Production and perception of tones by Dutch learners of Mandarin

Zou, T.; Zou T.

Citation

Zou, T. (2017, September 28). *Production and perception of tones by Dutch learners of Mandarin*. *LOT dissertation series*. LOT, Utrecht. Retrieved from https://hdl.handle.net/1887/52977

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/52977

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University
Repository

The handle http://hdl.handle.net/1887/52977 holds various files of this Leiden University dissertation.

Author: Zou, T.
Title: Production and perception of tones by Dutch learners of Mandarin
Issue Date: 2017-9-28

PRODUCTION AND PERCEPTION
OF TONES
BY DUTCH LEARNERS
OF MANDARIN

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

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

Cover illustration: The developmental path. The picture is from http://www.spa-francorchamps.be/ (edited by the author and reprinted with permission).

ISBN: 978-94-6093-255-7

NUR 616

Copyright © 2017: Zou Ting. All rights reserved.

PRODUCTION AND PERCEPTION OF TONES BY DUTCH LEARNERS OF MANDARIN

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 28 september 2017
klokke 10.00 uur

door

ZOU TING

geboren te Tianjin, P.R. China
in 1986

Promotor:	Prof. dr. Vincent J. van Heuven
Co-promotores:	Dr. Johanneke Caspers Dr. Yiya Chen
Promotiecommissie:	Prof. dr. Carlos Gussenhoven (Radboud Universiteit Nijmegen) Prof. dr. Pierre Hallé (Université Sorbonne Nouvelle Paris) Dr. Nivja de Jong Prof. dr. Niels O. Schiller

Contents

Acknowledgments

This thesis would not have been possible without the help and support from a large number of people. First of all, I would like to express my gratitude to my supervisors: Yiya Chen, Johanneke Caspers and Vincent van Heuven. I would like to thank Yiya for her guidance throughout all stages of my PhD research. Thank you for your constant encouragement and support, your invaluable feedback on my journal article and thesis, as well as your insightful advice on both research and career. I am grateful to Johanneke for her continuous support in these years and her kind help at various stages of my research. I owe special thanks to Vincent for his extensive comments with wit and insight on an early draft of this thesis. Thank you for being my promotor. I really benefited a lot from your critical mind and immense knowledge. I am extremely grateful to all my supervisors.

This PhD project received great help from many people. I am grateful to Jos Pacilly for all his timely and tireless support during the conduction of my eye-tracking experiment. I would like to thank Lotte Bijloo, Myrthe Kroon, Joren Pronk, Edie Corver and Ziyi Bai who lent their time generously for my pilot studies and provided supportive feedback. I am grateful to Jip Bouman, Björn Ooms, Ying-Ting Wang, Aurelie van 't Slot, Tjerk Loenhout, Raven Salemink and Anna de Rooij-van Broekhuizen for their help with participant recruitment and experimental materials. My deepest thanks go to all the participants of my experiments. A special word of thanks is due to dr. Kurt Debono from SR Research for his timely technical support on Experiment Builder.

I cherish the time I spent in the phonetics lab at the Leiden University Centre for Linguistics and owe thanks to all my colleagues for their help at various stages of my research. Special thanks go to Qian Li, Min Liu and Yifei Bi for all the inspiring discussions and helpful experience-sharing. I would also like to give thanks to Han Hu, Yang Yang, Junru Wu, Menghui Shi, Daan van de Velde, Jessie Nixon and my office mate Elisabeth Mauder.

My thanks also go to the Chinese friends I met in the Netherlands: Zenghui Liu, Anqi Yang, Mengru Han, Xin Li, ShuangShuang Hu, Zhaole Yang and Mengchen Wu. Thanks for your encouragement and support during these years.

I would like to express my gratitude to prof. JinSong Zhang for all the invaluable advice during these years.

Special thanks to Séverine Cirlande and the Spa circuit for the nice picture of the legendary Eau Rouge bend.

Last but not least, I owe my thanks to my parents for their support and understanding.

Leiden, August 2017

Chapter one

Background

1.1 Introduction

Pitch movements (cued mainly via fundamental frequency or f0 acoustically) have different functions across languages. For non-tone language speakers, pitch information is mainly used at the post-lexical level to convey linguistic and paralinguistic meanings. Linguistically, it can be used to indicate prominence at the word and sentence levels (Birch & Clifton, 2002; Welby, 2003; Xu & Xu, 2005), delimit prosodic constituents (such as intonation phrase, utterance, paragraph and so on) (Cole, 2015; Cutler, Dahan, & Van Donselaar, 1997; Snedeker & Trueswell, 2002), and mark sentence types, such as statements and interrogatives (Pierrehumbert & Steele, 1989; Van Heuven & Haan, 2002). Some paralinguistic information, such as emotion (anger, surprise, joy, fear, and so on) and attitude (politeness, uncertainty, irony, dejection, and so on) can also be (partly) encoded with pitch movements (Chen, Gussenhoven, & Rietveld, 2004; Chen, 2005; Luthy, 1983; also see Shattuck-Hufnagel & Turk, 1996 for a detailed review).

Tone-language speakers, on the other hand, primarily employ pitch information to convey lexical meanings, while at the same time, in a much more complex and sometimes subtle way, signal various post-lexical information comparable to that in non-tone languages (e.g., Chen, 2000, 2012; Chen & Gussenhoven, 2008; Cole, 2015; Gussenhoven, 2004; Xu, 2001; Yip, 2002).[1]

Mandarin is a tone language. Fundamental frequency (f0) has been demonstrated as the primary acoustic correlate of tones (Howie, 1976; Xu & Wang, 2001; Yip, 2002), although other acoustic parameters (e.g., intensity and temporal properties such as duration and position of pitch turning point in some contour tones) can also be used to mark tonal contrast (Hallé, 1994; Moore & Jongman, 1997; Xu, 2009).[2] In Mandarin Chinese, tonal information is an integral part of a word and the meaning of the segments is associated with the pitch contour superimposed on them. Mandarin Chinese has four main tones, in addition to a neutral tone. Tone 1 is a high-level tone; Tone 2 is a mid-rising tone; Tone 3 is a low tone; Tone 4 is a high-falling tone. When produced in prepausal position or in isolation, Tone 3 is realized with a dipping contour.

---

[1] For some African and Asian languages, tone can be used to signal grammatical information. In this thesis, "tone language" refers to languages in which tones are solely used to convey lexical meaning.

[2] Mandarin is used here to refer to Standard Chinese, the official language spoken in Mainland China, which is based on Beijing Mandarin. Only when referring to other Mandarin dialects, will we identify the specific dialect within the Mandarin dialect family. "Mandarin speakers", without mentioning a specific Mandarin dialect, then, is used to refer to speakers of Standard Chinese who may or may not speak a Mandarin dialect other than Standard Chinese.

This tone also has two variants in connected speech: it becomes a low falling tone preceding Tone 1, Tone 2, Tone 4 and neutral tone, and is realized with a rising contour similar to Tone 2 preceding another Tone 3. The four full tones are demonstrated in Figure 1.1. The neutral tone always comes at the end of a word or phrase, associated with a weak syllable. It has a static and mid target, but the target is realized with more pitch variation compared with lexical full tones: the pitch of a syllable with neutral tone is substantially influenced by the tone in the preceding syllable (Chen & Xu, 2006).

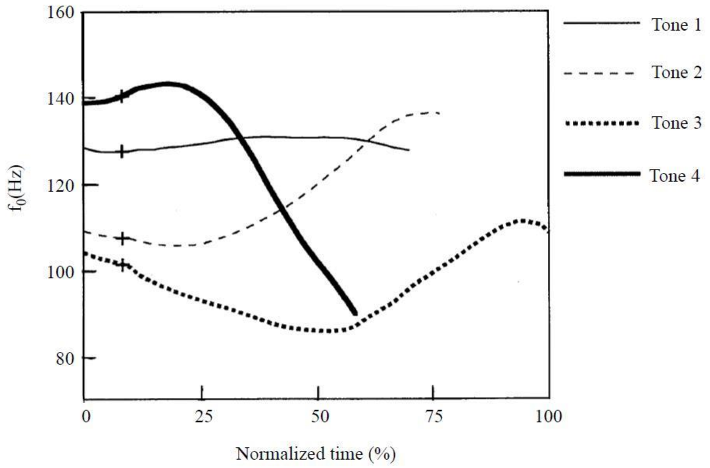

Figure 1.1. *Mean f0 contours of the four Mandarin tones in the monosyllable /ma/ produced in isolation (Xu, 1997). Averaged over 48 tokens produced by eight male native Mandarin speakers.*

When produced in isolation, different tones are realized with stable and distinctive pitch contours. However, when produced in context, tones can be influenced by adjacent tones and undergo substantial acoustic variation, leading to coarticulated tonal realization which are different from the canonical contours. As for the perception of tones, prior studies show a high level of interdependency in the processing of segmental and tonal dimensions by native Mandarin speakers (Choi, Tong, Gu, Tong, & Wong, 2017; Lin & Francis, 2014; Repp & Lin, 1990; Tong, Francis, & Gandour, 2008). That is, the segmental and tonal dimensions are integral and processed simultaneously by Mandarin native speakers.

With regard to perceptual identification and discrimination of lexical tones, previous research showed that tones can be perceived in a categorical fashion by native speakers (Francis & Ciocca, 2003; Hallé, Chang, & Best, 2004). In terms of the role of tone in word recognition, some previous studies suggest that tone serves as a weaker cue compared to segmental information. However, recent studies using online measures such as eye-tracking and event-related potentials (ERP) show parallel processing of segments and tones in word recognition, arguing that the role of tonal information is comparable to that of segmental information. For example, Schirmer, Tang, Penney, Gunter, and Chen (2005) used ERPs to investigate the role of tone and segmental information in Cantonese word processing. Comparing the ERPs elicited by semantically congruous words and by tonally and segmentally induced semantic violations, they found that both segments and tones were accessed at a similar point in time and elicited an N400-like negativity. Malins and Joanisse (2012) offered further support for the comparable roles of segments and tones, showing that both segmental and tonal information could be accessed and used as soon as they become available during word processing. Taken together, the existing literature suggests that tonal information is exploited in spoken word recognition. It plays an early constraining role in lexical activation, and words with non-matching tone would not be activated as candidates. This effect can be captured and revealed more readily in online measurements with tasks more similar to real communication situations.

Speakers of tone and non-tone languages have been reported to tune their auditory systems to the same acoustic stimuli differentially due to their first language (L1) experience. Both behavioral and neuroscientific studies have suggested that speakers of tone and non-tone languages process pitch information differently.

Using multidimensional scaling, Gandour (1983) investigated the influence of different language backgrounds on tonal perception. Speakers of tone languages (Cantonese, Mandarin, Taiwanese, Thai) and a non-tone language (English) were asked to judge the dissimilarity of paired tones. The result showed that listeners from a non-tone language (English) attached more importance to pitch height and gave less weight to pitch direction than did listeners from most of the tone language group. It is suggested that the absence of lexical contrastive tones in monosyllabic words can account for the low saliency of the direction dimension in English listeners' dissimilarity judgments. Bent, Bradlow, and Wright (2006) investigated the influence of long-term linguistic experience on identifying non-speech rising, falling and flat pitches. The results showed that the rising and falling stimuli were treated in the same way by English listeners, but Mandarin listeners more often misidentified flat and falling pitch contours than the English listeners, in a manner that could be related to specific features of pitch contours of Mandarin lexical tones. The authors argued that, in Mandarin, the pitch range of the falling tone (Tone 4) is larger than that of the rising tone (Tone 2), so that Mandarin listeners might have a different criterion for the distinction between falling and rising contours and were more reluctant to label stimuli as falling than to label them as rising. Thus, it appears that listeners' perception of pitch movements can be shaped by the way pitch information is used in their native language. Neurophysiological studies also showed differences in the hemispheric specialization of pitch processing by tone and non-tone language speakers: tonal contrasts are processed mainly in the left hemisphere by tone-language speakers, but in the right hemisphere or bilaterally by non-tone language speakers (Gandour, Dzemidzic, Wong, Lowe, Tong, & Hsieh, 2003; Gandour, Tong, Wong, Talavage, Dzemidzic, & Xu, 2004; Krishnan, Xu, Gandour, & Cariani, 2005; Zatorre & Gandour, 2008).

Since non-tone language speakers are not familiar with pitch information which conveys lexical meaning, tones can present a great difficulty to them. Such difficulty in tone production and perception experienced by beginning learners of Mandarin has been tested in a number of studies. In terms of tone production, Wang, Jongman, and Sereno (2003) tested tone production in monosyllables by beginning learners of Mandarin and showed that only 57 percent of the learners' tone productions was correctly identified by native Mandarin listeners. After a short-term training, 78 percent of their tone production can be correctly identified. This shows that beginning learners can be trained to improve their production accuracy in monosyllables. Other studies also showed that tones in monosyllabic words can be produced correctly by second language (L2) learners of Mandarin (e.g., Hao, 2012).

As for the perception of tones, Wang, Spence, Jongman and Sereno (1999) showed that, in a tone identification task, beginning English learners of Mandarin showed an identification accuracy rate of 69 percent with a prevailing Tone 2-Tone 3 confusion. This study also showed that learners could be trained to significantly improve their tone production and identification accuracy. After a two-week perceptual training, English learners of Mandarin improved their Mandarin tone identification accuracy by 18 percentage points.

Taken together, lexical tone plays an important role in Mandarin. Tones and segments are processed in an integral manner by native speakers and the role of tones in spoken word recognition is comparable to that of segments. For non-tonal L2 learners, establishing new tonal categories can be challenging since they do not make extensive use of suprasegmental features in lexically contrastive way in their L1. Previous studies demonstrated that beginning L2 learners of Mandarin can correctly produce tones in monosyllables and perceive tones with high accuracy in simple identification and discrimination tasks. These findings have presented a promising start, but the process of tone acquisition is more demanding. To produce natural and native-like words, learners need to learn coarticulation patterns of tones in addition to the canonical tonal contours. To process tonal information effectively, learners are also required to attach more perceptual weight to the previously ignored suprasegmental dimension. Furthermore, they need to learn to attune to the most reliable phonological tonal information for word form identification in connected speech. Last but not least, effective exploitation of tones is of great importance in spoken word recognition. Therefore, the crucial research questions are: (1) to what extent can advanced learners achieve these goals; and (2) what mechanisms underlie the development of their L2 tone acquisition? Few studies in the existing literature have addressed these issues. To fill this research gap, this dissertation examines the production and perception of Mandarin tones by both beginning and advanced Dutch learners of Mandarin, with native Mandarin speakers as a control group. Specifically, four experimental studies are reported in this dissertation, viz. on L2 tonal coarticulation patterns in disyllabic tone production (Chapter 2), attention redistribution in L2 tone processing (Chapter 3), L2 tone processing at the phonological level (Chapter 4), and the role of tones in lexical access for L2 learners (Chapter 5). The following sections provide the backgrounds of these topics.

1.2 Tonal coarticulation by native Mandarin speakers and L2 learners

Coarticulation refers to the influence of one sound on a neighboring sound in speech production, since speech is produced as "a sequence of sounds flow to articulatory movements" and "there is 'blurring of the edges' of segmental articulations as the vocal tract moves from one articulation configuration to the next" (Bell-Berti, Krakow, Gelfer, & Boyce, 1995). A speech sound is influenced by both the preceding sound (i.e., carryover effect) and the subsequent sound (i.e., anticipatory effect). Bidirectional coarticulatory effects have been reported in vowels and consonants, e.g., carryover effect as shown in Recasens (1984) and Beddor, Harnsberger, & Lindemann (2002); and anticipatory effects as shown in Martin & Bunnell (1981) and Grosvald (2009). Whalen's (1990) findings demonstrate that the carryover effect is a result of physiological constraints in realizing some motor program, since this effect remains robust when cognitive planning is constrained. The anticipatory effect, on the other hand, would be a reflection of speech planning, in that it decreases when the participants' planning mechanism is inhibited.

Previous research focusing on coarticulation patterns in Asian tone languages shows that tonal coarticulation is also bidirectional, which is in parallel with the cases of vowel and consonant coarticulation. The carryover effect is generally strong in terms of the magnitude and temporal domain, while the anticipatory effect is generally weaker (e.g., Thai: Abramson, 1979; Gandour, Potisuk, & Dechongkit, 1994; Vietnamese: Brunelle, 2009; Han & Kim, 1974). The patterns of Mandarin tonal coarticulation generally agree with the findings from other tone languages. Xu (1997) examined tonal coarticulatory patterns in disyllabic non-words /mama/ with all possible tonal combinations in Standard Chinese produced by native Beijing Mandarin speakers. The results showed that the carryover effect exhibits an assimilatory nature; a high offset of the first tone can raise the onset of the following tone, while a low offset lowers the onset of the following tone. The anticipatory effect, however, is largely dissimilatory. Xu (1997) showed that, in Standard Chinese, the anticipatory effect is mainly on the maximum f0 of the preceding tone.

In terms of L2 production of Mandarin tones, considerable evidence indicates that L2 learners are able to correctly produce lexical tones in isolation (e.g., Hao, 2012; Wang, Jongman, & Sereno, 2003). Producing tones in connected speech, however, does present a great challenge, evident in the higher error rates and decreased intelligibility of L2 speech (Hao, 2012; Shen, 1989; Sun, 1998; Yang, 2011, 2016). However, the acquisition of fine-grained tonal coarticulation patterns has received less research attention. Recently, Brengelmann, Cangemi and Grice (2015) tested tonal coarticulation in disyllabic sequences by German learners of Mandarin and found much f0 variation in the last 20 percent of the tone contours on the first syllable, which suggested a strong but non-native-like anticipatory effect. The extent to which the pattern is general among learners of non-tonal languages is an interesting issue to investigate.

As reviewed earlier, tonal coarticulation has been mainly investigated in three aspects: the directionality (carryover or anticipatory), the nature (assimilatory or dissimilatory), as well as the magnitude and temporal extent of the effects. Thus far, the underlying mechanism and source of the tonal coarticulatory effect for native speakers, as well as L2 acquisition of tonal coarticulation have been under-investigated. Therefore, Chapter 2 sets out to investigate these issues using a disyllabic tone production task with a high cognitive load. By testing native speakers, we aim to shed further light on the mechanisms underlying tonal coarticulation. By recruiting both beginners and advanced Dutch learners of Mandarin, we will investigate the developmental trajectory and mechanisms of tonal coarticulation that underlie the ultimate attainment of acquisition in a tone language by non-tonal L2 learners.

1.3 Attention redistribution and segment-tone integration in L2 acquisition

When learning a foreign language, learners are often confronted with difficulties in both low-level auditory processing and in the phonological processing of non-native segmental and suprasegmental contrasts. Different theoretical models have been proposed to account for such difficulties. The Speech Learning Model (SLM) holds that L2 learners perceive non-native sounds by referring to the phonetic categories of their L1 sound system (Flege, 1995). The mechanisms involved in L1 acquisition, such as category formation, remain intact through one's life and can also be used in L2 learning, although this ability tends to decrease as the learner's age of learning increases. PAM-L2 (Best & Tyler, 2007), based on the Perceptual Assimilation Model (PAM) (Best, 1994), assumes that a listener's perceptual system will automatically assimilate non-native speech sounds to the nearest categories in the L1 sound system, and the discrimination of non-native contrasts can be predicted from the way in which they are assimilated. For the case of L2 acquisition of Mandarin tones, both SLM and PAM-L2 suggest that a novel L2 speech contrast can potentially be acquired by learners.

