeners in word learning can be attributed to the inventory of pitch contrasts at the utterance level. German has a richer inventory of pitch accent types than Japanese and French. French and Japanese listeners relied little on pitch variation to signal post-lexical contrasts (Abe, 1998; Post, 2000; Turco et al. 2012; Asano & Braun, 2011), which can account for their limited ability to encode non-native tonal contrasts in the mental lexicon. Braun et al. (2014) argue that it is the prosodic features at the *utterance level* in the native language rather than the *word level* that can be beneficial for building non-native tone-to-meaning mapping at the lexical level.

Previous studies, as discussed above, have underscored the influence of native prosody at the word level and/or at the utterance level on the encoding of non-native pitch contrasts in building mental representations in sound-to-meaning mapping. However, these studies have mainly focused on the encoding of non-native tones in monosyllables when building sound-to-meaning associations. Comparatively, another word prosodic cue, *position*, has been given less attention. How will listeners with the use of positional marking in the native language perform in encoding non-native lexical pitch (tones) occurring in different positions in word learning? What about listeners whose native language does not use positional marking but does make use of lexical pitch?

Studies on the encoding of position at the lexical level have mainly investigated spoken word recognition, including investigation into whether listeners who use positional marking, e.g., Japanese and Dutch listeners, can exploit positional cues in native word recognition (for Japanese: Otake & Cutler, 1999; Sekiguchi & Nakajima, 1999; Shibata & Hurtig, 2008; Goss, 2015; for Dutch: Van Heuven, 1985; Van Heuven, 1988; Jongenburger & Van Heuven, 1995a; Jongenburger & Van Heuven, 1995b; Van Leyden & Van Heuven, 1996; Jongenburger, 1996; Van Kuijk et al., 1996; Quené & Koster, 1998; Cutler & Koster, 2000; Koster & Cutler, 1997; Cutler & Van Donselaar, 2001; Van Donselaar et al., 2005) and how native prosody (with or without positional marking) affects the processing of non-native positional cues (e.g., lexical stress) in L2 word recognition (Archibald, 1997; Cooper et al., 2002; Zhang & Francis, 2012; Connell et al., 2018). Findings from these studies show that listeners with lexical positional marking, i.e., Dutch and Japanese listeners, can exploit positional marking in lexical activation in their native languages (Dutch: Van Heuven, 1988; Cutler & Donselaar, 2001; Japanese: Otake & Cutler, 1999; Sekiguchi & Nakajima, 1999). Dutch listeners were found to benefit from the native positional marking to recognize words signaled by non-native positional cues (i.e., English) in L2 studies (Cooper et al., 2002). Moreover, there is evidence showing that listeners who do not use position but do use tones lexically in the native language, e.g., Mandarin listeners, were able to employ positional information (stressed vs. unstressed, without vowel reduction) in L2 word recognition in English (Connell et al., 2018). These studies have centred on L2 encoding of positional cues in word recognition, and mainly on L2 learners who have already achieved an advanced level which can be regarded as an “end-state” of acquisition. However, it remains unclear how naive listeners with or without positional cues encode non-native tones in different positions in building lexical representations in word learning. More specifically, it is unknown what role native word prosody plays in the lexical encoding of non-native tones in different positions, i.e., whether listeners with the use of positional marking will draw upon positional sensitivity to encode non-native tones at specified positions and whether listeners without the use of positional marking but with lexical pitch will be impeded in the encoding of non-native tones occurring in different positions.

Taken together, the goal of the present study is to explore, from a cross-linguistic perspective, to what extent native word prosody (and/or intonation prosody if there is any) comes into play in the encoding of non-native tones occurring in different positions in sound-to-meaning mapping in word learning by addressing the following questions:

1) Will native word prosody (and/or native intonation system if there is one) facilitate the encoding of non-native WPCs in building lexical representations in word learning? How will the three groups differ in word learning?

2) What language specific patterns will Mandarin, Japanese, and Dutch listeners show with respect to the encoding of non-native WPCs (non-native tones occurring in different positions position) in word learning? Will they share commonalities in sound-to-meaning mapping?

3) Will position influence, i.e., interfere or enhance the encoding of, non-native WPCs in word learning for each language group? To what extent will position interact with the processing of non-native pitch contrasts in word learning?

To that end, a picture selection task will be used in which listeners have to first learn sets of words (non-native tones, i.e., level or contour tones occurring on the initial or final syllable in disyllabic non-words) paired with fantasy pictures and later choose the correct pictures corresponding to the words they hear. Empirical predictions were made by and large according to the same hypotheses used in the previous two chapters (perception of non-native WPCs at the acoustic and phonological levels, respectively). That is, the Functional Pitch Hypothesis (Schaefer & Darcy, 2014), the Feature Hypothesis (McAllister et al., 2002), and an approach based on perceptual weighting of pitch features in long-term memory storage (Francis et al., 2008; Lee et al., 1996; Xu et al., 2006) were proposed not only to account for non-native speech perception but also for L2 tone acquisition which entails that these hypothesis may be carried over to processing involving phonological and lexical information (see §3.1 for details). Moreover, the above-mentioned hypotheses were proposed and attested in terms of the native phonology, specifically, word prosody. However, Braun et al. (2014) proposed that native intonation prosody, rather than native word prosody, would influence the encoding of non-native tones in building lexical representations in word learning. For this reason, we will make predictions based on these hypotheses in light of both native word prosody and native intonation prosody. Corresponding to the research questions, the predictions for processing non-native tones occurring in disyllables are as follows:

1a) Based on the Feature Hypothesis (McAllister et al., 2002), Japanese listeners (who use position of pitch accent lexically) will overall outperform Dutch listeners (who use position of lexical stress) and Mandarin listeners (who use lexical tones) in the encoding of non-native tones in establishing sound-to-meaning mapping in word learning.

1b) Alternatively, based on the Functional Pitch Hypothesis (Schaefer & Darcy, 2014), Mandarin listeners and/or Japanese listeners will encode non-native tones better than Dutch listeners in word learning.

1c) The influence of intonation prosody (Braun et al., 2014) proposes that a larger repertoire of pitch contrasts at the utterance level will predict better encoding of non-native tones in word learning. It is notable that cross-linguistic comparison among the size of pitch contrasts in intonation in Braun et al. (2014) is made on the grounds of ToBI (Tone and Break Indices) (Pierrehumbert & Beckman, 1986). Hence the prediction made here will follow the ToBI convention.

Table 4.1 shows the pitch contrasts at the utterance level in the three language groups. According to Dutch-ToBI (ToDI) (Gussenhoven, 2005), the Dutch intonation system has a rich inventory of pitch contrasts. It has three initial boundary tones (%L, %H, and H%L), three final boundary tones (L%, H%, and %), and eight pitch accents (nuclear tones) (H*L, !H*L, H*, H*!H, L*H, L*, L*HL and L*!HL) (Gussenhoven, 2005). In contrast, Japanese uses one initial boundary tone (%L), two final boundary tones (L%, HL%), and one pitch accent (H*L) (Pierrehumbert & Beckman, 1988). Note that pitch contrasts in intonation in Braun et al. (2014) include pitch accent types, initial boundary tones, final boundary tones, initial phrase accents, and final phrase accents. ToDI differs from the two-phrase intonation structure (intermediate phrase and intonational phrase) proposed in Beckman & Pierrehumbert (1986) and J-ToBI for Japanese (Pierrehumbert & Beckman, 1988). It uses only intonational phrase structure (see Gussenhoven & Rietveld, 1992; Van den Berg et al., 1992; Gussenhoven, 2005 for details) and thus the pitch contrasts of initial phrase and final phrase boundary tones are not compared here.

The ToBI for Mandarin is complex, with more structures (eight tiers instead of two-phrase structures) (Li, 1997; Li et al., 1999; Li, 2002; Peng et al., 2005) compared with ToBI. Despite some controversy regarding M-ToBI (Liu, 2009), according to M-ToBI (Peng et al., 2005) Mandarin has two final boundary tones (L%, H%), four initial boundary tones (%reset, %q-raise, %e-prom, %compressed),[23] and the nuclear tones are four lexical tones plus one neutral tone. Comparing the inventory of pitch contrasts in intonation among the three language groups, Mandarin and Dutch each has a larger size than Japanese. This as a result predicts that Mandarin and Dutch listeners will encode non-native tones occurring in different positions better than Japanese listeners.

Table 4.1. *Pitch contrasts in intonation in each language.*

	Dutch (ToDI) (Gussenhoven, 2005)	Japanese (J-ToBI) (Pierrehumbert & Beckman, 1988)	Mandarin (M-ToBI) (Peng et al., 2005)
Pitch accent types (nuclear tones)	8 (H*L, !H*L, H*, H*!H, L*H, L*, L*HL, L*!HL)	1 (H*L)	5 (H, LH, L, HL, neutral tone)

[23] In Li (2002), Mandarin has three initial boundary tones, %d, %r, %l (cf. Li, 1997; Li et al., 1999).

Initial boundary tones	3 (%L, %H and H%L)	1 (%L)	4 (%reset, %q-raise, %e-prom, %compressed)
Final boundary tones	3 (L%, H%, %)	2 (L%, HL%)	2 (L%, H%)

2) Based on the proposal of relative perceptual weighting of tonal features (Francis et al.; Lee et al., 1996; Xu et al., 2006), listeners store long-term memory representations of pitch-based phonological categories (tone and/or intonation) in the native language. These stored pitch representations in the native language determine listeners' assignment of relative weight to specific perceptual features of tones in cross-linguistic tone perception and tone acquisition. Accordingly, Mandarin listeners are predicted to be better at encoding non-native contour tones than level tones in word learning since Mandarin uses contrastive contour tones (three contour tones) but not contrastive level tones (only one level tone). In contrast, Dutch does not use lexical tones. The pitch-based phonological categories Dutch listeners store long-term are intonation patterns. Specifically, Dutch listeners may have stored intonational nuclear tones consisting of H and L tones (Gussenhoven, 2004, 2005), which may cause them to be more sensitive when encoding non-native level tones compared to non-native contour tones in word learning. Japanese listeners are predicted to encode non-native level tones better than contour tones because the accentual patterns in Japanese prosodic words (determined by the presence/absence and the location of the pitch accent) consist of H and L tones (Poser, 1984; Vance, 1987; Kubozono, 2008; Kawahara, 2015).

3) Mandarin does not use contrastive positional cues but does use lexical tones at the word level. Due to a lack of positional marking in the native language, Mandarin listeners may solely make use of native tonal categories to encode non-native tones in word learning. Thus, Mandarin listeners are predicted to show no preference for any position when lexically encoding non-native tones.

In contrast, Dutch does not use lexical tones but does use word stress for lexical contrast. The default position for lexical stress in Dutch is prefinal, which corresponds with the initial position in disyllabic words (Kager, 1989; Trommelen & Zonneveld, 1999). However, final stress occurs in Dutch disyllables as well. Thus, Dutch listeners are predicted to prefer word-initial position to word-final position in processing non-native tones in word learning.

Japanese uses the presence/absence and the location of pitch accent at the word level. *n*-mora words theoretically have *n+1* possible accentual patterns (McCawley, 1968; Akinaga, 1985; Haraguchi, 1999; Uwano, 1999). It is observed that all accentual patterns exist in bimoraic words, among which the initial-accent is prevalent in 2-mora words in Japanese (Sato, 1993; Kitahara, 2001). Due to the influence of native accentuation, Japanese listeners are predicted to favour word-initial position over word-final position in word learning.

4.2. Method

4.2.1 Subject

The participants were the same as in Chapters 2 and 3. Dutch and Mandarin participants participated in the present study nearly nine months after the phonological experiment in Chapter 3. Japanese participants participated in the study nearly one and a half months after the previous one.

4.2.2 Stimuli

In Chapter 3, we used Cantonese tonal contrasts T3-T6 (pitch level contrast), T4-T5 (pitch contour contrast), and T4-T6 (pitch contour vs. pitch level) (see Figure 4.1 for Cantonese tones) to investigate the perception of non-native pitch contrasts in different positions at the phonological level. The present study will continue to use T3, T4, T5, and T6, two contrastive level tones and two contrastive contour tones.

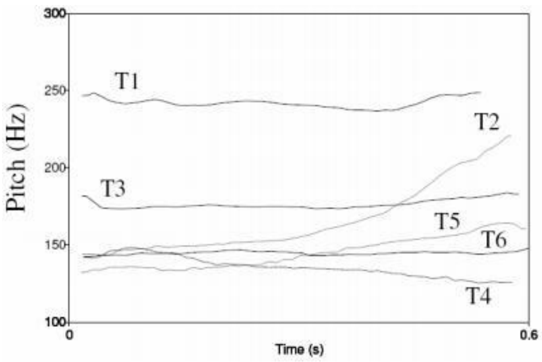


Figure 4.1. *Cantonese tones (Qin & Mok, 2013).*

Two monosyllables /ku/ and /pi/ carrying the Cantonese tones T3, T4, T5 and T6 were produced in isolation in Praat (Boersma & Weenink, 2008) six times each by a female native speaker of Cantonese who is a well trained phonetician. Three tokens of the best quality for each tonal syllable were selected and all of the tokens were adjusted to 300 ms with 70 dB in Praat. The monosyllables were concatenated to the disyllabic non-word /kupi/ (625 ms in total, 25ms interval of silence, see Appendix A for naturalness rating), a word that does not exist in any of the three languages. T4, T5, and T6 were the target tones to be investigated in the study. T3 (mid level tone) was used as the contextual tone as in Chapters 2 and 3, since T1 (high level tone) as the contextual tone was more acoustically salient than the other tones (Qin & Mok, 2013; Wong, 2019) while T6 (low level tone) as the contextual tone caused perceptual confusion when the target tone was T5 as reported in Chapter 2. All of the T4, T5, and T6 tokens occurred in word-initial (first syllable) and word-final (second syllable) positions. In total, the non-word /kupi/ carried six target tonal patterns: T4T3, T5T3, T6T3, T3T4, T3T5, and T3T6. In addition, /kupi/ with a T3T3 tonal pattern was used as a "baseline" word since it contained the same two level tones, which had been shown as the easiest to learn in a pilot study. Moreover, T3T3 could form pitch level contrast with T6T3 or T3T6, i.e., T3-T6 contrast in word-initial and -final position, respectively. It could also form pitch level vs. pitch contour contrast with T4T3, T5T3, T3T4 and T3T5, i.e., T3-T4 contrast and T3-T5 contrast in word-initial and -final position, respectively.

The stimuli design is shown in Table 4.2. The seven words were contrastive either in pitch or in position and each word was paired with a picture of a fantasy object, as shown in Table 4.3. We didn't use pictures of real objects so as to avoid interference (if any) from words representing the real objects in the native languages.

Table 4.2. *Stimuli: Non-words in tonal patterns paired with fantasy objects*

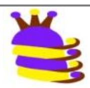
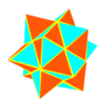
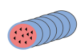
word /kupi/	**tonal pattern**	**presented with fantasy object**
Word 1	T4T3	
Word 2	T5T3	
Word 3	T6T3	
Word 4	T3T4	
Word 5	T3T5	
Word 6	T3T6	
Word 7	T3T3	

Table 4.3. *Non-native pitch contrasts and positional contrasts formed by stimuli*

<table>
<tr><th>Pitch contrasts</th><th>Tonal pattern</th><th>Positional contrasts</th><th>Tonal pattern</th></tr>
<tr><td rowspan="2">Pitch contour contrast</td><td>T4T3 vs. T5T3</td><td rowspan="12">Word-initial vs. word-final</td><td rowspan="12">T4T3 vs. T3T4
T5T3 vs. T3T5
T6T3 vs. T3T6</td></tr>
<tr><td>T3T4 vs. T3T5</td></tr>
<tr><td rowspan="8">Pitch contour vs. pitch level contrast</td><td>T4T3 vs. T6T3</td></tr>
<tr><td>T4T3 vs. T3T3</td></tr>
<tr><td>T5T3 vs. T3T3</td></tr>
<tr><td>T5T3 vs. T6T3</td></tr>
<tr><td>T3T4 vs. T3T3</td></tr>
<tr><td>T3T4 vs. T3T6</td></tr>
<tr><td>T3T5 vs. T3T3</td></tr>
<tr><td>T3T5 vs. T3T6</td></tr>
<tr><td rowspan="2">Pitch level contrast</td><td>T6T3 vs. T3T3</td></tr>
<tr><td>T3T6 vs. T3T3</td></tr>
</table>

4.2.3 Procedure

The experiment was programmed and conducted in ZEP (Veenker, 2012) on an experiment laptop. Dutch and Chinese listeners participated in the experiment in the phonetics lab at UiL-OTS. Japanese listeners participated in a sound-attenuated room in the Education and Research Centre in International Christian University in Tokyo.

The experiment contained five sessions: a learning session, a practice session, a pre-test session, a test session, and a post-test session. During the learning session, participants were instructed to learn seven "new words" (non-words) in a foreign language that differed only in "melody". The instructions were the same for all listeners and were presented in both English and the native language (See Appendix F for details.) The listeners were instructed to learn the words and the paired pictures together to form word-picture associations. Seven buttons, from 1 to 7, representing the seven words were shown on the screen, as shown in Figure 4.2. Each time a participant clicked on a button (using a mouse) the corresponding picture popped up. Participants could listen to the seven words as many times as they wanted until they were certain that they had learned the words with the paired pictures and were ready to proceed to the practice session. The learning session was self-paced. There was no time pressure in the learning session.

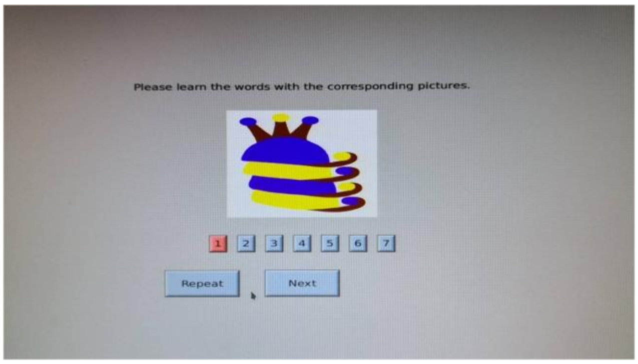


Figure 4.2. *Layout with stimuli presented in the learning session.*

In the practice session, the participants heard a word while simultaneously seeing a picture. They were required to judge whether the word and the picture matched or not. The practice session consisted of 21 trials, three repetitions of each word. All of the trials were randomized. The participants had to meet the criterion of making at least 13 out of 21 correct responses in total (over 60% accuracy rate) to proceed to the pre-test session. If they failed to reach this criterion, they were required to go back to the learning session to learn the words again (for a maximum of 6 times). The total number of times that each participant was trained in the practice session was recorded.

In the pre-test session, the participants heard a word while simultaneously seeing four pictures on the screen. They were required to choose the picture that corresponded to the word they heard by pressing the corresponding button on a four-button button box. They were told that they should respond only once they were certain instead of hurrying to respond. After the participants made a response, they automatically proceeded to the next trial. Feedback on whether or not the response was correct appeared immediately after each response. The correct answer was signalled by a check mark. If a participant responded incorrectly, the incorrect answer was signalled by a cross and the correct answer (the picture) would be shown in a red circle. The pre-test session contained 4 trials in total. The aim of the practice session was to give listeners an idea about the test session, and thus the four pictures displayed on the screen included one target word (the correct answer) and three other randomized pictures.

Following the pre-test session, the test session was administered. The procedure of the test session was the same as the pre-test session. Participants had to choose the correct picture (out of four) that paired with the word they heard simultaneously. However, no feedback was provided after their response in the test session. They were told to make the correct response as soon as possible. The four pictures were always presented to contain one target picture (the target tone) representing the correct answer, one tonal competitor (contrastive in pitch), one positional competitor (contrastive in position), and one distracter that was neither contrastive in position nor in pitch. Take the target word /kupi/ carrying tonal pattern T3T4 (T4 on the word-final position), for example. The picture representing T3T4 could appear together with one tonal competitor (T3T5, contrastive in pitch contour with T3T4), one positional competitor (T4T3, contrastive in position with T3T4), and one distracter (T6T3, neither contrastive in position nor pitch), as displayed in Figure 4.3 (see Appendix E for detailed stimuli list). The order of the pictures placed on the screen was counterbalanced. The seven non-words were tested three times in a semi-randomized order with exclusion of identical sequences, resulting in 21 trials in total.

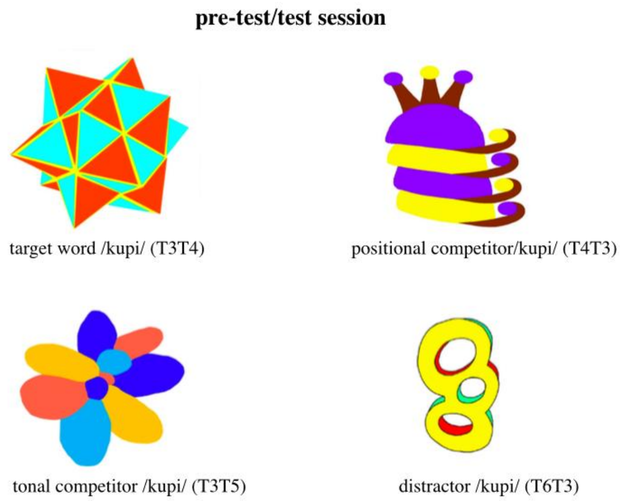


Figure 4.3. *Layout with stimuli presented in the pre-test/test session.*

In the post-test session, a judgment task similar to the one used in the practice session was administered. The participants heard a word and simultaneously a picture popped up on the screen. They were required to judge whether the word they heard matched the picture or not, without time pressure. The aim of the post-test was a "double-check" to make sure that they had managed to learn the words.

In addition, two types of working memory capacity tests, a backward digital span test and a phonological memory test, were conducted immediately after the experiment. Both the tests were presented auditorily. The digital span test consisted of two sets of numbers from a string of 2 digits to 8 digits (see Appendix D for details). Participants were required to repeat the string of the digits in backward order. The phonological memory test contained two sets of syllables from a string of two to seven syllables (see Appendix D for details). Participants were required to repeat the syllables in the string in the correct order. All of the numbers and syllables were the same across participants. All of the numbers were recorded in the participants' native language by three female native speakers of Dutch, Mandarin, and Japanese, respectively. The interval between each number and each syllable was

1s. The syllables were produced with a flat tone by a female native speaker of Cantonese. All of the speakers are well-trained phoneticians.

4.3 Results and discussion

4.3.1 Performance across groups in the picture selection task

4.3.1.1 Performance of word identification

Correct and incorrect responses were coded as 1 and 0, respectively. To compare word identification performance across the three groups, a generalized Linear Mixed Model (GLMM) was conducted in SPSS 25. Word (7 levels: 7 words) and Language (3 levels) were taken as fixed factors. Participants and items were added to the model as random effects (random intercepts). Moreover, the number of times needed by the participants in the practice session to proceed to the pre-test session was recorded in Table 4.4. It can be seen that the number of times to meet the criterion by Dutch participants varied from one to four times. Half of the Dutch participants needed two times to proceed to the test phase, while the majority of Japanese and Mandarin participants needed only one time. Performances on the working memory test, the backward digital span test, and the phonological memory test, respectively, were calculated in each language group, as reported in Table 4.5. We were not interested in how the language groups differed with respect to the memory test or the time spent on the practice session, but instead whether the two factors, i.e., the number of times necessary to proceed to the pre-test session and the results of the memory test, correlated with their performance on the word identification task. Therefore, the number of times needed in the practice session and the result of the memory capacity test (both backward digit span and phonological memory) were taken as covariates in the GLMM.

Table 4.4. *Number of times needed in the pre-test to proceed to the test phase in each language group*

Number of participants in each language group	Number of pre-test times needed to proceed to the test phase			
	one	two	three	four
Dutch	9	15	5	1
Japanese	22	6	2	
Mandarin	26	4		

Table 4.5. *Working memory capacity test for each language group.*

	Backward digit span		
Language group	Mean accuracy	Std. deviation	Std. Error
Dutch	82.13%	0.079	0.145
Japanese	81.77%	0.112	0.020
Mandarin	94.47%	0.066	0.012
	Phonological memory		
Language group	Mean accuracy	Std. deviation	Std. Error
Dutch	72.33%	0.087	0.016
Japanese	85.90%	0.060	0.011
Mandarin	80.53%	0.094	0.017

The F-tests showed that both Word (F (6, 1,869) = 6.694, p < 0.001) and Language (F (2, 1,869) = 15.369, p < 0.001) had significant main effects, as presented in Table 4.6, indicating there were differences across words and across groups. No interaction between Word and Language was found (F (12, 1,869) = 1.478, p = *0.125*), suggesting that Word performance did not differ depending on the Language groups.

Table 4.6. *Parameters of fixed effects and random effects in word identification.*

Fixed effects		F	df	p
Language group		15.369	2	<0.001
Word		6.694	6	<0.001
Language group * Word		1.478	12	0.125
Covariates				
Backward digit span		0.007	1	0.314
Phonological memory		1.014	1	0.936
Number of pre-test time needed to proceed to the test phase		0.185	1	0.668
Random effects	Est.	Std. Error	z	p
1 | Participant	1.045	0.219	4.769	<0.001
1 | Item	0.018	0.066	0.278	0.781

Figure 4.4 displays the performance of each language in each word. All three groups performed above chance level (1/4, 25%). Overall, Dutch listeners performed significantly worse than Mandarin (F (2, 1,869) = 14.479, p < *0.001*) and Japanese

listeners (F (2, 1,869) = 14.479, $p < 0.001$), while the latter two groups did not differ from each other (F (2, 1,869) = 1.957, $p = 0.838$). Word 7 (T3T3) showed an overall significantly better accuracy than the other words (F (6, 1,869) = 7.082, all p values <0.05) except for in the comparison to Word 4 (T3T4) (F (6, 1,869) = 1.313, $p = 0.102$). Despite there was no significant interaction found between Language and Word, GLMM analysis was conducted for each language group for exploratory purposes to examine how the groups performed regarding specific words. Dutch listeners was found to perform significantly better at identifying Word 4 (T3T4) (nearly 58.2% accuracy) and Word 7 (T3T3) (71.25% accuracy) compared to the other words, with accuracy around 40% to 50% (F (6, 623) = 5. 929, all $p < 0.05$). Japanese listeners showed a similar performance pattern to Dutch listeners. They had comparable performance identifying Word 4 (85.9% accuracy) and Word 7 (89.0% accuracy), which was significantly better than the other words (F (6, 623) = 6.504, all $p < 0.05$) (except for Word 2, F (6, 623) = 1.917, p = 0.122).

Mandarin listeners showed an overall performance with around 80% accuracy learning all words except for Word 6 (T3T6), for which accuracy dropped to 65.2%. Unlike Dutch and Japanese listeners, Mandarin listeners had more difficulty identifying Word 6 (T3T6) than the other words (F (6, 623) = 4.763, all $p < 0.05$), except for Word 3 (T6T3) which was just as troublesome as Word 6 (F (6, 623) = 1.153, p = 0.184). Word 7 (T3T3) was only advantageous over Word 6 (T3T6) (F (6, 623) = 3.431, $p = 0.022$) to Mandarin listeners when learning sound-to-meaning associations.

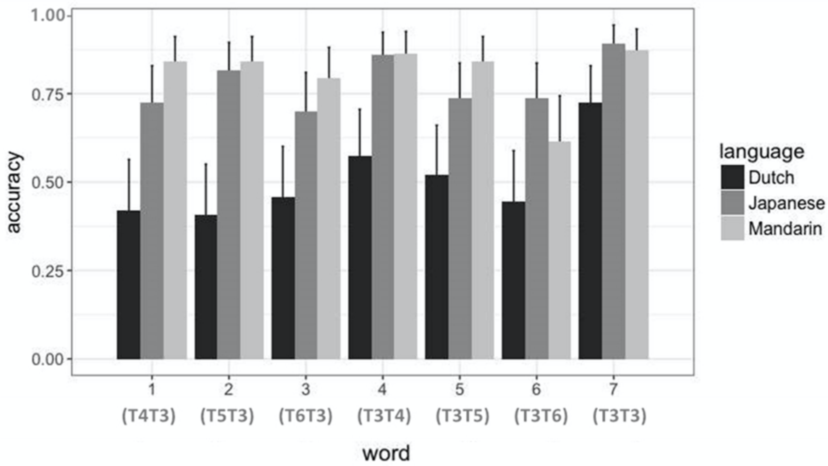


Figure 4.4. *Performance of each language group in word identification. Error bar: 95% CI.*[24]

[24] The error bar in the dissertation refers to 95% CI.

Moreover, the backward digital span test (F (1, 1,881) = 0.007, $p = 0.314$) and the phonological memory test (F (1, 1,881) = 1.014, $p = 0.936$) were both found not significant. Language did not interact with either the phonological memory test (F (2, 1,881) = 0.322, $p = 0.725$) or the backward digital span test (F (2, 1,881) = 0.006, $p = 0.994$), suggesting that the working memory capacity did not influence listeners' performance in word learning. The number of practice session times that the listeners attended was found insignificant (F (1, 89.88) = 0.185, p = 0.668) and it did not interact with Language (F (3, 89.88) = 0.281, p = 0.286). This indicates that the number of times participants attended the practice session to meet the passing criterion did not correlate with their performance in word learning.

To summarize, the three languages groups overall differed in their word identification performance. Mandarin and Japanese listeners performed comparably well and both outperformed Dutch listeners with respect to identifying novel words. Overall, Word 7 (T3T3) was the easiest to learn. The three groups also displayed specific learning patterns based on analysis of exploratory purposes. Mandarin listeners seemed struggle when learning Word 6 (T3T6) and Word 3 (T6T3), while they were good at identifying all the other words. In contrast, Dutch and Japanese listeners displayed a similar pattern in that both groups were better at identifying Word 7 (T3T3) and Word 4 (T3T4) compared to other words. In addition, neither the number of times they spent in the practice session before proceeding to the pre-test session nor their memory capacity had any correlation with their word identification performance. Note that the current results about the specific learning pattern of the words should be taken with caution, considering the interaction between word and language group was not significant.

4.3.1.2 Error types in word identification

To further disentangle the performance of each language group in word identification, we recorded the errors each participant made for each word and categorized the errors listeners made into three types: tonal error (choosing the wrong word contrastive in pitch), positional error (choosing the wrong word contrastive in position) and irrelevant error (choosing the wrong word that is irrelevant to the target response), as shown in Table 4.7. The transcribed error type of each participant for each word was analyzed in a generalized linear multinomial analysis in SPSS 25. Language (3 levels) was taken as fixed factor and error type was computed as dependent target in the model. The F-tests showed that Language was found significant (F (9, 1,887) = 39.921, $p < 0.001$), suggesting that language groups differ from each other in error types.

Table 4.7. *Error type in word learning.*

Error type		**Example** (e.g., target response T3T4)
Tonal error	Choosing the wrong word contrastive in pitch	T3T5
Positional error	Choosing the wrong word contrastive in position	T4T3
Irrelevant error	Choosing the wrong word that is irrelevant to the target response	T6T3

As shown in Figure 4.5, Mandarin listeners displayed 12.97% tonal errors, 4.92% positional errors, and 1.22% irrelevant errors (among a total error rate of 19.12%). Japanese listeners had 13.40% tonal errors, 5.02% positional errors, and 1.40% irrelevant errors (among a total error rate of 19.82%). The two groups did not differ significantly from each other in all the error types (tonal error type: F (2, 1,887) = 9.686, $p = 0.867$; positional error type: F (2, 1,887) = 2.183, $p = 0.960$; irrelevant error type: F (2, 1,887) = 3.144, $p = 0.729$). In contrast, Dutch listeners made more errors than Mandarin and Japanese listeners, with 30.81% tonal errors, 9.64% positional errors and 5.05% irrelevant errors (with a total error rate of 45.51%). Specifically, Dutch listeners were observed to make significantly more tonal errors than Mandarin (F (2, 1,887) = 9.686, $p < 0.001$) and Japanese listeners (F (2, 1,887) = 9.686, $p < 0.001$), suggesting that they were more confused by the non-native pitch contrasts and had more trouble encoding non-native tones, compared with Japanese and Mandarin listeners. Moreover, Dutch listeners made significantly more irrelevant errors than Mandarin (F (2, 1,887) = 3.144, $p = 0.013$) and Japanese listeners (F (2, 1,887) = 3.144, $p = 0.016$). This indicates that, in comparison with Mandarin and Japanese listeners, they showed more tendency to choose the wrong word that was irrelevant (neither contrastive in tone nor position) to the correct answer. The three groups didn't show any difference with respect to positional errors (F (2, 1,887) = 2.183, $p = 0.051$). Note that since the p value was marginal, we are unable to draw the conclusion that the three groups differed with respect to making positional errors.

Furthermore, as can be seen in Figure 4.5, within each language group, tonal errors were made significantly more than the other two error types (Dutch listeners: F (2, 629) = 7.311, all $p <0.05$; Mandarin listeners: F (2, 629) = 3.358, all $p < 0.05$; Japanese listeners: F (2, 629) = 3.050, all $p < 0.05$), suggesting that each language group was more confused by the tonal competitor than the positional competitor or distracter when identifying words.

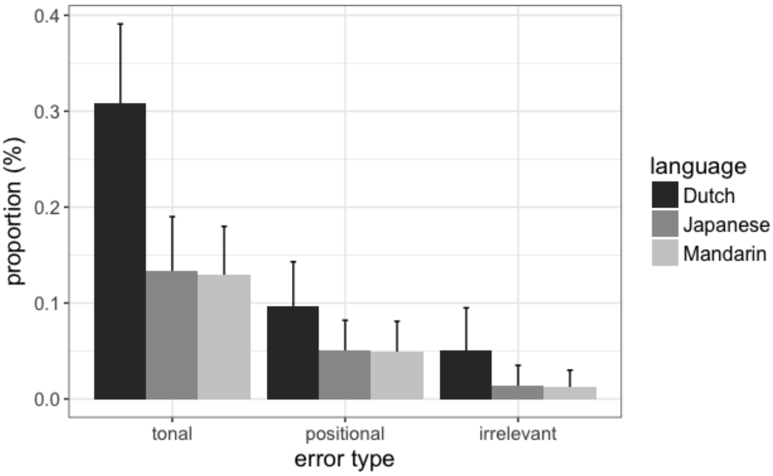


Figure 4.5. *Error types for each language group in word learning.*

Table 4.8 additionally provides a detailed scenario of the confusion in word identification for Dutch, Mandarin, and Japanese listeners based on their error responses. In each language group, the leftmost column contains the target words (the correct response); the top row refers to the words chosen as the wrong answer (error responses). The figure in each cell represents the proportion of error responses among all errors made for the word.

As can be seen in Table 4.8, in general, Dutch listeners tended to misidentify Word 1 (T4T3) as Word 3 (T6T3) and Word 4 (T3T4) as Word 6 (T3T6). They mostly confused Word 2 (T5T3) with Word 1 (T4T3), Word 5 (T3T5) with Word 1 (T3T4), and Word 3 (T6T3) with Word 7 (T3T3), while they misidentified Word 5 (T3T5) mostly as Word 2 (T5T3) and Word 6 (T3T6) as Word 3 (T6T3). From the misidentified words we see that Dutch listeners were most confused by the non-native tonal contrast T4-T6, T4-T5, and T3-T6 or otherwise the position where the non-native tones occurred (e.g., T3T4 vs. T4T3). Moreover, Table 4.8 provides details regarding the irrelevant error types Dutch listeners made. Apart from the tonal and positional errors they made, the proportion of the error response within the irrelevant errors seems to be random.

Unlike Dutch listeners, Mandarin listeners made very few irrelevant error types. Among the overall low error rates for each word (around 20%) presented in Figure 4.5, Mandarin listeners were mostly confused by the words that were contrastive either in pitch or in position. For instance, they mostly misidentified Word 1 (T4T3) as Word 3 (T6T3), Word 2 (T5T3) as Word 1 (T4T3), Word 5 (T3T5) as Word 4 (T3T4), Word 3 (T6T3) as Word 7 (T3T3) (tonal confusion), and Word 2 (T5T3) as Word 5 (T3T5) (positional confusion).

Japanese listeners showed a similar pattern to Mandarin listeners with respect to word confusion. They mostly misidentified the word as the counterpart contrastive in pitch contrast or in position. For instance, they misidentified Word 1 (T4T3) as Word 3 (T6T3) and Word 2 (T5T3), Word 2 (T5T3) as Word 1 (T4T3), Word 3 (T6T3) as Word 7 (T3T3), Word 5 (T3T5) as Word 4 (T3T4) (tonal error), Word 5 (T3T5) as Word 2 (T5T3) and Word 1(T4T3) as Word 4 (T3T4) (positional error).

In conclusion, among the errors the listeners made in word identification, all of the groups show more errors in tonal error type than positional and irrelevant error types. That is, they all tended to misidentify the target word as the counterpart that was contrastive in pitch. This suggests that for each language group, pitch was a relatively dominant factor that influenced listeners in word learning. For instance, Mandarin listeners were likely to misidentify Word 6 (T3T6) as Word 7 (T3T3); Dutch listeners showed tendency to misidentify Word 1 (T4T3) as Word 2 (T5T3) or Word 3 (T6T3); Japanese listeners confused Word 1 (T4T3) with Word 3 (T6T3).

Differences in error types were also found across language groups. Mandarin and Japanese listeners did not differ significantly from each other with respect to error types. In contrast, Dutch listeners made more tonal errors than Mandarin and Japanese listeners, suggesting that they were less capable of encoding non-native tones when establishing lexical representations. They were observed to produce more irrelevant errors than Mandarin and Japanese listeners as well, which indicates that, compared with Mandarin and Japanese listeners, they showed a stronger preference for a distracter that was neither contrastive in pitch nor in position to the target word when confused. This suggests that Dutch listeners encountered greater difficulty encoding non-native tones in building sound-to-meaning associations in word learning. Note that we are unable to conclude that Dutch listeners made more positional errors than Mandarin and Japanese listeners as the difference was marginal (p= 0.051).

Table 4.8. *Word identification confusion matrices for each language group.*

Dutch listeners

	Error response (%)						
Target	Word1 T4T3	Word 2 T5T3	Word 3 T6T3	Word 4 T3T4	Word 5 T3T5	Word 6 T3T6	Word 7 T3T3
Word 1 (T4T3)	correct	21.57	39.21	15.69	1.97	15.69	5.89
Word 2 (T5T3)	36.53	correct	30.77	5.77	26.92	1.92	0
Word 3 (T6T3)	25.50	17.31	correct	3.92	11.77	19.61	23.08
Word 4 (T3T4)	17.95	2.56	7.69	correct	38.46	25.64	7.69
Word 5 (T3T5)	9.30	27.91	9.30	30.23	correct	23.25	0

Word 6 (T3T6)	9.68	6.45	25.81	32.25	22.58	correct	3.22
Word 7 (T3T3)	36.36	9.09	18.18	9.09	9.09	18.18	correct

Mandarin listeners

Target	**Error response** (%) Word1 T4T3	Word 2 T5T3	Word 3 T6T3	Word 4 T3T4	Word 5 T3T5	Word 6 T3T6	Word 7 T3T3
Word 1 (T4T3)	correct	22.38	66.67	20.51	0	6.67	0
Word 2 (T5T3)	47.06	correct	0	0	52.94	0	0
Word 3 (T6T3)	0	0	correct	0	0	61.90	38.10
Word 4 (T3T4)	24	0	26	correct	10	40	0
Word 5 (T3T5)	0	50.82	0	49.18	correct	0	0
Word 6 (T3T6)	6.9	3.45	10.34	10.34	0	correct	68.97
Word 7 (T3T3)	0	0	50	0	0	50	correct

Japanese listeners

Target	**Error response** (%) Word1 T4T3	Word 2 T5T3	Word 3 T6T3	Word 4 T3T4	Word 5 T3T5	Word 6 T3T6	Word 7 T3T3
Word 1 (T4T3)	correct	17.86	64.29	10.71	0	7.14	0
Word 2 (T5T3)	52.63	correct	31.58	5.26	10.53	0	0
Word 3 (T6T3)	19.23	23.07	correct	0	11.53	19.23	26.92
Word 4 (T3T4)	38.46	0	15.38	correct	7.69	30.77	7.69
Word 5 (T3T5)	3.85	38.46	0	26.9	correct	30.77	0
Word 6 (T3T6)	4	4	24	20	20	correct	28
Word 7 (T3T3)	0	0	44.44	0	0	55.56	correct

4.3.1.3 Influence of tone and position on word identification

To better understand the influence of tone and position on performance among the three groups, the target tones in Words 1 to 6 were T4, T5, and T6, which can be categorized into 3 Tone categories. The target tones occurred word-initially and word-finally, hence positions where the target tones occurred can be categorized into 2 Position categories. Tone (3 levels: T4, T5, and T6), Position (2 levels: word-initial vs. word-final), and Language (3 levels) were taken as fixed factors, with participants and items as random effects (random intercepts) in GLMM. Note that T3 was not taken into the Tone category for analysis because the mid level tone T3 was not a target tone investigated in the current study but rather used as a contextual tone in the initial or final position of disyllabic non-words. T3 in the "baseline" word T3T3 could form a pitch level contrast (i.e., with low level tone T6 as T3-T6) or pitch level vs. pitch contour contrast (i.e., with contour tones T4, T5 as T3-T4, T3-T5).

Table 4.9 reports the parameters for fixed factors and random effects (intercepts). The F-tests show that Tone (F (2, 1,612) =39.557, $p<0.001$) and Language (F (2, 1,612) =16.694, $p<0.001$) are significant, indicating that overall performance with respect to tones differed across language groups. Position had no significant main effect (F (1, 1,612) =0.957, p=0.252), suggesting that there was no difference in overall performance with respect to word-initial and word-final position. Moreover, no interaction between Position and Language (F (2, 1,612) =2.181, $p=0.113$) or between Tone and Language (F (4, 1,612) =1.147, $p=0.333$) was found significant. A significant interaction was found between Tone and Position (F (2, 1,612) =3.998, $p=0.019$), suggesting there were differences across the identification of tones depending on positions.

Table 4.9. *Parameters of fixed factors and random effects with regard to tone, position, and language in word learning.*

Fixed effects	F	df	p	
Language group	16.694	2	<0.001	
Tone	39.557	11	<0.001	
Position	0.957	2	0.252	
Tone * Position	3.998	2	0.019	
Language * Tone	1.147	4	0.333	
Language * Position	2.181	2	0.113	
Random effects	Est.	Std. Error	z	p
1 | Participant	0.280	0.116	2.412	0.016
1| Item	0.563	0.250	2.253	0.024

Although there was no significant interaction among Language, Position and Tone, for exploratory purposes, GLMM analysis was conducted for each language with Tone (3 levels) and Position (2 levels) taken as fixed factors and participant and item as random effects. Figure 4.6 displays the performance of Dutch listeners identifying each tone in each position. As can be seen, Dutch listeners showed a comparable performance encoding non-native tones in each position, with accuracy around 40% to 50%. Tone (F (2, 534) = 0.42, $p=0.658$) and Position (F (1, 534) = 2.227, $p=0.134$) were found not significant. No interaction between Tone and Position was observed in Dutch listeners (F (2, 534) = 1.304, $p=0.272$), suggesting that position did not influence the encoding of non-native tones. Note that in Figure 4.6, accuracy identifying T4 in word-final position rose up to 57.4% from that in word-initial position with 42.5%. However, statistically speaking, the difference was marginal (p=0.055) and thus we are unable to draw a conclusion that Dutch listeners preferred word-final position over word-initial position when encoding T4.

Unlike Dutch listeners, Mandarin listeners showed a different pattern in encoding non-native tones in different positions, as shown in Figure 4.7. Tone (F (2, 534) = 3.716, $p = 0.025$) was found significant, suggesting that participants performed differently across non-native tones. To be specific, Mandarin listeners were worse at identifying T6 than T4 (F (2, 536) = 3.152, $p = 0.005$) and T5 (F (2, 536) = 3.091, $p=0.007$), while they had comparably good performance identifying T4 and T5, both with around 80% accuracy (F (2, 536) = 0.372, $p = 0.886$). Position was found insignificant (F (1, 534) = 0.730, $p = 0.393$) and did not interact significantly with Tone (F (2, 534) = 0.526, $p = 0.591$), suggesting that position did not influence Mandarin listeners' encoding of non-native tones lexically.

Figure 4.8 presents the performance of Japanese listeners' identification of each tone in each position. Different from Dutch and Mandarin listeners, for Japanese listeners Tone (F (2, 534) = 1.303, $p = 0.272$) and Position were found insignificant (F (1, 534) = 0.336, $p = 0.336$). However, Tone had a significant interaction with Position (F (2, 534) = 4.236, $p = 0.015$). Overall, Japanese listeners showed comparable performance identifying the non-native tones. Nonetheless, when identifying T4, the non-native falling tone, Japanese listeners preferred word-final position over word-initial position, with accuracy from 71% rocketing to 87%.

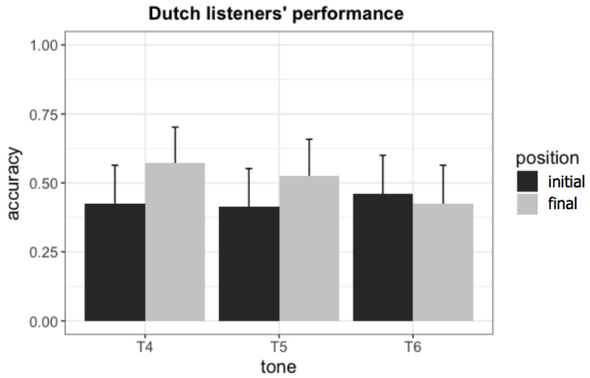


Figure 4.6. *Identification of each tone in each position by Dutch listeners.*

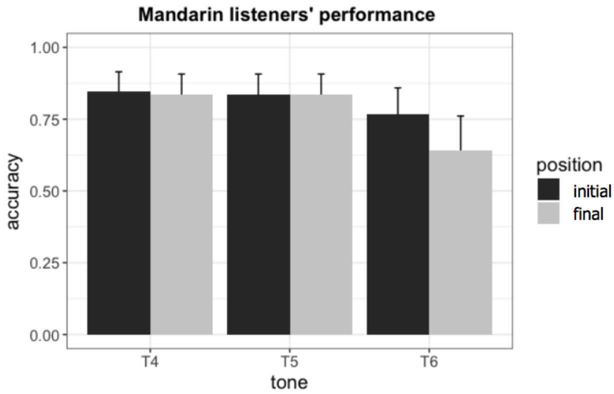


Figure 4.7. *Identification of each tone in each position by Mandarin listeners.*

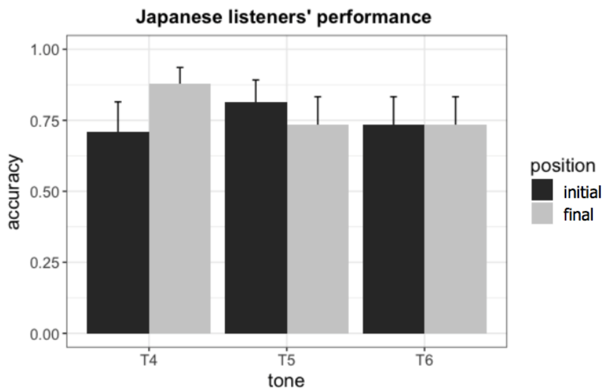


Figure 4.8. *Identification of each tone in each position by Japanese listeners.*

To sum up, Mandarin and Japanese listeners were overall better than Dutch listeners at encoding non-native tones occurring in different positions in word identification in the picture selection task. The three groups seemed to show specific patterns with regard to tone and position. Dutch listeners demonstrated comparable performance encoding all the non-native tones and they were not influenced by the position where the non-native tones occurred. Mandarin listeners were better at encoding the non-native contour tones T4 and T5 than the level tone T6. They were not influenced by position either. Differently, Japanese listeners showed a preference for position when encoding non-native tones lexically. That is, they more accurately identified T4 (the non-native falling tone) when it occurred word-finally than word-initially, while they performed comparably when identifying T5 and T6 in each position. Note that the current results with regard to the pattern of each language group should be taken cautiously, considering the interaction among tone, position and language group was not significant.

4.3.1.4 Reaction time

Participants made responses via a four-button button box, which allowed us to record the reaction time (RT) to evaluate performance making correct and incorrect answers across language groups. RT was calculated as the time from the end of the stimulus to the time when the response was made. Figure 4.9 depicts the RT for correct and incorrect responses for each language group. The average reaction times for a correct response in Dutch, Mandarin, and Japanese groups were 4856.13 ms, 4482.57 ms, and 3672.41 ms, respectively, while Dutch, Mandarin, and Japanese listeners took 5428.24 ms, 6538.60 ms, and 4711.94 ms, respectively, to make an incorrect response.

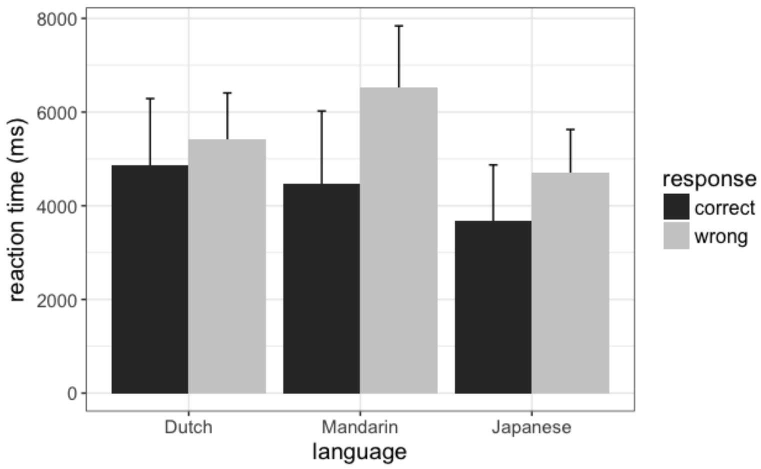


Figure 4.9. *Reaction time of each language group making correct and incorrect responses.*

Since the distribution of raw RT data was not normally distributed, we converted it logarithmically to get normally distributed data and analyzed them in a linear mixed effects regression model with Language (3 levels) and Result (2 levels: correct and wrong) as fixed factors and subject and item added as random effects. The F-tests showed that Language (F (2, 89.670) = 3.570, $p = 0.032$) and Result (F (1, 89.670) = 5.715, $p < 0.001$) had significant main effects, as reported in Table 4.10. Language interacted with Result (F (2, 89.670) = 23.001, $p < 0.001$), indicating that the differences between RTs of correct and incorrect responses depended on the language groups.

Table 4.10. *Parameters of fixed factors and random effects for reaction time.*

Fixed effects	F	df	p	
Language group	3.570	2	0.032	
Result	5.715	1	<0.001	
Language group * Result	23.001	2	<0.001	
Random effects	Est.	Std. Error	z	p
1 | Participant	0.163	0.028	5.844	<0.001
1| Item	0.227	0.008	26.579	<0.001

Figure 4.10 presents the RT of correct and incorrect responses for each language group (in logit). Post-hoc analyses indicate that Japanese listeners responded significantly faster than Dutch listeners (F (2, 89.628) = 4.315, $p = 0.011$) but comparable to Mandarin listeners (F (2, 89.628) = 3.570, $p = 0.072$) when making a correct response, while Mandarin listeners did not differ from Dutch listeners (F (2, 89.628) = 1.805, $p = 0.417$). In making an incorrect response, Mandarin listeners reacted significantly more slowly than Japanese listeners (F (2, 76.791) = 3.817, $p = 0.033$). There were no differences between Mandarin and Dutch listeners (F (2, 76.791) = 2.601, $p = 0.089$) or between Japanese listeners and Dutch listeners (F (2, 76.791) = 1.235, $p = 0.597$).

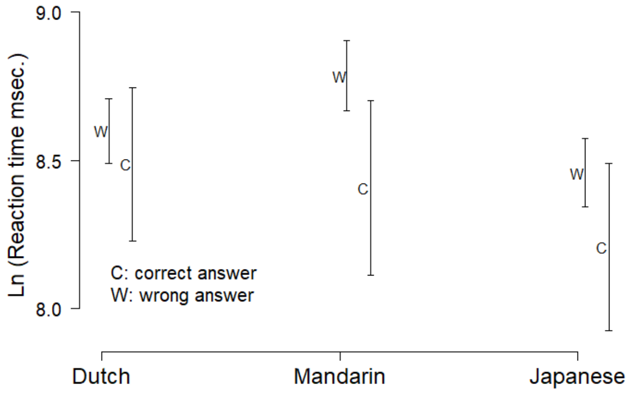


Figure 4.10. *Reaction time for correct and incorrect responses for each language group (in logit).*

The RTs when making correct responses observed for each language group were by and large in accordance with their performance reported in §4.3.1.1. Mandarin and Japanese listeners encoded non-native WPCs better than Dutch listeners in word identification with faster response times, while Dutch listeners had a lower accuracy and spent more time to make a correct response. This seems to suggest that lexical pitch in the native language facilitates Japanese and Mandarin listeners in building non-native pitch representations for word meanings. Notably, Mandarin listeners showed longer reaction time than Japanese listeners when making incorrect answers. Recall from the previous sections that the errors made by Mandarin listeners were mostly in identifying words contrastive in pitch level (T3-T6). Longer incorrect response time supports the conclusion that they experienced difficulty encoding non-native pitch level contrasts in building lexical representations.

4.3.2 Post-test judgment task

A post-test judgment task was administered so as to double check the word learning performance of the listeners. A GLMM analysis was conducted in SPSS 25. Language (3 levels) and Word (7 levels) were taken as fixed factors, with subject and item added as random effects. The F-tests found Language to be significant (F (2, 1,239) = 12.854, $p < 0.001$), as shown in Table 4.11, indicating overall performance differences across language groups. In general, Mandarin listeners (86.9% accuracy (F (2, 1,239) = 9.060, $p < 0.001$) and Japanese listeners (86.7% accuracy) (F (2, 1,239) = 7.985, $p < 0.001$) were significantly better than Dutch listeners (70% accuracy) at judging whether the word they heard matched the picture.

A main effect of Word (*F* (6, 1,239) = 12.822, *p = 0.091*) was found not significant. There was no interaction between Word and Language (*F* (12, 1,239) = 1.730, *p = 0.055*), indicating there were comparable differences across words and language groups that do not differ depending on the words.

Figure 4.11 displays the performance of the three language groups in the judgment task. It is notable that the accuracy of each word in each language group reached above chance level (50%), indicating that all groups were able to distinguish whether the sounds and meanings matched or not.

Table 4.11. *Parameters of fixed effects and random effects in the judgment task.*

Fixed effects	F	df	p	
Language group	12.854	2	<0.001	
Word	1.822	6	<0.001	
Language group * Word	1.730	12	0.055	
Random effects	Est.	Std. Error	z	p
1 | Participant	0.314	0.125	2.506	0.012
1| Item	0.462	0.290	1.593	0.111

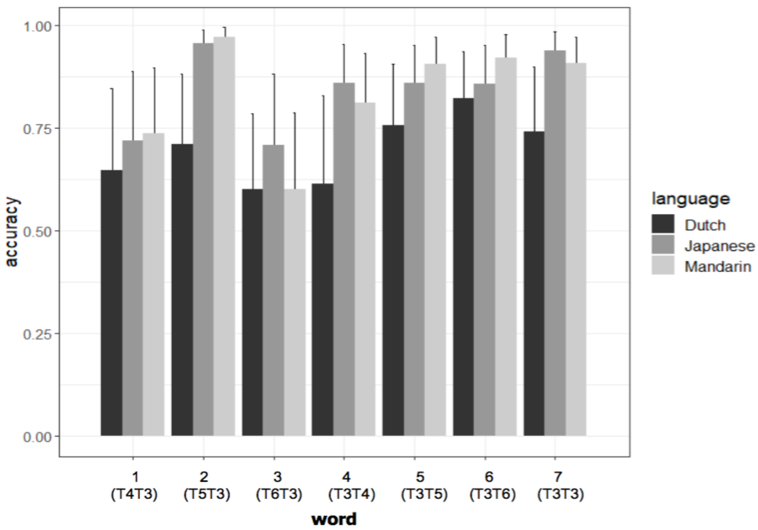


Figure 4.11. *Post-test performance of each language group in the judgment task.*

4.4 General discussion

The present study applied a picture selection task to investigate the sound-to-word mapping in word learning by Dutch, Mandarin, and Japanese listeners. It attempted to further understand how and to what extent the native language could influence the encoding of non-native WPCs (non-native tones occurring word-initially and word finally) in the mental lexicon from a cross-linguistic perspective, by addressing 1) how the three language groups differ from each other with respect to encoding non-native WPCs in word learning, 2) whether the three groups display language specific patterns or share commonalities when encoding non-native WPCs in word learning, and 3) to what extent position interacts with the encoding of non-native tones in each language group.