While these models of L2 acquisition have focused on whether new L2 categories can be acquired, much less has been investigated on how they are acquired. As discussed in the previous section, past research has shown that the same pitch movements can be attended to differentially by tone and non-tone language speakers. Braun and Johnson (2011) showed that Mandarin speakers were attentive to the rising and falling pitch contours on both the initial and final syllables in a disyllabic non-word. These contours signal two different lexical tones in Mandarin. Dutch speakers, in contrast, were much more sensitive to pitch movements in the final position than in the initial position, possibly because a Dutch final pitch movement serves as a salient cue for non-lexical meanings, such as question vs. statement.

Moreover, prior studies suggest that the processing of segmental and tonal dimensions by native Mandarin speakers is more interdependent than by non-tone language speakers. The segment-tone integration has been revealed in some studies testing the so-called Garner interference. That is, there is an increase in reaction time due to the inclusion of irrelevant information during perceptual processing (Garner, 2014). For example, Tong, Francis and Gandour (2008) tested the interactions between segmental and suprasegmental dimensions of Mandarin Chinese by asking participants to attend to one dimension while ignoring the other. Their results suggested that variations in the segmental dimension interfered more with tone classification than *vice versa*. While in non-tone languages, like English, the two dimensions are much less integrated, and therefore listeners are able to tune their attention to only one dimension and suppress interference from the other (Lin & Francis, 2014). This integrality of tones and segments in tone languages such as Mandarin Chinese and Cantonese has also been found in recent neuroscientific studies (Choi et al., 2017; Gao, Hu, Gong, Chen, Kendrick, & Yao, 2012; Tong, McBride, Lee et al., 2014). For example, Choi et al. (2017) tested perceptual integration of vowels and tones in native Cantonese speakers using the passive oddball paradigm. Tone-MMN, vowel-MMN and double-MMN were elicited. The results showed that double-MMNs were significantly smaller in amplitude than the sum of single feature MMNs, suggesting the perceptual integration of tones and vowels at the phonological level.

Therefore, the issues we address here are: during the course of their acquiring a tonal system, can Dutch learners of Mandarin learn to redistribute their attention to segmental and tonal information like native speakers and can they develop a more integral processing of these two dimensions? Chapter 3 investigates these questions by examining how beginners and advanced Dutch learners of Mandarin process tonal information in an ABX matching-to-sample task, compared to both native Mandarin speakers and native Dutch speakers without any tone language experience.

1.4 Phonological processing of tone contrasts by L2 learners

The automatic selective perception (ASP) model (Strange, 2011), which has been developed to characterize L1 and L2 speech perception, highlights the role of attention in language acquisition. It further differentiates between a phonological mode and a phonetic mode of perception. The phonological mode is employed by native listeners, in which automatic selective perception routines are used to detect phonologically contrastive information for identifying word forms. This automatic processing is shaped by language experience, and therefore costs little cognitive effort. The phonetic mode is employed by native speakers to detect fine-grained allophonic details, and requires more cognitive effort. It is hypothesized that at the beginning stage of L2 learning, the phonetic mode of perception has to be used when processing novel contrasts. The L2 learning process involves the development of new selective perception routines that optimize the attunement to information that is reliable for word-form recognition. The role of the task is also emphasized by the ASP model: in tasks with a high memory load and phonetic variability, L2 listeners are less likely to detect fine-grained phonetic details, and therefore have to use the phonological mode of processing; in less demanding tasks with simple stimuli, the phonetic mode can be used.

The problem of Japanese listeners' discrimination of the English /r/-/l/ contrast (Strange & Dittmann, 1984) is a good example of acquisition difficulty that can be accounted for by the ASP model. The L2 listeners showed a good performance in basic identification and discrimination tasks in which the phonetic mode of processing could be used. In a more demanding task with complex stimuli asking for the phonological mode of processing their performance was poor, since the selective perception routines of English had not been established yet. Likewise, for perception of a non-native /e/-/ɛ/ contrast as predicted by the ASP model, the level of difficulty is a function of task and stimulus factors (Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés & Soto-Faraco, 1999; Sebastián-Gallés, Echeverria, & Bosch, 2005). These findings also suggest the important role of task demands and stimulus complexity in the assessment of participants' processing ability of non-native segmental and suprasegmental contrasts at the phonological level.

Furthermore, learning to use tonal information in lexical access is another crucial issue in L2 tonal acquisition. However, to our knowledge, no systematic empirical research has been done to investigate tone processing in lexical access by L2 learners. To examine the developmental trajectory of the Dutch learners' phonological processing of tonal contrasts and the use of lexical tones in lexical activation, Chapter 4 adopts a cognitively demanding sequence recall task and a lexical decision task, testing both beginning and advanced Dutch learners of Mandarin and native Mandarin speakers as a control group.

1.5 Segmental versus tonal information in lexical access by L2 learners

As the studies investigating real-time spoken word recognition accumulate, it is becoming increasingly clear that as the input unfolds, lexical candidates are activated immediately with receipt of a minimal amount of acoustic information (McMurray, Clayards, Tanenhaus, & Aslin, 2008); the activation is updated incrementally (Dahan & Gaskell, 2007; Shen, Deutsch, & Rayner, 2013); multiple words are activated in parallel and compete with each other during the recognition process.

To capture these characteristics of spoken word recognition, current models (such as TRACE: McClelland & Elman, 1986 and Shortlist: Norris, 1994; Norris & McQueen, 2008) assume that, as speech input unfolds, the incoming sound can be mapped onto phonemic and lexical representations in the mental lexicon, and then a set of lexical candidates compete for recognition. Since these models were developed using non-tone languages, they do not encode lexical tones. As suggested by recent online studies on lexical tone processing (e.g., Malins & Joanisse, 2010), it is increasingly clear that tone plays a comparable role as segments in constraining lexical activation. Tones have been incorporated into the TRACE model in a recent simulation of monosyllabic spoken word recognition of Mandarin Chinese (Shuai & Malins, 2017), based on the finding and suggestions from previous studies (Malins & Joanisse, 2010; Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011).

There has also been an abundance of studies that have tested the perception and production in beginning Mandarin L2 learners. For example, Wang et al. (1999) show that English learners of Mandarin improved their tone identification accuracy in monosyllabic words from 69% to 90% after a two-week training. The training-induced improvement also generalized to new words and speakers. In addition to tones in isolated syllables, the perception of longer stimuli also has been tested. Hao (2012) found that both English and Cantonese learners of Mandarin performed better in monosyllabic tonal identification than in disyllabic identification. Both learner groups showed better Mandarin tone mimicry than tone identification and reading. The former task only involved low-level auditory perception and articulation while the latter task required a more abstract representation of tones. This suggests that the main difficulty in tone learning is the establishment of robust associations between pitch contours and tone categories.

More recently, learning to use lexical tone information in word recognition by naive non-native speakers of Mandarin have been tested in several training studies. The sound-to-word learning paradigm, which trains participants to associate minimal tone pairs with different meanings, has been employed in these studies to examine the contribution of individual variability in cue weighting in tone learning (Chandrasekaran, Sampath, & Wong, 2010), the effect of individual musical experience (Wong & Perrachione, 2007), as well as the influence of tonal context in tone learning (Chang & Bowles, 2015). Some studies also found training-induced changes in the participants' neural system (Wong, Chandrasekaran, Garibaldi, & Wong, 2011; Wong, Perrachione, & Parrish, 2007). Although the focus varied across these studies, the convergent result is that naive non-native speakers of Mandarin can be trained to use pitch information lexically.

While much research effort has been devoted to learning lexical tones by naive non-native Mandarin speakers and beginning learners, the processing of tones and segments by advanced L2 learners of Mandarin and the developmental trajectory have not been studied before. Moreover, L2 processing of tonal information has not been investigated using on-line methods. Therefore, Chapter 5 sets out to examine the role of tones and segments in auditory spoken word recognition using the Visual World Paradigm by monitoring the eye movements of both beginners and advanced Dutch learners of Mandarin. Native Mandarin speakers were also tested, as a control group.

Chapter two

Effect of cognitive load on tonal coarticulation: Evidence from native Mandarin speakers and Dutch learners of Mandarin

2.1 Introduction

In Mandarin Chinese, a lexical tone language, pitch movements (cued mainly via fundamental frequency) are used to convey lexical meaning. When produced in isolation, different tones are realized with stable and distinctive pitch contours. However, when produced in connected speech, tones can be influenced by the preceding and following tones and undergo substantial acoustic variation, leading to coarticulated f0 realizations which are different from the canonical contours. Such deviated f0 shapes make it a great challenge for adult non-tonal learners of Mandarin to achieve native-like tone production. It is evident in the literature that such anomalous tonal coarticulation patterns can be the cause of quite a part of the foreign accent of less proficient Mandarin speakers (Hao, 2012; Lee, Vakoch, & Wurm, 1996; Wang, Jongman, & Sereno, 2003).

Previous research on tonal coarticulation has mainly focused on the directionality (carryover or anticipatory), the nature (assimilatory or dissimilatory), and the magnitude of contextual effects on tonal production by native speakers (see Chen, 2012, for a review). In contrast, the underlying mechanisms of tonal coarticulation and the acquisition of coarticulated patterns by second language (L2) learners of Mandarin have remained much less-understood. This study was therefore designed to examine tonal coarticulation by both native and learners of Mandarin, with a particular focus on the effect of cognitive load on tonal coarticulation and the developmental trajectory with regard to the acquisition of tonal coarticulation by learners of non-tonal languages.

2.1.1 Tonal coarticulation in Mandarin Chinese

Coarticulation, the influence of one sound on a neighboring sound in speech production, is an issue that has been extensively studied. The traditional view is that it is a universal phenomenon caused by speech physiology, but it has become clear that both the pattern and the degree of coarticulation can be language-specific (Baumotte & Dogil, 2008; Beddor, Harnsberger, & Lindemann, 2002; Choi & Keating, 1991; Gandour, 1994; Hardcastle & Hewlett, 2006; Manuel, 1990; Oh, 2008).

A speech sound is influenced by both the preceding sound (i.e. carryover effect) and the subsequent sound (i.e. anticipatory effect). Such bidirectional coarticulatory effects have been reported in vowels and consonants, e.g., carryover effect as shown in Recasens (1984) and Beddor, Harnsberger, and Lindemann (2002); and anticipatory effects as shown in Martin and Bunnell (1981) and Grosvald (2009). Whalen (1990) has proposed that the carryover effect is a result of physiological constraints in realizing some motor program, since this effect remains robust when cognitive planning is constrained. The anticipatory effect, on the other hand, would be a reflection of speech planning, in that it decreases when the participants' planning mechanism is inhibited.

Different from vowel and consonant coarticulation, the realization of pitch target in tone production relies on a single articulator, the larynx. Therefore, adjacent tones with opposing pitch targets have to compromise with each other, as overlap in the timing of different gestures, which is common for coarticulation in segments (Browman & Goldstein, 1986), is less feasible in tonal coarticulation (Xu, 1994; DiCanio, 2014). Understanding tonal coarticulation, in comparison to segmental coarticulation, is thus important for research on coarticulation in general.

Most experimental studies on tonal coarticulation have been based on Asian contour-tone languages and the findings, as mentioned earlier, have been mainly in three aspects: the directionality (carryover or anticipatory), the nature (assimilatory or dissimilatory), as well as the magnitude and temporal extent of the articulatory effects. Similar to vowel and consonant coarticulation, previous findings show that tonal coarticulation can also be bidirectional, with assimilatory carryover effect and dissimilatory anticipatory effect. The carryover effect is generally strong and its influence can extend to the first half of or even the entire following syllable. The magnitude and temporal extent of the anticipatory effect is generally smaller compared to the carryover effect (e.g., Thai: Abramson, 1979; Gandour, Potisuk, & Dechongkit, 1994; Gandour, Potisuk, Ponglorpisit, Dechongkit, Khunadorn, & Boongird, 1996; Potisuk, Gandour, & Harper, 1997; Vietnamese: Brunelle, 2009; Han & Kim, 1974), although more recent experimental studies suggest that the anticipatory effect can also be quite salient (see, e.g. Chang & Hsieh, 2012; Li & Chen, 2016).

The patterns of Mandarin tonal coarticulation generally agree with the findings from the above reported patterns for other tone languages. Xu (1997) examined tonal coarticulatory patterns in disyllabic non-words /mama/ with all possible tonal combinations in Standard Chinese produced by native Beijing Mandarin speakers. The results showed that both carryover and anticipatory effects exist in Mandarin. The carryover effect exhibits an assimilatory nature; a high offset of the first tone can raise the onset of the following tone, while a low offset lowers the onset of the following tone. This effect shows a strong influence on the initial and middle part of the final syllable. The anticipatory effect, however, is largely dissimilatory. Xu (1997) showed that in Standard Chinese, the anticipatory effect is mainly on the maximum f0 of the preceding tone. Specifically, the low tone (T3) in the second syllable showed a raising effect on the initial part of the falling tone (T4) and the final part of the rising tone (T2). The magnitude of the anticipatory effect, however, is much smaller compared to the carryover effect. (It is to be noted that in a closely related Mandarin dialect, Tianjin Mandarin, Li & Chen (2016) showed that anticipatory raising may manifest as the raising of the whole tonal contour. More experimental studies are therefore needed to understand the full range of tonal coarticulatory effects in different dialects within the Mandarin dialect family.)

Thus far, what has remained under-investigated is the underlying mechanism and source of the tonal coarticulatory effects. Among the limited number of studies focusing on this issue, Xu (2001) argued that the carryover effect in Mandarin tone is most likely caused by articulatory constrains (i.e., the maximum speed of pitch change). For the anticipatory effect, Tilsen (2009, 2013) reported that native Mandarin speakers tended to dissimilate tones that were planned contemporaneously, which led him to suggest that the dissimilatory effect may result from an inhibitory speech planning mechanism between articulatory targets planned in parallel, with the goal to maintain and maximize phonemic contrasts.

Recently, Franich (2015) took a step further in this line of investigation by introducing the effect of cognitive load. Since motor planning in speech production is believed to recruit central processing resources (Gathercole & Baddeley, 2014; Meyer & Gordon, 1985), increase of cognitive load (Mattys & Wiget, 2011) is therefore expected to introduce the reduction of processing resources for articulatory planning. To introduce cognitive load, a dual-task paradigm was used in Franich's (2015) study. Native Mandarin speakers were asked to read disyllabic Mandarin non-words while being told that they would need to recall the two-digit numbers given before the reading of the non-words. A robust carryover effect was found in both normal and cognitive load conditions. Furthermore, dissimilatory anticipation effect was found to increase under high cognitive load, especially on the high tone (Tone 1) and the low tone (T3). This result is puzzling, however, given the earlier finding that inhibited planning should lead to decrease in anticipatory effect (Whalen, 1990). A possible reason is, as argued by Franich, that anticipatory coarticulation carries important linguistic function (maintaining and maximizing contrasts between phonemic categories) and may therefore have a dedicated cognitive mechanism for its realization even under high cognitive load. This then predicts that native and non-native speakers (especially beginning learners) may show differential effects of cognitive processing constraint on tonal coarticulation, if the cognitive mechanism is developed as a consequence of mastering the native language.

2.1.2 Tonal coarticulation patterns by L2 learners of Mandarin

Thus far, although existing studies on L2 Mandarin learners show that learners are able to correctly produce lexical tone in isolation (e.g., Hao, 2012; Wang, Jongman, & Sereno, 2003), their production of tones in connected speech is greatly challenged, evident in the higher error rates and decreased intelligibility of L2 speech (Hao, 2012; Shen, 1989; Sun, 1998; Yang, 2011, 2016). He and Wayland (2010) found that when producing coarticulated tones in disyllabic words, more experienced American learners of Mandarin are more accurate than less experienced learners. Brengelmann, Cangemi and Grice (2015) examined anticipatory tonal coarticulation in disyllabic sequences in German learners of Mandarin. Compared to native Mandarin speakers, German learners showed that for all four lexical tones, the influence of the following tone was mainly on the final part of the initial syllable. Furthermore, they also produced more f0 variations in the last 20 percent of the tone contours on the first syllable, with much of the variability due to non-native like anticipatory coarticulation. What remains to be learned is to what extent the observed pattern is general among learners of non-tonal languages. Their results motive more systematic studies of the acquisition of tonal coarticulation. A follow-up question is the developmental trajectory and mechanisms of tonal coarticulation that underlies the ultimate attainment of tonal acquisition by non-tonal second language learners.

To address these questions, the current study adapted the paradigm used in Franich (2015) and tapped further into the cognitive mechanisms of tonal coarticulation by both native and non-native Mandarin speakers. We intended to replicate the findings of Franich (2015) for native Mandarin speakers. Furthermore, we also tested beginning and advanced Dutch learners of Mandarin under cognitive load, aiming to reveal the developmental path of tonal coarticulation acquisition, which, we hope, can help to shed further light on the general mechanisms underlying tonal coarticulation.

2.2 Methods

2.2.1 Participants

Twelve Mandarin control participants and 22 Dutch learners of Mandarin participated in the experiment (10 beginning learners and 12 advanced learners). The native Mandarin control group had 3 males and 9 females (age: M = 26.3, SD = 3.0). All were from the Northern part of China and spoke standard Mandarin on a daily basis and fluently. Four were native speakers of Beijing Mandarin and the other eight speakers spoke standard Chinese as their dominant language, but they could speak another northern Mandarin dialect. All Dutch learners of Mandarin received formal Chinese training from the Chinese Studies program at Leiden University. The beginning group consisted of 4 males and 6 females (age: M = 20.6, SD = 2.5). Their Mandarin learning and speaking experience varied between 0.5 and 2 years (mean = 1.2, SD = 0.5), and they had never lived in China. The other 12 participants (4 males and 8 females; age: M = 24.0, SD = 3.6) were advanced Mandarin learners, who had Mandarin experience between 3 and 14 years (M = 4.8, SD = 3.1), and had spent at least one year in China.

2.2.2 Material and procedure

The stimuli, following the design of Xu (1997) on tonal coarticulation, were disyllabic non-word /mama/ with each syllable bearing one of the four Mandarin tones: the high tone (T1); the rising tone (T2), the low tone (T3) and the falling tone (T4) (Chen & Gussenhoven, 2008; Duanmu, 2000). When produced in the second syllable, T3 was expected to show a dipping contour just like its canonical form. It would be realized as a variant with low falling contour preceding T1, T2 and T4. According to the sandhi rule, T3 would be realized with a rising contour, similar to T2 preceding another T3. All 16 possible tonal combinations were tested with four repetitions in three conditions: no-cognitive-load, low-cognitive-load and high-cognitive-load condition. The cognitive-load conditions were manipulated following the paradigm of Lavie, Fockert and Viding (2004), with a minor change of using two-digit numbers as memory material in the low-cognitive-load condition instead of one-digit number. The participants were recorded individually in the Leiden University Phonetics Lab using E-prime (44.1 kHz, 16 bit) with a Sennheiser MKH416T microphone. The three groups of participants were asked to read the sequences given in pinyin, with instructions in their respective native languages (i.e. Chinese for the native Mandarin speakers and Dutch for the learners).

In the control no-cognitive-load (NCL) condition, a fixation point ("+") was first presented on the screen for 2s at the beginning of each trial. After that, a disyllabic pinyin with tone marks appeared on the screen. The participants were asked to simply read them aloud. They had 2.5s for each trial, and after that the next trial proceeded automatically.