4.4.1 Overall performance across language groups in word learning

To start with, it was found that Mandarin and Japanese listeners encoded non-native tones better than Dutch listeners in building sound-to-meaning associations, while the former two groups displayed comparably good performance. The advantage Mandarin and Japanese listeners have over Dutch listeners in word learning is supported by prediction (1b), but not (1a) or (1c). Hypothesis (1a) was based on the Feature Hypothesis (McAllister et al., 2002) that if a phonetic or phonological dimension is not used to differentiate lexical meanings in the native language, it will be difficult for naive listeners to perceive and acquire. Accordingly, Japanese listeners (use of lexical pitch accent and positional marking) would be expected to show advantage over Mandarin listeners (who use lexical tones) and Dutch listeners (who use lexical stress) in encoding non-native WPCs lexically, which was rejected by the current finding. (1c) predicted that Dutch and Mandarin listeners would outperform Japanese listeners in word learning. This was in light of the influence of prosody at the utterance level (Braun et al., 2014) that the larger inventory of pitch contrasts in intonation will facilitate listeners to encode non-native tones in sound-to-meaning mapping. Accordingly, Dutch enjoys three initial boundary tones, three final boundary tones, and seven pitch accent (nuclear tones) types (Gussenhoven, 2005). Mandarin has four initial boundary tones, two final boundary tones, and five lexical tones as nuclear tones (Peng et al., 2005). In contrast, Japanese has one initial boundary tone, two final boundary tones, and one nuclear tone (lexical pitch accent) (Pierrehumbert & Beckman, 1988). The larger repertoire of pitch contrasts in Dutch and Mandarin intonation prosody would predict an advantage of Dutch and Mandarin listeners over Japanese listeners in the encoding of non-native tones in word learning, which is contradicted by the present findings.

Instead, prediction (1b) held, based on the Functional Pitch Hypothesis that the functionality of pitch used for lexical contrasts will determine non-native tone perception and tone acquisition (Schaefer & Darcy, 2014). According to this hypothesis, the higher the functionality of pitch in the native language, the better the performance will be in non-native tone perception and acquisition. Mandarin, Japanese, and Dutch differ in degree of functionality of lexical pitch from maximal, high-intermediate, to low-intermediate, which shapes the performance of the three groups in a gradient order. Note that strictly speaking, the Functional Pitch Hypothesis (Schaefer & Darcy, 2014) would predict a performance hierarchy of Mandarin listeners >> Japanese listeners >> Dutch listeners. However, the current findings (Mandarin = Japanese listeners >> Dutch listeners) seem to suggest that when pitch is used lexically in the languages, even though its degree of functionality varies, it can still benefit listeners with lower degree of pitch functionality to be as good as listeners with higher degree of pitch functionality in building non-native tones associated with meanings. The findings may also support Hyman's typology (2009) in which languages with canonical lexical tones (e.g., Mandarin) and languages with lexical pitch accent (e.g., Japanese) are both regarded as tone languages, based on the definition in which "pitch enters into the lexical realization of at least some morphemes" (Hyman, 2006:229).

Furthermore, Dutch listeners' lesser capability of encoding non-native tones in word learning was also shown by the larger number of tonal errors and irrelevant errors they made when identifying words, compared with Mandarin and Japanese listeners. This suggests that the lack of lexical pitch (lower functionality of pitch) in their native language may bring Dutch listeners more confusion in building non-native tones paired with meanings.

4.4.2 Language-specific patterns in word learning

In addition to the overall performance differences across language groups, language-specific patterns were observed in encoding non-native tones in word learning in terms of analyses of exploratory purposes. Mandarin listeners attended more to the non-native contour tones T4 and T5 than the level tone T6 in sound-to-meaning mapping. Japanese and Dutch listeners displayed a similar pattern in that they performed comparably in encoding all of the non-native tones lexically. Note that despite the similar pattern, Japanese listeners showed equally good performance encoding all the non-native tones (all with around 80% accuracy), while the performance of Dutch listeners was generally poor across all the non-native tones (all with around 50% accuracy).

The language-specific patterns in the current findings were not fully supported by the predictions based on the proposal that the long-term stored representations of pitch-based phonological categories (tones and/or intonations) determine the weight listeners give to non-native pitch features in cross-linguistic tone perception or acquisition (Francis et al., 2008; Lee et al. 1996; Xu et al., 2006). Only the pattern of the Mandarin listeners supports this prediction. Mandarin uses three contour tones and one level tone; it uses pitch contour contrasts but not pitch level contrasts to differentiate word meanings. Hence, Mandarin listeners may have stored representations of tonal categories, i.e., categories of contrastive contour tones, in long-term memory, which may benefit them in encoding non-native contour tones in word learning. However, due to a lack of contrastive level tones in Mandarin, they were relatively vulnerable when encoding non-native level tone T6 which they especially confused with mid level tone T3. For instance, they tended to misidentify words with the target level tone T6 as another level tone T3 (e.g., confusion between T6T3 vs. T3T3).

As for the Dutch and Japanese listeners, the relative perceptual weighting proposal failed to predict their learning patterns. According to the proposal, Dutch listeners, due to a lack of lexical tones in the native language, may have stored long-term representations of intonational nuclear tones consisting of basic structure H and L tones (Gussenhoven, 2004), which may thus facilitate to give attention to the non-native level tones compared to the non-native contour tones in word learning. On the contrary, Dutch listeners were found to show no preference for non-native contour or level tones and performed comparably poor in the lexical encoding of all of the non-native tones. This can be accounted for by a lack of lexical tones in the native language. Dutch listeners were found good at discriminating or identifying non-native tonal contrasts at the acoustic level, which could be due to a benefit from a rich inventory of intonation patterns (Chen et al., 2016; Liu et al., 2017). When perceiving non-native tones in acoustic tasks, listeners may store the surface phonetic forms of pitch contrasts in short-term memory (Xu, 1991). However, when encoding non-native tones in word learning, the phonetic forms at the acoustic level may decay, and the phonological representations stored in long-term memory may kick in. Due to a lack of lexical tones, Dutch listeners do not store long-term representations of lexical tones. The pitch-based phonological representations they may have stored are intonation patterns (nuclear tones), which may not be readily available for re-use in word learning. This suggests that while native intonation patterns may facilitate acoustic pitch discrimination in non-tone language listeners (i.e., Dutch listeners) (Chen et al., 2016; Liu et al., 2017), only long-term storage of lexical tone patterns may facilitate listeners in building non-native tones associated with meanings in word learning. A lack of lexical tones in the native language has been reported to hinder the processing of non-native tones in word learning in previous studies. For example, Ramachers et al. (2017) found that Dutch listeners were not sensitive to Limburgian tonal contrasts in word identification.

As for Japanese listeners, it was predicted that they would better encode non-native level tones than contour tones because Japanese tonal patterns are mapped from the accentual patterns of prosodic words consisting of H and L tones. This prediction was contradicted by the current finding that Japanese listeners performed comparably well in all the non-native tones. No preference for a specific tone or a tone category was displayed by Japanese listeners. The current finding could be explained by the influence of accentual patterns in Japanese words as an entity. According to the native accentuation rules (McCawley, 1968; Haraguchi, 1999;

Uwano, 1999), disyllables can carry accentual patterns as H'L (e.g., H'L in an initial-accented bimoraic word), LH'[25] (e.g., LH' in a final-accented bimoraic word), HH' (e.g., HH' in a LHH' final-accented trimoraic word), and LL (e.g., LL in a H'LL initial-accented trimoraic word). Japanese listeners may have stored long-term representations of accentual patterns HH', H'L, LH', and LL as phonological categories, instead of the basic elements H versus L tones of which the tonal patterns consist, which may guide them to show comparable sensitivity to encode non-native tones T3 (mid level tone), T4 (low falling tone), T5 (low rising tone), and T6 (low level tone) in word learning. More specifically, Francis et al. (2008) further developed a proposal in terms of relative perceptual weighting to features of tones in long-term memory (Lee et al., 1996; Xu et al., 2006) and proposed that when there is a clear correspondence between native and non-native category representations, the non-native categories will be assimilated to the native categories (although Francis et al. (2008) did not define "clear correspondence"). When there is no clear correspondence, the relative weight will be distributed to specific perceptual features that are set by the native tone categorization. Accordingly, the non-native contour tones T4 and T5 may be assimilated to H'L and LH' accentual categories in Japanese, respectively. The non-native level tones T3 and T6 may be assimilated to HH' and LL accentual categories, respectively. Clear assimilation between the non-native tones and the native tonal patterns may account for Japanese listeners' comparably good performance encoding non-native tones in establishing lexical representations.

With respect to correspondence between native category and non-native category representations, Best & Tyler (2007) extended the Perceptual Assimilation Model (PAM) (Best, 1994) to PAM-L2 such that if the non-native phonemic contrast (here, non-native tones) is mapped onto the same category in the native language, it will yield poor perception and impede learning. If the non-native phonemic contrast is mapped onto two different categories in the native language, it will lead to good perception and acquisition. Based on PAM-L2 (Best & Tyler, 2007), the non-native T3, T4, T5, and T6 could be mapped onto four different accentual categories in Japanese, namely LL, H'L, LH', and HH, and thus Japanese listeners would be good at encoding these non-native tones in word learning. Previous studies by and large support the mapping between non-native tonal categories and Japanese accentual categories. For example, So (2010) found that Japanese listeners were able to categorize the Mandarin tones they heard onto native accentual categories, that is, they associated Mandarin high level tone, rising tone, and falling tone with the native accentual patterns HH, LH', and H'L, respectively. Nonetheless, the LL pattern was not investigated in So (2010).

[25] Note that it could also be LH, an unaccented pattern, as final-accented and unaccented words in Japanese have similar tonal patterns they appear in isolation (Vance 1995; Warner 1997; Sugiyama, 2006). However, a contrast between final-accented and unaccented words merges when followed by a gramatical particle (such as the nominative *ga*). Final-accented words exhibit a pitch fall on the final syllable while unaccented words do not.

Another account of the overall good performance of Japanese listeners in word learning could be due to benefit from the diverse pitch patterns in Japanese multimora words. In Japanese one mora is usually around 100-150 ms in natural speech (Hoequist, 1983). The disyllables (625ms) used in the study might have been perceived as words with 4-5-mora instead of two moras. Words with longer moras in Japanese exhibit diverse pitch patterns such as rising, falling, level high and rising+falling contour (Kubozono, 2008). The overall contours of the non-words could be assimilated to the various pitch patterns in Japanese words with longer moras. Note that assimilation between the non-native Cantonese tones and Japanese accentual categories or pitch patterns in the current study is in nature speculative. It is necessary to conduct overt studies to investigate correspondences between the native categories (e.g., tonal, accentual, or intonational categories) and the non-native Cantonese tonal categories as studies have done on the mapping between Mandarin tones and Japanese accentual patterns (So, 2010) and between Mandarin tones and Australian English and French intonation patterns (So & Best, 2014). Also, note that the current discussion regarding language-specific patterns should be taken caution considering the interaction among tone and language was insiginicant.

4.4.3 Influence of position on the lexical encoding of non-native tones

The target non-native tones occurred in word-initial and -final positions in the study. It was observed that position played different roles in each language group in the encoding of non-native tones in word learning, based on analyses of exploratory purposes. Position did not influence Dutch listeners and Mandarin listeners. Neither of the two groups showed any preference for either position. In contrast, position interacted with Japanese listeners' encoding of the non-native tone T4. They encoded T4 better when it occurred word-finally than word-initially.

The current findings are in line with the predictions made for Mandarin listeners, but not those for Dutch and Japanese listeners. Based on the native word phonology, Mandarin does not use positional marking but rather lexical tones in word prosody. Mandarin listeners may solely draw upon their tonal sensitivity in building mental representations of non-native tones paired with meanings, and thus they were predicted to show no preference for positions when encoding non-native tones lexically.

For Dutch listeners, it was predicted that they would favour word-initial over word-final position in that stress is usually placed on the first syllable in disyllabic Dutch words (Trommelen, 1983; Kager, 1989; Trommelen & Zonneveld, 1999). Such a prediction fails to account for the current findings. The non-preference for any position observed in the Dutch listeners' lexical encoding of non-native tones could instead be accounted for by a lack of lexical tones in the native language. As shown in previous studies (Cooper & Wang, 2013; Ramachers et al., 2017), listeners without the use of lexical tones in the native language may be hindered the encoding of non-native tones in sound-to-meaning mapping. Due to a lack of lexical tones in their native language, Dutch listeners may encounter difficulty when processing non-native tones associated with word meanings so that the position marked in the native phonology did not have an opportunity to take effect.

Japanese listeners were found to show a preference for word-final over word-initial only when perceiving the falling tone T4. They were predicted to be more sensitive to word-initial than word-final position overall when encoding non-native tones lexically, which is rejected by this finding. The original prediction was based on Japanese native word phonology in which pitch accent is dominantly placed on the initial mora in bimoraic Japanese words, although all accentual patterns are theoretically possible (McCawley, 1968; Akinaga, 1985; Haraguchi, 1999; Uwano, 1999). Native accentuation fails to account for Japanese listeners' preference for final position (T3T4) over initial position (T4T3). Instead, such positional preference for T4 could be attributed to an influence of intonation. The falling tone T4 in final position in the tonal pattern T3-T4 may resemble a falling contour at the end of an utterance that signals assertive meaning in Japanese (Bolinger, 1978, 1982; Sugito, 1981; Poser, 1984; Pierrehumbert & Beckman, 1988; Cruttenden, 1994). Japanese listeners may associate the falling tone in the final position with a terminal fall in a statement and thus it might be distinctive enough to be differentiated from T4 in the initial position. However, such association between non-native tone categories and native intonation categories is in nature speculative. It would be necessary to conduct a true study on the overt mapping between native intonation categories and non-native tone categories (for instance, as is done for the mapping between Mandarin tonal categories and intonation categories in Australian English and French in So & Best, 2014). Moreover, the terminal fall of an utterance is generally observed across languages including Mandarin (Shih, 1997, 2000; Lai et al., 2014) and Dutch (Gussenhoven & Rietveld, 1988). However, such a positional preference for T4 might not be observed in Mandarin and Dutch listeners (though it was marginal, $p=0.055$, in Dutch listeners). This matter needs further investigation in the future. Note that the current discussion should be taken with caution, given that the interaction of language group, tone and position in the analysis was not significant.

4.5 Conclusion

The present study investigated the encoding of non-native tones occurring in different positions in word learning by Dutch, Mandarin, and Japanese listeners. By applying a picture selection task, it attempted to explore how and to what extent the native word prosody and/or intonation prosody influence listeners in the encoding of non-native tones with positions when building sound-to-meaning mapping.

First of all, Mandarin and Japanese listeners performed comparably well, and they both were found to encode non-native tones occurring in different positions better than Dutch listeners did in word identification. Secondly, language-specific patterns were observed in the encoding of non-native tones in different positions in word learning, based on exploratory purposes. Mandarin listeners were better at encoding non-native contour tones than level tones, while both Japanese and Dutch listeners showed no preference for any non-native tones. Thirdly, position was found to interact with the encoding of non-native tones differently in each language group, based on analyses of exploratory purposes. Both Mandarin listeners and Dutch listeners seemed not in favor of any positions. In contrast, Japanese listeners preferred word-final position to word-initial position only when encoding the non-native falling tone T4.

These observations were discussed in terms of different theoretical frameworks, e.g., the Feature Hypothesis (McAllister et al., 2002), the proposal of relative perceptual weighting to tone features in long-term memory (Francis et al., 2008; Lee et al., 1996; Xu et al., 2006), and PAM-L2 (Best & Tyler, 2007). Moreover, we extended the frameworks that are mostly discussed in cross-linguistic tone perception to the encoding of non-native tones in sound-to-meaning mapping. The differences in performance between the language groups at the lexical, phonological, and acoustic level will be discussed in Chapter 5 (general discussion).

Chapter 5
General discussion and conclusion

5.1 Dissertation summary

The goal of this dissertation was to better explore the role of native word prosody in the cross-linguistic perception of non-native word prosodic cues (WPCs) (i.e., non-native pitch contrasts and position) at three processing levels: the auditory-acoustic level, the phonological level which taps into abstract representations (Dupoux et al., 2001, 2008), and the lexical level where phonological knowledge is encoded to integrate sound-to-meaning associations in word learning (Wong & Perrachione, 2007; Wang & Cooper, 2013; Braun et al., 2014). Mandarin, Japanese,[26] and Dutch listeners were selected since these three languages not only differ in their use of WPCs (Mandarin: lexical tones, Japanese: lexical pitch accent, Dutch: lexical stress) but also share commonalities (Mandarin and Japanese both use lexical pitch; Japanese and Dutch both use positional marking cues). More specifically, the dissertation sought to identify language-specificities and language-commonalties in the processing of non-native WPCs at the acoustic, phonological, and lexical levels.

To investigate the influence of native word prosody on the perception of non-native WPCs, three general issues were considered at each processing level: 1) how the three language groups differed from each other (i.e., comparing differences in overall performances across language groups), 2) what specific patterns each language group displayed, and 3) whether position played a role in the perception of non-native pitch contrasts (i.e., the interaction between pitch contrasts and position). Three empirical chapters, Chapters 2, 3, and 4, addressed these questions at the acoustic, phonological, and lexical levels, respectively.

Chapter 2 adopted AX discrimination tasks and used Cantonese tones (also used in the following studies) to investigate the perception of non-native WPCs at the acoustic level by Dutch, Mandarin, and Japanese listeners. It addressed to what extent native word prosody influenced the perception of non-native WPCs from two perspectives: perception of non-native pitch contrasts that occurred in different positions (specifically in trisyllables, e.g., T4T3T3 vs. T5T3T3, T3T4T3 vs. T3T5T3) and perception of non-native tones contrastive in positions (word-initial vs. -medial position, word-initial vs. -final position, and word-medial vs. -final position, e.g., T4T3T3 vs. T3T4T3).

When perceiving non-native pitch contrasts in different positions, the three groups showed comparably good performance when perceiving the acoustically

[26] "Japanese" in the dissertation is restricted to Tokyo Japanese.

salient ones (tones paired with T1 and T3), but they differed in their perception of non-native pitch contrasts that share similarity in acoustic space to some extent (T4-T5, T4-T6, and T5-T6). Overall, Mandarin listeners showed an advantage over Japanese and Dutch listeners when discriminating non-salient non-native pitch contrasts. This can be accounted for by the Functional Pitch Hypothesis (Schaefer & Darcy, 2014) which states that the higher functionality of pitch used in the native language, the better the listeners can be in non-native tone perception. The findings suggest that both acoustic salience and the use of lexical pitch and/or the degree to which lexical pitch is used in the native language can influence the perception of non-native pitch contrasts.

Moreover, language specific patterns were observed in the acoustic discrimination of non-native pitch contrasts in trisyllables. Dutch listeners were relatively sensitive to non-native pitch contrasts that were acoustically salient (tones paired with T1 and T3), while they were worse at acoustically discriminating non-salient pitch contrasts (T4-T5, T4-T6 and T5-T6). Japanese listeners were good at discriminating most of the non-native pitch contrasts except for one specific non-salient non-native contrast (T5-T6). Mandarin listeners perceived all the non-native contrasts equally well.

In addition, position did not interact with the perception of non-native pitch contrasts in trisyllables. All three groups showed the same pattern, viz. they performed better when the non-native pitch contrasts occurred word-finally than word-initially and word-medially, which could be due to a recency effect (Demany & Semal, 2008). Their discrimination of non-native pitch contrasts did not depend on the position where the contrasts occurred.

In the perception of non-native tones contrastive in positions (positional tones), all the three groups overall achieved nearly ceiling performance. However, a perceptual gradient order was still observed: Mandarin listeners >> Japanese listeners >> Dutch listeners. Specifically, in terms of analyses of exploratory purposes, Japanese listeners and Dutch listeners might have difficulty discriminating T5 contrasting in position with the contextual tone T6 (low level tone) (i.e., T5T6T6 vs. T6T5T6 vs. T6T6T5), which could be due to their general confusion between T5 and T6 instead of an inability to perceive non-native positional tones.

Chapter 3 examined how and to what extent native prosody (word and/or intonation) came into play in the processing of non-native pitch contrasts that occurred word-initially, -medially, and -finally at the phonological level by Dutch, Mandarin, and Japanese listeners. A sequence recall task (Dupoux et al., 2001) was applied where listeners had to learn a pair of words and recall the order of the words in different sequence lengths. The results showed that, first, the three language groups performed in a hierarchical order: Mandarin listeners >> Japanese listeners >> Dutch listeners, which could be attributed to the degree of functionality of pitch used in the native language (Schaefer & Darcy, 2014).

Second, language-specific patterns were found with different weighting given to non-native pitch contrasts. Dutch listeners were more sensitive to perceive non-native pitch level contrast than pitch contour and pitch level vs. pitch contour contrast at the phonological level. Japanese listeners were more sensitive to non-native pitch contour contrast and pitch level vs. pitch contour contrast than pitch level contrast. Mandarin listeners displayed a gradient order favouring non-native pitch contour contrast >> non-native pitch level vs. pitch contour contrast >> non-native pitch level contrast. The observed language-specific patterns were accounted for in terms of the long-term memory representations of pitch-based phonological categories (tones and/or intonation) (Xu et al., 2006) and assimilation between the native categories and non-native categories (So & Best, 2010, 2011, 2014; Francis et al., 2008).

Thirdly, position was found to interact with the perception of non-native pitch contrasts at the phonological level in each language group. On the one hand, a language-general pattern was found, that is, when perceiving non-native pitch contour (falling tone) vs. pitch level T4-T6, all three language groups preferred word-final position to word-initial and/or word-medial position. On the other hand, language-specific patterns with respect to position were observed. Dutch listeners were only influenced by position when perceiving non-native pitch contour vs. pitch level contrast T4-T6. They preferred word-final to word-initial and word-medial position when perceiving T4-T6. In contrast, when perceiving the non-native pitch contour contrast T4-T5 and the pitch level contrast T3-T6, they were not affected by position at all. That is, they showed comparably poor performance in each position with respect to these two pitch contrasts at the phonological level. Unlike Dutch listeners, Japanese listeners preferred word-medial position to word-initial and word-final position when perceiving all of the non-native pitch contrasts. As for Mandarin listeners, they showed no preference for position when perceiving non-native pitch contour contrast. However, they favoured word-medial position to word-initial and word-final position when perceiving non-native pitch level contrast T3-T6. The different roles position played in each language group in the perception of non-native pitch contrasts at the phonological level was explained in light of the influence of the native phonology (word prosody and intonation).

Chapter 4 explored the role of native prosody (word and/or intonation) in the encoding of non-native pitch contrasts with position in sound-to-meaning mapping in word learning. It constructed non-native tones that occurred word-initially and word-finally in disyllabic non-words. Using a picture-selection task, listeners heard a word and were required to identify the corresponding picture from four pictures that consisted of the target, one tonal competitor (contrastive in pitch), one positional competitor (contrastive in position), and one distracter (neither contrastive in position nor in pitch). The results revealed that first of all, Mandarin and Japanese listeners displayed comparably overall good performance and they both encoded non-native pitch contrasts better than Dutch listeners in word identification, which could be accounted for by the employment of lexical pitch in the native language of Mandarin and Japanese listeners. Moreover, Dutch listeners tended to misidentify tonal competitors and distracters as the correct response more often than Mandarin and Japanese listeners, suggesting that they were less capable of encoding non-native tones in sound-to-meaning associations.

Second, each language group manifested specific patterns in the encoding of non-native tones in word learning based on analyses of exploratory purposes. Mandarin listeners were found better at encoding non-native contour tones than level tones, while both Japanese and Dutch listeners showed no preference for any non-native tones.

Third, the influence of position seemed to be robust only in Japanese listeners, but not in Dutch and Mandarin listeners, based on analyses of exploratory purposes. Specifically, Japanese listeners were in favour of word-final position over word-initial position only when encoding the non-native falling tone T4, presumably due to the influence of the native intonation pattern. In contrast, both Mandarin listeners and Dutch listeners showed no preference for any position when encoding non-native tones in word learning.

5.2 Research questions revisited

This dissertation addresses an overarching question: how and to what extent does native word prosody play a role in the processing of non-native WPCs? This general issue is considered more specifically by addressing the following questions:

(1) How will Mandarin, Japanese, and Dutch listeners differ when processing non-native WPCs?

(2) What perceptual patterns will Mandarin, Japanese, and Dutch listeners display when processing non-native WPCs?

(3) To what extent will position interact with non-native pitch contrast in each language group in the encoding of non-native WPCs?

These three questions were investigated not only at the acoustic level but also at higher linguistic levels, namely the phonological level and the lexical level. The acoustic level is a processing level at which listeners can relatively easily access acoustic information (e.g., fine-grained acoustic cues) with less memory load imposed for acoustic discrimination, revealed by tasks such as the AX discrimination task (Logan & Pruitt, 1995; Strange & Shafer, 2008). The phonological level refers to a relatively abstract processing level at which phonological representations are highlighted, revealed by a short-term memory task, the sequence recall task (Dupoux et al., 2001, 2008). The lexical level refers to a relatively high level of linguistic processing at which listeners use phonological knowledge to integrate novel sound (here, non-native WPCs) with paired meanings in word learning (Wong & Perrachione, et al., 2007; Braun et al., 2014).

Different hypotheses were considered to make predictions corresponding to the three research questions. More specifically, predictions for research question (1) were based on the Feature Hypothesis (McAllister et al., 2002) and Functional Pitch Hypothesis (Schaefer & Darcy, 2014). The Feature Hypothesis (McAllister et al., 2002) proposes that if a phonetic or phonological dimension is not used for lexical contrast in the native language, it will be difficult for naive listeners to perceive that non-native feature. The Functional Pitch Hypothesis (Schaefer & Darcy, 2014), derived from the Feature Hypothesis, specifically focuses on non-native tone perception and acquisition and states that the higher functionality pitch has in signalling lexical contrasts in the native language, the better performance listeners will display in non-native tone perception. Predictions for research question (2) were based on an approach regarding the relative perceptual weighting to pitch features in long-term storage (Francis et al., 2008; Lee et al., 1996; Xu et al., 2006) proposing that listeners store long-term memory representations of pitch-based phonological categories (i.e., tonal and/or intonational). Pitch representations stored for the native language cause listeners to assign relative weight to specific perceptual features of pitch in cross-linguistic tone perception. Predictions for research question (3) were based on the influence of the native phonology.

Since none of these hypotheses specifically or strictly tease apart acoustic perception, perception at the phonological level, or processing at the lexical level, we applied the abovementioned hypotheses at all three processing levels. The goal was to elucidate non-native tone perception and provide “a baseline for L2 tone acquisition” (e.g., in Schaefer & Darcy, 2014: 514), which entails that the predictions from these hypotheses may carry over to perception involving phonological information or lexical information. Moreover, we considered an additional hypothesis in terms of the influence of native intonation prosody on word learning (Braun et al., 2014) to predict the outcomes (1) at the lexical level. It is proposed that a richer inventory of pitch contrasts at the utterance level predicts better performance in the encoding of non-native tones in sound-to-meaning mapping.

The predictions made for the processing of non-native WPCs at each processing level were by and large identical, as follows:

(1) Based on the Feature Hypothesis (McAllister et al., 2002), it was hypothesized that Japanese listeners (who use the position of pitch accent lexically) would show an overall advantage over Mandarin listeners (who use lexical tones) and Dutch listeners (who use the position of lexical stress) when perceiving non-native pitch contrasts that occurred in different positions at each processing level. Alternatively, according to the Functional Pitch Hypothesis (Schaefer & Darcy, 2014), at each processing level, Mandarin listeners were predicted to show overall better performance than Japanese and Dutch listeners, while Japanese listeners would outperform Dutch listeners in the processing of non-native pitch contrasts occurring in different positions.

Additionally, based on the proposal that considers intonation's influence on word learning (Braun et al., 2014), Mandarin and Dutch listeners would encode non-native WPCs better than Japanese listeners in building sound-to-meaning associations in word learning (i.e., at the lexical level).

Note that at the acoustic level, we also investigated the perception of non-native tones contrastive in positions. As for the perception of non-native tones contrastive in positions, the Feature Hypothesis (McAllister et al., 2002) yielded two divergent predictions. a) Japanese listeners (who use position for lexical pitch accent) would outperform Dutch listeners (who use position for lexical stress) and Mandarin listeners (who use lexical tones). b) If perception of non-native positional tones is fundamentally driven by the perception of position, it was predicted that Japanese and Dutch listeners will outperform Mandarin listeners. c) Alternatively, if perception of non-native positional tones is driven primarily by the perception of tone, it was predicted based on Functional Pitch Prominence (Schaefer & Darcy, 2014) that Mandarin listeners would outperform Dutch and Japanese listeners.

(2) At each processing level, each language group was predicted to show different perceptual patterns based on the selective perceptual weighting given to specific pitch features in long-term storage (Lee et al., 1996; Xu et al., 2006; Francis et al., 2008). Accordingly, Mandarin listeners were predicted to be more sensitive to non-native pitch contour contrasts than pitch level contrasts due to contrastive contour tonal categories in the native language. Dutch listeners would give more weight to non-native pitch level contrasts than pitch contour contrasts because the rich inventory of nuclear tones in the Dutch intonation system consists of H and L tones. Japanese listeners would show more sensitivity to non-native pitch level contrasts than pitch contour contrasts because the accentual patterns at word prosody are composed of H and L tones.

(3) The influence of position on the perception of non-native pitch contrasts occurring in different positions was also hypothesized in light of native word phonology. Mandarin does not employ positional cues in word prosody. Dutch uses lexical stress which is generally placed in a three-syllable window at the right-hand word edge, but the final syllable is not a canonical position (Kager, 1989). Japanese, on the other hand, uses the position of lexical pitch accent. Based on the native accentuation rule, initial accent is prevalent in 2-mora words while an unaccented form is dominant in three mora words (Sato, 1993; Kitahara, 2001). According to the native word phonology, at the acoustic and phonological levels where non-native pitch contrasts in trisyllables[27] were investigated, Mandarin listeners were predicted to show no preference for position. Dutch listeners may prefer the prefinal position (i.e., medial position in trisyllables) for non-native pitch contrasts. Japanese listeners were predicted to show no preference for any position when perceiving non-native pitch contrasts. At the lexical level where non-native pitch contrasts in disyllables

[27] Note that while trisyllables were used at the acoustic and phonological levels, this dissertation used CV disyllabic (bimoraic) words at the lexical level.

were investigated, Mandarin would show no preference for position. Dutch listeners would prefer word-initial to word-final position when processing non-native tones in word learning since prefinal is the default position for lexical stress in Dutch, corresponding to the initial position in disyllabic words (Kager, 1989; Trommelen & Zonneveld, 1999). Japanese listeners were predicted to prefer word-initial to word-final position in word learning due to the fact that initial accent is prevalent in 2-mora words in Japanese native accentuation (Sato, 1993; Kitahara, 2001).

5.3 General discussion

This section will discuss the findings at the acoustic, phonological, and lexical levels from three perspectives: 1) how native word prosody influences the perception of non-native WPCs (i.e., overall differences across language groups), 2) what specific or general patterns the language groups manifest, and 3) the influence of position on the processing of non-native pitch contrasts.

5.3.1 Overall differences across language groups in non-native WPC processing across acoustic, phonological, and lexical levels

To begin with, the three language groups were examined at the acoustic level via a task requiring the discrimination of non-native pitch contrasts in monosyllables (as baseline) and trisyllables. The latter results were in line with the former results; hence discussion will mainly focus on the findings on non-native pitch contrasts in trisyllables. It was observed that all three groups showed comparable ceiling performance when discriminating non-native pitch contrasts that were located in high-mid register (tonal pairs with T1 and T3). However, they differed significantly when discriminating non-native pitch contrasts in mid-low register (T4-T5, T4-T6, and T5-T6), that is, Mandarin listeners displayed perceptual advantage over Japanese and Dutch listeners. The findings were in discrepancy with the predictions that Mandarin listeners would perform better overall compared to Japanese listeners and Dutch listeners (based on the Functional Pitch Hypothesis) or that Japanese listeners would overall outperform Dutch and Mandarin listeners (based on the Feature Hypothesis). Mandarin listeners' advantage over Japanese and Dutch listeners was not an overall advantage as predicted but was instead limited to the discrimination of non-native pitch contrasts in a mid-low register. The predictions seemed to underestimate the influence intrinsic acoustic features of the non-native pitch contrasts could have on the perception of non-native pitch contrasts.

All three groups were good at discriminating non-native pitch contrasts in high-mid register (i.e., contrasts paired with T1 and T3), which could be attributed to the intrinsic acoustic salience of T1 and T3. T1 is located in the high register, far from the other tones in the acoustic space, which may make it distinctive enough to be distinguished from other tones as has been observed in previous studies (Qin &

Mok, 2013; Qin, 2014; Mok & Wong, 2010; Wong, 2019). T3, the mid level tone located in the high-mid register area, is relatively distinctive from low tones in acoustic perceptual space (Francis et al., 2008; Qin & Mok, 2013). Hence, contrasts paired with T1 and T3 were easy to discriminate for all three groups. However, when it comes to the non-native pitch contrasts in mid-low register (T4-T5, T4-T6, and T5-T6), Mandarin listeners displayed more strength than Japanese and Dutch listeners in discrimination. T4 (low falling tone), T5 (low rising tone), and T6 (low level tone) are located in a relatively crowded acoustic space. They share similarity in their starting point and differ only towards the ending point (Qin & Mok, 2013). Their acoustic similarity may cause more difficulty to Japanese and Dutch listeners than to Mandarin listeners in that Japanese and Dutch have lower functionality of pitch than Mandarin, according to the Functional Pitch Hypothesis (Schaefer & Darcy, 2014). In contrast, the maximal-high functionality of pitch in Mandarin may facilitate Mandarin listeners when discriminating non-native pitch contrasts that are not acoustically salient. The differences observed across the three language groups also suggest that when the non-native tonal contrasts are not acoustically salient, Mandarin listeners may approach the discrimination task more phonologically than acoustically in that they may perceive the non-native tones based on their similarity to native tonal categories (cf. PAM (So & Best, 2010)) and hence they may not need to focus on memorizing detailed acoustic information when making the same-or-different judgement.

One notable finding is that Japanese listeners discriminating T4-T6 achieved as good performance as Mandarin listeners did. This can be due to the influence of native word prosody in Japanese. Japanese uses lexical pitch accent (a pitch fall from H to L tone in two successive moras) to differentiate lexical items (Poser, 1984; Kubozono, 1993, 2008). Japanese listeners have been found sensitive to the location of pitch fall in Japanese in previous studies (Ayusawa et al., 1998; Vance, 1995; Yoneyama, 2002). This sensitivity to pitch fall in the native language may be transferred and benefit Japanese listeners when discriminating a non-native falling tone from a level tone in the mid-low register.

When discriminating non-native tones that contrasted in position (positional tones), a perceptual hierarchy was observed such that Mandarin listeners >> Japanese listeners >> Dutch listeners, despite the fact that all three groups achieved nearly ceiling performance. The result supports predictions based on the Functional Pitch Hypothesis and suggests that although Mandarin listeners do not use lexical positional marking, the high functionality of pitch in the native language may assist them in discriminating non-native tones contrastive in positions. In contrast, both Japanese and Dutch employ positional cues in word prosody. However, they differ in the use of lexical pitch. Japanese enjoys a higher functionality of pitch than Dutch does, which may account for the advantage Japanese listeners have over Dutch listeners when discriminating non-native positional tones. The findings seem to suggest that lexical pitch is a more dominant factor than positional cues in determining the cross-linguistic perception of non-native positional tones.

The results of the AX discrimination task at the acoustic level suggest that native word prosody, specifically the role of lexical pitch, together with intrinsic acoustic features of the non-native pitch contrasts themselves cast influence on the acoustic discrimination of non-native WPCs. In a more cognitively demanding task, the sequence recall task that tapped into the phonological level of representations, the three language groups showed an overall gradient pattern, Mandarin listeners >> Japanese listeners >> Dutch listeners, when perceiving the non-native pitch contrasts in trisyllables. The Functional Pitch Hypothesis successfully predicted this observation while the Feature Hypothesis failed, which indicates that the use of lexical pitch and to what degree lexical pitch functions in the native language influence the perception of non-native pitch contrasts at the phonological level.

When encoding non-native pitch contrasts in sound-to-meaning mapping, Mandarin and Japanese listeners both performed better than Dutch listeners with comparably good performance. This follows the Functional Pitch Hypothesis, but not the Feature Hypothesis or the approach in terms of intonation phonology (Braun et al., 2014). According to the Functional Pitch Hypothesis, the degree of pitch functionality in Mandarin is regarded as maximal while in Japanese it is intermediate-high (Schaefer & Darcy, 2014). The functionality of pitch in Dutch is at an intermediate-low degree since pitch is one of the acoustic correlates of lexical stress, not exclusively used for lexical contrasts. The higher functionality of pitch in the native language, the better the performance will be in non-native tone perception and acquisition. Note that strictly speaking, the Functional Pitch Hypothesis would predict a hierarchy of performance such that Mandarin listeners >> Japanese listeners >> Dutch listeners. However, the current finding for word learning, Mandarin = Japanese listeners >> Dutch listeners, seems to suggest that when pitch is used lexically in a language, listeners with relatively lower degrees of pitch functionality (i.e., Japanese listeners) may benefit just as much as listeners with higher degree of pitch functionality in their native language (i.e., Mandarin listeners). This also indicates that the grouping together of “canonical tone” languages and “pitch accent” languages as in Hyman’s typology (2009) is on the right track.

Taken together, findings at each of the three processing levels show that the use of pitch and the extent to which pitch is used for lexical contrasts in the native language may overall facilitate the perception of non-native pitch contrasts at the acoustic level, especially when the non-native pitch contrasts were not acoustically salient, both at the phonological level and the lexical level.

5.3.2 Language specificities in non-native WPC processing at the acoustic, phonological, and lexical levels

5.3.2.1 Acoustic discrimination

First of all, language specific patterns in the discrimination of non-native pitch contrasts in trisyllables were found at the acoustic level. Note that for each language group, the findings regarding perceptual patterns with respect to trisyllables were by and large in accordance with those for monosyllables (the baseline) and thus the discussion will mainly focus on perceptual patterns in trisyllables. When discriminating non-native pitch contrasts in trisyllables at the acoustic level, it was found that Dutch listeners did not show a preference for non-native pitch level contrasts or contour contrasts. Instead, their perceptual performance was dependent on where the contrasts were located in acoustic space. They were more sensitive to non-native pitch contrasts that are acoustically salient in acoustic space (contrasts paired with T1 and T3) than to those with non-acoustic salience (T4-T5, T4-T6, and T5-T6). This perceptual pattern for Dutch listeners deviated from the prediction based on the influence of nuclear tones in intonation prosody that they would attend more to non-native pitch contrasts than to contour contrasts. Their disadvantage when acoustically discriminating non-salient non-native pitch contrasts could be due to a lack of lexical tones in the native language. When non-native pitch contrasts share (partial) similarity in acoustic space, the detailed acoustic information may not be sufficient enough for Dutch listeners to rely on. However, despite their relative difficulty with the acoustic discrimination of non-acoustically salient non-native pitch contrasts, Dutch listeners were still capable of discriminating them (with accuracy above chance level). This could be attributed to a benefit from the diverse patterns of nuclear tones in Dutch intonation and/or the use of lexical stress in word prosody, as suggested in previous studies (Leather, 1987, 1990; Braun & Johnson, 2011; Liu et al., 2017; Chen et al., 2015). Note that T5-T6 was the exceptional contrast that Dutch listeners seemed unable to discriminate, with around 14% correct, much below chance level. Difficulty discriminating T5-T6 could be due to the psychoacoustic similarity of the two tones to non-tone language listeners, as reported in Qin & Mok (2013) who found that non-tone language listeners perceived T5 and T6 closer at the ending point than T4-T6 and T4-T5 in perceptual space.

In contrast, Mandarin listeners were found to attend well to both non-native pitch contour contrasts and level contrasts, which was in discrepancy with the prediction that they would be more sensitive to non-native pitch contour contrasts than level contrasts. Their relatively good performance discriminating among all the non-native pitch contrasts could be due to an overall facilitation from the use of lexical pitch in the native language. However, it is noteworthy that in the baseline scenario (perception of non-native pitch contrasts in isolation), Mandarin listeners were more sensitive to discriminate non-native pitch contour contrasts and pitch level vs. pitch contour contrasts than pitch level contrasts (T1-T3 and T3-T6). The differences between their discrimination patterns for non-native pitch contrasts in monosyllables and trisyllables suggests that Mandarin listeners may rely on tonal context when perceiving the non-native level contrasts to achieve good performance. The contextual tones surrounding the target level contrasts were other level tones (e.g., T1-T1-T3 vs. T1-T1-T6). Presumably due to a lack of contrastive level tones in their native word prosody, Mandarin listeners may need contextual tones to build tonal references when discriminating non-native pitch level contrasts. Such reliance on tonal context was also reported in previous studies where tone language listeners (e.g., Mandarin listeners) depended more on context when perceiving level tones than when perceiving contour tones (Fox & Qi, 1990; Moore & Jongman, 1997; Francis et al., 2003; Wong & Diehl, 2003; Xu et al., 2006). Thus, Mandarin listeners' comparably good performance in discriminating non-native pitch contour contrasts and level contrasts in trisyllables could benefit from the native lexical tones as well as from assistance from the contextual tones.

Similar to the perceptual pattern observed in Mandarin listeners, Japanese listeners paid comparable attention to all of the non-native pitch contrasts. However, they showed difficulty in discriminating T5-T6 with performance below chance level (around 30% correct), while they showed equally good performance for the other non-native pitch contrasts. The findings contradicted the prediction that they would attend more to non-native pitch level contrast than to contour contrast. The finding that Japanese listeners were sensitive to all non-native pitch contrasts except for T5-T6 suggests that they might be facilitated by the use of lexical pitch accent in the native language. However, the restricted tonal system in the native language may not be sufficient enough for Japanese listeners to rely on in order to discriminate the problematic psycho-acoustically similar T5-T6 contrast (Qin & Mok, 2013; Wong, 2019).

To sum up the perceptual patterns in acoustic discrimination, Dutch listeners were influenced by intrinsic acoustic features of non-native pitch contrasts as well as native word prosody (i.e., the lack of lexical tones). In contrast, Mandarin listeners, who showed comparable sensitivity to both non-native pitch contour contrasts and level contrasts, were influenced by the use of lexical pitch in general and the presence of contextual tones. Japanese listeners' equal sensitivity to all non-native pitch contrasts except T5-T6 suggests a native language influence of lexical pitch fall, which, however, was restricted to some extent since Japanese listeners were perceptually confused when the non-native pitch contrasts were psycho-acoustically similar.

5.3.2.2 Phonological processing

The phonological level highlights phonological representations that can be revealed by the sequence recall task, a short-term memory task with phonetic variability (Dupoux et al., 2001, 2008). It is a relatively abstract processing level at which fine-grained acoustic information is not easily accessible for listeners to consciously rely on. To investigate processing at the phonological level, the pitch level contrast T3-T6, the contour contrast T4-T5, and the contour vs. level contrast (T4-T6) were selected to examine the cross-linguistic processing of non-native pitch contrasts in trisyllables. Specific patterns were observed in each language group. The perceptual patterns of Dutch and Mandarin listeners were largely as predicted in terms of selective perceptual weighting given to non-native pitch features (Xu et al., 2006; Francis et al., 2008). Dutch listeners were more sensitive to non-native pitch level contrast T3-T6 than the contour contrast T4-T5 or level vs. contour contrast T4-T6. This preference for non-native pitch level contrast could be due to the influence of nuclear tones that are composed of H and L tones in the Dutch intonation system. Dutch listeners may have stored long-term representations of nuclear tones consisting of H and L categories, which may benefit them, making them more attentive to non-native pitch level contrasts than other contrasts in perception at the phonological level.

Mandarin listeners, as predicted, were found to attend more to non-native pitch contour contrast than pitch level contrast. Mandarin has three contour tones and one level tone; it uses pitch contour contrasts but not pitch level contrasts to signal lexical meanings. Mandarin listeners may have stored pitch representations of tonal categories, i.e., categories of contrastive contour tones, in long-term memory, which may facilitate them in perceiving non-native pitch contour contrasts at the phonological level. However, due to a lack of contrastive level tones in Mandarin, they struggled to process non-native pitch level contrasts at the phonological level.

Unlike Dutch and Mandarin listeners, Japanese listeners were better at perceiving non-native pitch contrasts paired with a falling tone (T4-T6 and T4-T5) than the non-native pitch level contrast (T3-T6), which was different from the prediction in terms of the influence of the surface tonal patterns that consist of H and L tones in prosodic words. Rather, this finding can be accounted for by the influence of native lexical pitch fall used to contrast lexical items in Japanese. Japanese listeners may have stored long-term memory representations of pitch-based phonological categories, i.e., lexical pitch fall, which may have aided Japanese listeners to distinguish a falling pitch contour from a non-native pitch contrast. The strength in detecting the lexical pitch fall in the native language at the phonological level was reported in Ustugi et al. (2010) who used a sequence recall task. This sensitivity to pitch fall in the native language might be transferred to aid in the perception of a non-native falling tone.

As seen in the findings, native word prosody and/or intonation prosody played a robust role in shaping the selective perceptual weighting listeners gave to a specific pitch feature at the phonological level. More specifically, lexical tonal categories in Mandarin may cause Mandarin listeners to be more sensitive to non-native pitch contour contrasts than non-native pitch level contrasts. Lexical pitch fall in Japanese may lead to Japanese listeners' preference for non-native pitch contour contrasts (all of which happened to be paired with the non-native falling tone) over the non-native pitch level contrast. In comparison, Dutch listeners, i.e., non-tone language listeners, were unable to rely on lexical tones since Dutch does not employ lexical pitch to contrast word meanings. Yet, long-term stored representations of intonational nuclear tones in the native language could be resorted to and may cause Dutch listeners to be more sensitive to non-native pitch level contrasts than other contrasts at the phonological level.

5.3.2.3 Word learning

When it comes to the encoding of non-native pitch contrasts in disyllables in sound-to-meaning mapping at the lexical level (i.e., in word learning), based on analyses of exploratory purposes, Mandarin listeners, as predicted, were better able to encode the non-native contour tones than level tones. Mandarin uses pitch contour contrasts (three contour tones) but not pitch level contrasts (only one level tone) to differentiate word meanings. Hence, Mandarin listeners may have stored representations of tonal categories, i.e., categories of contrastive contour tones, in long-term memory, which may benefit them in encoding non-native contour tones in word learning. However, due to a lack of contrastive level tones in Mandarin, they were relatively vulnerable in encoding non-native level tones lexically and struggled to distinguish between the non-native level tones in word learning.

Unlike Mandarin listeners, Japanese listeners were found to not prefer any non-native pitch contrasts when encoding non-native tones in their building of sound-to-meaning associations. They displayed comparably good performance encoding all of the non-native pitch contrasts. This finding is not compatible with the prediction that they would be better at encoding non-native level tones based on an influence of the surface tonal patterns consisting of H and L tones. The pattern observed in Japanese listeners could be presumably explained by a correspondence between non-native tones and native accentual patterns in word prosody. In Japanese, the tonal pattern of a word is determined by the presence/absence and location of the lexical pitch accent (Poser, 1984; Haraguchi, 1999). According to native accentual rules (Kubozono, 1993, 2008; Haraguchi, 1999; Uwano, 1999), patterns for disyllables can be H’L (the initial mora accented in a bimoraic word), LH’ (the final mora accented in a bimoraic word), HH’ (e.g., HH’ in a LHH’ final-accented trimoraic word) and LL (e.g., LL in a H’LL initial accented trimoraic word). Japanese listeners may have stored representations of tonal patterns, i.e., H’L, LH’, HH’, and LL, as phonological categories instead of the basic structure H versus L tones consisting of tonal patterns in long-term memory. According to PAM-S proposed by So & Best (2010, 2011, 2012, 2014), if a non-native phonemic contrast (here, tone) is mapped onto two different categories in the native language, it will lead to good perception (cf. Francis et al., 2008). Accordingly, Japanese listeners may assimilate the non-native tones T4 (falling tone), T5 (rising tone), T3 (mid level tone), and T6 (low level tone) onto the representations of H’L, LH’, HH’, and LL native accentual categories, respectively. Such a clear correspondence between the non-native tones and the native accentual categories may cast a beneficial effect on Japanese listeners when encoding non-native tones in word learning. Note that although So (2010) supported the mapping between non-native tonal categories (Mandarin tones) and Japanese accentual categories, the mapping between the Japanese accentual patterns and non-native tones (Cantonese tones) in the current study is speculative in nature. At best, the Japanese listeners' comparable sensitivity to encode all the non-native tones could be presumably given credit to the diverse accentual patterns lexically stored in Japanese, which may allow Japanese listeners to tap into the phonological abstractness of tonal patterns to build mental representations in word learning.

Dutch listeners showed no preference for any non-native pitch contrasts in word learning. Prima facie, Dutch listeners displayed a similar pattern to Japanese listeners. However, unlike Japanese listeners with comparably good performance, Dutch listeners were comparably poor at encoding all the non-native tones. This finding is in conflict with the prediction that Dutch listeners would be more sensitive to non-native pitch level contrasts than contour contrasts in terms of the influence of intonational nuclear tones. Dutch does not use lexical tones and thus the long-term memory representations of pitch-based phonological categories stored in Dutch listeners may be intonational nuclear tones consisting of a basic structure of H and L tones (Gussenhoven, 2004). This was hypothesised to facilitate Dutch listeners to better encode non-native level tones in word learning. However, the findings seem to suggest that long-term memory representations of intonation pattern storage are unavailable for application or re-use in the encoding of non-native tones in mental representations in word learning.

To summarize the word learning results, language specific patterns seemed to be determined by the native word prosody, more specifically, whether or not there are lexical tones in the native language. For Mandarin listeners, tonal categories, i.e., three contrastive contour tones and one level tone, are lexically stored in the native language, which may cause these listeners to show more sensitivity to non-native contour tones than level tones in building sound-to-meaning associations. Japanese, a restricted type of tone language (Hyman, 2009), uses pitch accent for lexical contrasts. The presence/absence and the location of the pitch accent determine the tonal patterns in prosodic words. Japanese tonal patterns are composed of a diverse repertoire of rising (LH'), falling (H'L), and level (HH'/HH, LL) patterns, which may enhance Japanese listeners to be sensitive to encode non-native contour and level tones in word learning. In contrast, Dutch listeners as non-tone language listeners struggled to encode all of the non-native tones lexically due to a lack of lexical tones in the native language.

5.3.2.4 Integration of the three processing levels

To put all the pieces together, the findings suggest that acoustic sensitivity to non-native WPCs does not always predict successful processing of the non-native WPCs at the phonological level and at the lexical level. Moreover, the native prosody played different roles in influencing the listeners at differet processing levels. Specically, the acoustic level, there were several factors that had an influence on the listeners' perceptual patterns: not only native word prosody, but also other factors, i.e., intrinsic acoustic features of the non-native pitch contrasts and contextual tones. More specifically, their native tonal experience (three contrastive contour tones and one level tone) caused Mandarin listeners to be more sensitive to non-native pitch contour contrasts than pitch level contrasts in the discrimination of non-native pitch contrasts in isolation, which was observed in the current study as well as in previous studies (Gandour, 1983; Francis et al., 2008; Qin & Mok, 2013). However, due to facilitation by contextual tones in trisyllables, Mandarin listeners may be able to build level tonal references that allow them to discriminate both non-native pitch level contrasts as well as non-native pitch contour contrasts. The restricted tonal system of Japanese may equip Japanese listeners with a general ability to discriminate most non-native pitch contrasts, with the exception of T5-T6 which is composed of two psycho-acoustically similar tones. Dutch listeners, without the use of lexical pitch, were more successful in discriminating acoustically salient non-native pitch contrasts than contrasts that were less acoustically salient. The findings at the acoustic level showed that native word prosody was not the only factor to influence the listeners. Moreover, native word prosody did not seem robust enough to direct the listeners towards perceptual preferences for a specific pitch feature, contrary to the relative perceptual weighting approach (Francis et al., 2008; Xu et al., 2006; Lee et al., 1996).

However, at the phonological level, native word prosody and even native intonation prosody seemed to take a salient effect in determining the listeners' selective attention to a specific pitch feature (pitch contour or pitch level) when processing the non-native pitch contrasts in trisyllables. Specifically, at the phonological level listeners may not be able to rely on fine-grained acoustic information, but rather have to tap into abstract representations to process the non-native pitch contrasts. Hence, the long-term representations of pitch-based phonological categories (i.e., tone and/or intonation) stored in the native language kick in. More specifically, the long-term memory representations of lexical tonal categories in Mandarin, lexical pitch fall in Japanese, and nuclear tones consisting of H and L tones in Dutch intonation cast influence on the Mandarin, Japanese, and Dutch listeners, respectively, leading them to give language-specific selective perceptual weighting to non-native pitch contour contrasts or pitch level contrasts.

Phonological representations are relevant at both the phonological level (in a sequence recall task) and lexical level (in word learning). The phonological level requires the abstract representations of items (without meaning involved), while the lexical level requires listeners to draw upon abstract representations to integrate non-native tones (sound) with meanings, which may be more cognitively demanding than the processes involved in a sequence recall task. At the lexical level, it was found that native word prosody, particularly the use of lexical tones, played a pivotal role in determining the perceptual weighting assigned to non-native pitch contrasts in word learning for the three language groups. Specifically, Mandarin listeners were better at encoding non-native contour tones than level tones in building mental representations, which could be attributed to a lack of level tones in Mandarin (which has three contour tones and only one level tone). The diversity of tonal patterns in Japanese prosodic words may benefit Japanese listeners to show comparable sensitivity to both non-native contour tones and level tones. In contrast, Dutch listeners were comparably insensitive to both non-native contour and level contrasts, presumably due to a lack of native lexical tones and inability to reuse native intonational nuclear tones to encode non-native tones in word learning.

In conclusion, it was found that not only native word prosody but also acoustic factors influenced cross-linguistic perception of non-native pitch contrasts at the acoustic level. Tone language (Mandarin and Japanese) and non-tone language (Dutch) listeners did not display any preference for a particular pitch feature (pitch contour/pitch level) in acoustic discrimination. Comparatively, at the phonological and lexical levels native word prosody had a more robust and active influence on listeners' selective perceptual weighting on a specific pitch feature. At the phonological level, Mandarin and Japanese listeners were affected by lexical tones and lexical pitch fall to distribute selective attention to the features of pitch, while the attention of non-tone language listeners, i.e., Dutch listeners, was determined by intonational nuclear tonal categories. However, the native intonational nuclear tones seemed unavailable for Dutch listeners to make use of in word learning, while Mandarin and Japanese listeners were profoundly influenced by the native word prosody. This seems to suggest that only long-term storage of lexical tone patterns facilitates listeners' selective weighting on pitch features in word learning. Note that only Mandarin listeners, canonical tone language listeners, showed consistent perceptual patterns at the phonological and lexical level in that they were more sensitive to non-native pitch contour contrasts than pitch level contrasts. In contrast, Japanese listeners, with a restricted tonal system, revealed different patterns at the phonological level and lexical level. They were guided by the lexical pitch fall at the phonological level so that they were more attentive to non-native pitch contrasts paired with the non-native falling tone. However, they were comparably sensitive to encode both non-native pitch contour contrasts and pitch level contrasts in word learning, which was presumably due to the diversity of tonal patterns (accentual patterns) in their native word prosody. The differences between the consistency of Mandarin listeners' behaviour and inconsistency of Japanese listeners' seem to indicate differences in the actual status of lexical tones and lexical pitch accent in these languages.