For the two cognitive load conditions, the reading task was presented in the retention interval of a short-term memory task. In each trial, the reading task was preceded by memory material, and followed by memory testing material. In the low-cognitive-load (LCL) condition, the memory material was two one-digit numbers and in the high-cognitive-load (HCL) condition, it was six one-digit numbers. For both conditions, each trial started with a 2s presentation of a fixation point ("+") in the center of the screen. After that, a row of two digits were presented equally spaced (horizontally) for 500 ms in the LCL condition. In the HCL condition, a row of six digits were presented for 2s. During the presentation of the memory material, the participants were asked to try their best to remember the digits. Then, the memory digits were replaced by masking arrays with a 500-ms display of two asterisks for the LCL condition, and one 1s display of 6 asterisks for the HCL condition. The masking array was then followed by the presentation of pinyin with tone marks. A time window of 2.5s was provided for participants to read the pinyin aloud. After that, a green digit was presented as the memory testing material. The participants were required to decide whether this digit was present or not in the preceding memory material by pressing "j" (indicated by a green sticker with "yes" on the keyboard) or "k" (indicated by a red sticker with "no" on the keyboard). After the participants responded, the next trial followed automatically.

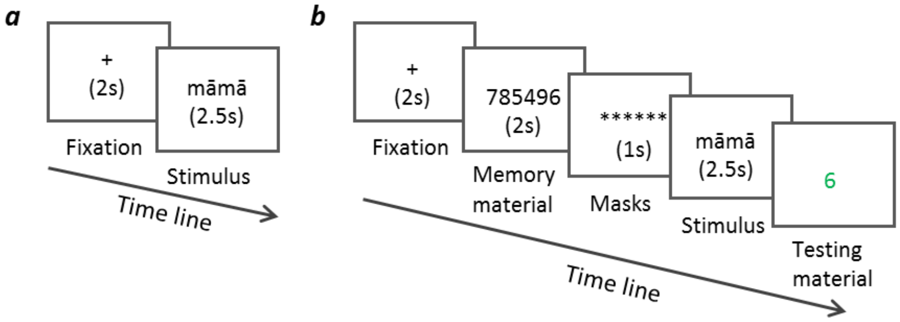

Figure 2.1. *The procedures of the no-cognitive-load condition (panel a) and high-cognitive-load condition (panel b).*

The digits in the memory set were selected from 1 to 8. Each digit was equally likely to appear in each position in the memory set of both conditions. The order of the two digits in the LCL condition was random, and the two digits for the same trial were always different from each other. The order of the six digits in the HCL condition was also random, under the condition that the same digit never appeared more than twice in a trial, and no more than two digits appeared in sequential order. For both conditions, the memory testing digit was equally likely to be present or absent in the memory material. If the digit was present in the memory material, it was equally likely to appear in any possible two or six positions in the memory sequences (see Figure 2.1).

Each condition consisted of four repetitions of the 16 disyllabic non-words. The 64 trials were presented as two blocks in different random orders to each participant. Furthermore, the order of the three conditions was also randomized across participants. In total, there were 192 trials (16 disyllabic combinations × 4 repetitions × 3 conditions) for each participant.

2.2.3 Pre-processing of the data

For all three participant groups, the coarticulation patterns in the LCL condition were not apparently different from the NCL control condition. So, only the results from the NCL control condition and the HCL condition are reported here. In the HCL condition, the error rate in the memory test was low across three groups (2.3% for NM; 7.5% for BL; 5.7% for AL). A total of 238 trials (out of 4,352) in which participants failed to respond accurately in the memory test were excluded. Furthermore, for both conditions, 34 trials were excluded in which participant had failed to read the stimulus within the assigned time window (2 trials for NM; 23 trials for BL; 12 trials for AL). In all, the recordings from 4,077 trials (out of 4,352) were included in the next step.

All 4,077 recordings were evaluated by a native Mandarin speaker in a tone identification task. For native Mandarin speakers, 2.9% trials could not be correctly identified. The production error rate for advanced learners was 7.7% and even higher for the beginning learners (13.8%). Only the recordings that were correctly identified by the native Mandarin listener (3,767 recorded disyllabic sequences) were used for the final f0 analysis.

2.2.4 F0 analysis

The boundaries of the vowels and nasal consonants were manually labelled in Praat (Boersma & Weenink, 2016) using a custom-written script (Chen, 2011).[3] The f0 extraction was also done in Praat (time step = 0.01s; pitch floor = 75 Hz). The f0 contours were obtained by taking 20 equidistant points for vowels, and 10 points for nasal consonants using the same custom-written script. To normalize the individual differences in f0 range, each participant's raw f0 data was transformed to participant-specific z-scores.

---

[3] Chen, Y. (2011). Generate norm F0.praat (praat script).

2.2.5 Statistical analysis

In order to give a more systematic and detailed report on coarticulatory effect, we examined the overall f0 height, the slope, as well as the steepness of the f0 contour of target tones.

Given the time-varying nature of the tonal contours, we adopted the growth curve analysis (GCA) (Mirman, 2014) with linear mixed model in R. GCA is a multilevel regression method using orthogonal polynomials to fit non-linear time course data. It is powerful in quantifying and analyzing the shapes of time course curves (e.g. f0 data). In the present study, second-order orthogonal polynomials were used with three parameters representing a curve's characteristics, that is, $y = a + bx + cx^2$. The intercept a refers to the overall mean of the f0 curves; the linear term b indicates the direction of the f0 curves (rise or fall); the quadratic term c refers to the steepness of the curvature. Two different f0 contours should expect at least one statistical significance among the three aspects.

Models were built for each target tone in both the first-syllable and second-syllable positions. The fixed effects consisted of the Linear Term, the Quadratic Term, and the experimental conditions, which include the Tonal Context (i.e. the preceding and following tones), Cognitive Load Condition (i.e. NCL and HCL), the Participant Group (i.e. NM, BL and AL), in addition to their interactions. Repetition and interaction between Repetition and Linear and Quadratic Terms were also included as fixed effects. For random effects, we had intercepts for Subjects, as well as by-Subject random slopes for the Cognitive Load Condition and Tonal Context.

Duration of the first and second vowel in the /mama/ sequence were also tested with linear mixed modeling. A model was built with Participant Group, Tone of the 1st syllable, Tone of the 2nd syllable, Cognitive Load, the interactions of these factors, and Repetition as fixed effects. Intercepts for Subjects was used as a random effect.

The significance of main effects in models of f0 contours and vowel durations were obtained via likelihood ratio comparisons with the change in log-likelihood distributed as χ^2. The degrees of freedom equaled the number of parameters added. The results of the main effects are presented in Appendices A1 and A2.

For both models of f0 contours and vowel duration, post-hoc comparisons were conducted using the *glht* function in the Multcomp package with Bonferroni adjustment in R (Hothorn, Bretz & Westgall, 2008). More specifically, for models of f0 contours, we compared the influence of each pair of contextual tones with contrastive offsets or onsets on the target tones for each participant group and cognitive load condition in the post-hoc comparison. Specifically, for the carryover effect, we compared all pairs of high-ending tones versus low-ending tones (T1 vs. T3, T1 vs. T4, T2 vs. T3 and T2 vs. T4). For the anticipatory effect, tonal pairs with contrastive onsets were compared (T1 vs. T2, T1 vs T3, T2 vs. T4 and T3 vs. T4).

2.3 Results

The fixed effect of Tonal Context was significant in all models of target tones in the first- and second-syllable positions (all p values < 0.05). The main effect of Cognitive Load was significant for all the models of tones in the first-syllable position (all p values

< 0.05), while for models of tones in the second syllable, this effect was only significant for T4 [$\chi^2(1) = 7.42$, $p < 0.05$]. The effect of Participant Group was found significant only for T2 and T3, both in the first syllable position [T2: $\chi^2(2) = 32.47$, $p < 0.001$; T3: $\chi^2(2) = 10.64$, $p < 0.01$] and in the second syllable position [$\chi^2(2) = 16.55$, $p < 0.01$; $\chi^2(2) = 15.08$, $p < 0.001$]. The three-way interaction of Tonal Context, Cognitive Load and Participant Group was significant in all models for target tones on both the first and the second syllable positions (all p values < 0.05) (see Appendix A1 for more detailed results).

For models of vowel duration in first- and second-syllable positions, the effects of Participant Group, Tone of the 1st Syllable, Tone of the 2nd Syllable and Cognitive Load were significant (all p values < 0.05). For the model of vowel duration in the first syllable, other significant effects were the interaction between Participant Group and Tone in the 1st Syllable [$\chi^2(6) = 249.96$, $p < 0.001$], the interaction between Participant Group and Tone in the 2nd Syllable [$\chi^2(6) = 27.05$, $p < 0.001$], as well as the three-way interaction of Participant Group, Tone in the 1st Syllable and Cognitive Load [$\chi^2(8) = 31.97$, $p < 0.001$]. For the model of vowel duration in the second syllable, significant interaction was found between Participant Group and Tone in the 2nd Syllable [$\chi^2(6) = 423.95$, $p < 0.001$].

In the following, we will present figures and statistical analyses of the carryover (§ 2.3.1) and anticipatory effects (§ 2.3.2). In both sections, the figures of tonal contours will be presented first. Subsequently, the post-hoc results of interaction of Participant Group, Tonal Context and Cognitive Load in each model will be presented for discussion of the fine-grained details in f0-contour. Finally, the results of the target tones' duration will be reported.

2.3.1 Carryover effect

2.3.1.1 Native Mandarin speakers

Figure 2.2 presents the carryover effect of the preceding tones on the contours of the following tone in /mama/ sequences produced by native Mandarin speakers without cognitive load. At the syllable boundary, the f0 onset of the second syllable was considerably influenced along the same direction by the offset of the first tone. Specifically, the high offset of the preceding tone (T1 and T2) raised the f0 of the initial nasal part of the following tone, and the low offsets (in T3 and T4) lowered the nasal part of the following tone. Furthermore, the influence of the preceding tone decreased over time: the f0 contour varied enormously during the initial nasal part; the f0 contours also differed at the beginning part of the vowel, and remained sizeable at the vowel offset for T1 and T3. In Xu (1997), the target non-words were produced in carrier sentences, therefore T3 in the second syllable was only realized as a low falling contour. Since the non-words were presented in isolation in our study, T3 in the second syllable showed a dipping contour. The general observations of carryover effect in the current study are similar to the results in Xu (1997).

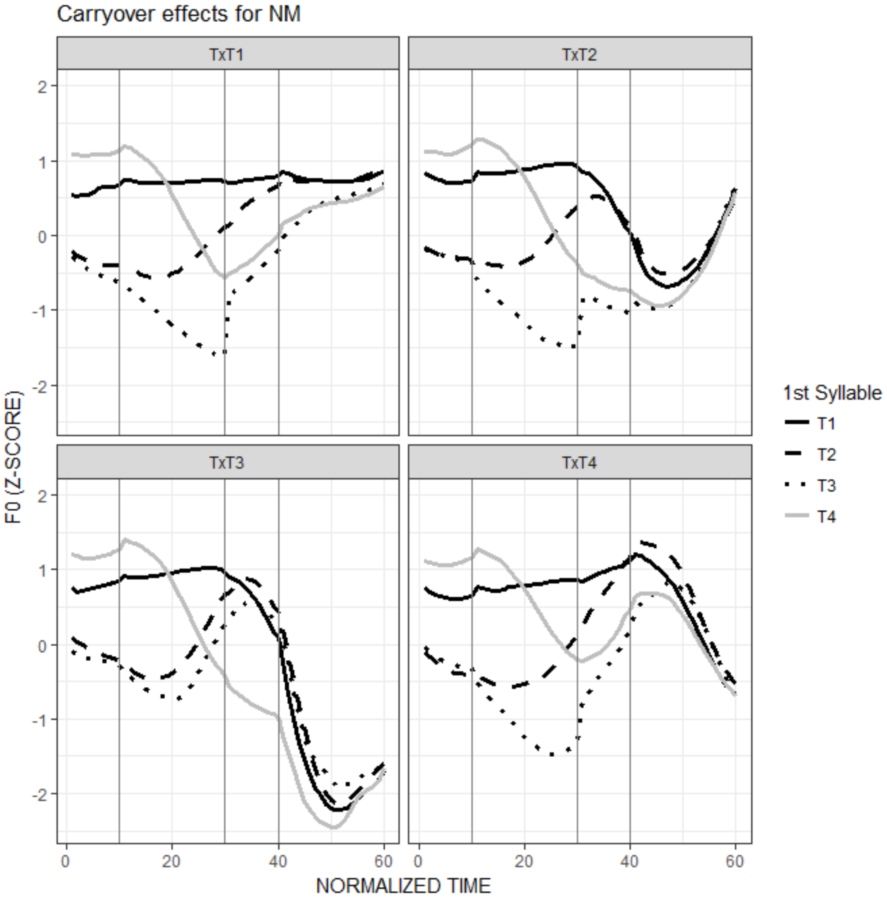

Figure 2.2. *Carryover effects for native Mandarin speakers in the NCL control condition. In each panel, the tone in the second syllable is held constant (T1-T4) and the tone in the initial syllable varies.*

To examine the effect of cognitive load on carryover coarticulation, we plotted in Figure 2.3 the f0 contours over the vowel portion of the second syllable, as a function of different preceding tones under different cognitive load conditions (left: the NCL condition and right: the HCL condition).

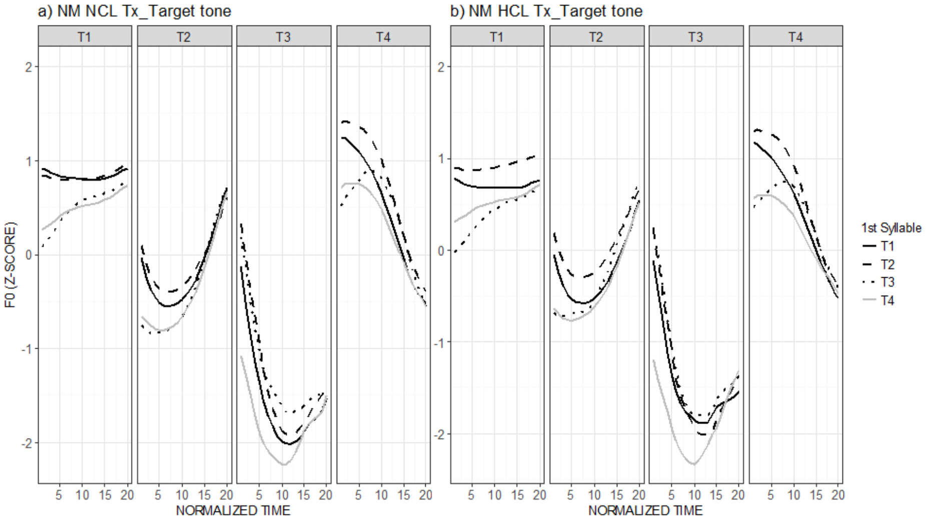

Figure 2.3. *F0 contours (over the vowel part) of the four target tones when preceded by different tones produced by native Mandarin speakers. Normalized f0 contours averaged across participants.*

Figure 2.3*a* shows that in the NCL condition, the whole contour of T1 was lowered by the low offset of the preceding T3 and T4, resulting in an initial rising contour. Both the overall f0 height and f0 slope of the second T1 were significantly different following tones with high offsets vs. low offsets, presenting an assimilatory pattern (Table 2.1*a*). Similar to T1, the initial part of T2 in the second syllable was also significantly affected by the offsets of the preceding tone in the first syllable in the NCL condition (Table 2.1*b*). The contour of T3 (Figure 2.3*a*) in the second syllable was significantly lower when following T4 than following T1 and T2. It should be noted that, according to the phonological rule, when T3 is followed by another T3, the first T3 is realized with a rising contour, similar to the lexical rising tone (T2) with high offset. So the contour of T3 did not show significant difference when following T3 vs. following T1 and T2 (Table 2.1*c*). T4 in the second syllable was affected by the preceding tone (Table 2.1*d*), showing significant difference in at least one parameter in all tone pairs with contrastive offsets (T1 vs. T3 and T4; T2 vs. T3 and T4). Overall, the statistical analysis of the carryover coarticulatory effect for native Mandarin speakers is in line with that in Xu (1997).

In the HCL condition, similar assimilatory carryover effect was found for all tones in the second syllable, as suggested by the similar significant effect of tonal context in the two conditions (Figure 2.3*b*, Table 2.1).

Table 2.1. *Pairwise comparison of results for adapted contours on the vowel part of the second tone due to preceding tones with high offsets (T1 and T2) vs. low offsets (T3 and T4) for native Mandarin speakers.*

***a.* T1 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	P	Est.	z	p	Est.	z	p	Est.	z	p
intercept	−0.32	−4.15	<.001	−0.32	−4.16	<.001	−0.33	−4.22	<.001	−0.33	−4.24	<.001
slope	0.87	8.41	<.001	0.56	5.46	<.001	0.71	6.91	<.001	0.41	3.96	<.01
quadratic	0.07	−5.75	<.001	−0.22	−3.01	<.05	−0.38	−5.29	<.001		n.s.	
HCL	**after T1 vs. after T3**			**after T1 vs. after T4**			**after T2 vs. after T3**			**after T2 vs. after T4**		
	Est.	z	P	Est.	z	p	Est.	z	p	Est.	z	p
intercept	−0.30	−3.93	<.01			n.s.	−0.49	−6.31	<.001	−0.39	−5.02	<.001
slope	0.92	8.83	<.001	0.49	4.70	<.001	0.69	6.54	<.001		n.s.	
quadratic	−0.36	−4.86	<.001			n.s.	−0.31	−4.06	<.001		n.s.	

***b.* T2 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	P	Est.	z	p	Est.	z	p	Est.	z	p
intercept			n.s.			n.s.	−0.26	−3.39	<.05	−0.28	−3.72	<.01
slope	0.88	4.45	<.001	0.62	3.15	<.05	1.16	5.85	<.001	0.90	4.56	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	**after T1 vs. after T3**			**after T1 vs. after T4**			**after T2 vs. after T3**			**after T2 vs. after T4**		
	Est.	z	P	Est.	z	p	Est.	z	p	Est.	z	p
intercept		n.s.			n.s.		−0.26	−3.39	<.05	−0.39	−5.22	<.001
slope	1.03	5.18	<.001	0.61	3.11	<.05	1.09	5.47	<.001	0.68	3.42	<.05
quadratic	−0.32	−3.03	<.05		n.s.			n.s.			n.s.	

Table 2.1. *(Continued)*

***c.* T3 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.		−0.51	−4.52	<.001
slope		n.s.		1.20	3.43	<.05		n.s.		1.76	5.03	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		−0.43	−3.80	<.01		n.s.		−0.48	−4.27	<.001
slope		n.s.		1.47	4.19	<.001		n.s.		1.89	5.33	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	

***d.* T4 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.		−0.56	−4.78	<.001
slope	1.05	5.33	<.001	0.77	3.90	<.01	1.07	5.46	<.001	0.80	4.04	<.001
quadratic	−0.87	−6.91	<.001		n.s.		−0.56	−4.42	<.001		n.s.	
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.		−0.51	−4.35	<.001
slope	1.14	5.74	<.001	0.89	4.44	<.001	1.28	6.38	<.001	1.02	5.09	<.001
quadratic	−0.57	−4.40	<.001		n.s.			n.s.			n.s.	