Last but not least, the language specific patterns observed at different processing levels in this dissertation provide some support for the approach suggesting selective perceptual weighting on pitch features in long-term storage (Francis et al., 2008; Xu et al., 2006; Lee et al., 1996). Such an approach seems to be more tenable when the processing of non-native pitch contrasts involves abstractness (at the phonological level and the lexical level), compared with acoustic discrimination.

5.3.3 Influence of position on cross-linguistic pitch processing at the acoustic, phonological, and lexical levels

5.3.3.1 Influence of position at the acoustic level

To start with, position was found to play no role in the acoustic discrimination of non-native pitch contrasts in trisyllables. All three language groups showed better performance with respect to word-final position compared to word-initial and word-medial positions, which was not compatible with the predictions based on the native phonology. According to the predictions, Mandarin listeners would not prefer specific positions due to a lack of lexical positional marking; Dutch listeners would be more sensitive to non-native contrasts that occurred word-medially than those word-initially and/or word-finally according to the tendency for lexical stress to fall at the right-hand word edge in Dutch; Japanese listeners would not prefer any specific positions due to native accentuation in Japanese, as the unaccented pattern is dominant in Japanese trimoraic words. Contrary to these predictions, an observed preference for word-final position over word-initial and/or medial position in all three groups was found. This could be attributed to a recency effect leading listeners to recall acoustic information in the final position better than in the initial position since the acoustic information of the offset retains longer than that of the onset (Demany & Semal, 2008).

5.3.3.2 Influence of position at the phonological level

Different from its role at the acoustic level, at the phonological level position was observed to influence the perception of non-native pitch contrasts in trisyllables for each language group. Previously it was predicted that Mandarin listeners and Japanese listeners would show no preference for any position, while Dutch listeners would favour the word-medial position. However, the current findings were not in accordance with these predictions. For one, a language-general pattern was observed, which was not predicted. All three groups displayed a preference for word-final position over word-initial position and/or word-medial position when perceiving non-native pitch contour vs. pitch level contrast T4-T6. This could not be parsimoniously accounted for by the recency effect, since if it were due to the recency effect, the preference for word-final position rather than word-initial and word-middle position should also be observed in the perception of the other contrasts. This, however was not found in the current study. Instead, the influence of intonation contours may explain the observed language-general pattern. To be more specific, it is generally accepted that languages use intonation patterns to signal post-lexical linguistic (pragmatic and paralinguistic) meanings. For instance, a falling contour at the end of an utterance is cross-linguistically associated with assertive meaning (Ladd, 1996; Cruttenden, 1981, 1997), which is observed in Mandarin (Shih, 1997, 2000; Lai et al., 2014), in Japanese (Sugito, 1981; Poser, 1984; Pierrehumbert & Beckman, 1988), and in Dutch (Gussenhoven & Rietveld, 1988). Listeners may associate the word-final falling tone with an intonational terminal fall signalling assertiveness and thus it might be distinctive enough to be differentiated from the word-final low level tone T6, compared with this T4-T6 contrast on word-initial and word-medial position. However, such perceptual association between the non-native tone categories and intonation categories is speculative. It would be necessary to conduct a true study on overt mapping between the native intonation categories and the non-native tone categories (as was done for the mapping between Mandarin tonal categories and intonation categories in So & Best, 2014).

Language-specific patterns were also found with respect to the influence of position. Dutch listeners relied on position in the perception of specific non-native pitch contrasts. They preferred the final position over word-initial and -medial positions only when perceiving T4-T6. However, they showed comparable confusion with T4-T5 and T3-T6 (still above chance level) regardless of the positions they occurred in a word. Presumably due to a lack of lexical tones in the native language, Dutch listeners may refer to intonation patterns signalling post-lexical meanings when perceiving the non-native pitch contrasts. When the non-native tonal contrast belongs to the same category, i.e., both T4 and T5 are contour tones or both T3 and T6 are level tones, Dutch listeners seemed unable to rely on the native intonation contours. It seems they could rely on the native intonation only when the two tones in the contrast were in different tonal categories, i.e., pitch contour vs. pitch level T4-T6, and when such contrast occurred in the final position. It is worth noting that Braun & Johnson (2011) reported that Dutch listeners were more sensitive to non-native Mandarin falling tone versus the rising tone word-finally than word-initially in AXB speeded tasks. The authors ascribed such a preference to the resemblance of the final pitch patterns with intonation patterns signalling statement vs. question in Dutch. However, the current findings seem to suggest that Dutch listeners may be able to associate the intonation patterns with non-native contrastive contour tones at the acoustic level, but such associations may be weak when they need to incorporate tonal patterns into word meaning.

As for Mandarin listeners, position played a complex role in their perception of non-native pitch contrasts. On the one hand, Mandarin listeners were not influenced by position when perceiving the non-native pitch contour contrast T4-T5 at the phonological level. They were good at perceiving T4-T5 in all positions, which could be attributed to facilitation from the native contrastive contour tones (Mandarin rising vs. falling tone). On the other hand, unlike their overall strong perception of T4-T5 regardless of position, Mandarin listeners preferred word-medial position over word-initial and word-final position when perceiving the non-native pitch level contrast T3-T6. Presumably due to a lack of pitch level contrasts in the native word prosody (only one level tone in Mandarin), Mandarin listeners seemed to require tonal references to successfully perceive non-native pitch level contrasts. More specifically, the T3-T6 contrast that occurred word-medially (T3-T3-T3 vs. T3-T6-T3) enjoys the benefit of two neighbouring level tones. The two contextual level tones could “highlight” the T3-T6 contrast in the medial position to be more salient for Mandarin listeners, compared with when T3-T6 occurred in the initial or final position with only one contextual tone.

Unlike Mandarin and Dutch listeners, Japanese listeners always favoured word-medial position to word-initial and word-final position. The prediction that Japanese listeners would not prefer any position was based on the assumption that they would perceive trisyllabic non-words as 3-mora words. From the observed pattern, however, it can be inferred that Japanese listeners perceive trisyllabic CVCVCV non-words not as 3-mora words, but as words of six moras or even longer, considering the duration of the word (Hoequist, 1983).[28] According to the native accentuation (Sato, 1993; Kubozono, 1998, 2006; Suzuki, 1995; Kitahara, 2001), the dominant position where pitch accent falls in 5- to 7-mora words is the antepenultimate mora, which is the medial position in this case.

To summarize the role of position at the phonological level, it was shown that position interacts closely with the perception of non-native pitch contrasts. All groups of listeners perceived the non-native pitch contour vs. pitch level contrast (T4-T6, falling tone vs. level tone) more accurately in word-final than word-initial or word-medial position, which can be presumably attributed to the influence of intonation across languages. The native word prosody and intonation prosody influenced each language group to display specific patterns depending on position when processing non-native pitch contrasts at the phonological level.

5.3.3.3 Influence of position at the lexical level

In the most cognitively demanding task, namely word learning, position was found to influence each language group differently, based on analyses for exploratory purposes. It affected Japanese listeners but not Mandarin and Dutch listeners when encoding non-native tones lexically. More specifically, Japanese listeners were

[28] One mora in Japanese is usually around 100-150 ms in normal tempo in natural speech (see Hoequist, 1983 for details). The trisyllabic non-words used in the study were each 950 ms, much longer than the phonetic duration of 3-mora words to Japanese ears. Thus, they may have perceived the trisyllabic non-words as consisting of six or more moras.

observed to prefer word-final to word-initial position only in the case of the falling tone T4, different from the prediction that they would be overall more sensitive to word-initial than word-final position based on the native word prosody. The fact that the influence of position only emerged in the encoding of T4 can perhaps be accounted for by intonation. The falling tone T4 in final position in a tonal pattern T3-T4 may resemble a falling pitch contour at the end of an utterance that signals assertive meaning (Bolinger, 1978, 1982; Cruttenden, 1981; Sugito, 1981; Poser, 1984; Pierrehumbert & Beckman, 1988). Japanese listeners may thus associate a falling tone in final position with a terminal fall in a statement, which may provide final T4 with salience and differentiate it from T4 in initial position. Note that proposing perceptual equivalence between non-native tone categories and native intonation categories is admittedly speculative.

Mandarin listeners, as predicted, showed no preference for any positions in the encoding of non-native tones in disyllabic non-words in word learning. Their comparably good performance in encoding the non-native tones occurring either word-initially or word-finally could be due to the use of lexical tones in the native language. Although Mandarin does not have lexical positional marking, the native lexical tones may allow them to make use of the tonal sensitivity to establish mental representations of non-native tones associated with meanings.

Although Dutch has a default right-hand word edge position for lexical stress, Dutch listeners were not predicted to display a bias for initial position in the encoding of non-native tones in word learning. Rather, they performed relatively poorly when encoding non-native tones in both initial and final position. Presumably due to a lack of lexical tones in the native language, they had difficulty building mental representations for non-native tones in sound-to-meaning mapping, which may have obscured the contrastive role of position in the native language and prevented it from affecting their word learning. This shows that positional perception of pitch at the lexical level seems to depend on pitch perception accuracy.

5.3.3.4 Integration of the influence of position at three processing levels

Altogether, position played different roles at different processing levels. In acoustic discrimination, position did not influence listeners when discriminating non-native pitch contrasts in trisyllables. Listeners all preferred word-final position to word-initial and medial positions, attributable to the recency effect. The influence of position was not robust in that listeners may discriminate the pitch information through an acoustic approach. However, when the acoustic level of representations is inaccessible (at the phonological level), position may jump into the foreground and take part in the perception of non-native pitch contrasts in trisyllables. Position, on the one hand, had a general influence on all groups of listeners in the perception of the non-native pitch contour vs. pitch level contrast (T4-T6). All three groups manifested more sensitivity in perceiving T4-T6 word-finally than word-initially and/or word-medially, which was attributable to the influence of intonation across languages.

On the other hand, position also exerted an effect on each language group differently. For Mandarin listeners, canonical tone language listeners, perception of specific non-native pitch contrasts depended on certain positions and such dependence was determined by the features of native tones. For instance, the contrastive contour tones in Mandarin benefited Mandarin listeners to perceive a non-native pitch contour contrast regardless of the position in which it occurred. However, presumably due to a lack of native contrastive level tones, they needed to depend on the word-medial position to build tonal references in order to achieve good performance when perceiving non-native level contrast. For Japanese listeners, who use the position of lexical pitch accent, the effect of position was licensed by the native pattern of lexical pitch accent, that is, the dominant accentual pattern (i.e., where pitch accent dominantly falls in words) determined their positional preference in the perception of non-native pitch contrasts. In contrast, for Dutch listeners, non-tone language listeners, the influence of position was active with a limited “licensing” from intonation. Although Dutch uses positional marking, the language lacks the use of tones. Position affected Dutch listeners only in the perception of the non-native pitch contour vs. pitch level contrast T4-T6, which may be due to the influence of intonation contours. However, the terminal intonation contour seemed to be unavailable to facilitate Dutch listeners’ perception of the non-native pitch contour contrast T4-T5 and the level contrast T3-T6 in word-final position. Although Braun & Johnson (2011) argued that Dutch listeners may link the non-native rising tone vs. falling tone in word-final position with sentential assertive and interrogative meaning, respectively, in AXB speeded tasks, the observed pattern of Dutch listeners in the current study indicates that such association might be unavailable at the phonological level. Moreover, the observed pattern seems to suggest that the mental representations of intonational categories are weaker than those of tone categories in the perception of non-native pitch contrasts at the phonological level.

In the encoding of non-native tones in sound-to-meaning mapping in disyllables, position might play a role only in Japanese listeners but not Dutch or Mandarin listeners, based on analyses of exploratory purposes. For Japanese listeners, position seemed influential only for the encoding of non-native falling tone T4, that is, Japanese listeners showed better performance with respect to word-final position than word-initial position. This could be due to the word-final falling tone being associated with the terminal falling contour in intonation, making it more easily differentiated compared to when it occurred in word-initial position. Mandarin listeners, despite their lack of positional cues in word prosody, were nevertheless adept at encoding non-native tones occurring either word-initially or word-finally. Their tonal experience in the native language benefited their encoding of non-native tones as they were able to do so robustly without being affected by the position in which they occurred in a word. In contrast, Dutch listeners, who do not use lexical tones in the native language, may have needed to refer to another pitch-based category, intonation. However, it seems that access to intonational categories was not available for Dutch listeners building mental representations of non-native tones. The positional perception of pitch at the phonological and lexical levels seems to depend on pitch perception overall. Dutch listeners never showed their predicted sensitivity to position as they likely could not perceive the tones in the first place.

To conclude, position had no influence on the perception of non-native pitch contrasts in trisyllables at the acoustic level for all three language groups. This seems to suggest that for the purpose of non-native pitch discrimination, the acoustic information in trisyllables may have sufficed for listeners without a need for additional information about where the pitch contrasts occurred. However, when processing taps into more abstract representations, i.e., at the phonological and lexical levels, the position where the non-native pitch contrasts occurred mattered. More specifically, the influence of position seems to be determined by the native word prosody. For instance, Mandarin does not use contrastive level tones. When perceiving non-native pitch level contrast T3-T6 in trisyllables at the phonological level, Mandarin listeners preferred word-medial position over word-initial and word-final position. Presumably due to a lack of contrastive level tones, Mandarin listeners may need tonal references to perceive non-native pitch level contrasts. The two contextual level tones may highlight T3-T6 in the medial position, making it more salient for Mandarin listeners, compared with when T3-T6 occurred in the initial and final positions with only one contextual tone. Furthermore, the finding that Dutch listeners were not influenced by position as was predicted at the phonological and lexical levels seems to indicate that the influence of position is licensed by pitch perception. If the listeners could not achieve good perception of non-native tones in the first place, the assistance of positional information seems to not be able to compensate at abstract processing levels.

5.3.4 Influence of native intonation prosody on the processing of non-native WPCs

From the findings it can be concluded that native intonation prosody plays a role, to some extent, in the processing of non-native WPCs at the acoustic, phonological, and lexical levels. I will now discuss the influence of native intonation prosody based on the overall performance across groups and the patterns observed within each language group, respectively.

To begin with, according to the overall performance at each processing level, Dutch non-tone language listeners showed a disadvantage compared to Mandarin and Japanese tone language listeners in the processing of non-native pitch contrasts in trisyllables and disyllables. This is presumably due to a lack of use of lexical pitch (i.e., low functionality of pitch) in their native language (Schaefer & Darcy, 2014). However, despite their comparatively worse performance, Dutch listeners' overall performance was still above chance level when processing non-native WPCs in acoustic discrimination, sequence recall, and in word learning. This could be due to the influence of native intonation prosody. Although Dutch does not have lexical tones, it has a rich repertoire of intonational nuclear tones (24 types) (Gussenhoven, 2004). This diversity of nuclear tonal patterns may allow them to perform better than chance when processing the non-native pitch contrasts at each level.

Regarding the patterns observed within each language group, at the acoustic level, Mandarin listeners showed comparably good performance discriminating non-native pitch contour and pitch level contrasts. Japanese listeners were good at discriminating most of the non-native pitch contrasts except for T5-T6. Finally, Dutch listeners were more sensitive to acoustically salient non-native pitch contrasts than acoustically non-salient ones. Their perceptual patterns in acoustic discrimination were accounted for by the influence of native use of lexical pitch and the acoustic salience of the non-native tones. It seems that at the acoustic level, intonation did not guide listeners' perceptual weighting on specific features of non-native tones.

However, intonation was found to play a role to some extent in determining listeners' (particularly, Dutch listeners') perceptual pattern at the phonological level. It was found that Mandarin listeners were better at processing non-native contour contrasts than level contrasts. Japanese listeners were more sensitive to non-native pitch contrasts paired with a non-native falling tone than other contrasts. Dutch listeners showed better performance when processing non-native pitch level contrasts than contour contrasts. The language specific patterns of Mandarin and Japanese listeners were presumably guided by the native tonal system in word prosody, i.e., the long-term storage of pitch-based representations of the contrastive contour tones in Mandarin listeners, lexical pitch fall in Japanese listeners (Francis et al., 2008; Lee et al., 1996; Xu et al., 2006. In contrast, due to a lack of lexical tones in the native language, Dutch listeners may have to draw upon native representations of pitch-based categories, specifically intonation patterns, to process non-native pitch contrasts. The nuclear tones in intonation consist of H and L tones (Gussenhoven, 2004), which may determine Dutch listeners' preference for non-native pitch level contrasts over contour contrasts.

Furthermore, intonation may have a general influence on tone language (Mandarin and Japanese) and non-tone language (Dutch) listeners when perceiving a specific non-native pitch contrast T4-T6 (low falling vs. low level) at the phonological level. All three groups preferred to perceive this contrast when it occurred word-finally rather than word-initially and/or word-medially. This could be presumably attributed to an influence of intonation. All three languages were observed to use falling contours at the end of an utterance to signal assertive meanings (Dutch: Gussenhoven & Rietveld, 1988; Japanese: Sugito, 1981; Poser, 1984; Pierrehumbert & Beckman, 1988; Mandarin: Shih, 1997, 2000; Lai et al., 2014). Listeners may perceptually equate the word-final falling tone T4 with a terminal falling contour signalling assertive intonation, rendering it distinctive enough to be differentiated from the word-final low level tone T6, compared with the same T4-T6 contrast occurring in word-initial and word-medial position. However, note that such perceptual equivalence between non-native tone categories and intonation categories is speculative since we did not conduct mapping experiments between non-native tones and intonation categories as was done in So & Best (2014).

At the lexical level, intonation seemed inactive. Mandarin listeners were more accurate in encoding non-native contour tones compared to level tones in sound-to-meaning mapping. Japanese listeners showed comparably good performance encoding both non-native contour tones and level tones in word learning. The learning patterns of Mandarin and Japanese listeners were accounted for by the influence of Mandarin lexical tonal categories and Japanese lexical accentual categories, respectively (see §4.4.3 and §5.3.2.3 for details). Nonetheless, Dutch listeners were comparably poor at encoding non-native contour tones and level tones lexically. That is, intonational nuclear tonal categories seem to be unavailable for Dutch listeners to reuse in word learning. This seems to suggest that the representations of intonational patterns are not as strong as those of lexical tonal categories with respect to listeners' integration of non-native tones with meanings in word learning.

To summarize, native intonation prosody may assist non-tone language listeners, i.e., Dutch listeners, in the processing of non-native pitch contrasts in trisyllables/disyllables at three processing levels, with overall performance above the chance level. However, presumably due to a lack of lexical tones (i.e., a low functionality of pitch) in the native language, Dutch listeners showed overall worse performance than tone language listeners, i.e., Mandarin and Japanese listeners, at all three processing levels. Moreover, it seems that neither native word prosody nor intonation prosody influenced listeners (of all three groups) to show relative sensitivity to features of non-native tones in acoustic discrimination. However, at more abstract phonological and lexical levels, the presence of tonal categories in Mandarin and accentual patterns in Japanese led these listeners to show relative preference for specific pitch contrasts. In contrast, Dutch listeners' perceptual pattern was guided by native intonation prosody but limited at the phonological level. Intonation prosody did not seem to cause Dutch listeners to show relative sensitivity in word learning.

5.4 General conclusion and suggestions for future research

This dissertation investigated how and to what extent native word prosody comes into play in the cross-linguistic processing of non-native pitch contrasts and position in acoustic discrimination (at the acoustic level), in a sequence recall task (at the phonological level), and in word learning (at the lexical level) by examining

Mandarin, Japanese, and Dutch listeners. It focused on how native word prosody in general affects the processing of non-native WPCs (overall performance across language groups), how language groups differ in their perceptual patterns with respect to specific pitch features (the specific patterns of each language group) and how position influences the processing of non-native pitch contrasts (interaction between position and non-native pitch contrasts). It addressed how native word prosody, sometimes with native intonation prosody, guides listeners from different language groups from a lower linguistic level, the acoustic level, to higher levels, the phonological and lexical levels, when processing non-native pitch contrasts and position, and provided evidence of both language-general and language-specific patterns in the processing of non-native WPCs.

Based on the findings, some suggestions can be made for future studies. First of all, this dissertation selected Cantonese tones (three contrastive level tones and two contour tones) as its stimulus materials. In an early stage of experimental design, Thai tones were considered as an option since Central Thai has five tones, viz. high vs. mid. vs. low level tones and rising vs. falling contour tones (Abramson, 1978; Burnham et al., 2015). However, considering that the sequence recall task in Chapter 3 requires different voices for the stimuli (three females and three males as in Dupoux et al., 2001, 2008), a large number of Thai native speakers were needed, more than could be found in our location in the Netherlands. Due to this practical issue, Cantonese was selected as it is, compared with Thai, spoken by a relatively large population in the Netherlands. Future studies can consider using Thai tones to investigate cross-linguistic processing of non-native WPCs at the three levels, provided enough Thai speakers can be recruited for the creation of stimuli. It is notable that tonal features *per se* are different in Cantonese and Thai. For instance, high level tone and mid level tone in Thai are not phonetically contrastive "high vs. mid". Thai high level tone shares similar onset with mid level tone and it rises towards the offset, which is in parallel to Thai rising tone (Abramson, 1978; Morén & Zsiga, 2006; Zsiga & Nitisaroj, 2007; Burnham et al., 2015) while the three level tones of Cantonese are relatively distinctive in acoustic space (Fok Chan, 1974; Bauer & Benedict, 1997; Qin & Mok, 2013).

Second, this dissertation investigated naive listeners of Mandarin, Japanese, and Dutch who employ different combinations of prosodic cues for lexical contrast and found that native word prosody (sometimes together with the native intonation prosody) played a pivotal role in the processing of non-native WPCs at different levels. To further understand the role of native word prosody from a typological perspective, it would be illuminating for future studies to include naive listeners whose native language does not use any word prosodic cues, such as French (Rossi, 1980; Di Cristo, 1988; Vaissière, 1991) or Seoul Korean listeners (Jun, 2007). This may provide a more comprehensive view of how the native word prosody and/or native intonation prosody together influence listeners' encoding of non-native WPCs, for instance, whether and/or to what extent listeners without the native use of WPCs (e.g., French or Seoul Korean listeners) are affected by their native intonation prosody when processing non-native WPCs acoustically, phonologically and lexically.

Third, this dissertation focused on naive listeners without any exposure to non-native tones and thus provided a baseline for the role native word prosody plays in listeners at the earliest stage of acquisition in the processing of non-native WPCs at different processing levels. It would be interesting to examine listeners with experience with the non-native tones (i.e., L2 learners with different proficiency levels) and investigate how and to what extent the native language can influence the encoding of non-native WPCs in the three processing levels by L2 learners in comparison with naive listeners.

Last but not least, Chapter 4 applied a picture selection task to investigate the encoding of non-native WPCs at the lexical level, revealing for each language group specific patterns of integrating non-native tones in disyllables to meanings in word learning. The findings were based on the listeners' responses, i.e., correct or incorrect and the type of errors they made. It would be interesting to conduct an online processing experiment, for example an eye-tracking study, which can probe the real-time dynamics of word learning in each language group.

Bibliography

Abe, I. (1998). Intonation in Japanese. In D. Hirst & A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press, pp. 360-375.

Abramson, A. S. (1975). Thai tones as a reference system. *Haskins Laboratories: Status Report on Speech Research*, 44, 127-136.

Abramson, A. S. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öhman (eds.), *Frontiers of Speech Communication Research*. New York: Academic Press. pp. 127-134.

Akinaga, K. (1985). Kyootuugo no akusento [Accent of standard Japanese]. In NHK (ed.), *Pronunciation and Accent Dictionary*. Tokyo: NHK.

Altmann, H. (2006). *The Perception and Production of Second Language Stress: A Cross-linguistic Experimental Study*. Ph.D. Dissertation, University of Delaware.

Altmann, H. & Vogel, I. (2002). L2 acquisition of stress: The role of L1. Paper presented at DGfS Annual Meeting “Multilingualism Today”, Mannheim, Germany, March 2002.

Archibald, J. (1995). The acquisition of stress. In J. Archibald (ed.), *Phonological Acquisition and Phonological Theory.* New Jersey: L. Erlbaum Associates. pp. 81-109.

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics*, 35, 167-181.

Asano, Y., & Braun, B. (2011). Integrating lexical and postlexical suprasegmental information in native and non-native Japanese. *NINJAL International Conference on Phonetics and Phonology (ICPP 2011)*. Kyoto, Japan.

Ayusawa, T. (2003). Acquisition of Japanese accent and intonation by foreign learners. *Journal of the Phonetic Society of Japan*, 7 (2), 47-58.

Ayusawa, T. & Nishinuma, Y. (1997). How do non-native speakers perceive Japanese pitch accent? Results from 10 language groups, *Travaux de l '1nstitut do Phonetique d'Aix*, 17, 185-199.

Ayusawa, T., & Odaka, K. (1998). 東京語アクセントの聞き取りテスト 21 言語グループの母語別・ 成績群別正答率. (Listening test for Tokyo Japanese accent: 21 language groups) 『*「国際社会における日本語についての総合的研究論文集 1』新プロ「日本語」総括班*, 57-71.

Ayusawa, T. et al. (1995). 鮎澤孝子･ 西沼行博･ 李明姫･ 荒井雅子･ 小高京子･ 法貴則子 「東京語アクセント聴取実験結果の分析 10 言語グループの結果一」(Perception of Tokyo Japanese accent: results from 10 language groups)『*「国際社会における日本語についての総合的研究」第 2 回研究報告会予稿集*』新プロ「日本語」研究成果報告書, 25- 32.

Ayusawa, T. et al. (1997a). 鮎澤孝子・ 西沼行博・ 楊立明・ 小高京子「北京語を母語とする日本語学習者の東京語アクセントの知覚」(Perception of Tokyo Japanese accent by Beijing L2 learners of Japanese)『*日本語教育論文集—小出詞子先生退職記念—*』東京: 凡人社, 69- 81.

Ayusawa, T. et al. (1997b). 鮎澤孝子・ 楊立明・ 磯村一弘・ 西沼行博・ 小高京子「北京語母語話者による東京語アクセントの知覚」(Perception of Tokyo Japanese accent by Beijing Mandarin speakers)『21 世紀の日本語音声教育に向けて』新プロ「日本語」研究成果報告書, l3-21.

Bao, H., & Lin, M. (2014). *Introduction of Experimental Phonetics*. Beijing: Peking University Press.

Bauer, R. S. & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Berlin: Mouton de Gruyter.

Beckman, M. E. (1986). *Stress and Non-stress Accent*. Dordrecht: Foris.

Berg, R. van den, Gussenhoven, C., & Rietveld, T. (1992). Downstep in Dutch: Implications for a Model. In G. Docherty & D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press. pp. 335-367.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore, MD: York Press. pp. 171-204.

Best, C. & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. Munro (eds.), *Second-Language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege*. Amsterdam: John Benjamins. pp. 13-34.

Boersma, P., & Weenink, D. (2008). *Praat: Doing phonetics by computer*. Institute of Phonetic Sciences, University of Amsterdam.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language*, 37, 83–96.

Bolinger, D. (1978). Intonation across languages. In J. H. Greenberg (ed.), *Universals of Human Language. Vol. 2: Phonology*. Stanford University Press. pp. 471-523.

Booij, G. E. (1995). *The Phonology of Dutch*. Oxford: Oxford University Press.

Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning,* 66 (4), 774-808.

Braun, B., Galts, T. & Kabak, B. (2014). Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. *Second Language Research,* 30 (3), 323-350.

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, 39, 585-594.

Burnham, De & Francis, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. S. Abramson (ed.), *Southeast Asian*

Linguistic Studies in Honour of Vichin Panupong. Bangkok: Chulalongkorn University Press. pp. 29-46.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In H. Bunnell & W. Idsardi (eds), *Proceedings of the Fourth International Conference on Spoken Language Processing.* Philadelphia. pp. 2514-2517.

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda F., Attina, V., Rattanasone, N. X., Schwarz, I-C., Webster, D. (2015). Universality and language-specific experience and in the perception of lexical tone and pitch. *Applied Psycholinguistics*, 36, 1459-1491.

Byun, H-G. (2018). Tokyo Accent perception ability for native speakers from different areas. *Journal of the Phonetic Society of Japan*, 22 (2), 1-21.

Cao, J. (1992). On neutral-tone syllable in Mandarin Chinese. *Canadian Acoustics*, 20 (3), 49-50.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustic Society of America*, 128, 456-465.

Chao, Y. R. (1948). *Mandarin Primer.* Cambridge: Harvard University Press.

Chao, Y. (1979). *A Grammar of Spoken Chinese*. Beijing: The Commercial Press.

Chen, A., Liu, L. & Kager, R. W. J. (2016). Cross-domain correlation in pitch perception: The influence of native language. *Language, Cognition and Neuroscience*, 31 (6), 751-760.

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian Speakers. *Journal of Speech, Language, and Hearing Research*, 57, 1468-1479.

Cohen, A., Collier, R. & 't Hart, J. (1982). Declination: Construct or intrinsic feature of speech pitch? *Phonetica*, 39, 254-273.

Connell, K., Hüls, S., Martínez-García, M.-T., Qin, Z., Shin, S., Yan, H., & Tremblay, A. (2018). English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning*, 68, 635-668.

Cooper, N., Cutler, A. & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45, 207-228.

Cooper, A, & Wang, Y. (2010). Cantonese tone word learning by tone and non-tone language speakers. Paper presented at *Interspeech, the 11th Annual Conference of the International Speech Communication Association.* Chiba, Japan, September 26-30.

Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *Journal of the Acoustical Society of America*, 134, 133-139.

Cruttenden, A. (1981). Falls and rises: meanings and universals. *Journal of Linguistics*, 17 (1), 77-91.

Cruttenden, A. (1994). *Intonation.* 2nd edition. Cambridge: Cambridge University Press.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29, 201-220.

Cutler, A., & Koster, M. (2000). Stress and lexical activation in Dutch. In B. Yuan, T. Huang, & X. Tang (eds.), *Proceedings of the Sixth International Conference on Spoken Language Processing*, Vol. 1. Beijing. pp. 593-596.

Cutler, A., & van Donselaar, W. (2001). *Voornaam* is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, 44, 171–195.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W. J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany. pp. 1913-1916.

Demany, L., & Semal, C. (2007). The role of memory in auditory perception. In W. A. Yost, A. N. Popper & R. R. Fay (eds.), *Auditory Perception of Sound Resources*. New York: Springer. pp. 77-113.

Di Cristo, A. (1988). Intonation in French. In D. Hirst & A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press. pp. 195-218.

Dommelen, W. A. & Husby, O. van. (2009). The perception of Norwegian word tones by second language speakers. In A. S. Rauber, M. A. Watkins & B. O. Baptista (eds.), *New Sounds 2007: Proceedings of the 5th International Symposium on the Acquisition of Second Language Speech*. University of Santa Catarina, Florianopolis, Brazil. pp. 165-174.

Donselaar, W. van, Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, 58, 251-273.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. (2nd edition). New York, NY: Oxford University Press.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A distressing "deafness" in French? *Journal of Memory and Language*, 36 (3), 406-421.

Dupoux, E., Peperkamp, S., & Sebastian-Galle, N. (2001). A robust method to study stress "deafness". *Journal of the Acoustical Society of America*, 110 (3), 1606-1618.

Dupoux, E., Sebastian-Galle, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition,* 106 (2), 682-706.

Fok Chan, Y. Y. (1974). *A Perceptual Study of Tones in Cantonese*. Centre of Asian Studies publications, 18. Hong Kong: Centre of Asian Studies, University of Hong Kong.

Fox, R. A., & Qi, Y.-Y. (1990). Context effects in the perception of lexical tones. *Journal of Chinese Linguistics*, 18, 261-284.

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

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36, 268-294.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1, 126-152.

Fujisaki, H., Ohno, S., & Tomita, O. (1996). On the levels of accentuation in spoken Japanese. In H. T. Bunnell & W. Idsardi (eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing (ICSLP'96)*, Vol. 2. Philadelphia: IEEE. pp. 634-637.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11 (2), 149-175.

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

Gottfried, T. L & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25 (2), 207-231.

Goss, S. J. (2015). *The Effects of Internal and Experience-Based Factors on the Perception of Lexical Pitch Accent by Native and Nonnative Japanese Listeners*. Unpublished doctoral dissertation. The Ohio State University.

Grabe, E., Rosner, B. S., Garcia-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech*, 46 (4), 375-401.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. Munro (eds.), *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins. pp. 57-77.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. (2005). Transcription of Dutch Intonation. In S.-A., Jun (ed.), *Prosodic Typology and Transcription: A Unified Approach*. Oxford: Oxford University Press. pp. 118-145.

Gussenhoven, C. (2000a). On the origin and development of the Central Franconian tone contrast. In A. Lahiri (ed.), *Analogy, Leveling, Markedness. Principles of Change in Phonology and Morphology*. Berlin: Mouton de Gruyter. pp. 215-260.

Gussenhoven, C. (2000b). The lexical tone contrast of Roermond Dutch in optimality theory. In M. Horne (ed.), *Prosody: Theory and Experiment*. Amsterdam: Kluwer. pp. 129-167.

Gussenhoven, C., & Rietveld, A. C. M. (1988). Fundamental frequency declination in Dutch: Testing three hypotheses. *Journal of Phonetics*, 16 (3), 355-369.

Gussenhoven, C. & Rietveld, A. (1992). Intonation contours, prosodic structure and preboundary lengthening. *Journal of Phonetics*, 20 (3), 283-303.

Haan, J., Pacilly, J. J. A., & Van Bezooijen, R. (1997). An anatomy of Dutch question intonation. In J. Coers & H. de Hoop (eds.), *Linguistics in the Netherlands 1997*. Amsterdam: John Benjamins. pp. 97-108.

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.

Haraguchi, S. (1977). *The tone pattern of Japanese: An autosegmental theory of tonology*.Tokyo: Kaitakusha.

Haraguchi, S. (1991). *A Theory of Stress and Accent*. Dordrecht: Foris.

Haraguchi, S. (1999). Accent. In N. Tsujimura (ed.), *The Handbook of Japanese Linguistics*. Oxford: Blackwell. pp. 1-30.

Herman, R. (1996). Final lowering in Kipare. *Phonology*, 13 (2), 171-196.

Heuven, V. J. van. (2017). Prosody and sentence type in Dutch. *Nederlandse Taalkunde*, 22 (1), 3-29.

Heuven, V. J. van. (1985). Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America*, 78, s21.

Heuven, V. J. van. (1988). Effects of stress and accent on the human recognition of word fragments in spoken context: Gating and shadowing. In W. A. Ainsworth & J. N. Holmes (eds.), *Proceedings of Speech '88, 7th FASE Symposium*. Edinburgh. pp. 811-818.

Heuven, V. J. van, & Haan, J. (2002). Temporal development of interrogativity cues in Dutch. In C. Gussenhoven & N. Warner (eds.), *Papers in Laboratory Phonology 7*. Berlin: Mouton de Gruyter. pp. 61-68.

Heuven, V. J. van, & Kirsner, R. S. (2004). Phonetic or phonological contrasts in Dutch boundary tones? In L. Cornpis & J. Doetjes (eds.), *Linguistics in the Netherlands 2004*. Amsterdam: John Benjamins. pp. 102-113.

Heuven, V. J. van., & Sluijter, A.M.C. (1996). Notes on the phonetics of word prosody. In: R. Goedemans, H. van der Hulst & E. Visch (eds.), *Stress Patterns of the World, Part 1: Background*. The Hague: Holland Academic Graphics. pp. 233-269.

Heuven, V. J. van., & Sluijter, A.M.C. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100 (4), 2471-2485.

Ho, A. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33 (5), 353-367.

Hoequist Jr., C. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, 40, 203-237.

Howie, J. M. (1976). *Acoustic Studies of Mandarin Vowels and Tones*. Cambridge: Harvard University Press.

Hu, S. (2015). *Cross-linguistic Perception of Pitch Position*. Master Thesis. Utrecht University.

Hulst, H. van der & Kooij, J. G. (1992). Main stress and secondary stress: Two modes of stress assignment. In W. U. Dressler, M. Prinmzhorn & J. R. Rennison (eds.), *Phonologica. Proceedings of the 7th International Phonology Meeting*. Turin: Drosenberg & Sellier. pp. 108-114.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, 23 (2), 225-257.

Hyman, L. M. (2009). How (not) to do phonological typology: The case of pitch-accent. *Language Sciences*, 31, 213-238.

Ishihara, S., Fan, Y., Jalil, D., & Tsurutani, C. (2011). A comparative study on perception of foreign-accented Japanese by L2 Japanese listeners having different L1 backgrounds: English, Chinese and Indonesian. In W.-S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong. pp. 950-953.

Ishikawa, K. & Nomura, J. (2008). Word stress placement by native speakers and Japanese learners of English. In D. Burnham (ed.), *Proceedings of Interspeech, the 9th Annual Conference of the International Speech Communication Association*. Brisbane, Australia. pp. 1955-1958.

Jongenburger, W. (1996). *The Role of Lexical Stress during Spoken-word Processing*. Ph.D. dissertation, University of Leiden. The Hague: Holland Academic Graphics.

Jongenburger, W., & Van Heuven, V. J. (1995a). The role of linguistic stress in the time course of word recognition in stress-accent languages. In J. M. Pardo (ed.), *Proceedings of the 4th European Conference on Speech Communication and Technology*. Madrid. pp. 1695-1698.

Jongenburger, W., & Van Heuven, V. J. (1995b). The role of lexical stress in the recognition of spoken words: Prelexical or postlexical? In K. Elenius & P. Branderud (eds.), *Proceedings of the 13th International Congress of Phonetic Sciences*. Stockholm. pp. 368-371.

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL: International Review of Applied Linguistics in Language Teaching*, 28 (2), 95-117.

Jun, S.-A. (2005). Korean intonational phonology and prosodic transcription. In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. pp. 201-229.

Kaan, E., Barkely, C. M., Bao, M., & Wayland R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC Neuroscience 2008*, 9: 53.

Kager, R. (1989). *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: Foris.

Kalashnikova, M., Mattock, K., & Monaghan, P. (2015). The effects of linguistic experience on the flexible use of mutual exclusivity in word learning. *Bilingualism*, 18, 626-638.

Kawahara, S. (2015). The phonology of Japanese accent. In H. Kubozono (ed.), *The Handbook of Japanese Phonetics and Phonology*. De Gruyter Mouton. pp. 445-492.

Kijak, A. M. (2009). *How Stressful is L2 Stress? A Cross-linguistic Study of L2 Perception and Production of Metrical Systems*. Ph.D. dissertation, Utrecht University. LOT School Publications.

Kitahara, M. (2001). *Category Structure and Function of Pitch Accent in Tokyo Japanese*. Ph.D. dissertation. Indiana University. Bloomington, IN.

Koster, M., & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in Dutch. *Proceedings of Eurospeech 97*. Rhodes. pp. 2167-2170.

Krishnan, A. & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, 110, 135-148.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brain stem. *Journal of Neurolinguistics*, 23, 81-95.

Krishnan, A., Swaminathan J., & Gandour, J. T. (2009). Representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, 21, 1092-1105.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25, 161-168.

Kubozono, H. (1993). *The Organization of Japanese Prosody*. Kuroshio, Tokyo.

Kubozono, H. (1998). Onin koozoo no fuhensee to kobetsusee [Universality and individuality of phonological structure]. In N. Minoru (ed.), *Onin Koozoo to akusento [Phonological Structure and Accent]*. Kuroshio, Tokyo. pp 2-108.

Kubozono, H. (2006). Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua*, 116, 1140-1170.

Kubozono, H. (2008). Japanese accent. In S. Miyagawa & M. Saito (eds.), *The Oxford Handbook of Japanese Linguistics*. Oxford: Oxford University Press. pp.165-191.

Kuijk, D. van. (1996). Does lexical stress or metrical stress better predict word boundaries in Dutch? *Proceedings of the Fourth International Conference of Spoken Language Processing*, Vol. 3. Philadelphia. pp.1585-1588.

Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology Yearbook*, 1, 53-74.

Ladd, R. D. (1996). *Intonational phonology*. New York: Cambridge University Press.

Lai, Y. (2008). *Acoustic realization and perception of English lexical stress by Mandarin learners*. Ph.D. dissertation. University of Kansas.

Lai, W., Li, Y., Che, H., Liu, S., Tao, J., & Xu, X. (2014). Final lowering effect in questions and statements of Chinese Mandarin based on a large-scale natural dialogue corpus analysis. In N. Campbell, D. Gibbon & D. Hirst (eds.), *Proceedings of the 7th International Conference on Speech Prosody*. Dublin, Ireland. pp. 653-657.

Leather, J. (1987). F0 pattern inference in the perceptual acquisition of second language tone. In A. James & J. Leather (eds.), *Sound Patterns in Second Language Acquisition*. Dordrecht: Foris. pp. 59-81.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (eds.), *New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech*. Amsterdam: University of Amsterdam. pp. 72-89.

Lee, Y., Vakock, D., & Wurm, L. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 125, 527-542.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Leyden, K. van, & Heuven, V. J. van. (1996). Lexical stress and spoken word recognition: Dutch versus English. In C. Cremers & M. den Dikken (eds.), *Linguistics in the Netherlands 1996*. Amsterdam: John Benjamins. pp. 159-170.

Li, A. (2002). Chinese Prosody and Prosodic Labelling of Spontaneous Speech. In S. Bel & I. Marlin (eds.), *Proceedings of Speech Prosody 2002 International Conference.* Aix-en-Provence, France. pp. 39-46

Li, A., Zu, Y., & Li, Z. (1999). A national database design and prosodic labelling for speech synthesis. Paper presented at the *1999 Oriental COCOSDA Workshop on East Asian Language Resources and Evaluation*. Taipei, Taiwan.

Li, B., & Shuai, L. (2011). Effects of native language on perception of level and falling tones. In W.-S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong. pp.1202-1205.

Li, C. N., & Thompson, S. A. (1979). The pragmatics of two types of yes-no questions in Mandarin and its universal implications. In W. F. H. Paul, R. Clyne & C. L. Hofbauer (eds.), *Papers from the Fifteenth Regional Meeting of the Chicago Linguistic Society*. Chicago, Illinois: Chicago Linguistic Society. pp. 197-206.

Li, Z. (1997). A preliminary study of the prosodic transcription system for standard Chinese. Paper presented at *The Third National Conference of Intelligence Interface and Its Application*. Zhangjiajie, China

Liberman, M. Y. & Pierrehumbert, J. B. (1984). Intonational invariance under changes in pitch range and length. In M., Aronoff, & R. Oehrle (eds.), *Language Sound Structure.* Cambridge, Massachusetts: MIT Press. pp. 157-233.

Lin, C. Y., Wang, M., Idsardi, W. J. & Xu, Y. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition*, 17, 316-346.

Liu, Fang. (2009). *Intonation Systems of Mandarin and English: A Functional Approach*. Ph. D. dissertation, University of Chicago.

Liu, L., Chen, A. & Kager, R. W. J. (2017). Perception of tones in Mandarin and Dutch adult listeners. *Language and Linguistics*, 18 (4), 622-646.

Lin, M.C. (1988). 普通话声调的声学特性和知觉征兆 [Tone characteristics and perception of Mandarin Chinese]. *中国语言 Zhongguo Yuyan*, 3, 182-193.

Logan & Pruitt, (1995). Methodological issues in training listeners to perceive nonnative phonemes: Linguistic experience and the "perceptual magnet effect." In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Baltimore: York Press. pp. 121-154.

Lu, J., & Wang, J. (2005). On defining "Qingsheng (neutral tone)". *Contemporary Linguistics*, 7, 107-112.

Maeda, S. (1976). *A Characterization of American English Intonation*. Ph.D. dissertation. Massachusetts Institute of Technology.

Maniwa, K. (2002). Acoustic and perceptual evidence of complete neutralization of word-final tonal specification in Japanese. *Kansas Working Papers in Linguistics*, 26, 93-112.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14, 57-77.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain meanings of words. *Cognitive Psychology*, 20, 121-157.

McCawley, J. D. (1968). *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton.

McCawley, J. D. (1977). Accent in Japanese. In L. Hyman (ed.), *Studies in Stress and Accent. Southern California Occasional Papers in Linguistics 4*. University of Southern California, Los Angeles, CA. pp. 261-302.

McCawley, J. D. (1978). What is a tone language? In V. A. Fromkin (ed.), *Tone: A Linguistics Survey*. New York: Academic Press. pp. 113-131.

Michaux, M. C., Brognaux, S., & Christodoulides, G. (2014). The production and perception of L1 and L2 Dutch stress. In N. Campbell, D. Gibbon & D. Hirst (eds.), *Proceedings of the 7th International Conference on Speech Prosody*. Dublin, Ireland. pp. 462-466.

Mochizuki-Sudo, M. and Kiritani, S. (1991). Production and perception of stress-related durational patterns in Japanese learners of English. *Journal of Phonetics*, 19, 231-248.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102, 1864-1877.

Morton, J., Crowder, R., & Prussin, H. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology*, 91, 169-190.

Morton, J., Marcus, S., & Ottley, P. (1981). The acoustic correlates of ‘‘speechlike’’: A use of the suffix effect. *Journal of Experimental Psychology: General*, 110, 568-593.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, 21, 81-108.

Nagano-Madsen, Y. (2003). Phonetic realization of the HL and LH accents in Japanese. In S. Kaji (ed.), *Proceedings of the Symposium “Cross-linguistic Studies of Tonal Phenomena”*. Tokyo: Tokyo Press. pp. 243-264.

Nishinuma, Y. (1994). How do the French perceive tonal accent in Japanese? Experimental evidence. In *Proceedings of the 3rd International Conference on Spoken Language Processing.* Yokohama, Japan. pp. 1739-1742.

Nguyen, A.-T. T., & Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, 39, 309-319.

Nguyen, A.-T. T., Ingram, J., & Pensalfini, R. (2008). Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, 36, 158-190.

Otake, T. & Cutler, A. (1999). Pitch accent in spoken word recognition in Japanese. *Journal of the Acoustical Society of America*, 105, 1977-1988.

Obler, L. (1989). Exceptional second language learners. In S. Gass, S. Madden, D. Preston & L. Selinker (eds.), *Variation in Second Language Acquisition: Volume II: Psycholinguistic Issues*. Clevedon: Multilingual Matters. pp. 141-159.

Orie, O. Ọ. (2006). L2 acquisition and Yoruba tones: Issues and challenges. In O. F. Arasanyin & M. A. Pemberton (eds.), *Selected Proceedings of the 36th Annual Conference on African Linguistics*. Somerville, MA: Cascadilla Proceedings Project. pp. 121-128.

Peng, S., Chan, M. K. M., Tseng, C-Y., Huang, T., Lee, O. J. & Beckman, M. (2005). Toward a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. pp. 230-270.

Peperkamp, S. & Dupoux, E. (2002). A typological study of stress “deafness”. *Laboratory Phonology*, 7, 203-240.

Peperkamp, S., Vendelin, I. & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, 38, 422-430.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America*, 130, 461-472.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, 66 (2), 363-369.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese Tone Structure.* Cambridge, MA: MIT Press.

Pierrehumbert, J., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, 46, 181-196.

Poltrock, S., Chen, H., Kwok, C., Cheung, H., & Nazzi, T. (2018). Adult learning of novel words in a non-native language: consonants, vowels, and tones. *Frontiers in Psychology*, 9:1211

Poser, W. J. (1984). *The Phonetics and Phonology of Tone and Intonation in Japanese*. Ph.D. dissertation. Massachusetts Institute of Technology.

Post, B. (2000). Pitch accents, liaison and the phonological phrase in French. *Probu*s, 12, 127-164.

Qin, Z., Chien, Y., Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Linguistics*, 38 (3), 541-570.

Qin Z. & Mok, P. (2013). Discrimination of Cantonese tones by speakers of tone and non-tone languages. *Kansas Working Papers in Linguistics*, 34, 1-25.

Quam, C., & Swingley, D. (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, 62, 135-150.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20 (3), 331-350.

Quené, H., & Koster, M. L. (1998). Metrical segmentation in Dutch: Vowel quality or stress? *Language and Speech*, 41, 185-202.

Ramachers, S., Brouwer, S., & Fikkert, P. (2017). How native prosody affects pitch processing during word learning in Limburgian and Dutch toddlers and adults. *Frontiers in Psychology*, 8:1652.

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception & Psychology*, 77, 571-591.

Rietveld, T., Kerkho, J., & Gussenhoven, C. (2004). Word prosodic structure and vowel duration in Dutch. *Journal of Phonetics*, 32, 349-371.

Rossi, M. (1980). Le français, langue sans accent? [French: A non-stressed language?]. In I. Fónagy, P. R. Léon, & F. Carton (eds.), *L'accent en français contemporain* [Stress in modern French]. Paris: Didier. pp. 13-51.

Sato, H. (1993). Kyootsuugo akusento no seiin bunseki [Accent analysis of common Japanese words]. *Nihon Onkyoo Gakkai Shi [Journal of the Acoustic Society of Japan]*, 49, 775-784.

Schwab, S. & Llisterri J. (2011). Are French speakers able to learn to perceive lexical stress contrasts? In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong. pp. 1774-1777.

Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistics*, 28, 439-454.

Shen, G., & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism: Language and Cognition*, 22 (2), 253-265.

Schaefer, V., & Darcy, I. (2014). Lexical function of pitch in the first language shapes cross-linguistic perception of Thai tones. *Laboratory Phonology*, 5 (4), 489-522.

Shibata, T. (1994). Gairaigo-ni okeru akusentokaku-no iti [The osition of accent nucleus in loanwords]. In K. Sato (Ed.), *Gendaigo/hougen-no kenkyuu* [The study of the modern Japanese language and dialects]. Tokyo: Meiji-Shoin. pp. 388-418.

Shibata, T., & Hurtig, R. (2008). Prosody acquisition by Japanese learners. In Z. Han (ed.), *Understanding Second Language Process.* Clevedon, UK: Multilingual Matters. pp. 176-204.

Shibata, T. & Shibata, R. (1990). Akusento wa doo'ongo o donoteido benbetsu shiuruno ka?: Nihongo, eigo, chuugokugo no baai [How much can accent distinguish homophones? Cases of Japanese, English and Chinese]. *Mathematical Linguistics*, 17, 317-327.

Shih, C. (1997). Declination in Mandarin. In G. Botinis, G. Kouroupetroglou, & G. Carayiannis (eds.), *Proceedings of ESCA Workshop on Intonation: Theory, Models and Applications*. Athens, Greece. pp. 293-296.

Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (ed.), *Intonation: Analysis, Modeling and Technology*. Amsterdam: Kluwer. pp. 243-268.

Showalter, C. E., & Hayes-Harb, R. (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, 29 (2), 185-200.

Shport, I. (2008). Acquisition of Japanese pitch accent by American learners. In P. Heinrich & Y. Sugita (eds.), *Japanese as Foreign Language in the Age of Globalization*. München: Iudicium Verlag. pp. 165-187.

Shport, I. (2011). *Cross-linguistic Perception and Learning of Japanese Lexical Prosody by English Listeners*. Ph.D. dissertation. University of Oregon, Eugene.

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science*, 17, 94-109.

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, 33, 47- 68.

Slowiaczek. M. (1991). Stress and context in auditory word recognition. *Journal of Psycholinguistic Research*, 20, 465-481.

Sluijter, A. M. C. (1995). *Phonetic Correlates of Stress and Accent*. Holland Institute of Generative Linguistics Dissertations, 15. The Hague: Holland Academic Graphics.

Sluijter, A. M. C & Heuven., V. J. van. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In A. M. C. Sluijter & V. J. van Heuven (eds.), *Proceedings of the International Congress of Spoken Language Processing*. Philadelphia. pp. 630-633.

Sugito, M. (1982). *Nihongo Akusento no Kenkyuu* [Studies on Japanese accent].Sanseido, Tokyo.

So, C. K. (2006). *Effects of L1 Prosodic Background and AV Training on Learning Mandarin Tones by Speakers of Cantonese, Japanese, and English.* Unpublished Ph.D. dissertation. Simon Fraser University.

So, C. K. (2010). Categorizing Mandarin tones into Japanese pitch-accent categories: The role of phonetic properties. In *Proceeding of Interspeech 2010 Satellite Workshop on Second Language Studies*. Tokyo, Japan. O1-3.

So , C. K. (2012). Cross-language categorization of monosyllabic foreign tones: Effects of phonological and phonetic properties of native language. In T. Stolz , N. Nau , & C. Stroh (eds.), *Monosyllables: From Phonology to Typology*, Berlin: Akademie Verlag. pp. 55-69.

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

So, C. K. & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics*, 47, 133-145.

So, C. K. & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36, 195-221.

Sohn, H.-M. (1999). *The Korean Language*. Cambridge: Cambridge University Press.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39, 456-466.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. H. Edwards & M. L. Zampini (eds.), *Phonology and Second Language Acquisition.* Amsterdam: John Benjamins. pp. 193-218.

Strik, H. & Boves, H. (1992). Control of fundamental frequency, intensity and voice quality in Speech. *Journal of Phonetics*, 20 (1), 15-25.

Sugahara, M. (2011). Identification of English primary stress and bias toward strong word-initial syllables: Native vs. Japanese listeners. In *Proceeding of the 17th International Congress of Phonetics Sciences*. Hong Kong. pp. 1918-1921.

Sugahara, M. (2016). Is Japanese listeners' perception of English stress influenced by the antepenultimate accent in Japanese? Comparison with English and Korean listeners. *Doshisha Studies in English*, 96, 61-111.

Sugiyama, Y. (2006). Japanese pitch accent: Examination of final-accented and unaccented minimal pairs. *Toronto Working Papers in Linguistic*s, 26, 73-88.

Sugiyama, Y. (2008). *The Nature of Japanese Pitch Accent: An Experimental Study*. Unpublished Ph.D. Dissertation. The State University of New York at Buffalo.

Suzuki, H. (1995). Minimal words in Japanese. In D. Audra, H. Rachel, L. Barbara, N. Barbara & P. Sheri (eds.), *CLS 31*, *The Main Session*, *Vol. 1*. Chicago Linguistic Society, Chicago, IL. pp. 448-463.

't Hart, J. & Cohen, A. (1973). Intonation by rule: A perceptual quest. *Journal of Phonetics*, 1, 309-327.

Trommelen, M., & Zonneveld, W. (1986). Dutch morphology: Evidence for the Righthand Head Rule. *Linguistic Inquiry*, 17, 147-169.

Trommelen, M., & Zonneveld, W. (1999). Word stress in English and Dutch. In H. van der Hulst (ed.), *Word Prosodic Systems of the Languages of Europe*. Berlin: Mouton de Gruyter. pp. 478-515.

Tokuma, S. (2003). Perception of English lexical stress by English and Japanese speakers: Effect of duration and "realistic" intensity change. In *Proceedings of Eurospeech 2003*. Geneva. pp. 2122-2124.

Tsukada, K., Kondo, M. & Sunaoka, K. (2016). The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience. *Journal of Second Language Pronunciation*, 2 (2), 225-252.

Tsurutani, C., Tsukada, K., Ishihara, S. (2010). Comparison of native and non-native perception of L2 Japanese speech varying in prosodic characteristics. In *Proceedings of Australasian International Conference on Speech Science and Technology*. Melbourne. pp. 122-125.

Turco, G., Dimroth C., & Braun, B. (2012) Intonational means to mark verum focus in German and French. *Language and Speech*, 56 (4), 461-491.

Utsugi, A., Koizumi, M., & Mazuka, R. (2010). The perception of non-native lexical pitch accent by speakers of "accentless" Japanese dialects. In *Proceedings of the Fifth Speech Prosody 2010*. Chicago, Illinois: University of Illinois. Paper 090.