Table 2.2. *Mean duration (in ms) of second vowel in* /mama/ *sequence with four tones produced by native Mandarin speakers in the NCL and the HCL conditions.*

Cognitive load	Tone			
	T1	T2	T3	T4
NCL	251	272	273	162
HCL	233	267	264	152

In both NCL and HCL conditions, T4 in the second syllable was the shortest tone, significantly shorter than the other three tones (all *p* values < 0.01). T1 showed an intermediate duration, while T2 and T3 were the longest tones, significantly longer than T1 and T4 (all *p* values < 0.01). In the HCL condition, all tones exhibited a reduced duration. T1, T3 and T4 were significantly shorter than that in the NCL condition (all *p* values < 0.05). This result was different from that in Xu (1997), which found T3 the shortest one in the second syllable, and T1, T2 and T4 comparable in duration. Such divergence may be attributed to different recording procedures. In Xu (1997), the non-words were produced in carrier sentence, and the duration might be influenced by the context. In the current study, the non-words were recorded in isolation. So in the second syllable, the duration of four tones showed a pattern which was similar to that in monosyllabic production.

2.3.1.2 Beginning Dutch learners of Mandarin

Figure 2.4 presents the carryover effect of the preceding tones on the contours of the following tone in /mama/ sequences without cognitive load produced by beginning Dutch learners of Mandarin. For all tones, the general pattern was also assimilatory, but the magnitude was smaller compared to native speakers. For T1 and T4, the influence of the tonal context was shown on both nasal part and the vowel part. For T2 and T3, however, the assimilatory pattern was not apparent on the vowel part.

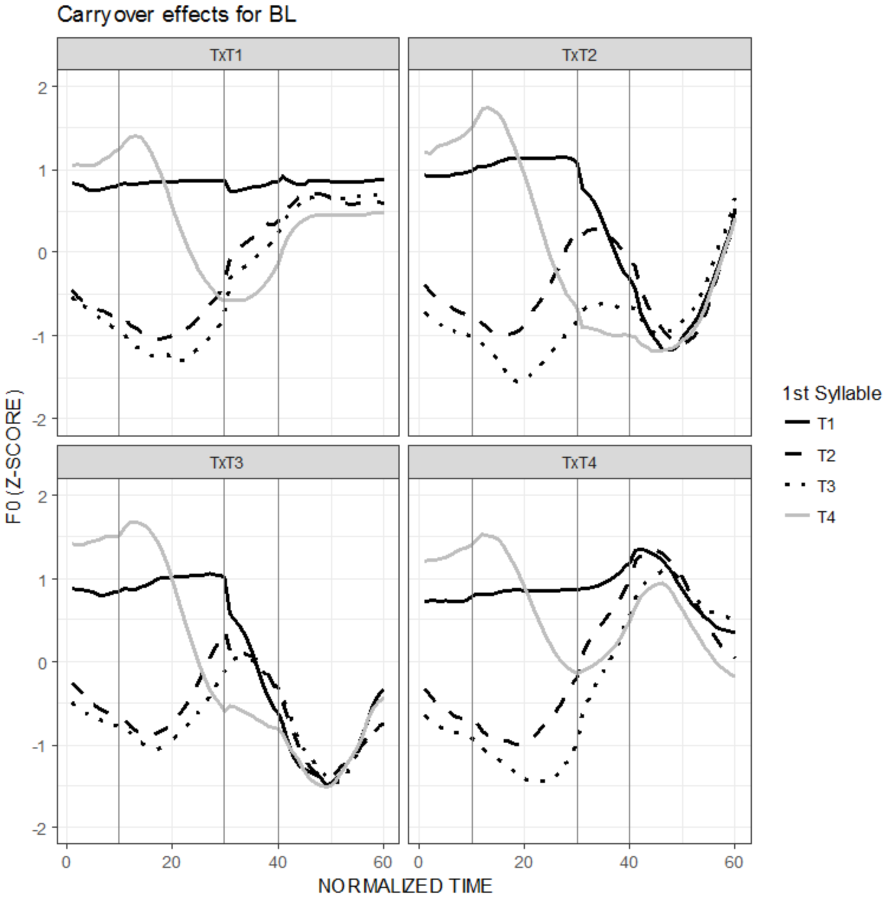

Figure 2.4. *Carryover effects for beginning Dutch learners of Mandarin in the NCL control condition. In each panel, the tone in the second syllable is held constant (T1-T4) and the tone in the initial syllable varies.*

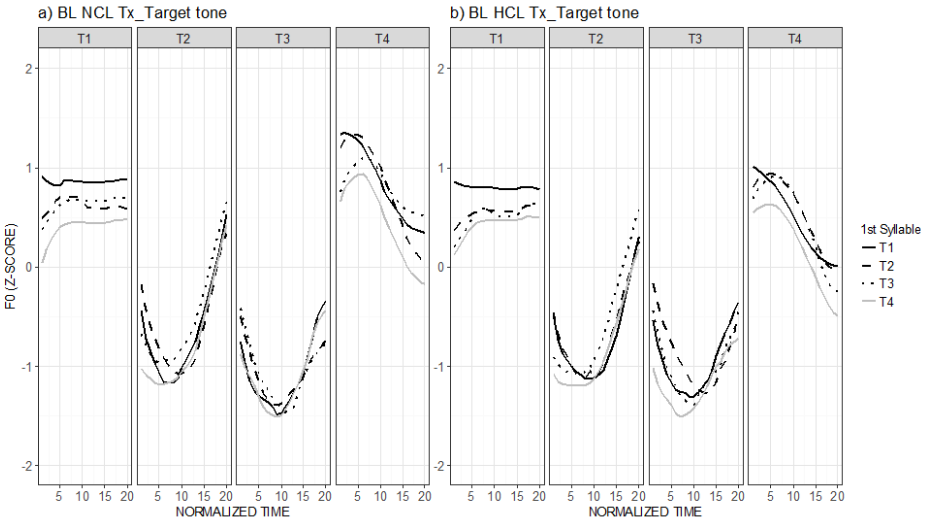

Figure 2.5. *F0 contours (over the vowel part) of the four target tones when preceded by different tones produced by beginning learners of Mandarin. Normalized f0 contours averaged across participants.*

In the NCL condition (Figure 2.5*a*), the f0 contour of target T1 was lower following T4 than the other three tones. Table 2.3*a* shows that overall f0 mean and quadratic term of T1 in the second syllable were significantly different following T1 and T4. As shown in the previous section, for Mandarin native speakers, the contours of target T1 on the second syllable clustered into two groups according to offsets of the initial tones. For beginning learners, this tendency was not clear. In the case of target T2 (Table 2.3*b*), the assimilatory effect on f0 height was not apparent. The influence of the preceding tone was mainly found in the timing of the turning point and sharpness of the contour. When following tones with low offsets (T3 and T4), target T2 was less concave and showed earlier turning point. Statistical significances in slope and sharpness were found when following T2 vs. T3 and T2 vs. T4. For target T3 in the NCL condition (Table 2.3*c*), the assimilatory pattern was not obvious. The contour of T4 (Table 2.3*d*) was influenced along the same direction by the offset of the first tone, showing significant difference in slope or steepness in three tone pairs with contrastive offsets (T1 vs. T3, T1 vs. T4 and T2 vs. T3).

The assimilatory influence from the preceding tone was slightly stronger for T1 in the HCL condition, with additional significant difference after T1 and T3. In case of T2, the influence from the initial tone was mainly on the slope and steepness of the contour, which was similar to that in the NCL condition. For T3, significant difference was found in slope following T2 vs. T4. In terms of T4, significant difference was only found for quadratic terms, showing a slightly weaker assimilatory pattern compared to the NCL condition.

Table 2.3. *Pairwise comparison of results for adapted contours on the vowel part of the second tone due to preceding tones with high offsets (T1 and T2) vs. low offsets (T3 and T4) for beginning learners.*

***a.* T1 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		−0.46	−5.46	<.001		n.s.			n.s.	
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.		−0.28	−3.50	<.01		n.s.			n.s.	
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	−0.35	−4.09	<.001	−0.44	−5.23	<.001		n.s.			n.s.	
slope	0.47	3.98	<.01	0.43	3.66	<.01		n.s.			n.s.	
quadratic		n.s.		−0.27	−3.16	<.05		n.s.			n.s.	

***b.* T2 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope		n.s.			n.s.		1.01	4.48	<.001	1.15	5.27	<.001
quadratic		n.s.			n.s.		−0.37	−3.05	<.05	−0.51	−4.35	<.001
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope		n.s.		0.67	3.02	<.05	0.78	3.31	<.05	0.82	3.68	<.01
quadratic		n.s.		−0.59	−4.72	<.001		n.s.		−0.46	−3.70	<.01

Table 2.3. *(Continued)*

***c.* T3 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope		n.s.			n.s.			n.s.		1.30	3.19	<.05
quadratic		n.s.			n.s.			n.s.			n.s.	

***d.* T4 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope	0.80	3.56	<.01		n.s.		1.17	4.95	<.001		n.s.	
quadratic		n.s.		−0.53	−3.64	<.01		n.s.			n.s.	
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope		n.s.			n.s.			n.s.			n.s.	
quadratic	−0.74	−4.66	<.001	−0.53	−3.53	<.01		n.s.			n.s.	

Table 2.4. *Mean duration (in ms) of second vowel in* /mama/ *sequence with four tones produced by beginning learners in the NCL and the HCL conditions.*

Cognitive load	Tone			
	T1	T2	T3	T4
NCL	305	368	386	177
HCL	293	348	370	172

For beginning learners, T4 was the shortest tone in both NCL and HCL conditions, significantly shorter than the other three tones (all *p* values < 0.01). T1 showed an intermediate duration, while T2 and T3 were significantly longer than the other two tones (all *p* values < 0.01). This result was similar to that of native speakers. In the HCL condition, all tones became shorter. T2 and T3 were significantly shorter in the HCL condition compared to the NCL condition.

2.3.1.3 Advanced learners of Mandarin

Figure 2.6 presents the carryover effect of the preceding tones on the contours of the following tone in /mama/ sequences produced by advanced Dutch learners of Mandarin without cognitive load. A similar assimilatory carryover effect was found, but the magnitude was smaller compared to native speakers.

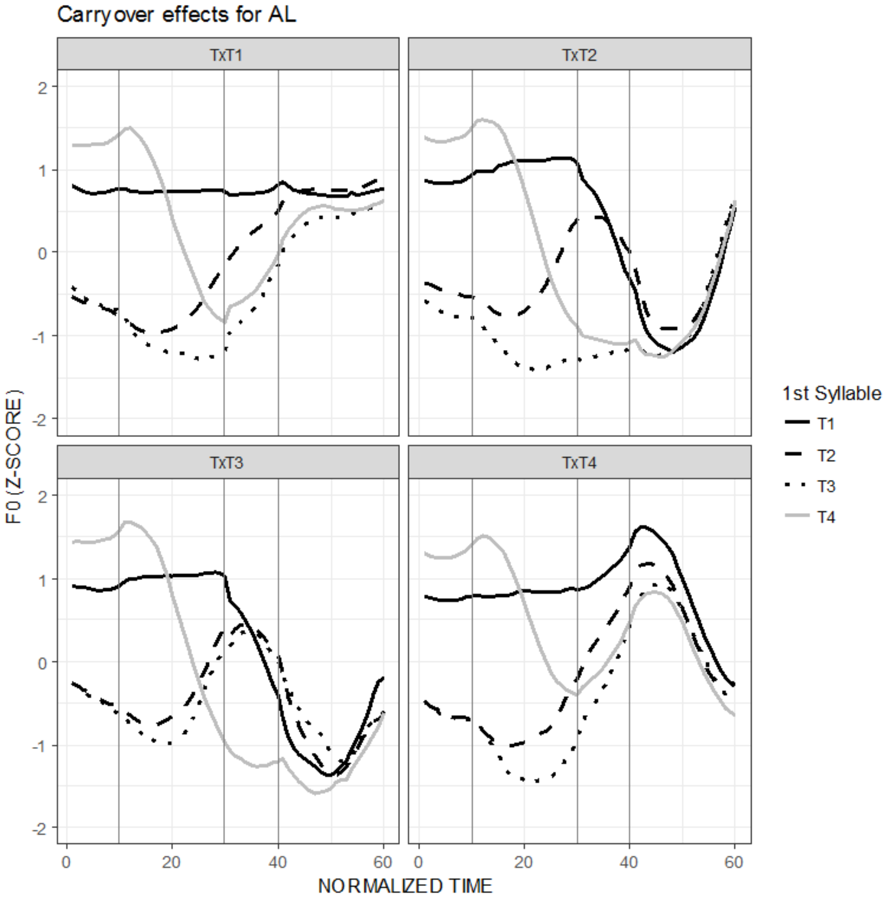

Figure 2.6. *Carryover effects for advanced Dutch learners of Mandarin in the NCL control condition. In each panel, the tone in the second syllable is held constant (T1-T4) and the tone in the initial syllable varies.*

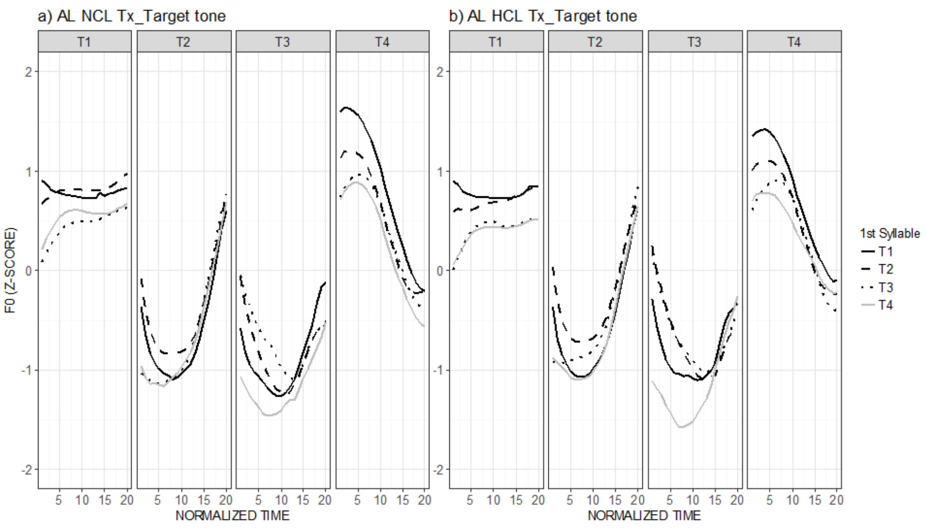

Figure 2.7. *F0 contours (over the vowel part) of the four target tones when preceded by different tones produced by advanced learners of Mandarin. Normalized f0 contours averaged across participants.*

In the NCL condition (Figure 2.7*a*), T1 showed divergent f0 contours following tones with contrastive offsets, presenting an assimilatory pattern. The T1 contours following all tone pairs with contrastive offsets showed statistical significance in at least one parameter (Table 2.5*a*), which was in line with the pattern of native speakers. T2 in the NCL condition also demonstrated assimilatory (Table 2.5*b*) carryover effect exerted by the offsets of the preceding tones, showing significant differences when following all tone pairs with contrastive offsets. A clear assimilatory tendency was also found for T3 (Figure 2.7*a*), however, the significant differences was only found for slope following T1 vs. T3 and T2 vs. T4 (Table 2.5*c*). The significant difference exerted by T1 vs. T3 in the first syllable was not expected. According to the tone sandhi rule, the offset of initial T3 was high when followed by another T3, and therefore T1 and T3 in the initial syllable (both with high offsets) should exert similar carryover effect on the following T3. In terms of T4, its contours was raised or lowered by the high or low offsets of the initial tones, showing significant difference when following all tone pairs with contrastive offsets except for T2 vs. T4 (Table 2.5*d*).

In the HCL condition, the assimilatory effect was maintained for target T1. For T2 in the HCL condition, the assimilatory effect became weaker. The assimilatory carryover effect was generally maintained in the HCL condition for T3 and T4.

Table 2.5. *Pairwise comparison of results for adapted contours on the vowel part of the second tone due to preceding tones with high offsets (T1 and T2) vs. low offsets (T3 and T4) for advanced learners.*

a. **T1 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	−0.32	−4.18	<.001	−0.23	−3.04	<.05	−0.30	−3.49	<.01		n.s.	
Slope	0.62	6.07	<.001	0.42	4.10	<.001	0.36	3.09	<.05		n.s.	
quadratic	−0.28	−5.29	<.001	−0.41	−5.71	<.001		n.s.		−0.26	−3.18	<.05
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	−0.36	−4.59	<.001	−0.36	−4.69	<.001	−0.29	−3.41	<.05	−0.30	−3.50	<.01
slope	0.46	4.37	<.001	0.45	4.31	<.001		n.s.			n.s.	
quadratic	−0.54	−7.18	<.001	−0.45	−6.04	<.001	−0.35	−3.92	<.01		n.s.	

b. **T2 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope	0.90	4.53	<.001	0.74	3.77	<.01	1.13	5.72	<.001	0.98	4.96	<.001
quadratic	−0.38	−3.63	<.01	−0.47	−4.51	<.001	−0.37	−3.51	<.01	−0.46	−4.40	<.001
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.		−0.33	−4.37	<.001
slope		n.s.			n.s.		0.88	4.41	<.001	0.85	4.26	<.001
quadratic	−0.46	−4.17	<.001		n.s.		−0.43	−3.91	<.01		n.s.	

Table 2.5. *(Continued)*

***c.* T3 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope	−1.39	−3.92	<.01		n.s.			n.s.		1.08	3.06	<.05
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		−0.36	−3.20	<.05		n.s.		−0.48	−4.16	<.001
slope		n.s.			n.s.			n.s.		2.09	5.75	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	

***d.* T4 in the second syllable**

NCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	−0.42	−3.57	<.01	−0.53	−4.54	<.001		n.s.			n.s.	
slope	0.79	3.95	<.01	0.71	3.60	<.01		n.s.			n.s.	
quadratic		n.s.			n.s.		−0.40	−3.00	<.05		n.s.	
HCL	after T1 vs. after T3			after T1 vs. after T4			after T2 vs. after T3			after T2 vs. after T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	−0.38	−3.19	<.05	−0.43	−3.61	<.01		n.s.			n.s.	
slope		n.s.		0.73	3.68	<.01		n.s.			n.s.	
quadratic	−0.55	−4.05	<.001		n.s.		−0.44	−3.02	<.05		n.s.	

Table 2.6. *Mean duration (in ms) of second vowel in* /mama/ *sequence with four tones produced by advanced learners in the NCL and the HCL conditions.*

Cognitive load	Tone			
	T1	T2	T3	T4
NCL	278	308	267	176
HCL	257	288	252	168

Different from native speakers, in both NCL and HCL conditions, T2 was the longest one, significantly longer than the other three tones (all p values < 0.01); T1 was the second longest tone, significantly longer than T3 and T4 (all p values < 0.01). T4 was the shortest one. In the HCL condition, all tones except T4 showed significantly shorter duration (all p values < 0.05).

2.3.1.4 Summary of carryover effect

For native Mandarin speakers, a strong and robust assimilatory carryover effect was found. The increase of cognitive load (Mattys & Wiget, 2011) is expected to introduce a reduction of processing resources for articulatory planning. However, the carryover effect was not obviously influenced by the high cognitive load, which indicated a lack in speech planning for this effect. This result was in line with previous findings (Franich, 2015; Whalen, 1990).