Vaissière J. (1991). Rhythm, accentuation and final lengthening in French. In J. Sundberg, L. Nord & R. Carlson (eds.) *Music, Language, Speech and Brain. Proceedings of an International Symposium at the Wenner-Gren Center, Stockholm, 5–8 September 1990*. Wenner-Gren International Symposium Series, 59. pp. 108-120.

Vance, T. J. (1987). *An Introduction to Japanese Phonology*. Albany, NY: SUNY Press.

Vance, T. J. (1995). Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics*, 23 (4), 487-499.

Vance, T. J. (2008). *The Sounds of Japanese*. Cambridge: Cambridge University Press.

Veenker, T.J.G. (2012). *The Zep Experiment Control Application.* Utrecht Institute of Linguistics OTS, University University.

Uwano, Z. (1999). Classification of Japanese accent systems. In S. Kaji, (ed.), *Cross-linguistic Studies of Tonal Phenomena, Tonogenesis, Typology, and Related Topics*, Tokyo: ILCAA. pp. 151-186.

Wang, J. (1997). The representation of the neutral tone in Chinese Putonghua. In J. Wang & N. Smith (eds.), *Studies in Chinese Phonology*. Berlin: Mouton de Gruyter. pp. 154-183.

Wang, Q. (2008). *Perception of English Stress by Mandarin Chinese Learners of English: An Acoustic Study*. PhD. dissertation. University of Victoria.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone production before and after perceptual training. *Journal of the Acoustical Society of America*, 113, 1033-1042.

Wang, Y., Sereno, J. A. & Jongman, A. (2003). FMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience*, 15 (7), 1019-1027.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, A. J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649-3658.

Warner, N. (1997). Japanese final-accented and unaccented phrases. *Journal of Phonetics*, 25, 43-60.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54, 681-712.

Wayland, R. P., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics*, 38, 654-662.

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language and Hearing Research*, 46, 413-421

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28, 565-585.

Wong, P. C. M., 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, Y-P. (2019). Perception of Cantonese lexical tones by Japanese non-learners and learners of Mandarin. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia. pp. 3867-3871.

Wu, X. H., & Lin, H. (2008). Perception of Mandarin tones by Mandarin and English listeners. *Journal of Chinese Language and Computing*, 18 (4), 175-187.

Wu, X. H., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, 46, 86-100.

Wu, X., Tu, J. Y., & Wang, Y. (2012). Native and nonnative processing of Japanese pitch accent. *Applied Psycholinguistics*, 33, 623-641.

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 Acoustic Society of America*, 120, 1063-1074.

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

Yoneyama, K. (2002). *Phonological Neighborhoods and Phonetic Similarity in Japanese Word Recognition*. Unpublished Ph.D. dissertation. The Ohio State University.

Yu, V. Y., & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research*, 39, 323-344

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society*, 363, 1087-1104.

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, 38, 260-271.

Zhu, H. Y. (2009). Neutral tone's character and its regulative principles. *Applied Linguistics*, 2, 34-41

Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27, 3-42.

Zsiga, E., & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech*, 50, 343-383

Zou, T. (2017). *Production and perception of tones by Dutch learners of Mandarin*. Ph.D. dissertation. Leiden University. LOT School Publication.

Appendices

Appendix A: Rating for "naturalness" of non-words

The /bababa/ sequence (in §2.2.2) was presented to five Cantonese native speakers, four Mandarin native speakers, two Japanese native speakers and four Dutch native speakers in four conditions: no silence inserted among each syllable in /bababa/, 15 ms, 25 ms, 35ms and 50ms silence inserted among each syllable in /bababa/, respectively. They were required to score to what extent /bababa/ sounded like a natural word on a 5-point scale (where 1 means 'not word-like' and 5 'extremely word-like'). Table A.1 shows the number of participants in each language group who rated the words.

Table A.1. Rating scores for the naturalness of the word /bababa/.

		Rating for the naturalness of the word /bababa/				
Silence types	**Number of participant**	1	2	3	4	5
No silence	Cantonese	1	3	1		
	Mandarin	2	1	1		
	Dutch	2	2			
	Japanese	1	1			
15ms	Cantonese	1	2	2		
	Mandarin		3	1		
	Dutch	1	2	1		
	Japanese		1	1		
25ms	Cantonese			3	2	
	Mandarin			1	2	1
	Dutch			2	1	1
	Japanese			1	1	
35ms	Cantonese	1	2	2		
	Mandarin	1	1	2		
	Dutch	1	2	1		
	Japanese		1	1		
50ms	Cantonese	2	2	1		
	Mandarin	1	2	1		
	Dutch	1	1	2		
	Japanese		2			

The /kupi/ sequence (in §4.2.2) was presented to five Cantonese native speakers, four Mandarin native speakers, two Japanese native speakers and four Dutch native speakers in four conditions: no silence inserted among each syllable in /kupi/, 15 ms, 25 ms, 35ms and 50ms silence inserted among each syllable in /kupi/, respectively. They were required to score to what extent /kupi/ sounded like a natural word on a 5-point scale (where 1 means 'not word-like' and 5 'extremely word-like'). Table A.2 shows the number of participants in each language group who rated the words.

Table A.2. Rating scores for the naturalness of the word /kupi/.

		Rating for the naturalness of the word /kupi/				
Silence types	**Number of participant**	1	2	3	4	5
No silence	Cantonese	1	2	2		
	Mandarin	1	2	1		
	Dutch	2	1	1		
	Japanese	1	1			
15ms	Cantonese	1	2	2		
	Mandarin	1	2	1		
	Dutch	1	1	2		
	Japanese	1		1		
25ms	Cantonese		1	2	1	1
	Mandarin			1	2	1
	Dutch			2	1	1
	Japanese			1	1	
35ms	Cantonese	2	2	1		
	Mandarin	1	1	2		
	Dutch	2	1	1		
	Japanese		1	1		
50ms	Cantonese	2	3			
	Mandarin	2	1	1		
	Dutch	3		1		
	Japanese	1	1			

Appendix B: Parameter estimates

Table B.1: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AX discrimination task in Experiment 2 (perception of non-native pitch contrasts in trisyllables) in Chapter 2.

Fixed effects	**Est.**	**Std. Error**	**t**	**Sig.**	**95% CI**	
					Lower	**Upper**
Intercept	2.513	.689	3.649	.000	1.163	3.863
lan=D	-2.837	.375	-7.560	.000	-3.573	-2.101
lan =J	-2.281	.376	-6.062	.000	-3.018	-1.543
con=1	.876	.974	.900	.368	-1.033	2.785
con=2	3.649	1.409	2.590	.010	.888	6.410
con =3	5.124	1.681	3.049	.002	1.829	8.418
con =4	2.105	1.126	1.870	.062	-.102	4.311
con =5	1.983	1.028	1.929	.054	-.032	3.998
con =6	2.261	1.014	2.231	.026	.275	4.248
con =7	.039	.944	.042	.967	-1.812	1.891
con =8	2.159	.968	2.232	.026	.262	4.056
con =9	-.061	.927	-.066	.948	-1.877	1.756
pos=1	-.735	.914	-.804	.421	-2.526	1.056
pos=2	-.451	.918	-.491	.623	-2.249	1.348
[lan=D]*[pos=1]	.348	.269	1.295	.195	-.179	.875
[lan=D]*[pos=2]	.195	.285	.684	.494	-.364	.754
[lan=J]*[pos=1]	-.172	.285	-.604	.546	-.732	.387
[lan=J]*[pos=2]	-.617	.300	-2.056	.040	-1.205	-.029
[lan=D]*[con =1]	2.575	.422	6.100	.000	1.748	3.403
[lan=D]*[con=2]	.312	1.092	.286	.775	-1.828	2.452
[lan=D]*[con =3]	.528	1.107	.477	.634	-1.643	2.698
[lan=D]*[con =4]	1.607	.733	2.193	.028	.171	3.044
[lan=D]*[con =5]	1.810	.484	3.739	.000	.861	2.759
[lan=D]*[con =6]	.528	.460	1.150	.250	-.372	1.429
[lan=D]*[con =7]	2.842	.346	8.214	.000	2.164	3.520
[lan=D]*[con =8]	-.478	.404	-1.182	.237	-1.270	.315
[lan=D]*[con =9]	1.690	.318	5.321	.000	1.067	2.312
[lan=J]*[con =1]	3.265	.468	6.973	.000	2.347	4.183
[lan=J]*[con =2]	.565	1.107	.510	.610	-1.604	2.734
[lan=J]*[con =3]	.862	1.120	.770	.441	-1.332	3.057
[lan=J]*[con =4]	2.947	.955	3.086	.002	1.075	4.820
[lan=J]*[con =5]	1.767	.480	3.680	.000	.826	2.708
[lan=J]*[con =6]	1.273	.472	2.695	.007	.347	2.198
[lan=J]*[con =7]	3.227	.354	9.120	.000	2.533	3.921
[lan=J]*[con =8]	-.313	.403	-.775	.438	-1.103	.478
[lan=J]*[con =9]	2.501	.325	7.696	.000	1.864	3.139

[pos=1]*[con=1]	-.138	1.310	-.105	.916	-2.705	2.429
[pos=2]*[con =1]	.289	1.322	.218	.827	-2.302	2.879
[pos=1]*[con =2]	-.053	1.376	-.039	.969	-2.750	2.644
[pos=2]*[con=2]	1.710	1.539	1.111	.267	-1.308	4.727
[pos=1]*[con =3]	-1.417	1.669	-.849	.396	-4.688	1.853
[pos=2]*[con =3]	-1.116	1.683	-.663	.507	-4.414	2.182
[pos=1]*[con =4]	.549	1.387	.396	.692	-2.170	3.268
[pos=2]*[con =4]	2.305	1.673	1.378	.168	-.974	5.583
[pos=1]*[con=5]	-.379	1.334	-.284	.776	-2.995	2.236
[pos=2]*[con=5]	.214	1.353	.158	.874	-2.437	2.866
[pos=1]*[con =6]	-.730	1.304	-.559	.576	-3.287	1.827
[pos=2]*[con =6]	.284	1.322	.215	.830	-2.307	2.874
[pos=1]*[con =7]	-.902	1.285	-.702	.483	-3.421	1.617
[pos=2]*[con =7]	-.081	1.294	-.063	.950	-2.618	2.456
[pos=1]*[con =8]	-.669	1.273	-.526	.599	-3.164	1.826
[pos=2]*[con =8]	-1.095	1.271	-.862	.389	-3.587	1.397
[pos=1]*[con =9]	-.805	1.267	-.636	.525	-3.289	1.679
[pos=2]*[con =9]	-.463	1.269	-.365	.715	-2.950	2.024

Note: Mandarin group is the reference language group; tonal contrast 10, position 3 (word-final position) are the reference categories.

lan: language; pos: position; con: tonal contrast

D: Dutch group, J: Japanese group.

Position 1: word-initial position, Position 2: word-medial position, Position 3: word-final position.

Tonal contrast 1 to 10: T1-T3, T1-T4, T1-T5, T1-T6, T3-T4, T3-T5, T3-T6, T4-T5, T4-T6, T5-T6, respectively.

Table B.2: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the AX discrimination task in Experiment 3 (perception of non-native positional tones in trisyllables) in Chapter 2.

Fixed effects	**Est.**	**Std. Error**	**t**	**Sig.**	**95% Confidence Interval**	
					Lower	**Upper**
Intercept	2.045	1.173	1.744	.081	-.254	4.344
lan=D	-1.674	1.185	-1.413	.158	-3.997	.649
lan=J	-.193	1.346	-.143	.886	-2.832	2.447
T=1	2.578	1.510	1.708	.088	-.382	5.538
T=3	2.010	1.334	1.506	.132	-.606	4.626
T=4	1.902	1.270	1.498	.134	-.588	4.392
T=5	.464	1.157	.401	.689	-1.805	2.733
poscon=1	.763	.736	1.037	.300	-.680	2.207
poscon=2	.136	.636	.213	.831	-1.111	1.383
[poscon=1]*[lan=D]	-.653	.787	-.829	.407	-2.195	.890
[poscon=2]*[lan=D]	.396	.704	.563	.574	-.985	1.777
[poscon=1]*[lan=J]	-.838	.868	-.966	.334	-2.540	.863
[poscon=2]*[lan=J]	-.345	.784	-.440	.660	-1.883	1.192
[T=1]*[lan=D]	.353	1.552	.227	.820	-2.691	3.397
[T=3]*[lan=D]	.716	1.359	.527	.598	-1.948	3.381
[T=4]*[lan=D]	.095	1.258	.076	.940	-2.371	2.562
[T=5]*[lan=D]	-.873	1.176	-.743	.458	-3.179	1.432
[T=1]*[lan=J]	1.888	1.135	.013	.990	-2.577	3.353
[T=3]*[lan=J]	.699	1.596	.438	.661	-2.430	3.829
[T=4]*[lan=J]	.409	1.457	.281	.779	-2.448	3.266
[T=5]*[lan=J]	-.968	1.319	-.734	.463	-3.554	1.618
cont=1	4.086	.927	4.410	.000	2.269	5.902
cont=3	2.610	.412	6.332	.000	1.802	3.418
[T=1]*[cont=3]	-.911	1.201	-.758	.448	-3.265	1.443
[T=3]*[cont=1]	-3.045	1.171	-2.600	.009	-5.342	-.749
[T=4]*[cont=1]	-3.006	1.093	-2.751	.006	-5.149	-.863
[T=4]*[cont=3]	-1.893	.670	-2.823	.005	-3.207	-.578
[T=5]*[cont=1]	-.350	1.044	-.335	.738	-2.397	1.697

Note: Mandarin group is the reference language group; Tone 6, positional contrast 3 (word-medial vs. word-final), contextual tone T6 are the reference categories.
lan: language; T: tone; poscon: positional contrast; cont: contextual tone
D: Dutch group, J: Japanese group.

Positional contrast 1: word-initial vs. word-medial position, Position 2: word-initial vs. word-final position, Position 3: word-initial vs. word-final position.
Contextual tone 1: T1; Contextual tone 3: T3; Contextual tone 6: T6

Table B.3: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the sequence recall task in Chapter 3.

Fixed effects	**Est.**	**Std. Error**	**t**	**Sig.**	**95% CI**	
					Lower	**Upper**
Intercept	.681	.250	2.724	.006	.191	1.171
lan=D	-.638	.2840	-2.248	.025	-1.195	-.082
lan=M	.720	.289	2.493	.013	.154	1.285
con=1	2.285	.212	10.790	.000	1.870	2.700
con =2	1.350	.229	5.898	.000	.902	1.799
con =3	1.969	.231	8.509	.000	1.516	2.423
[con=1]*[lan=D]	-.040	.122	-.324	.746	-.280	.200
[con=2]*[lan=D]	-1.150	.109	-10.546	.000	-1.364	-.936
[con=3]*[lan=D]	-1.100	.110	-10.030	.000	-1.315	-.885
[con=1]*[lan=M]	-.485	.152	-3.190	.001	-.783	-.187
[con=2]*[lan=M]	.188	.1420	1.328	.184	-.090	.467
[con=3]*[lan=M]	-.437	.130	-3.368	.001	-.692	-.183
pos=1	-.573	.224	-2.560	.010	-1.012	-.134
pos=2	1.705	.231	7.394	.000	1.253	2.157
[pos=1]*[lan=D]	.532	.099	5.396	.000	.339	.725
[pos=2]*[lan=M]	-1.413	.114	-12.347	.000	-1.637	-1.188
[pos=1]*[lan=D]	1.346	.124	10.885	.000	1.104	1.589
[pos=2]*[lan=M]	-.563	.134	-4.187	.000	-.826	-.299
[con=1]*[pos=1]	-.769	.254	-3.027	.002	-1.267	-.271
[con=1]*[pos=2]	-1.120	.245	-4.563	.000	-1.601	-.639
[con=2]*[pos=1]	-.618	.298	-2.074	.038	-1.202	-.034
[con=2]*[pos=2]	-.829	.305	-2.717	.007	-1.427	-.231
[con=3]*[pos=1]	-1.473	.309	-4.774	.000	-2.078	-.868
[con=3]*[pos=2]	-2.111	.311	-6.791	.000	-2.721	-1.502

Note: Japanese group is the reference language group; contrast 4 (T3-T6), position 3 (word-final position) are the reference categories.

lan: language; pos: position; con: contrast

D: Dutch group, M: Mandarin group.

Position 1: word-initial position, Position 2: word-medial position, Position 3: word-final position.

Contrast 1 to 4: segmental, T4-T5, T4-T6, T3-T6, respectively.

Table B.4: Parameter estimate, standard error, *t*-value, *p*-value, and 95% confidence interval of the predictors for the picture selection task in Chapter 4.

Fixed effects	**Est.**	**Std. Error**	**t**	**Sig.**	**95% CI**	
					Lower	**Upper**
Intercept	.939	.284	3.304	.001	.382	1.497
lan=D	-1.997	.378	-3.171	.002	-1.938	-.457
lan=J	-.338	.388	-.871	.384	-1.098	.422
T=4	.773	.300	2.582	.010	.186	1.360
T=5	.353	.289	1.221	.222	-.214	.919
pos=1	.171	.257	.667	.505	-.332	.674
[pos=1]*[T=4]	-.793	.289	-2.747	.006	-1.359	-.227
[pos=1]*[T=5]	-.266	.285	-.933	.351	-.826	.293
[lan=1]*[pos=1]	-.144	.267	-.539	.590	-.666	.379
[lan=2]*[pos=1]	.426	.290	1.469	.142	-.143	.995
[T=4]*[lan=D]	-.145	.344	-.421	.674	-.819	.530
[T=5]*[lan=D]	-.117	.339	-.346	.730	-.782	.547
[T=4]*[lan=J]	.418	.373	1.120	.263	-.314	1.150
[T=5]*[lan=J]	.538	.368	1.462	.144	-.184	1.260

Note: Mandarin group is the reference language group; Tone 6, position 2 (word-final position) are the reference categories.

lan: language; pos: position; T: tone

D: Dutch group, M: Mandarin group.

Position 1: word-initial position, Position 2: word-final position.

Appendix C: Data in pilot study in Chapter 3

Chapter 3 (Processing of non-native WPCs at the phonological level) conducted a pilot study with regard to the selection of pairs that were optimally and sub-optimally discriminated in AX discrimination task in Chapter 2. The data are presented in the following tables.

Table C.1.Summary of generalized linear mixed effect models for optimally pairs in acoustic discrimination

Fixed effects	**F**	**df**	**p**	
Language group	0.165	2	<0.001	
Contrast	0.004	3	0.992	
Position	0.002	2	0.998	
Sequence	6.775	2	0.001	
Language group * Contrast	0.724	4	0.485	
Language group * Position	0.992	4	0.440	
Contrast * Position	0.913	4	0.455	
Random effects	**Est.**	**Std. Error**	**z**	**p**
1 | Participant	0.125	0.068	1.838	0.066
1 | Item	0.377	0.105	3.599	<0.001

Table C.2. Performance of each language group in optimal pairs.

Pairs optimally perceived at the acoustic level			
contrast	language group	accuracy	Std. Error
T1-T3	Dutch	0.873	0.024
	Japanese	0.901	0.017
	Mandarin	0.896	0.031
T1-T4	Dutch	0.850	0.033
	Japanese	0.869	0.027
	Mandarin	0.967	0.011
T1-T5	Dutch	0.902	0.022
	Japanese	0.903	0.022
	Mandarin	0.908	0.021
T1-T6	Dutch	0.925	0.018
	Japanese	0.930	0.010
	Mandarin	0.917	0.008

T3-T4	Dutch	0.887	0.024
	Japanese	0.893	0.024
	Mandarin	0.956	0.002
T3-T5	Dutch	0.871	0.029
	Japanese	0.846	0.023
	Mandarin	0.953	0.019

Table C.3 Summary of generalized linear mixed effect models for suboptimal pairs.

Fixed effects	**F**	**df**	**p**
Language group	10.498	2	<0.001
Contrast	10.368	2	<0.001
Position	31.843	2	<0.001
Sequence	15.177	2	<0.001
Language group * Contrast	13.723	4	<0.001
Language group * Position	3.338	4	0.001
Contrast * Position	15.254	4	<0.001

Random effects	**Est.**	**Std. Error**	**z**	**p**
1 | Participant	1.018	0.440	2.294	0.022
1 | Item	0.129	0.062	2.085	0.037

Table C.4 Performance of each language group in sub-optimal pairs.

Pairs sub-optimally perceived at the acoustic level			
contrast	language group	accuracy	Std. Error
T4-T5	Dutch	0.565	0.116
	Japanese	0.827	0.035
	Mandarin	0.924	0.013
T4-T6	Dutch	0.475	0.110
	Japanese	0.879	0.045
	Mandarin	0.927	0.034
T3-T6	Dutch	0.573	0.114
	Japanese	0.811	0.074
	Mandarin	0.882	0.051

Appendix D: Working memory capacity test

Below presents the working memory capacity test (the backward digital span and phonological memory). All the participants were tested on the four sets.

Backward digital span test

Set 1

1 6

5 3 7

2 1 5 4

7 8 3 5 2

4 3 9 6 7 1

6 0 1 2 5 3 9

2 4 3 1 0 7 8 5

Set 2

5 9

2 1 8

7 3 4 6

3 8 6 2 5

4 0 3 5 8 1

1 6 5 7 3 9 4

8 3 1 9 7 5 6 2

Phonological memory test

Set 1

ba gu

mi ku ga

do ke ta pi

be mu to ki ma

ma ko le di pu no

ki pa be go mi tu ga

Set 2

tu pi

ta bi ku

bu pi ge da

li gu pa bo me

ho bi ka te ni pu

ba ku pi ga ko tu pe

Appendix E: Stimuli list

E.1 Stimuli in detail in picture selection task in word learning (Chapter 4).

	Word /kupi/ (tonal patterns)		
Target	**Tonal competitor**	**Positional competitor**	**Distracter**
T4T3	T5T3	T3T4	T3T6
T4T3	T6T3	T3T4	T3T5
T4T3	T3T3	T3T4	T3T6
T5T3	T4T3	T3T5	T3T6
T5T3	T6T3	T3T5	T3T4
T5T3	T3T3	T3T5	T3T6
T6T3	T4T3	T3T6	T3T5
T6T3	T5T3	T3T6	T3T4
T6T3	T3T3	T3T6	T3T5
T3T4	T3T5	T4T3	T6T3
T3T4	T3T6	T4T3	T6T3
T3T4	T3T3	T4T3	T3T5
T3T5	T3T4	T5T3	T4T3
T3T5	T3T6	T5T3	T6T3
T3T5	T3T3	T5T3	T4T3
T3T6	T3T4	T6T3	T4T3
T3T6	T3T5	T6T3	T4T3
T3T6	T3T3	T6T3	T5T3
Baseline		**Tonal competitor**	
T3T3	T6T3	T3T4	T5T3
T3T3	T3T6	T4T3	T3T5
T3T3	T3T6	T6T3	T4T3

Appendix F: Instructions of experiments

F.1 Experiments in Chapter 2 (Perception of non-native WPCs at the acoustic level)

(AX discrimination task)

English version

Dear participant,

Thank you for participating in this experiment.

The experiment has three blocks. In each block, you are going to hear sounds from the speakers. The sounds differ in “melody” and come in pairs. You will hear two sounds each time.

Your task is to judge whether the two sounds are the same or different as soon as possible. To respond, hit the left button on the button-box for “Same” and the right button for “Different".

If you have any questions, please ask the experimenter.

Now you are going to do Block 1. Hit any button on the button-box when you are ready to start.

In the first block, you are going to hear all the sounds in monosyllables.

First we will do a short practice. Hit any button on the button-box when ready to start.

You can take a break for 3 minutes between each block.

After the first break, you are going to do the next two blocks where you are going to hear all the sounds in trisyllables.

Dutch version:

Beste deelnemer,

Bedankt voor uw deelname aan dit experiment.

Het experiment heeft drie blokken. In elk blok hoort u geluiden uit de luidsprekers.

De geluiden verschillen in “melodie” en komen in paren. U hoort elke keer twee geluiden. Het is uw taak om zo snel mogelijk te beoordelen of de twee geluiden hetzelfde of verschillend zijn.

Om te reageren, drukt u op de linkerknop op het knoppen paneel voor " Same " en de rechterknop voor "Different".

Als u vragen heeft, stel deze dan aan de onderzoeker.

Nu gaat u Blok 1 doen.

In het eerste blok zijn alle geluiden één lettergreep.

We zullen eerst een korte oefening doen. Druk op een willekeurige knop op het knoppen paneel wanneer u klaar bent om te starten.

Je kunt tussen elk blok 3 minuten pauze nemen. Na de eerste pauze gaat u het volgende twee blokken doen, waar alle geluiden drie lettergrepen hebben.

Druk op een willekeurige knop op het knoppen paneel wanneer u klaar bent om te starten.

Mandarin version:

感谢您来参加实验。

本次实验有三个部分。在每个部分，您每次将听到两个声音。您需要尽可能快地判断这两个声音是否“音律”相同。相同请按左侧按钮（标有“same 相同”），不同请按右侧的按钮（标有“Different 不同”）。

如您有任何问题，请询问实验者。

现在您将进行第一部分。请按任意的按钮开始。

在第一部分中，您将听到的单音节的声音。首先您将进行一个练习。请按任意的按钮开始。

在每个环节结束后，您可以休息三分钟。

第一部分结束后，在接下来的第二和第三部分中，您将听到三音节的声音。

Japanese version:

研究協力者のかたへ、

私の実験にご協力くださり、ありがとうございます。実験のご説明をいたします。

この実験は 3 つのセッションから構成されています。各セッションの手順は同じです。

ヘッドフォンから単語が 2 つ、続けざまに流れますので、2 つの単語の「メロディー」が同じかどうかを判断してください。

２つの単語が同じだと思った場合は Same「同じ」と書かれたをボタンを押してください。

2 つの単語が異なると判断した場合は Different「異なる」と書かれたボタンを押してください。２つの単語が同じかどうかの判断は、語が流れたあとできるだけ速やかに行ってください。

ここまでで何かご質問がありましたら、お聞きください。

セッション 1 では、単音節の単語が聞こえます。"

最初にまず練習問題を行います。

練習を始める用意ができたら、ボタンを押してください。どのボタンでも結構です。各セッションの間に 3 分間の休憩をとることができます。

最初の休憩の後、次のセッションを行います。次のセッションでは、三音節の単語が聞こえます。

F.2 Experiment in Chapter 3 (Processing of non-native WPCs at the phonological level)

(sequence recall task)

The 12 pairs of words in the sequence-recall task in Chapter 3 were divided into 6 pairs each time. Participants were required to come to the lab two times. The procedure of each pair was identical.

English version:

Dear participant,

Thank you for participating in this experiment.

In today's experiment, you will learn 6 pairs of words. The procedure in learning each pair is the same.

Now you are going to learn the two new words A and B in Pair 1. The two words differ either in "melody" or in segment. You need to follow the instructions on the screen through the whole experiment.

If you have any questions, please ask the experimenter.

If you are ready, please press any button to start the experiment.

Please press button A and B to learn word A and B. You can press and listen to the two words as many times as you wish until you are sure you have learned them.

You have already learned word A and word B. Now you are going to do a practice.

In the practice, you will hear a word, and you need to differentiate whether the word is A or B by pressing the corresponding button.

You can only press the button after you see "please press the button" on the screen.

A sound of "OK" will follow your response. You will get a feedback of correct or wrong after your response.

You will proceed to the next stage after you make 15 times correct response. Otherwise, you have to restart to do it again until you reach the criteria.

Well done in the first practice! Now you are going to do a warm up!

You will hear word A and B in order.

For instance, if you hear word B followed by word A, press the corresponding button B and A in such a B-A order.

If you hear word A followed by word A, press the corresponding button A and again A in such an A-A order.

If you hear word A firstly, word B secondly, word A thirdly, press button A firstly, button B secondly and button A lastly in such an A-B-A order.

You will get a feedback of correct or wrong after your response.

Attention: You can only press the button after you see ´please press the button´ on the screen.

You have finished the warm-up!

Now you are going to do the test! The procedure is the same as in the warm up. In the test, you will hear word A and B in two-word, three-word and four-word sequence length and recall their sequence.

<u>*Dutch version:*</u>

Beste deelnemer,

Bedankt voor uw deelname aan dit experiment.

In het experiment van vandaag leert u zes woord paren. De procedure om elk paar te leren is hetzelfde.

Nu gaat u de twee nieuwe woorden A en B in Paar 1 leren. De twee woorden verschillen in "melodie" of in segmenten. U moet de instructies op het scherm gedurende het hele experiment volgen.

Als u vragen heeft, stel deze dan aan de onderzoeker.

Als u klaar bent om te beginnen, drukt u op een willekeurige knop om het experiment te starten.

Druk op knop A en B om de twee woorden te leren. U kunt zo vaak als u wilt op de twee woorden drukken en deze beluisteren totdat u zeker weet dat u ze heeft geleerd.

Druk op “continue” als u ze heeft geleerd.

Je hebt woord A en woord B al geleerd. Nu gaat u een oefening doen.

In de oefening hoort u een woord en moet je differentiëren of het woord A of B is door op de overeenkomstige knop te drukken.

U kunt pas op de knop drukken nadat u ´drukt u alstublieft op de knop´ op het scherm ziet.

Na uw reactie hoort u OK. Na uw reactie krijgt u een goed of fout feedback.

U gaat door naar de volgende fase nadat u 15 keer correct hebt gereageerd.

Anders moet u herstarten om het opnieuw te doen totdat u aan de criteria voldoet.

Goed gedaan in de eerste training! Nu gaat u een warming-up doen!

U hoort de worden A en B in volgorde.

Als u bijvoorbeeld knoppen B en A in een B-A-volgorde, drukt u op de overeenkomstige woord B gevolgd door woord A hoort.

Als u woord A gevolgd door woord A hoort, druk dan op de overeenkomstige knop A en nogmaals op A in een A-A-volgorde.

Als u ten eerste woord A hoort, ten tweede woord B, en ten derde woord A, druk dan eerst op knop A, vervolgens op knop B en ten slotte op knop A in een A-B-A volgorde.

Na uw reactie krijgt u een goed of fout feedback.

Let op: U kunt pas op de knop drukken als u ́please press the button ́op het scherm ziet.

Je bent klaar met de warming-up!

Nu gaat u de test doen!

De procedure is hetzelfde als bij het opwarmen. In de test hoort u woord A en B in een reeks van twee woorden, drie woorden en vier woorden en moet u hun volgorde invoeren.

Mandarin version:

感谢您来参加实验。

本次实验中您将学习六组词语。每组词的学习步骤相同。

现在您将开始学习第一组词，词语 A 和词语 B。这两个词在“音律”或者一个音节上不同。

在实验过程中，请全程根据屏幕上的指令进行。如果您已准备就绪，请按任意一个按钮开始试验。如您有任何问题，请询问实验者。

请分别按下按钮 A 和 B 学习这两个词语。您可以多次按这两个按钮来学习词语 A 和 B。当您确定您已记住了词语 A 和 B 之后，请按“continue”（继续）按钮。

您已经学习并记住了词语 A 和 B。现在您将进行一个练习。在练习中，您会听到词语 A 或 B，您需要判断您所听到的词语是 A 还是 B。当您听到的词语是 A 时，请按按钮 A，当您听到的词语是 B 时，请按按钮 B。

当您看到屏幕上“请按下按钮”时，您才能进行回答（按下按钮）。每一道题后，您会听到“OK”，并且屏幕上会显示您的回答是正确还是错误。

在练习中，您需要答对 15 道题才能进入下一个环节。否则，您将会回到最初的词语学习环节进行再一次学习，直到您能达到练习标准。

您已完成了练习，真棒！现在您将进入热身环节！在热身环节中，您将同时听到词语 A 和 B。比如，您先听到 B，再听到 A，当您听完后，请依次按下按钮 A 和按钮 B（B-A 顺序）。比如，您先听到 A，再听到 A，当您听完后，请依

次按下按钮 A 和按钮 A （A-A 顺序）。比如，您先听到 A，再听到 B，之后又听到 A，当您听完后，请按顺序依次按下按钮 A、B、A（A-B-A 顺序）。

在您每次回答完之后，屏幕上会显示您的回答是正确还是错误。

注意：当您看到屏幕上“请按下按钮”时，您才能进行回答（按下按钮）。

您已完成了热身！接下来您将进入测试环节！

测试环节的内容与热身环节相同。您将在测试环节中听到词语 A 和 B 以两个词、三个词、四个词的方式出现，请您按顺序依次按下按钮做出回答。

<u>*Japanese version:*</u>

研究協力者のかたへ、

私の実験にご協力くださり、ありがとうございます。実験のご説明をいたします。今日の実験で、6 ペアの単語を覚えていただきます。

それぞれのペアには、2 つの新しい単語が含まれます。2 つの単語が「メロディー」と「セガメント」での発音は違います

今度はペア 1 の新しい単語を覚えていただきます。実験中の画面の指示に従ってください。

ここまでで何かご質問がありましたら、お尋ねください。

準備ができたら、 ボタンを、どれでも良いので、押して下さい。

A と B のボタンを押して、A と B の単語を学習してください。2 つの単語を何度でも押して、聞くことができます。

単語 A と B を覚えましたか?

これから練習をしてみましょう。

練習中に、単語が流れます。この単語が A か B かを判断してください。画面上のに「ボタンを押してください」と表示されているときだけ、ボタンを押すことができます。

ボタンを押したら、 「OK」が流れます。応答のボタンを押した後、正解（√ ）または不正解（×）が画面に表示されます。

15 回正解をした後、次のパートに進みます。 それ以外の場合は、単語を最初からもう一度覚えていただき、基準に達するまで練習をお願いします。

おつかれさまでした！ 次はウォームアップです。

単語 A と B を順番に聞きます。

例えば、単語 B のあとに単語 A が聞こえたら、その順番で、ボタン B を最初に押し、次にボタン A を押してください。

例えば、単語 A のあとに単語 A が聞こえたら、その順番で、ボタン A を最初に押し、次にボタン A を押してください。

例えば、単語 A が最初に、単語 B が 2 番目に、単語 A が 3 番目に流れた場合、ボタンを A-B-A 順番で押してください。

応答のボタンを押した後、正解（√ ）または不正解（×）が画面に表示されます。

画面上のに「ボタンを押してください」と表示されているときだけ、ボタンを押すことができます。

ウォームアップが終わりました！

次は実際の実験に入ります。手順はウォームアップと同じです。実験の中に、単語 A と B は、二単語、三単語または四単語の長さで表せるので、順番を覚えてください。

F.3 Experiment in Chapter 4 (Processing of non-native WPCs in word learning)
(picture selection task)

English version:

You are going to learn 7 new words, corresponding to 1 to 7 on the screen. Each word is represented by a picture. You will hear the word and see the corresponding picture. Please remember the words and their corresponding pictures.

You will learn from Word 1 to Word 7 one by one. After you have learned the 7 words one by one, please click “next” and then you can click any button of Button 1 to 7 (representing Word 1 to 7, respectively) to strengthen your learning.

When you are ready to move on to practice, you can click “continue"

You have learned the words. Now let’s do a practice! In the practice, you will see a picture and hear a sound at the same time. Your task is to press the left button (yes) if the sound and the picture match correctly, or press the right button (no) if they don’t match.

You will proceed to the next stage after you make 13 times correct response. Otherwise, you have to restart to do it again until you reach the criteria.

You have finished the practice! Now let’s start a warm up!

You will see four pictures on the screen. When you hear the word, please choose the correct picture that represents the word you heard. The feedback will be shown on the screen.

You have learned and practiced the words. Now you are going to do the test.

The procedure in the test part is the same as in the Warm up. But there will be no feedback in the Test.

In the final part, you'll see a picture and hear a sound. Your task here is to press the left button (yes) if the sound and the picture match correctly, or press the right button (no) for a mismatch.

Dutch version:

Je gaat 7 nieuwe woorden leren, die overeenkomen met 1 tot 7 op het scherm. Elk woord wordt vertegenwoordigd door een afbeelding. U hoort het woord en ziet de bijbehorende afbeelding. Onthoud de woorden en de bijbehorende afbeeldingen.

U leert Woord 1 tot Woord 7 een voor een. U kunt op "repeat" (herhalen) klikken om het woord herhaaldelijk te horen. U kunt op "volgende" klikken om het volgende woord te leren. Nadat u de 7 woorden een voor een heeft geleerd, klikt u op "volgende" en vervolgens kunt u op een willekeurige knop van Knop 1 tot 7 (respectievelijk voor woord 1 tot 7) klikken om uw kennis te versterken.

Als u klaar bent om te beginnen met oefenen, kunt u op "continue" klikken.

Je hebt de woorden geleerd. Laten we nu beginnen met oefenen!

Ziet u een plaatje en hoort u een geluid. Het is uw taak hier om op de linkerknop “yes” (ja) te drukken als het geluid en het beeld overeenkomen, of op de rechterknop “no” (nee) te drukken voor een mismatch.

U gaat door naar de volgende fase nadat u 13 keer correct hebt gereageerd.

Anders moet u herstarten om het opnieuw te doen totdat u aan de criteria voldoet.

Nu gaat u een warming-up doen!

U ziet vier afbeeldingen op het scherm. Als je het woord hoort, kies dan de juiste afbeelding die het woord dat je hebt gehoord vertegenwoordigt. De feedback wordt na uw reactive op het scherm getoond.

Je hebt de woorden geleerd en geoefend. Nu ga je de test doen.

De procedure in het testgedeelte is hetzelfde als tijdens het oefene. Maar er zal geen feedback zijn in de test.

In het laatste deel, ziet u een plaatje en hoort u een geluid. Het is uw taak hier om op de linkerknop (ja) te drukken als het geluid en het beeld overeenkomen, of op de rechterknop (nee) te drukken voor een mismatch. "

Mandarin version:

感谢您来参加实验。

本次实验中，您将学习七个新词。这七个词语分别对应屏幕上的数字 1 到 7。每个词语对应一个图片。您听到词语的同时，屏幕上会展示相对应的图片。请记住词语与它相对应的图片。

您将从词语 1 开始学习。您可以用鼠标点击“repeat”（重复）来多次学习词语。点击“next”(下一个)学习下一个词语。在您依次学完七个词语后，请按“next”（继续），之后您将任意点击按钮 1 到 7 来学习这七个词语。

当您确定您已记住了这七个词语及其对应的图片后，请点击“继续”进入到练习环节。

现在您将开始练习。在练习环节中，您将听到一个词语，同时屏幕上会展示一张图片，您需要判断所听到的词与图片是否对应一致。请按“yes”(是)表示一致，“no”（否）表示不一致。在练习中，您需要答对 13 道题才能进入下一个环节。否则，您将会回到最初的词语学习环节进行再一次学习，直到您能达到练习标准。

您已完成了练习！接下来，您将进入热身环节。在热身环节中，您将听到词语，同时屏幕上会出现四张图片。您需要从四张图片中选择出与您所听到的词语相对应的片。在您回答之后，屏幕上会显示您的回答是正确还是错误。

您已经完成了热身！接下来将进入到测试环节。测试环节的内容与热身环节相同。您将听到词语，同时屏幕上会出现四张图片。您需要从四张图片中选择出与您所听到的词语相对应的片。在您回答之后，屏幕上将不会显示您回答的结果。

测试的最后，您将听到一个词语，同时屏幕上会展示一张图片，您需要判断所听到的词与图片是否对应一致。请按“yes”(是)表示一致，“no”（否）表示不一致。

Japanese version:

研究協力者のかたへ、

私の実験にご協力くださり、ありがとうございます。実験のご説明をいたします。

画面に表示される 1 から 7 の単語を覚えてください。各単語は画像で表示されます。

単語の音声が流れ、対応する画像が表示されます。単語とそれに対応する画像を覚えてください。

初めに単語 1 の音声が流れ、画像が表示されます。「繰り返す」ボタンをクリックして、この単語を繰り返すことができます。「次へ」ボタンをクリックすると、次の単語に進みます。ボタン 1〜7 の任意のボタンをクリックして、もう一度覚えることができます。

練習の準備ができたら、「続ける」をクリックしてください。

練習を始めましょう。

練習では、画像が表示され、単語の音声が流れます。単語と画像が一致している場合は、左のボタンを押して「一致」を選んでください。単語と画像が一致しない場合は、右のボタンを押して「一致しない」を選んでください。応答のボタンを押した後、正解（√ ）または不正解（×）が画面に表示されます。

13 回正解をした後、次のパートに進みます。 それ以外の場合は、単語を最初からもう一度覚えていただき、基準に達するまで練習をお願いします。

おつかれさまでした！ 次はウォームアップです。

ウォームアップでは、画面に 4 枚の画像が表示されます。単語の音声が流れるので、単語に対応する画像を選んでください。選んだ後、正解は緑色で表示されます。

単語を覚えたので、ウォームアップは終わりです。これから本実験に入ります。

本実験の手順は練習と同じですが、本実験では正解の表示はありません。

最後のパートでは、画像が表示され、単語の音声が流れます。単語と画像が一致している場合は、左のボタンを押して「一致」を選んでください。単語と画像が一致しない場合は、右のボタンを押して「一致しない」を選んでください。

Summary

Languages differ in the use of prosodic cues to signal lexical meanings. For instance, Mandarin uses lexical tones. Dutch uses lexical stress. Japanese uses lexical pitch accent. Yet languages also share certain common properties in the use of WPCs. For instance, on the one hand, Mandarin and Japanese use lexical pitch (Mandarin: pitch variations; Japanese: pitch accent). On the other hand, Japanese and Dutch use the abstract feature of positional marking (Japanese: position of pitch accent; Dutch: position of stress). Previous studies have focused on how listeners with the use of different prosodic cues perceive non-native pitch contrasts (lexical tones) in isolated syllables. However, comparatively, it is not fully understood the role of native prosody in the perception of non-native positional marking from a cross-linguistic perspective. Furthermore, previous studies investigating the processing of non-native pitch contrasts have mainly centered at the acoustic level. However, prosodic processing does not only involve sensory-auditory processing (i.e., processing at the acoustic level) but also involves higher linguistic levels such as the phonological level, a more abstract level than the acoustic level where acoustic traces that the listeners might rely on are inaccessible and the lexical level where phonological knowledge is encoded to build lexical representations in word learning.

The dissertation examines the cross-linguistic processing of non-native word prosodic cues at three processing levels, the auditory-acoustic level, the phonological level and the lexical level within one single study. Mandarin, Japanese and Dutch listeners are selected in the study given that the three languages differ yet share common properties in the use of word prosodic cues. It attempts to address language-specificities but also language-commonalities in processing non-native WPCs at the three processing levels, which may helps to develop a fuller understanding of the dynamic function of WPCs used in human languages from low to high linguistic processing levels and to what extent the native language come into play at the three processing levels.

Chapter 1 is a general introduction. It provides a literature review, elaborate methodologies applied in the dissertation, put forward research questions, and outline the subsequent dissertation chapters

Chapter 2 adopts AX discrimination tasks and used Cantonese tones (also used in the following studies) to investigate cross-linguistic perception of non-native WPCs at the acoustic level. It addresses the perception of non-native pitch contrasts in trisyllable and the perception of non-native tones contrastive in positions. It is found that the three groups show comparable performance in perceiving non-native pitch contrasts that are acoustically salient, but they differ in perceiving non-salient non-native pitch contrasts (contrasts, to some extent, that share similarity in acoustic space). Overall Mandarin listeners show advantage over Japanese and Dutch listeners in discriminating non-salient non-native pitch contrasts in trisyllable. The

findings suggest that both the acoustic salience and the use of lexical pitch and/or to what degree the lexical pitch is used in the native language can influence the perception of non-native pitch contrasts. Language-specific patterns are investigated on exploratory purposes. It is found that the native prosody, especially the use of lexical pitch, together with acoustic salience of tones may influence perception of non-native WPCs.

Chapter 3 applies the sequence-recall task (Dupoux et al., 2001) and examines how and to what extent the native prosody (word and/or intonation) plays a role in processing non-native pitch contrasts occurred word-initially, -medially and –finally at the phonological level. Firstly, an overall perceptual hierarchy is observed: Mandarin listeners >> Japanese listeners >> Dutch listeners. Secondly, language-specific patterns are found with different weighting given to non-native pitch contrasts. Dutch listeners attend more to non-native pitch level contrast than pitch contour and pitch level vs. pitch contour contrast. Japanese listeners are more sensitive to perceive non-native pitch contour and pitch level vs. pitch contour contrast. Mandarin listeners display a gradient order favouring non-native pitch contour contrast >> non-native pitch level vs. pitch contour contrast >> non-native pitch level contrast. The observed language-specific patterns are accounted for in terms of the long-term memory representations of pitch-based phonological categories (tones and/or intonation) (Xu et al., 2006) and assimilation between the native categories and non-native categories (So & Best, 2010, 2013; Francis et al., 2008). Thirdly, position is found to interact with the perception of non-native pitch contrasts depending on different positions. On the one hand, interestingly, a language-general pattern is unveiled that all the three language groups prefer word-final position than word-initial and/or word-medial position when perceiving non-native pitch level vs. pitch contour (falling tone), which seems to indicate the influence of intonation across languages. On the other hand, language-specific patterns are observed. Position played different roles in each language group in the perception of non-native pitch contrasts at the phonological level.

Chapter 4 explores the role of the native prosody (word and/or intonation) in the encoding of non-native pitch contrasts in sound-to-meaning mapping. By using a picture-selection, it reveals that first of all, Mandarin and Japanese listeners perform comparably well, and they both are found better at encoding non-native pitch contrasts occurring in different positions than Dutch listeners in word identification, which could be attributed to the employment of lexical pitch in the native language. Secondly, analyzed for exploratory purposes, each language group displays different patterns in the encoding of non-native tones in different positions in word learning. Mandarin listeners are found better at encoding non-native contour tones than level tones, while both Japanese and Dutch listeners show no preference for any non-native tones. Thirdly, position seems to influence the encoding of non-native tones differently in each language group. Both Mandarin listeners and Dutch listeners are not in favor of any positions. In contrast, Japanese listeners prefer word-final position than word-initial position only when encoding the non-native falling tone, presumably due to the influence of the intonation pattern.

Finally, **Chapter 5** recapitulates the research questions, summarizes the main findings in the three empirical chapters and makes a general discussion with integrating the findings at the three processing levels. This chapter also provides suggestions for future research.

Samenvatting

Talen verschillen in het gebruik van prosodische contrasten van toonhoogte en positie (Word Prosodic Cues: WPC’s) om lexicale betekenis aan te geven. Het Mandarijn Chinees gebruikt bijvoorbeeld lexicale tonen, het Nederlands gebruikt lexicale klemtoon, terwijl het Japans lexicale toonhoogteaccent gebruikt. Toch delen ze ook bepaalde eigenschappen in het gebruik van WPC's; enerzijds gebruiken Mandarijn en Japans lexicale toonhoogte (Mandarijn: toonhoogtevariaties; Japans: toonhoogteaccent). Anderzijds gebruiken het Japans en het Nederlands het abstracte kenmerk van positionele markering (Japans: positie van toonhoogteaccent; Nederlands: positie van klemtoon). Eerdere studies hebben onderzocht hoe luisteraars met het gebruik van verschillende prosodische signalen niet-native toonhoogtecontrasten (lexicale tonen) in geïsoleerde lettergrepen waarnemen. Vanuit een cross-linguïstisch perspectief is de rol van native prosodie in de perceptie van niet-native positionele markering echter nog niet volledig duidelijk. Bovendien richtten eerdere studies die de verwerking van niet-native toonhoogtecontrasten hebben onderzocht zich voornamelijk op het auditief-akoestische niveau. Prosodische verwerking omvat echter niet alleen sensorisch-auditieve verwerking (d.w.z. verwerking op akoestisch niveau), maar het omvat ook hogere linguïstische niveaus, zoals het fonologische niveau en het lexicale niveau. Het fonologische niveau is abstracter dan het akoestische niveau, omdat de akoestische sporen waar de luisteraars op zouden kunnen vertrouwen, niet toegankelijk zijn (Dupoux et al., 2001, 2008). Ook het lexicale niveau is abstracter, hierbinnen wordt namelijk de fonologische kennis gecodeerd om lexicale representaties op te bouwen bij het leren van woorden.

Het proefschrift onderzoekt, binnen één enkele studie, de cross-linguïstische verwerking van prosodische signalen van niet-native woorden op drie verwerkingsniveaus, het auditief-akoestische niveau, het fonologische niveau en het lexicale niveau. Luisteraars met als moedertalen Mandarijn Chinees, Japans en Nederlands zijn geselecteerd om deel te nemen aan het onderzoek, omdat deze drie talen zowel verschillende, als ook gemeenschappelijke, eigenschappen delen in het gebruik van prosodische aanwijzingen. Het proefschrift poogt taalspecifieke en taalalgemene patronen bij de verwerking van niet-native WPC's op de drie verwerkingsniveaus te identificeren. Dit kan ons helpen om de dynamische functie van WPC's die worden gebruikt in menselijke talen beter te begrijpen, zowel in lagere orde als in hogere orde taalverwerkingsniveaus, en om te begrijpen in hoeverre de moedertaal een rol speelt op de drie verwerkingsniveaus.

Hoofdstuk 1 is een algemene inleiding. Het omvat een literatuuroverzicht, uitgebreide beschrijving van de methodologie die in het proefschrift zijn toegepast, introduceert de onderzoeksvragen, en schetst de daaropvolgende dissertatiehoofdstukken.

Hoofdstuk 2 onderzoekt de cross-linguïstische perceptie van niet-native WPC's op akoestisch niveau middels AX-discriminatietaken waarbij Kantonese tonen gebruikt worden (ook gebruikt in de volgende studies). Het hoofdstuk bespreekt de perceptie van niet-native toonhoogtecontrasten in woorden met drie syllabes en de perceptie van niet-native toonhoogtecontrasten in verschillende posities. De drie groepen deelnemers laten vergelijkbare prestaties zien bij het waarnemen van niet-native toonhoogtecontrasten die akoestisch saillant zijn. Echter, verschillen ze in het waarnemen van niet-saillante niet-native toonhoogtecontrasten (de contrasten die, tot op zekere hoogte, overeenkomen in akoestische eigenschappen). Op groepsniveau laten luisteraars met Mandarijn als moedertaal een voordeel zien ten opzichte van Japanse en Nederlandse luisteraars bij het onderscheiden van niet-saillante, niet-native toonhoogtecontrasten in drielettergreepwoorden. Deze bevindingen suggereren dat zowel de akoestische saillantie, als wel het gebruik van lexicale toonhoogte en/of de mate waarin de lexicale toonhoogte wordt gebruikt in de moedertaal, de perceptie van niet-native toonhoogtecontrasten kan beïnvloeden. Daarnaast worden taalspecifieke patronen verkennend onderzocht. Er wordt geobserveerd dat de perceptie van niet-native WPC’s beïnvloed wordt door de native prosodie, met name het gebruik van lexicale toonhoogte, samen met de akoestische saillantie van tonen.

Hoofdstuk 3 gebruikt de *sequentiële herinneringstaak* (Dupoux et al., 2001) om te onderzoeken op welke manier, en in hoeverre de native prosodie (woord en /of intonatie) een rol speelt bij het verwerken van niet-native toonhoogtecontrasten die op initiële, mediale, en finale posities in het woord voorkomen, en op fonologisch niveau. Ten eerste, wordt een algemene perceptuele hiërarchie waargenomen: Mandarijnse luisteraars >> Japanse luisteraars >> Nederlandse luisteraars. Ten tweede, worden taalspecifieke patronen gevonden die in verschillende mate afhankelijk zijn van niet-native toonhoogtecontrasten. Nederlandse luisteraars besteden meer aandacht aan niet-native toonhoogtecontrast dan toonhoogtecontouren en toonhoogteniveau vs. toonhoogtecontourcontrast, terwijl Japanse luisteraars zijn gevoeliger voor het waarnemen van niet-native toonhoogtecontouren en toonhoogteniveau vs. toonhoogtecontourcontrast. Mandarijnse luisteraars daarentegen laten een voorkeur zien met de volgende gradiëntvolgorde: niet-native toonhoogtecontrast >> niet-eigen toonhoogte-niveau vs. toonhoogte-contourcontrast >> niet-eigen toonhoogte-contrast. De waargenomen taalspecifieke patronen worden verklaard in termen van de langetermijngeheugen-representaties van op toonhoogte gebaseerde fonologische categorieën (tonen en/of intonatie) (Xu et al., 2006) en assimilatie tussen de native categorieën en niet-native categorieën. (So & Best, 2010, 2013; Francis et al., 2008). Ten derde, blijkt dat de positie van het toonhoogtecontrast interacteert met de perceptie van niet-native toonhoogtecontrasten. Opvallenderwijs, wordt aan de ene kant een taal-algemeen patroon gevonden dat alle drie de groepen luisteraars de voorkeur geven aan de woordfinale positie boven de begin- en/of woord-mediale positie bij het waarnemen van niet-native toonhoogteniveau vs. toonhoogtecontour (dalende toonhoogte). Dit lijkt de invloed van intonatie tussen talen lijkt aan te geven. Aan de andere kant worden ook taalspecifieke patronen waargenomen. Positie speelt in elke taalgroep een andere rol in de perceptie van niet-native toonhoogtecontrasten op het fonologisch niveau.

Hoofdstuk 4 onderzoekt de rol van de native prosodie (woord en/of intonatie) bij het encoderen van niet-native toonhoogtecontrasten in de geluid-betekenis representatie. Bij gebruik van een *picture-selection* taak wordt duidelijk dat luisteraars met Mandarijn en Japans als moedertaal even goed presteren, en dat ze bij woordidentificatie beiden beter zijn in het encoderen van niet-native toonhoogtecontrasten die op verschillende posities voorkomen dan Nederlandse luisteraars, die kunnen worden toegeschreven op het gebruik van lexicale toonhoogte in de moedertaal. Ten tweede, is in een verkennende analyse gevonden dat tijdens het leren van nieuwe woorden elke taalgroep verschillende patronen vertoont bij het encoderen van niet-native tonen op verschillende posities. Mandarijnse luisteraars zijn beter in het encoderen van niet-native contourtonen dan niveautonen, terwijl zowel Japanse als Nederlandse luisteraars geen voorkeur hebben voor niet-native tonen. Ten derde lijkt de positie van het toonhoogtecontrast de codering van anderstalige tonen in elke taalgroep anders te beïnvloeden. Zowel Mandarijnse als Nederlandse luisteraars hebben geen voorkeur voor een bepaalde positie. Japanse luisteraars daarentegen hebben een voorkeur voor de woordfinale positie vergeleken met de woordinitiële positie, maar enkel bij het encoderen van de niet-native dalende toon, vermoedelijk vanwege de invloed van het intonatiepatroon.

Tot slot vat **Hoofdstuk 5** de onderzoeksvragen samen, beschrijft de belangrijkste bevindingen in de drie empirische hoofdstukken, en bevat een algemene discussie met integratie van de bevindingen op de drie verwerkingsniveaus. Dit hoofdstuk geeft ook suggesties voor toekomstig onderzoek.

Curriculum Vitae

Shuangshuang Hu was born on October 22 in 1988 in Shanghai, China. She attended Shanghai Normal University from 2007 until 2011 and obtained bachelor degree in Teaching Chinese as a Foregin Language. She started her first master programme at Shanghai Normal University in 2011 and obtained her master's degree in Teaching Chinese to Speakers of Other Languages in 2013. During the six years, she received systematic training in second language acquisition and language teaching. She studied her second master programme "Linguistics: Study of the Language Faculty (Research MA)" at Utrecht Institute of Linguistics in Utrecht University in 2013 and obtained her master's degree in Linguistics in 2015. During the two years, she received academic training in linguistics and developed her research interest in phonology, phonetics, speech perception and psycholinguistics. She began her PhD research on cross-linguistic processing of non-native word prosodic cues at Utrecht Institute of Linguistics in Utrecht University in 2015. This dissertation is the result of the research.