For beginning learners, the assimilatory carryover effect was found for T1, T2 and T4 in the NCL condition. The influence of cognitive load was weak with a slight increase in assimilatory effect for T1, T2 and T3, indicating the instability of this effect. Although these learners may be able to produce a lexical tone that fall into one of the four categories, correctly identified by native speakers, their assimilatory carryover effect was weaker than native speakers.

The advanced learners showed stronger carryover effect than beginners, exhibiting similar pattern with native Mandarin speakers. For T1, T2 and T4, strong assimilatory effect was found in the NCL condition. The assimilatory effect was weaker for T3. The carryover effect was generally maintained in the HCL condition.

2.3.2 Anticipatory effect

2.3.2.1 Native Mandarin speakers

Figure 2.8 presents variations in f0 contour of the first tones when followed by different tones in /mama/ sequences produced without cognitive load by native Mandarin speakers. Compared to the strong assimilatory effect exerted by the first ones on the second tones, the contours of the first tones showed less variability when followed by different tones. Moreover, the anticipatory was dissimilatory: tonal contours were higher when they were followed by tones with high onsets (T1 and T4) than followed by tones with low onsets (T2 and T3). The effect can be observed clearly for T1, T2 and T4 in the first syllable. For T1, the whole contour was raised by following tones with low onsets. For T2 and T4, this raising effect was strongest on the maximum f0 value in the tonal contours. The strongest effect was exerted by T3 in the following syllable, the maximum f0 of a tone preceding T3 was always higher than preceding other three tones. T2 in the following syllable also showed a similar raising effect for initial T1, T2 and T4. This general pattern was comparable to the findings of Xu's study (Xu, 1997), except that the whole contour of initial T1 in our study was clearly raised by tones with low onsets.

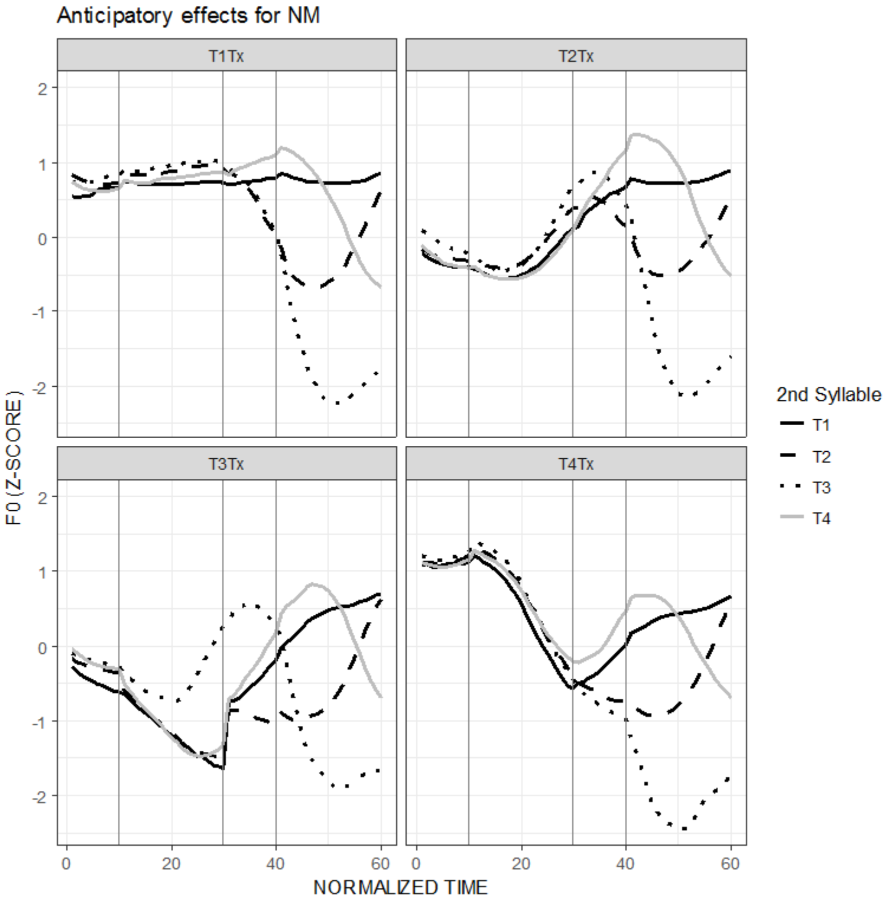

Figure 2.8. *Anticipatory effects for native Mandarin speakers in the NCL control condition. In each panel, the tone in the first syllable is held constant (T1-T4) and the tone in the second syllable varies.*

To examine the effect of cognitive load on anticipatory coarticulation, we plotted in Figure 2.9 the f0 contours over the vowel portion of the first syllable, as a function of different following tones under the NCL condition (panel *a*) and the HCL condition (panel *b*).

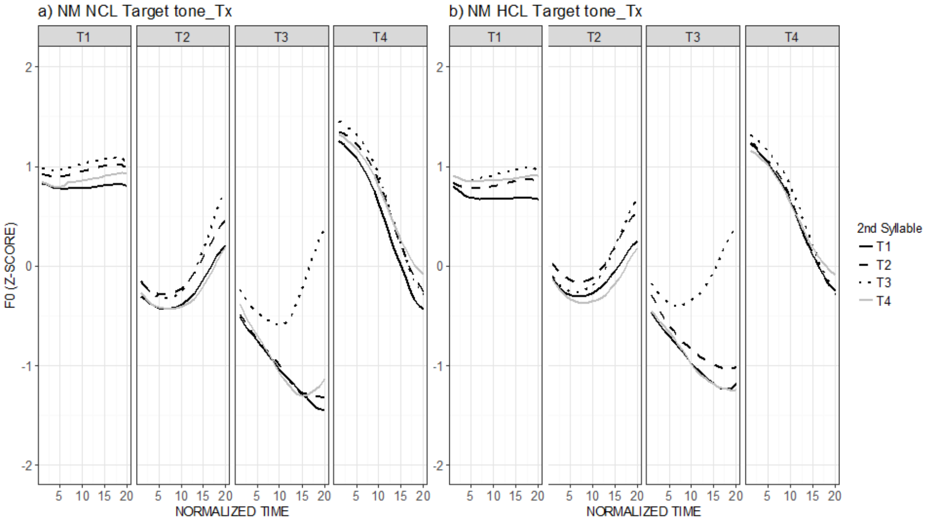

Figure 2.9. *F0 contours (over the vowel part) of the four target tones when followed by different tones produced by native Mandarin speakers. Normalized f0 contours averaged across participants.*

Figure 2.9*a* shows that in the NCL condition, the whole contour of T1 was raised by the low onsets of the following tones with significant differences when followed by all tone pairs with contrastive onsets except T2 vs. T4 (Table 2.7*a*). For T2, the whole contour was raised by the low onsets of the following tone. Significant difference was found in at least one parameter when T2 was followed by all tone pairs with contrastive onsets except T1 vs. T2 (Table 2.7*b*). According to the phonological rule, when T3 is followed by another T3, the first T3 is realized with a rising contour, similar to the lexical rising tone (T2). So the contour of T3 before another T3 shows different contour compared to T3 followed by the other tones. Other than this, the contours of T3 did not vary significantly when followed by tone pairs with contrastive onsets (when followed by T1 vs. T2 and followed by T2 vs. T4). This result was in line with the finding of Xu (1997), which also presented that the contour of T3 showed no statistical difference whether the following tone had a high offset of a low offset. The potential reason could be that the anticipatory raising effect mainly exerted on the maximum value of the initial tone. There was no high target in T3 therefore its contour was not sensitive to the anticipatory effect. In the case of T4, its contour was higher when followed by tones with low onsets than followed by tones with high onsets. The initial portion was raised the most, which also suggested that the anticipatory raising effect mainly affected the maximum f0 value. The comparison was significant when T4 was followed by T1 vs. T2 and T1 vs. T3 (Table 2.7*d*).

Table 2.7. *Pairwise comparison of results for adapted contours on the vowel part of the first tone due to following tones with high onsets (T1 and T4) vs. low onsets (T2 and T3) by native Mandarin speakers.*

***a.* T1 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	Z	*P*	Est.	z	*p*	Est.	z	*p*
intercept	0.16	3.10	<.05	0.23	4.39	<.001		n.s.		−0.16	−3.02	<.05
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	Z	*P*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.24	4.50	<.001		n.s.			n.s.	
slope		n.s.		0.25	3.41	<.05		n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	

***b.* T2 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	Z	*P*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.23	3.64	<.01	−0.20	−3.13	<.05	−0.25	−4.17	<.001
slope		n.s.		0.56	3.53	<.01		n.s.		−0.66	−4.18	<.001
quadratic		n.s.		0.31	4.82	<.001		n.s.		−0.26	−3.99	<.001
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	Z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.		−0.25	−3.91	<.01		n.s.	
slope		n.s.		0.60	3.74	<.01		n.s.		−0.75	−4.73	<.001
quadratic		n.s.		0.21	3.10	<.05		n.s.			n.s.	

Table 2.7. *(Continued)*

***c.* T3 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.71	7.14	<.001		n.s.		−0.66	−6.62	<.001
slope		n.s.		2.18	8.35	<.001		n.s.		−2.08	−8.17	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.77	7.77	<.001		n.s.		−0.83	−8.27	<.001
slope		n.s.		1.92	7.38	<.001		n.s.		−2.02	−7.68	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	

***d.* T4 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.19	3.00	<.05	0.23	3.69	<.01		n.s.			n.s.	
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	

In the HCL condition (Figure 2.9*b*), the dissimilatory anticipatory effect decreased for T1, with a significant difference only before T1 vs. T3. For T2 and T3, the dissimilatory anticipatory effect was maintained compared to the NCL condition. The anticipatory effect decreased for T4, with no significant difference found before tone pairs with contrastive onsets.

Table 2.8. *Mean duration (in ms) of first vowel in* /mama/ *sequence with four tones produced by native Mandarin speakers in the NCL and the HCL conditions.*

Cognitive load	Tone			
	T1	T2	T3	T4
NCL	176	188	184	176
HCL	174	187	188	171

In the NCL condition T2 was significantly longer than T1 and T4 (all p values < 0.05). In the HCL condition, the duration of T2 and T3 was significantly longer than T1 and T4 (all p values < 0.01). Compared to the NCL condition, the duration of the four tones in the HCL condition did not change significantly.

2.3.2.2 Beginning Dutch learners of Mandarin

As plotted in Figure 2.10, the tonal contours in the first syllable in the NCL condition also varied when followed by different tones for beginning Dutch learners of Mandarin. For most tones, the general pattern of the anticipatory effect was also dissimilatory.

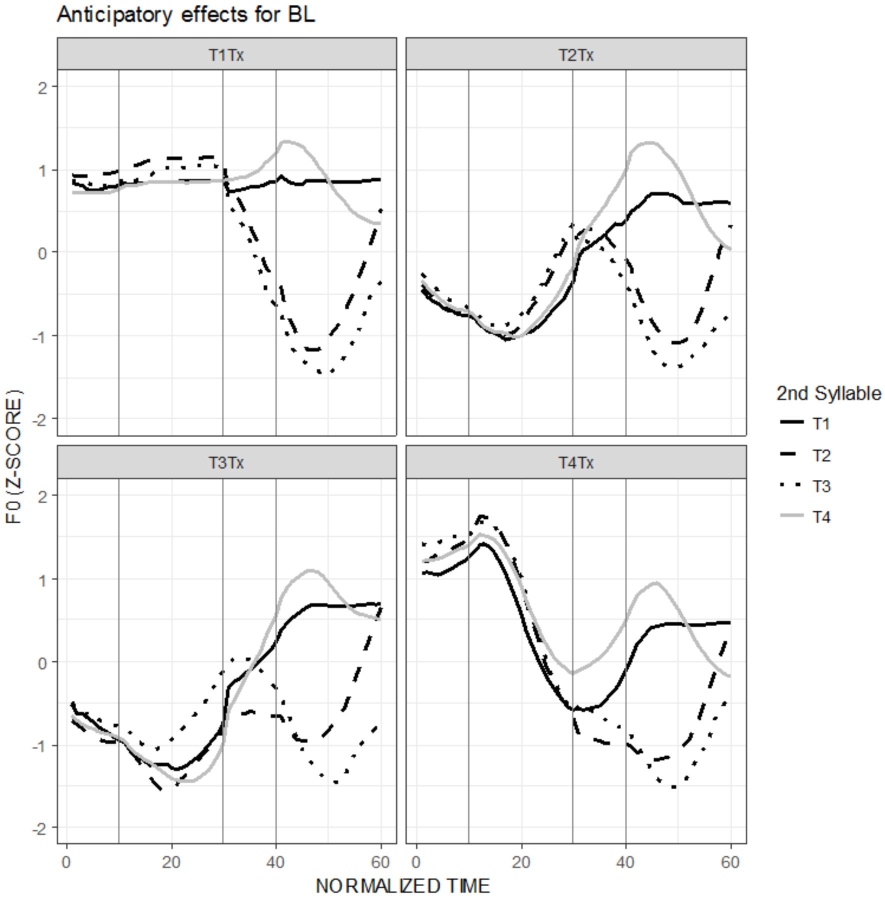

Figure 2.10. *Anticipatory effects for beginning learners of Mandarin in the NCL control condition. In each panel, the tone in the first syllable is held constant (T1-T4) and the tone in the second syllable varies.*

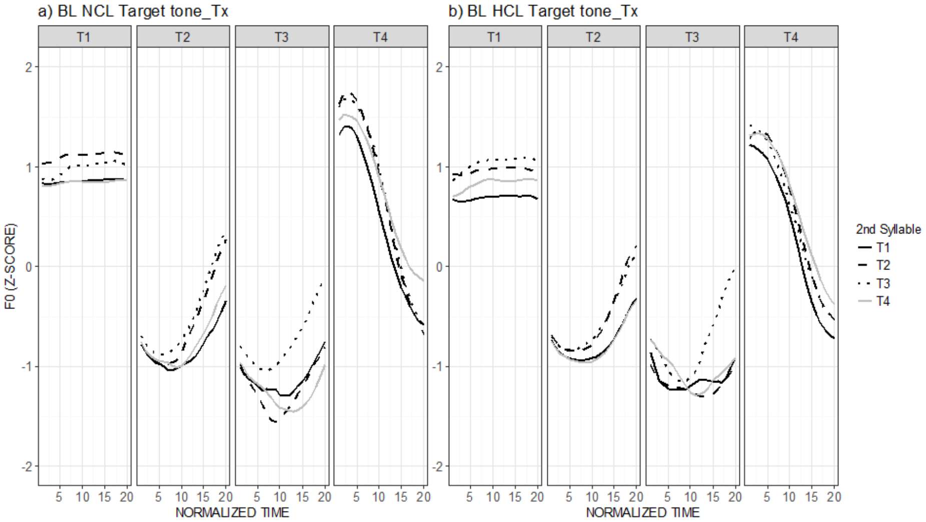

Figure 2.11. *F0 contours (over the vowel part) of the four target tones when followed by different tones produced by beginning learners of Mandarin. Normalized f0 contours averaged across participants.*

The contours of T1 were raised by the following tones with low onsets in the NCL condition (Figure 2.11*a*, Table 2.9*a*), showing significant difference before T1 vs. T2 and T2 vs. T4. T2 showed the strongest anticipatory effect among the four tones with significant differences when followed by all tone pairs with contrastive onsets. T3 in the NCL condition exhibited significantly different f0 contours when followed by T1 vs. T3 and T3 vs. T4 due to the phonological sandhi rule (see 2.3.1.1). The initial portion of T4 was raised by tones with low onsets, showing a dissimilatory pattern. The T4 contours also differed significantly followed by all tone pairs with contrastive onsets.

In the HCL condition, a similar pattern was found for T1, T2 and T3. The anticipatory effect became weaker in the HCL condition for T4, with only significant difference found in the overall f0 height when followed by T1 vs. T2.

Table 2.9. *Pairwise comparison of results for adapted contours on the vowel part of the first tone due to following tones with high onsets (T1 and T4) vs. low onsets (T2 and T3) by beginning learners.*

***a.* T1 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.27	4.57	<.001		n.s.		−0.26	−4.40	<.001		n.s.	
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.24	4.06	<.001	0.31	5.30	<.001		n.s.		−0.19	−3.29	<.05
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	

***b.* T2 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.29	4.09	<.001	0.40	5.58	<.001		n.s.		−0.26	−3.63	<.01
slope	1.22	6.66	<.001	1.10	5.98	<.001	−0.90	−4.92	<.001	−0.78	−4.25	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.26	3.59	<.01	0.27	3.67	<.01	−0.22	−3.09	<.05	−0.23	−3.17	<.05
slope	1.03	5.37	<.001	0.94	4.90	<.001	−0.77	−4.20	<.001	−0.68	−3.72	<.01
quadratic		n.s.			n.s.			n.s.			n.s.	

Table 2.9. *(Continued)*

c. T3 in the first syllable

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.35	3.22	<.05		n.s.		−0.41	−3.71	<.01
slope		n.s.			n.s.			n.s.		−1.01	−3.44	<.05
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.39	3.55	<.01		n.s.			n.s.	
slope		n.s.		1.04	3.52	<.01		n.s.		−1.11	−3.65	<.01
quadratic		n.s.			n.s.			n.s.			n.s.	

d. T4 in the first syllable

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.02	2.98	0.05		n.s.			n.s.			n.s.	
slope		n.s.			n.s.		1.18	5.59	<.001	1.06	5.06	<.001
quadratic		n.s.		−0.34	−3.28	<.05		n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.21	3.08	<.05		n.s.			n.s.			n.s.	
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	

Table 2.10. *Mean duration (in ms) of first vowel in* /mama/ *sequence with four tones produced by beginning learners of Mandarin in the NCL and HCL conditions.*

Cognitive load	Tone			
	T1	T2	T3	T4
NCL	272	274	294	215
HCL	256	265	274	209

In both conditions, T3 was the longest tone, significantly longer than the other three tones (all *p* values < 0.01). T1 and T2 had similar intermediate duration, while T4 was the shortest one, significantly shorter than the other three tones (all *p* values < 0.01). In the HCL condition, the duration of T1, T3 and T4 was significantly shorter than the NCL condition.

2.3.2.3 Advanced learners of Mandarin

The varied tonal contours in the first syllable due to different following tones produced without cognitive load by advanced learners of Mandarin in the NCL condition is plotted in Figure 2.12. The contours of the target tones were high when followed by tones with low onsets, showing a dissimilatory anticipatory pattern comparable to that of native Mandarin speakers.

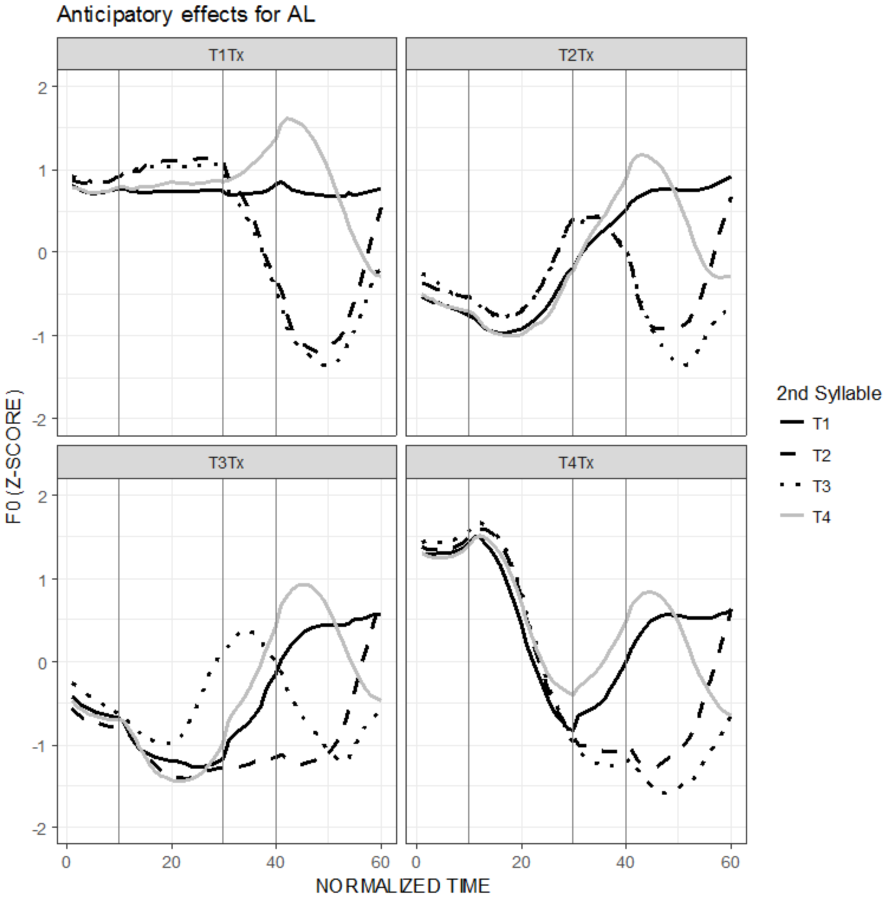

Figure 2.12. *Anticipatory effects for advanced Dutch learners of Mandarin in the NCL control condition. In each panel, the tone in the first syllable is held constant (T1-T4) and the tone in the second syllable varies.*

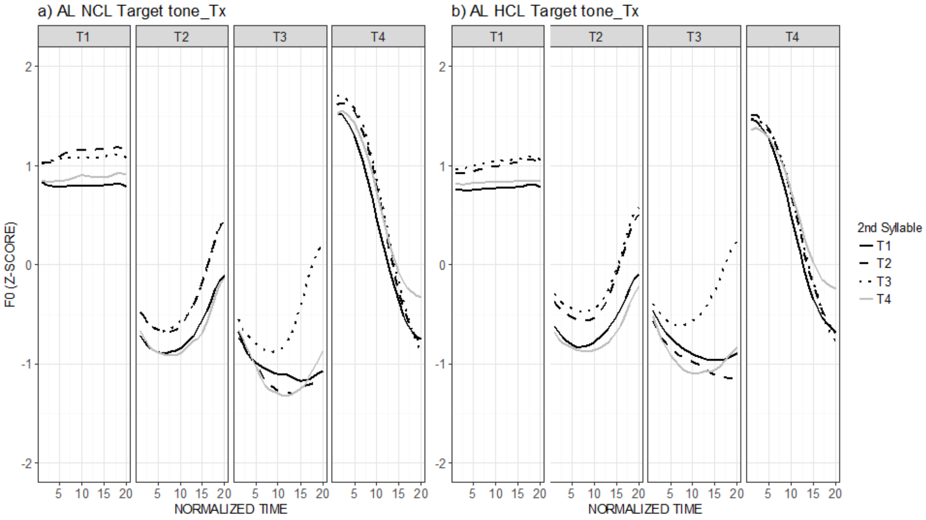

Figure 2.13. *F0 contours (over the vowel part) of the four target tones when followed by different tones produced by advanced learners of Mandarin. Normalized f0 contours averaged across participants.*

For advanced learners, the anticipatory effect for T1 in the NCL condition was strong (Figure 2.13*a*, Table 2.11*a*). When followed by tones with low onsets, the overall f0 of initial T1 was significantly higher than when it followed by tones with high onsets. Strong anticipatory effect was also found on initial T2. Contours of target T2 clustered into two groups according to the onsets of the second tone. T3 exhibited significantly different f0 contours when followed by T3 vs. T1 and T4 due to the phonological sandhi rule, which was similar to the performance of native Mandarin speakers. When followed by tones with contrastive onsets, the contours of T3 showed more variation than native Mandarin speakers, but the differences were not significant. The dissimilatory effect on initial T4 was also strong. The initial portion in the contours of the target tone was significantly raised by the low onsets of the following tones.

In the HCL condition, the strong effect on T1 and T2 maintained. Similar pattern was also found for T3 in the HCL condition. The effect also remained robust for T4, except for the insignificance when followed by T1 vs. T2.

Table 2.11. *Pairwise comparison of results for adapted contours on the vowel part of the first tone due to following tones with high onsets (T1 and T4) vs. low onsets (T2 and T3) by advanced learners.*

***a.* T1 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*P*	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*
intercept	0.33	6.25	<.001	0.28	5.26	<.001	−0.25	−4.74	<.001	−0.20	−3.75	<.01
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.24	4.56	<.001	0.28	5.17	<.001	−0.18	−3.38	<.05	−0.21	−3.97	<.01
slope		n.s.			n.s.			n.s.			n.s.	
quadratic		n.s.			n.s.			n.s.			n.s.	

***b.* T2 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.36	5.21	<.001	0.35	5.10	<.001	−0.38	−5.85	<.001	−0.37	−5.72	<.001
slope	0.65	3.70	<.01	0.74	4.17	<.001	−0.68	−4.17	<.001	−0.77	−4.68	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	*P*	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept	0.35	5.09	<.001	0.44	6.26	<.001	−0.41	−6.31	<.001	−0.49	−7.55	<.001
slope	0.69	3.86	<.01	0.66	3.64	<.01	−0.81	−4.83	<.001	−0.78	−4.57	<.001
quadratic	0.24	3.05	<.05		n.s.		−0.27	−3.68	<.01	−0.24	−3.06	<.05

Table 2.11. *(Continued)*

***c.* T3 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	P	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.48	4.78	<.001		n.s.		−0.53	−5.26	<.001
slope		n.s.		1.65	6.31	<.001		n.s.		−1.47	−5.58	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	P	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.47	4.67	<.001		n.s.		−0.57	−5.70	<.001
slope		n.s.		1.47	5.59	<.001		n.s.		−1.20	−4.49	<.001
quadratic		n.s.			n.s.			n.s.			n.s.	

***d.* T4 in the first syllable**

NCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	P	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.		0.23	3.65	<.01		n.s.			n.s.	
slope		n.s.			n.s.		0.80	4.25	<.001	0.90	4.79	<.001
quadratic	−0.43	−4.62	<.001	−0.52	−5.53	<.001	0.36	3.88	<.01	0.45	4.79	<.001
HCL	before T1 vs. before T2			before T1 vs. before T3			before T2 vs. before T4			before T3 vs. before T4		
	Est.	z	P	Est.	z	*p*	Est.	z	*p*	Est.	z	*p*
intercept		n.s.			n.s.			n.s.			n.s.	
slope		n.s.			n.s.		0.89	4.70	<.001	0.88	4.63	<.001
quadratic		n.s.		−0.43	−4.48	<.001		n.s.			n.s.	

Table 2.12. *Mean duration (in ms) of first vowel in* /mama/ *sequence with four tones produced by advanced learners of Mandarin in the NCL and the HCL conditions.*

Cognitive load	Tone			
	T1	T2	T3	T4
NCL	215	218	207	190
HCL	195	198	192	179

In both cognitive-load conditions, T2 was the longest tone, significantly longer than T3 and T4 (all *p* values < 0.01). T3 was realized with an intermediate duration, while T4 was the shortest. In the HCL condition, the duration of all four tones became significantly shorter (all *p* values < 0.01).

2.3.2.4 Summary of anticipatory effect

Compared to the carryover effect, the dissimilatory anticipatory effect was found in T1, T2 and T4 with smaller magnitude for native Mandarin speakers. This effect decreased under the influence of cognitive load. Compared to native speakers, the magnitude of dissimilatory effect was smaller for beginning learners, which decreased in the HCL condition (especially for T4). The advanced learners exhibited a dissimilatory effect similar to native Mandarin speakers in the NCL condition which remained robust in the HCL condition.

2.4 Discussion

In this study, we reexamined the directionality, the nature and the magnitude of tonal coarticulation for native Mandarin speakers using disyllabic non-words following the design in Xu (1997), and more important, we tapped into the underlying mechanism and source of the tonal coarticulatory effects by introducing the effect of cognitive load. Since L2 acquisition has been remained less investigated, we further tested the beginning and advanced Dutch learners of Mandarin to reveal the developmental trajectory and mechanisms of tonal coarticulation that underlies the ultimate attainment of tonal acquisition in by non-tonal second language learners.

2.4.1 Tonal coarticulation for native Mandarin speakers

Our results showed that, for native speakers, tonal coarticulation was bidirectional, with both carryover and anticipatory effects. The carryover effect exerted by the offset of the initial tones exhibited an assimilatory nature, which replicated the findings in Xu (1997). In the current study we examined the variability in target tones exerted by all pairs of initial tones with contrastive offsets. The assimilatory effect was found for all four tones when preceded by all pairs of tones with contrastive offsets. Although, the carryover effect decreased over time, the influence can still be seen at least two third into the vowel. The anticipatory effect exerted by the following tones on the tones in the initial position showed a dissimilatory nature and had a smaller magnitude compared to the carryover effect. These findings were also in line with Xu (1997). For T1, T2 and T4, the dissimilatory effect was found when followed by most pairs of tones with contrastive onsets. It should be noted that, different form Xu's (1997) finding, the raising effect of the following low onsets were not constrained to the maximum f0. The whole contour of T1 and T2 were raised by following tones with low onsets (T2 and T3) in the current experiment. There was a lack of anticipatory effect for initial T3, except for the phonological change when it was followed by another T3.

We further investigated the effect of cognitive load on the tonal coarticulatory effects for native speakers of Mandarin. The carryover effect was robust and was not affected by high cognitive load. This is in line with previous findings that carryover effect does not involve advance planning (Whalen, 1990), thereby supporting the view that it is mainly caused by physiological constraints (Xu, 2011). The anticipatory effect decreased with high cognitive load, which was in contrast to the findings of Franich (2015). Significant dissimilatory anticipatory effect was found for all tones except T3 in the NCL condition, while in the HCL condition, this effect on T1 and T4 became weaker. This result lends support to the view that anticipatory coarticulation is planned, and diminishes under the influence of concurrent mnemonic processing. This finding can be potentially accounted for by the model proposed by Tilsen (2009, 2013), which argued that an inhibitory speech planning mechanism was used for contemporaneously planned articulatory targets to maintain and maximize the contrasts of different phonemes. In the NCL condition of the present study, the inhibitory mechanism functioned well and led to a clear dissimilatory anticipatory effect, maximizing the contrast of the adjacent tones. In the HCL condition, however, such inhibitory mechanism was constrained and resulted in a decreased dissimilatory anticipatory effect.

The overlap of different articulatory gestures, which happens in vowel and consonant coarticulation, is less feasible in tonal coarticulation. However, the involvement of cognitive planning in anticipatory tonal coarticulation found in the current study is compatible with the planned anticipatory effect found in vowel and consonant coarticulation (Katsika, Whalen, Tiede, & King, 2015; Whalen, 1990) and may reflect a characteristic of coarticulation in general.

2.4.2 Tonal coarticulation for Dutch learners of Mandarin

For both groups of learners, tonal coarticulation was also bi-directional. For beginning learners, assimilatory carryover effect was found for T1, T2 and T4 in the second syllable in the NCL condition. The magnitude of this effect was smaller compared to that of native Mandarin speakers. The influence of cognitive load was weak. The carryover effect for advanced learners was also assimilatory in nature and substantial for T1, T2 and T4, with magnitude similar to that of native speakers. Comparing the patterns of carryover effect of the beginning vs. advanced learners of Mandarin, we observed a clear developmental path. These interesting and new findings suggest that although the carryover effect does not include advance planning and is mainly bases on physiological constraints in articulation, its acquisition is still a gradual learning process. Fine-tuned motor skills are required for native-like production of tonal sequences.

Different from the carryover effect, the anticipatory coarticulation was strong for Dutch learners of Mandarin. For beginning learners, the dissimilatory anticipatory effect was found on T1, T2 and T4 in the first syllable. This effect was robust for T1 and T2 with high cognitive load. Our advanced learners also showed anticipatory effect in the NCL condition like that of native Mandarin speakers. For T1, T2 and T4, the dissimilatory effect was found when followed by all pairs of tones with contrastive onsets. This strong effect could be seen for the whole contour for T1 and T2, and for the maximum f0 value for T4. It did not show obvious decrease for all tones under the influence of high cognitive load for advanced learners, showing a more robust pattern than native Mandarin speakers. That is, as suggested by Tilsen's model, the inhibitory mechanism was acquired by beginning and advanced learners as an effective way to maintain the contrast and ensure the perceptibility of different tonal categories in sequence.

Brengelmann, Cangemi and Grice (2015) reported results which revealed greater variability in L2 production in the final portion of the initial tone in disyllabic sequences than native speakers. More specifically, German learners of Mandarin were more likely to produce anticipatory coarticulation. However, only the final part of the initial tone was examined in that study and therefore it is not clear whether there is an anticipatory effect on the whole contour or maximum f0 value of the initial tone for German learners of Mandarin. More studies examining learners with different L1s are needed to shed light on the general pattern and language-specific characteristics among learners of non-tonal languages.

2.4.3 Conclusion

The results of the current study show that for native Mandarin speakers, the carryover effect in tonal coarticulation was assimilatory in nature and did not involve speech planning. The anticipatory coarticulation, on the other hand, was dissimilatory in nature and was planned. The carryover effect could be acquired gradually by L2 learners, suggested by a developmental path found in beginning and advanced Dutch learners of Mandarin. The anticipatory effect was strong for both beginning and advanced learners. The advanced learners showed a more robust anticipatory effect compared to native Mandarin speakers, since for them, this effect was not reduced in high cognitive load condition. The anticipatory effect was adopted by L2 learners as an effective way in maintaining and maximizing contrast of tonal categories.

Chapter three

Developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin

3.1 Introduction[4]

It is well-known by now that the function of vocal pitch (acoustically cued mainly by fundamental frequency or f0) varies across languages. For non-tone language speakers, pitch information is mainly used at the post-lexical level to signal sentential information such as pragmatic nuances and sentence modes, as well as to mark the grouping of words into larger units such as syntactic constituents and higher-level discourse units (see e.g., Cole, 2015; Cutler, Dahan, & Van Donselaar, 1997; Shattuck-Hufnagel & Turk, 1996 for detailed review). Tone language speakers, on the other hand, primarily employ pitch information to convey lexical meaning, while at the same time, in a much more complex and sometimes subtle way, to signal various post-lexical information comparable to that in non-tone languages (e.g., Cole, 2015; Chen, 2000; Chen, 2012; Chen & Gussenhoven, 2008; Gussenhoven, 2004; Xu, 2001; Yip, 2002).

Speakers of tone and non-tone languages have been reported to tune their auditory systems to the same acoustic stimuli differentially due to the different prosodic systems of their native languages. Behavioral studies have suggested that there are differences in the way tone and non-tone language speakers identify non-speech pitch contours (Bent, Bradlow, & Wright, 2006), and how they process both level and contour tones (Gandour, 1983). There are also neurophysiological studies showing differences in the hemispheric specialization of pitch processing in the brain: tonal contrasts are processed mainly in the left hemisphere by tone language speakers, but in the right hemisphere or bilaterally by non-tone language speakers (Gandour, Wong, Hsieh, Weinzapfel, Van Lancker, & Hutchins, 2000; Krishnan, Xu, Gandour, & Cariani, 2005; Wang, Sereno, Jongman, & Hirsch, 2003; Xu, Gandour, Talavage, Wong, Dzemidzic, Tong, & Lowe, 2006; Zatorre & Gandour, 2008). Braun and Johnson (2011) showed that Mandarin and Dutch listeners differentially attend to the same pitch movements with different locations on a segmental string. Mandarin speakers were attentive to the

---

[4] This chapter appeared as Zou, T., Chen, Y., & Caspers, J. (2016). The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones. *Bilingualism: Language and Cognition*, 1-13. DOI 10.1017/S1366728916000791

rising and falling pitch contours on both the initial and the final syllables in a disyllabic non-word. These contours signal two different lexical tones in Mandarin (i.e., the lexical Rising and Falling tone). Dutch speakers, in contrast, were much more sensitive to pitch movements in the final position than to pitch movements in the initial position, probably because a final pitch movement can reveal post-lexical meaning, such as finality vs. non-finality (e.g. Van Heuven & Kirsner, 2004; Van Heuven, 2017).

The issue that we address here is whether native speakers of a non-tone language such as Dutch can learn to effectively process the non-native Mandarin lexical tonal contrasts at the phonological level. A related issue is whether, during the course of their acquiring a tonal system, Dutch learners of Mandarin can learn to redistribute their attention to segmental and tonal information like native speakers and whether they adapt their processing of pitch movements in a lexically contrastive way similar to that of native tonal speakers. For both issues, it would be relevant to understand the developmental path by investigating learners with different levels of proficiency in Mandarin. The goal of this study was therefore to address these issues by examining how beginners and advanced Dutch learners of Mandarin process tonal information in an ABX task, compared to both native Dutch speakers (without any experience of learning a tone language) and native Mandarin speakers.

3.1.1 Phonetic and phonological processing of non-native contrasts

When learning a foreign language, adults are often confronted with difficulties in both low-level auditory processing and phonological processing of non-native segmental and suprasegmental contrasts (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Takagi & Mann, 1995). Different theoretical models have been proposed to account for such difficulties. The Speech Learning Model (SLM) holds that second language (L2) learners perceive non-native sounds by referring to the phonetic categories of their L1 sound system (Flege, 1995). The mechanisms and processes involved in L1 acquisition, such as category formation, remain intact throughout one's life and can be used in L2 learning, although this ability tends to decrease as learners' age of learning increases. PAM-L2 (Best & Tyler, 2007), based on the Perceptual Assimilation Model (PAM) (Best, 1994), assumes that a listener's perceptual system will automatically assimilate non-native speech sounds to the closest categories in their native language, and the discrimination of non-native contrasts can be predicted from the way in which they are assimilated into the native system, ranging from excellent discrimination if each sound of a non-native contrast can be assimilated to a different category in the native language, to relatively poor discrimination when both sounds are mapped onto a single native category (Best, 1994; Best & Tyler, 2007).

Both SLM and PAM-L2 suggest that a novel L2 speech contrast can potentially be learned by L2 learners. The L2 phonological acquisition model proposed by Brown (2000), on the other hand, holds that the phonological structure of the first language will hinder the proper acquisition of L2 features throughout adulthood, due to the direct mapping of these features onto the existing L1 categories. This consequently prevents learners from fine-tuning their perception of L2 contrasts even with prolonged exposure to the L2.

There have been an increasing number of studies examining vowel and consonant perception, lending support to PAM and PAM-L2 (Best, McRoberts, & Sithole, 1988; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Hayes-Harb & Masuda, 2008; Heeren & Schouten, 2008, 2010). Less effort, however, has been devoted to suprasegmental perception. Compared to segments, suprasegmental cues are more global and are always superimposed on a succession of segments. Furthermore, their functions vary across languages ranging from signaling lexical to post-lexical information. Recently, PAM-S has expanded the original PAM to include non-native suprasegmental perception. So and Best (2010, 2011, 2014) conducted a series of cross-linguistic studies which demonstrated that Australian English and French speakers could categorize Mandarin tones according to the given intonation categories ("statement", "question", "flat-pitch" and "exclamation"), although their discrimination of the lexical tones could not be fully accounted for by their assimilation patterns. Note that two earlier studies (Hallé, Chang, & Best, 2004, for French listeners and Wang, Spence, Jongman, & Sereno, 1999, for English listeners) have also shown that non-tone listeners could discriminate Mandarin tones adequately, lending additional support to the model. PAM-S is further supported by a study on another suprasegmental contrast, i.e., lexical stress (Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997), which showed that French participants can distinguish novel lexical stress contrasts in Spanish even though French is not a stress language.

The above-mentioned studies concerning suprasegmental perception only tested low-level auditory processing, using cognitively less demanding phonetic discrimination and identification tasks. As for phonological processing of non-native suprasegmental contrasts, Dupoux et al. (2008) tested the short-term storage and retrieval of lexical stress by French learners of Spanish, using a cognitively demanding sequence recall task. They found a persistent "stress deafness" at the phonological level, which is not predicted by the PAM-L2 model. This difficulty, however, can be better accounted for by Brown's model (2000), which states that non-native contrasts are perceived in terms of the features established in the learners' L1, and therefore the phonological processing of Spanish stress is predicted to be impossible for French listeners due to the absence of lexical stress in French.

Thus far, no study has tapped into the level of phonological processing of lexical tones by L2 non-tone learners. To fill in this gap, the present study set out to investigate the discrimination of tonal contrasts by Dutch learners of Mandarin at an abstract phonological level.

3.1.2 Attention redistribution and integration of perceptual dimensions in the acquisition of new categories

While earlier models of L2 category acquisition have focused much on whether new L2 categories can be acquired, much less has been investigated on how they are acquired. Francis and Nusbaum (2002) have provided the insight that the establishment of new L2 phonetic categories requires the redistribution of attention to different perceptual dimensions. In their study, English listeners were trained to perceive the three-member consonant contrasts in Korean (known as fortis, lenis, and aspirated; see Cho, Jun, & Ladefoged (2002) for more details). These contrasts employ acoustic cues (such as fundamental frequency and formant structures) in a different way from the English consonant contrasts. Their results showed that English native listeners learned to redistribute their attention to the acoustic cues after training and were then able to approximate the behavior of native Korean listeners in the post-test.

Their findings are consistent with the predictions of the generalized context model (GCM, Nosofsky, 1986) and Goldstone's model (1993, 1994) on categorical learning. These models emphasize a multidimensional structure of the categorization space, and suggest that perceptual learning of new categories involves developing perceptual acuity to new acoustic dimensions. In light of these studies, our goal is to examine the role of attention redistribution between the segmental and the suprasegmental dimension during the acquisition of lexical tones by non-tone speakers.

Prior studies also suggest that the processing of segmental and tonal dimensions by native Mandarin speakers is more interdependent than those of speakers of non-tone languages such as English and Dutch (Lin & Francis, 2014; Repp & Lin, 1990; Tong, Francis, & Gandour, 2008). Mandarin speakers attend to both segment and tone and these two dimensions are integrated and processed simultaneously. The two dimensions may intrude into each other, making it difficult for native listeners to attend to one dimension only while ignoring the other (Garner, 1976, 2014; Goldstone, 1994). In intonation languages like English, however, listeners seem to pay more attention to the segmental dimension and the two dimensions are much less integrated and, consequently, listeners are able to tune their attention to only one dimension and to suppress interference from the other. Of interest here is whether and how native speakers of an intonational language such as Dutch manage to retune their attention to both the segmental and tonal information in processing Mandarin in the process of acquiring Mandarin.

Note that a more recent model of first and second language speech, the automatic selective perception (ASP) model, also emphasizes the role of attention and it further differentiates the phonological mode and phonetic mode of perception (Strange, 2011). The phonological mode is employed by native listeners, in which automatic selective perception routines are used in order to detect phonologically contrastive information for identifying word forms. The phonetic mode, on the other hand, was employed to detect fine-grained allophonic details, which requires more cognitive effort in processing.

The literature reviewed above leads to the hypothesis that at the beginning stage of learning Mandarin tones by a non-tone language speaker, the phonetic mode of perception is used when learners process tonal contrasts and they have to make more effort to attend to the tonal dimension for reliable word-form recognition. During this stage, tonal and segmental information are much less integrated in processing. For more advanced learners, they are expected to develop a much more automatic perceptual routine for tonal and segmental processing in word recognition, which is facilitated by their redistributed attention to the tonal and the segmental dimensions. The development of such a new selective perception routine leads to more automatic integration and simultaneous processing of the segmental and tonal information. Another goal of the present study is therefore to test these predictions by examining the developmental characteristics of learners of Mandarin in terms of their redistribution of attention to lexical tones and segments as well as the integration of these two kinds of information during their phonological processing of lexical tones.

3.1.3 The present study

The research questions that this study examines are summarized as the following:

(1) Can Dutch learners of Mandarin successfully discriminate Mandarin lexical tones within a phonological mode of processing?
(2) Are they able to redistribute their attention to segments and tones and to develop a more integral processing of these two dimensions?
(3) What is the developmental trajectory of the Dutch learners' phonological processing of non-native contrasts and their segment-tone integration during the period of their acquisition of Mandarin?

In order to address these questions, both beginning and advanced Dutch learners of Mandarin were examined in their processing of segments and lexical tones. Native Mandarin and Dutch listeners were recruited as the control groups.

Among the four Mandarin tones, a non-final tonal contrast (a rising tone followed by a neutral tone versus a falling tone followed by a neutral tone) was selected as the stimulus. Braun and Johnson (2011) tested native Mandarin and Dutch listeners in a speeded ABX task, in which the target non-word could be classified according to either segmental or tonal information. They demonstrated that Mandarin listeners were attentive to such a non-final pitch rise and pitch fall, while Dutch listeners were much less attentive to them. Therefore, this pair of tonal sequences provides us with a good test case to explore the development of tone perception by Dutch learners of Mandarin.

A cognitively demanding ABX task was employed, which is commonly recognized as a good method that can be used to tap into the phonological mode of processing (Dupoux, Peperkamp, & Sebastián-Gallés, 2001). Participants were asked to classify the target X according to standard A and B. A and B can be similar to X along the dimensions of both segment and lexical tone, thereby creating four possible conditions: only segments shared with X (forced-segment condition), only tone shared with X (forced-tone condition), either segment or tone shared with X (segment-or-tone condition), or both segment and tone shared with X (segment-and-tone condition) The trials of the four condition were mixed in random order and blocked into four sessions in the experiment.

To investigate the first question, what is crucial is the comparison of the forced-segment and forced-tone conditions. Correct classification of the target in these two conditions requires a proper representation and short-term retention of tonal or segmental categories. According to PAM-L2, there are two possible assimilation scenarios for the tonal pair used in our experiment. First, as both Mandarin lexical tones may fall within the L1 Dutch intonational phonetic space, but neither fits any single L1 phonological category (i.e., the Both-Uncategorized scenario in PAM-L2), the discrimination of this tonal contrast can be expected to be good for naïve listeners, and relatively easy to learn by L2 learners.

The alternative possibility within PAM-L2 is that the tonal contrast in our study fits the Uncategorized-Categorized scenario. That is, the sequence of Tone 4 followed by a neutral tone may be mapped onto the "pointed hat" pitch accent (H*L), followed by a low boundary tone (H*L L%). This is the most neutral form of pitch accent in Dutch (Gussenhoven, 2005) and a contour used naturally for producing a one-word phrase in statements (Gussenhoven, Rietveld, Kerkhoff, & Terken, 2003).

The sequence of Tone 2 and a neutral tone, in contrast, is less likely to be mapped onto Dutch intonation category. It is not a question because the pitch goes down on the final syllable but not down to baseline, so no L%. It suggests paralinguistic uncertainty or hesitation. It is also reminiscent of the Limburgian way of ending a statement. If this is the case, the discrimination by native Dutch speakers is expected to be good, and we would further predict a better classification performance of the target word X with Tone 4 than the target word X with Tone 2 in the forced-tone condition by native Dutch listeners and Dutch learners of Mandarin.

To investigate the redistribution of attention between the segmental and tonal dimensions, we compared the segment-and-tone condition with the segment-or-tone condition, which measures the amount of attention implicitly attached to each dimension. We expect that native Mandarin listeners will be attentive to both dimensions. The classification by native Dutch listeners, however, is expected to be uniformly made along the segmental dimension only.

To tap further into the issue of integrality between segmental and tonal processing, we will examine the reaction time (RT) that listeners from each group need to perform the ABX task, as an index of the ease in separating the two dimensions when they make the judgments.

3.2 Methods

3.2.1 Participants

Fifteen Dutch control participants, 15 Mandarin control participants and 30 Dutch learners of Mandarin participated in the experiment. The native Dutch control group consisted of 4 males and 11 females (mean age = 20.6, SD = 1.3). The native Mandarin control group had 7 males and 8 females (mean age = 25.8, SD = 1.3). All were students at Leiden University from the Northern part of China and could speak Standard Mandarin. None of the native Mandarin control listeners had lived in a Dutch speaking environment for more than three years. All the Dutch learners of Mandarin were students of the Chinese Studies program at Leiden University. The beginner group consisted of 7 males and 8 females (mean age = 22.0, SD = 2.7). Their Mandarin learning and speaking experience varied from 8 to 20 months, and they had never lived in China. The other 15 participants (6 males and 9 females; mean age = 24.6, SD = 2.9) were advanced Mandarin learners, who had Mandarin learning experience from 3 to 14 years and who had spent at least one year living in China.

3.2.2 Stimuli

Nine pairs of CVCV non-words were selected with Mandarin Tone 2 (a pitch rise) or Tone 4 (a pitch fall) on the initial syllable, similar to the stimuli in Braun and Johnson (2011). The final syllable was always produced with a neutral tone. Figure 3.1 illustrates the pitch contours of these tonal combinations (i.e., Rising + Neutral tone in Fig. 3.1a and Falling + Neutral tone in Fig. 3.1b). The vowel set consisted of [a], [i], [u] and [o]. In the consonant set, there were three voiceless pairs of stops (labial: [p]-[p^h]; alveolar: [t]-[t^h]; velar: [k]-[k^h]), two voiceless fricatives (labial: [f], alveolar: [s]), and two nasals (bilabial: [m], alveolar: [n]). In each non-word pair, the vowels were constant, while the consonants in each syllable only differed in place of articulation (e.g., *kasu* vs. *tafu*). The full set of stimuli used in this study is provided in Appendix A3.

Multiple speakers were asked to produce the stimuli to increase phonetic variability and memory load, and thereby to further ensure that participants had to classify the target word based on a phonological level of representation.

The stimuli were recorded by three Beijing Mandarin speakers (two females and one male). Each item was recorded 12 times with a Sennheiser MKH416T microphone at the Leiden University Phonetics Lab (44.1 kHz, 16 bit). The speakers were asked to read aloud the disyllabic non-words presented in pinyin (a system for transliterating Mandarin Chinese in Roman letters) on a computer screen. The non-words were presented one by one, and the pace of reading was controlled by the experimenter. According to the pinyin system, the first syllable was presented with a tone label; the second syllable was presented without a tone label, which indicated it should be produced with a neutral tone, again following the pinyin marking system.

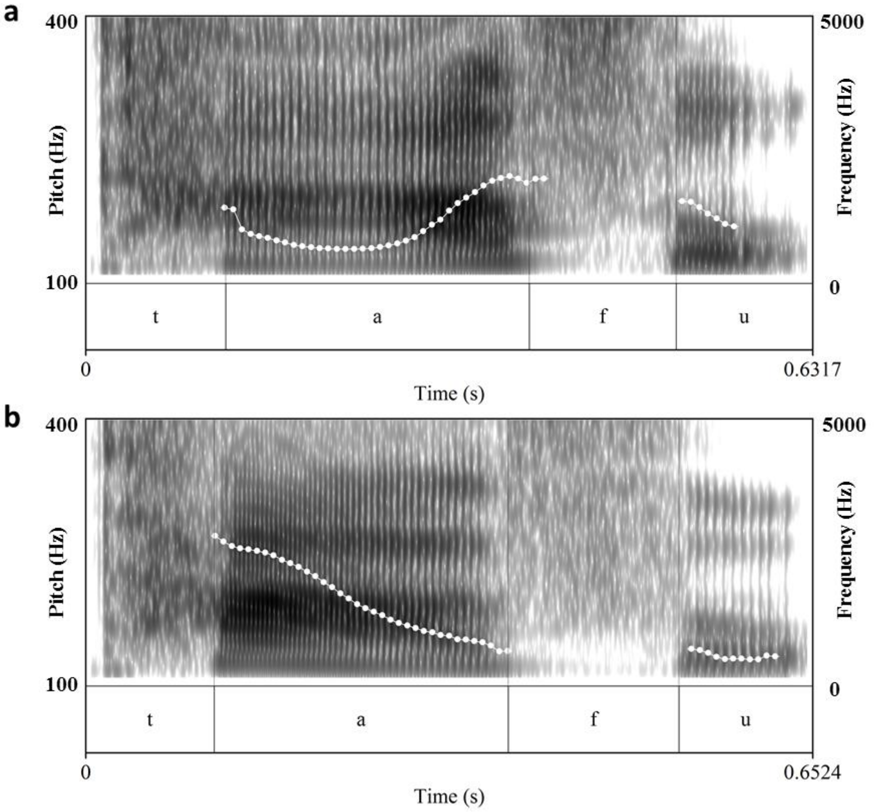

Figure 3.1. *Examples of the tonal sequences produced by a male speaker. Fig. 1a shows the pitch contour and the spectrogram of Tone 2 followed by a neutral tone. Fig. 1b shows the pitch contour and the spectrogram of Tone 4 followed by a neutral tone.*

For the male speaker, the mean f0-excursion of Tone 4 (pitch fall) in the first syllable was −126.6 Hz (SD = 25.2 Hz), larger than that of Tone 2 (M = 56.7 Hz, SD = 15.8 Hz). A similar pattern was also found for the two female speakers. The mean f0-excursion of Tone 4 (first female speaker: M = −116.6 Hz, SD = 22.8 Hz; second female speaker: M = −115.7 Hz, SD = 22.7 Hz) was larger than that of Tone 2 (first female speaker: M = 46.0 Hz, SD = 17.9 Hz; second female speaker: M = 47.3 Hz, SD = 15.2 Hz). The larger pitch excursions for the falls than for the rises in Mandarin tones have also been found in previous studies (e.g., Xu, 1994; Bent et al., 2006). They also correspond with the impressionistic tone transcriptions suggested by Chao (1930): 51 for Tone 4 against 35 for Tone 2. For all three speakers, the mean f0 of the second syllable neutral tone was lower when following Tone 4 (male speaker: M = 144.0 Hz, SD = 53.7 Hz; first female speaker: M = 195.7 Hz, SD = 53.7 Hz; second female speaker: M = 176.2 Hz, SD = 18.2 Hz) than following Tone 2 (male speaker: M = 186.2 Hz, SD = 40.5 Hz; first female speaker: M = 261.2 Hz, SD = 18.5 Hz; second female speaker: M = 225.4 Hz, SD = 10.0 Hz). Such acoustic features of the neutral tones have also been found in other studies (e.g., Chen & Xu, 2006).

Four types of ABX trials were included (see Table 3.1 for illustration). In the segment-and-tone condition, target word X matched either A or B along both the segmental and tonal dimensions. In the forced-segment and the forced-tone conditions, participants were forced to classify the target word X along, respectively, the segmental or tonal dimension. There is always a mismatch in the other dimension so that consistency along both the segmental and tonal dimensions was not available as a cue for classification. Therefore, correct classification of the target in these two conditions required a proper representation and short-term retention of tonal or segmental categories. In the segment-or-tone condition, target word X was matched along the segmental or the tonal dimension, which allowed participants to choose freely along either dimension. This condition thus measured the amount of attention implicitly attached to each dimension.

Table 3.1. *Sample stimuli for standard A, standard B and target X in four conditions.*

Condition	A	B	X
Forced-segment	gu2ta ka4su	du2ka ta4fu	gu4ta ka2su
Forced-tone	gu2ta ka4su	gu4ta ka2su	du2ka ta4fu
Segment-and-tone	gu2ta ka4su	du4ka ta2fu	gu2ta ka4su
Segment-or-tone	gu2ta ka4su	du4ka ta2fu	du2ka ta4fu

The main experiment consisted of 288 ABX trials with 72 trials for each condition. Take the forced-segment condition as an example. Within this condition, classification can only be made along the segmental dimension. There are four A-B combinations for each non-word pair: e.g., *ka2su-ta2fu*, *ka4su-ta4fu*, *ta2fu-ka2su*, *ta4fu-ka4su*. The target X always has the same segments as A or B, and this creates 8 items. The design of this condition is 4×2×9: four A-B combinations × congruency of A or B × nine non-word pairs. The items for the other three conditions were constructed in the same way.

The three stimuli in each trial were always produced by three different speakers. The order of these three voices was counterbalanced between the trials so participants could not predict the order of the voices in the coming trial. The 288 trials were blocked into four sessions and presented in random order. Trials of all four conditions were mixed in every session so the participants could not predict which dimension they had to focus on in the coming trial. At the start of the experiment, five familiarization trials (all segment-and-tone trials) were provided.

3.2.3 Procedure

Each participant was seated in front of a computer screen. The instructions were given in English, so all participants could understand. This also helped to avoid influence of their native languages. They were asked to listen to a group of three disyllables (ABX) and to decide whether the third word (X) was more similar to the first (A) or the second (B) by pressing "1" or "2" on the keyboard. Within each trial there was a 600-ms pause between A and B. The critical word (X) followed after a 900 ms pause (cf. Braun & Johnson, 2011). The interval between two consecutive trials was 1000 ms. The experiment was controlled using E-prime. The response buttons and reaction times (RTs) of the participants were recorded. The RTs were recorded from the beginning of the target X, and if the participant failed to respond within 7 seconds, then the next trial would proceed automatically.

3.2.4 Statistical analyses

Analysis of the response type (classification along the segmental or the tonal dimension) was performed with a mixed effects logistic regression model using R with lme4 package (Bates, Maechler, Bolker & Walker, 2015). The fixed factors of the model included Participant Group (i.e., Native Mandarin listeners, Beginning learners, Advanced learners, and Native Dutch listeners), Trial Type (i.e., forced-segment, forced-tone, segment-and-tone, and segment-or-tone), and their interactions. By-Participant intercept (60 levels) and By-Item intercept (9 levels) were included as random effects. In addition, we also included the factor Response Button (1 or 2) as a control variable. The initial model also included The Tone of The Target Word (Tone 2 vs. Tone 4) as a fixed effect, but it was removed since it was not significant and did not appear in significant interactions.

For reaction time, the raw RT data was natural-logarithmically transformed to achieve better normalcy. The analysis of RT was also performed with a linear mixed effect model using R with lme4 package (Bates et al., 2015), initially with a full model. Model comparisons showed a significant effect of the following fixed factors: Participant Group, Trial Type, Response Button and their interactions. With regard to random effects, both By-Participant intercept (60 levels) and By-Item intercept (9 levels) were included in the final model.

For both models of response type and RT, trials with residuals beyond 3 standard deviations of the mean were removed as outliers. R^2 values for both models were calculated with the MuMIn package (Bartoń, 2015) in R according to the method suggested by Nakagawa and Schielzeth (2013) with the marginal R^2 measuring the variance explained by the fixed effects and the conditional R^2 representing the variance explained by both fixed and random factors. Post-hoc comparisons of differences between different levels within each factor were conducted using the Multcomp package in R with Single-step adjustment (Hothorn, Bretz, & Westgall, 2008).

3.3 Results

Fifty trials (0.3%) were excluded because participants did not respond within 7 seconds after the target word was presented. So in total, 17,230 trials (out of 17,280) were analyzed.

Table 3.2. *Summary of mixed effects models for response type and reaction time (RT).*

Fixed effects	Response type			RT (log)		
	df	χ^2	p	df	χ^2	p
Group	3	0.16	0.98	3	6.65	0.08
Trial Type	3	940.99	< .001	3	1307.10	< .001
Response Button	1	3.81	0.05	1	0.18	0.67
Group : Trial type	9	703.08	< .001	9	188.59	< .001
Group : Response Button	3	7.93	0.05	3	2.90	0.41
Trial Type : Response Button	3	10.46	0.02	3	7.21	0.06
Group : Trial Type : Response Button	9	15.46	0.08	9	6.77	0.66
Random effects						
1|Subject	1	473.34	< .001	1	4366.10	< .001
1|Item	1	88.23	< .001	1	58.06	< .001
Marginal R^2		0.67			0.09	
Conditional R^2		0.71			0.33	

The statistical results for the models of response types and reaction times are presented in Table 3.2. The χ^2 and corresponding p values for the fixed and random effects were obtained from likelihood ratio tests.

For response type, there was a significant main effect of Trial Type [$\chi^2(3) = 940.99$, $p < 0.001$] as well as a significant interaction between the Participant Group and the Trial Type [$\chi^2(9) = 703.08$, $p < 0.001$]. There was also a significant main effect for the Response Button [$\chi^2(1) = 3.81$, $p = 0.05$). The interaction between the Trial Type and the Response Button was also significant [$\chi^2(3) = 10.46$, $p = 0.02$].

For RT, there was a significant main effect of Trial Type [$\chi^2(3) = 1307.10$, $p < 0.001$]. The interaction between Participant Group and Trial Type was also significant [$\chi^2(9) = 188.59$, $p < 0.001$). In the following, we will present a more detailed analysis of the interaction of Participant Group and Trial Type, according to the research questions we have posed. The classification types and reaction times in the forced-segment and forced-tone conditions will be discussed first (§ 3.3.1) since they reflect the performance in segmental and tonal processing. After that response type and RTs in the segment-and-tone and the segment-or-tone conditions will be discussed (§ 3.3.2), which reveal the distribution of attention between the segmental and the tonal dimensions. Finally, the comparison of RTs among conditions within each participant group will be presented, which shows the degree of perceptual integration of the segmental and the tonal dimensions (§ 3.3.3).

3.3.1. Phonological processing of tonal contrasts

The response types and RTs of the four participant groups are presented in Figure 3.2. The black line represents the mean percentage of correct classification in the forced-segment, forced-tone and segment-and-tone conditions. Note that since there is not a "correct" classification in the segment-or-tone condition, the black line in that condition represents the mean percentage of the segment-based classification.

In the forced-segment condition, the overall accuracy (segment-based classification) was high across all four participant groups (above 86.0%). The two learner groups scored a bit lower than the two native groups, but these differences were not statistically significant. For RT in this condition, native Dutch listeners and beginning learners responded significantly faster than advanced learners (AL vs. BL: $z = 2.85, p = 0.023$; AL vs. ND: $z = 3.00, p = 0.014$). This suggests that listeners with less Mandarin experience can ignore the tonal information more easily and focus their attention better on the segmental dimension.

In the forced-tone condition, the accuracy (classification along the tonal dimension) of native Mandarin listeners (NM) (87.2%) and advanced learners (AL) (82.0%) was significantly higher than that of the beginning Dutch learners (BL) (64.9%) and native Dutch listeners (ND) (58.5%) (NM vs. BL: $z = 5.78, p < 0.001$; NM vs. ND: $z = 7.28, p < 0.001$; AL vs. BL: $z = -4.23, p < 0.001$; AL vs. ND: $z = -5.74, p < 0.001$). Within each subgroup (ND and BL; AL and NM), there was no significant difference, but there was a slight trend of native Mandarin listeners performing better than advanced learners and beginning learners performing better than native Dutch listeners (all p values > 0.05). Although the accuracy was low for native Dutch listeners, one-tailed t-tests showed that their performance was significantly above the level of chance (50.0%) (data aggregated by subjects: $t(14) = 5.62, p < 0.001$; data aggregated by items: $t(71) = 4.87, p < 0.001$). The effect of the response button was not significant for all groups. In this condition, RT was not significantly different across the four participant groups, but advanced learners generally responded more slowly than the other three groups (all p values > 0.05).

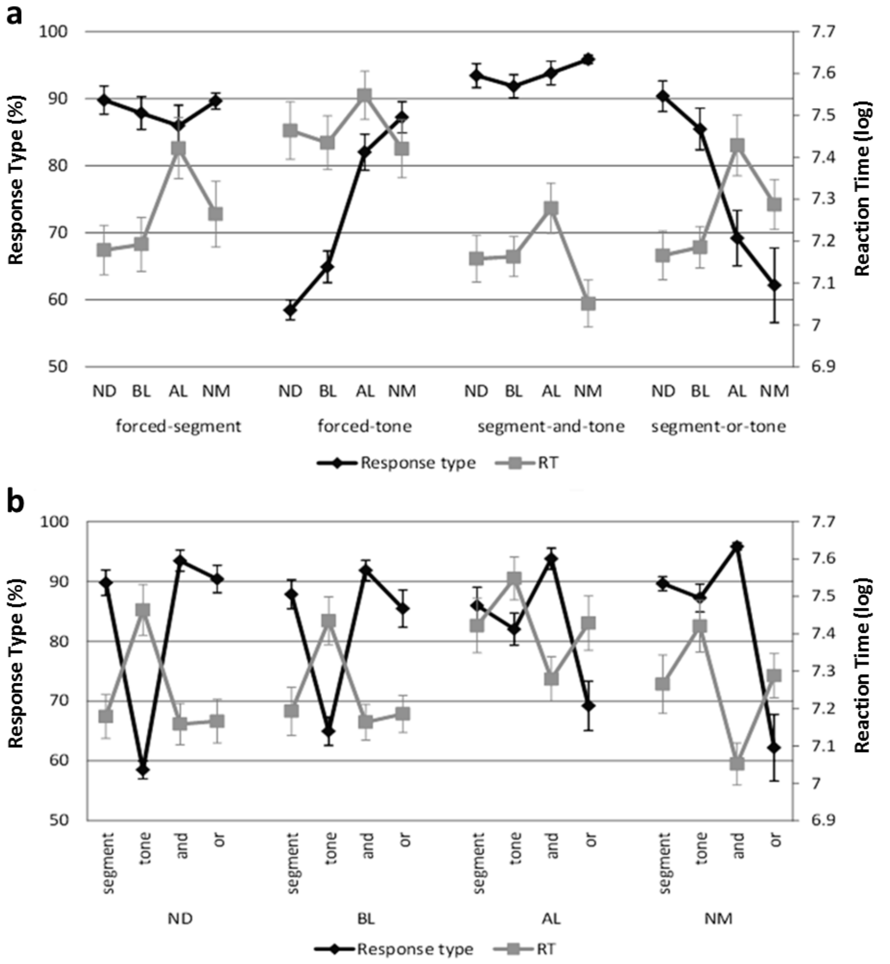

Figure 3.2. *Mean percentage of response types and RTs across participants with standard errors for four participant groups in four conditions. The black line shows percentage of segment-based, tone-based, segment-and-tone-based, and segment-based classification in forced-segment, forced-tone, segment-and-tone and segment-or-tone condition, respectively. The grey line shows natural-logarithmic RT. The four groups of participants are listed along the x-axis: native Dutch listeners without Mandarin experience (ND), beginning Dutch learners of Mandarin (BL), advanced Dutch learners of Mandarin (AL), and native Mandarin listeners (NM). The data was grouped in panel a by trial conditions and in panel b by participant groups.*

In order to further illustrate the similar processing patterns between the native Mandarin listeners and the advanced Dutch learners of Mandarin versus the similarity between the native Dutch listeners and the beginning learners of Mandarin, the difference scores of the segmental and the tonal classifications have been plotted in Figure

3.3. The difference score was defined as the percentage of correct classifications in the forced-segment condition minus the correct classifications in the forced-tone condition, following Dupoux et al. (2001).

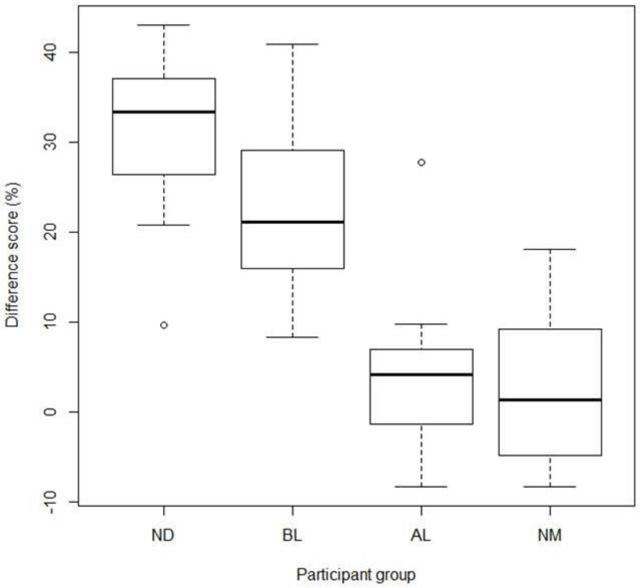

Figure 3.3. *Difference scores of the four participant groups. ND: native Dutch listeners without Mandarin experience; BL: beginning Dutch learners of Mandarin; AL: advanced Dutch learners of Mandarin, NM: native Mandarin listeners. The difference scores were calculated as the percentage of correct segmental classification in the forced-segment condition minus the percentage of correct tonal classification in the forced-tone condition. The median scores are indicated by the thick horizontal lines. The 25th and 75th percentiles correspond to bottom and top edges of the box. The whiskers extend 1.5 interquartile range from the boxes.*

The overlap between the groups refers to the percentage of participants whose difference scores were in the common area of the two groups. There is hardly any overlap between the native Dutch and Mandarin listener groups, which indicates that the task clearly reveals impairment in the native Dutch participants' tonal perception. The response type from this task therefore provides a robust criterion for differentiating between the two native listener groups. The large overlapping area between the native Dutch listeners and the beginning learners demonstrates their poor performance in tonal classification, while the similar distributions for advanced learners and native Mandarin listeners demonstrate their comparable performance in the tonal and segmental classifications.

3.3.2 Redistribution of attention to the segmental vs. the tonal dimension

In the segment-and-tone condition, the overall accuracy was very high across the four groups (over 91.0%), with the performance of native Mandarin listeners significantly better than that of the beginning learners (NM vs. BL: $z = 2.97$, $p = 0.016$) (Figure 3.2, panel a). There was an effect of the response button for native Mandarin listeners ($z = -3.57$, $p < 0.001$). For them, more errors were associated with the response button "2". In other words, the native Mandarin listeners mistakenly chose "2" more often when the answer should have been "1" than the other way around. Native Mandarin listeners responded very fast in this condition, hence the common bias toward the "B" standard in an ABX task (Macmillan & Creelman, 2004) became more obvious. A similar bias for the "2" response was also found for beginning learners ($z = -1.97$, $p = 0.048$). For RT in this condition, native Mandarin listeners responded faster than the other three groups, although only the RT difference between native Mandarin and advanced learners reached significance ($z = 2.79$, $p = 0.027$).

In the segment-or-tone condition, native Mandarin listeners (62.2%) and advanced learners (69.2%) classified the stimuli along the segmental dimension significantly less often than beginning learners (85.5%) and native Dutch listeners (90.4%) (NM vs. BL: $z = -6.00$, $p < 0.001$; NM vs. ND: $z = -7. 73$, $p < 0.001$; AL vs. BL: $z = 4.42$, $p < 0.001$; AL vs. ND: $z = 6.18$, $p < 0.001$). (Note that in Figure 3.2, the black line for this condition refers to the percentage of the segment-based classification.) Within each subgroup ({ND BL} versus {AL NM}), there was no significant difference, but there was a slight trend for native Mandarin listeners to be more attentive to the tonal dimension than advanced listeners, as well as for beginning learners to beg more attentive to the tonal dimension than native Dutch listeners (all p values < 0.1). For RT in this condition, advanced learners responded significantly slower than native Dutch listeners and beginning learners (AL vs. ND: $z = 3.23$, $p = 0.007$; AL vs. BL: $z = 3.00$, $p = 0.015$). Native Mandarin listeners were also slower than native Dutch listeners and beginning learners, but the differences were not statistically significant (all p values > 0.05).

3.3.3 Integrality of segmental and tonal information

Related to the redistribution of attention to tonal and segmental dimensions is the issue of integrality of segmental and tonal information in speech processing, which was further examined within each participant group by comparing the RTs of the forced-segment and forced-tone conditions against the RTs of the segment-and-tone condition (grey lines in Figure 3.2*b*).

Results showed that native Mandarin listeners responded significantly slower in both the forced-segment and the forced-tone conditions than in the segment-and-tone condition ($z = -12.64$, $p < 0.001$; $z = -21.86$, $p < 0.001$). This suggests that when it was required that participants direct their attention to either the segmental or the tonal dimension, native Mandarin listeners were slowed down by the mismatch in the other dimension. Furthermore, the RT in the forced-tone condition was longer than in the forced-segment condition ($z = -9.24$, $p < 0.001$), which indicates that the mutual integrality between these two dimensions is not symmetrical. The results showed that the segmental dimension interfered more with judgment in the tonal dimension than *vice versa*.

For native Dutch listeners, there was no significant difference between the RTs in the forced-segment and segment-and-tone dimensions. There was, however, a significant difference in the RTs between the segment-and-tone condition and the forced-tone condition ($z = -18.10$, $p < 0.001$). The longer RT in the forced-tone condition mainly resulted from the difficulty in phonological tonal processing (as evident from the accuracy of the responses). This suggests that the two dimensions were processed in a separate manner.

The pattern of the beginning learners was similar to that of the native Dutch listeners, with no significant difference in RTs between the forced-segment and segment-and-tone conditions. The significant difference in RTs between the forced-tone and segment-and-tone conditions ($z = -15.82$, $p < 0.001$) was also a result of difficulty in discriminating between tonal contrasts. Advanced learners have developed a stronger integration of the segmental and the tonal dimensions. Their responses in the forced-segment and forced-tone conditions were significantly slower than that in the segment-and-tone condition ($z = -8.61$, $p < 0.001$; $z = -16.08$, $p < 0.001$). The RTs in the forced-tone condition were significantly longer than those in the forced-segment condition ($z = -7.64$, $p < 0.001$), which indicates an asymmetry in the processing of these integral dimensions, similar to that of native Mandarin listeners.

3.4 Discussion and conclusion

This experiment was designed to investigate three research questions concerning the acquisition of new tonal categories, the redistribution of attention over segments and lexical tones, as well as the integration of the segmental and suprasegmental perceptual dimensions. In the following, we will discuss how the results of this experiment can shed light on the three research questions that we set out to investigate.

The first research question concerns the phonological discrimination of Mandarin tone categories, which was revealed in the forced-tone ABX condition. The performance of advanced learners was significantly better than that of both the native Dutch control group and beginning learners, and approximated that of the native Mandarin listeners. This suggests that although pitch movements are not used to convey lexical meaning in Dutch, Dutch learners can perceive tonal contrasts with proper practice. This counters the phonology-based model (Brown, 2000), which would predict a persistent impairment in tonal perception for Dutch learners of Mandarin, since tone contrasts are not phonemic in Dutch and should therefore not trigger acquisition.

SLM, on the other hand, does predict the learning of new tones. PAM-L2 further predicts that there are two possible assimilation scenarios for the tonal pair used in our experiment. First, the tonal contrast fits the Both-Uncategorized scenario, which predicts that naïve listeners would show a good discrimination performance and this contrast would be relatively easy to learn by L2 learners. The percentage of correct classifications by native Dutch listeners in the forced-tone condition was 58.5%, which is not good, but significantly above chance level. This shows that our native Dutch listeners were sensitive to the acoustic distinctions of this tonal pair to some extent, but they could not encode tonal information accurately in the ABX task with high memory load and phonetic variability, which requires a short-term retention and an abstract phonological level of representation. For the two groups of learners, correct tonal classifications increased with Mandarin learning experience, which is in line with the prediction of PAM-L2. Since Mandarin tones cannot be assimilated into an existing category by the learners' L1, new categories are expected to be established first at a phonetic level. As the L2 vocabulary expands, learners are also expected to become attuned to the phonological structure of L2, and the newly established tonal categories will be discriminated in a phonologically contrastive way.

The alternative possibility within PAM-L2 is that the tonal contrast in our study fits the Uncategorized-Categorized scenario. That is, the sequence of Tone 4 followed by a neutral tone may be mapped onto the "pointed hat" pitch accent (H*L), followed by a low boundary tone (H*L L%). The sequence of Tone 2 and a neutral tone, in contrast, is less likely to be mapped onto Dutch intonation category. In this case, we would expect an asymmetry between target non-words with T2 and T4 for naive listeners and Dutch learners of Mandarin. However, this tendency was not observed in the data The effect of Tone of The Target Word (Tone 2 vs. Tone 4) was not significant and it did not appear in significant interactions. This shows that, at least in an ABX task with high memory load, a similar intonation pattern in Dutch cannot help native Dutch listeners to fully discriminate this pair of tonal sequences.

The redistribution of attention between the segmental and the tonal dimensions (the second research question) was tested by comparing the performance in the segment-and-tone with the segment-or-tone condition. Native Mandarin listeners adopted both dimensions as possible classification criteria, while the control group of native Dutch listeners uniformly classified the target along the segmental dimension. These results were in line with the findings of Braun and Johnson (2011), which showed that only Chinese listeners – who use pitch information in a lexically contrastive way – classified target words in incongruent trials along the pitch dimension. The beginning Mandarin learners in our study were not yet very sensitive to tonal information, and showed a pattern similar to the Dutch control group, whereas the advanced learners behaved more similarly to the native Mandarin listeners. That is, the advanced learners were attentive to both dimensions. In the segment-and-tone condition, processing of both dimensions integrally requires little cognitive effort (as suggested by the short RT). In the segment-or-tone condition, however, native Mandarin listeners processed both dimensions efficiently but extra time is needed for the classification task (as suggested by the increased RT). The advanced learners of Mandarin were shaping new selective perception routines and optimizing the attunement to information reliable for word-form detection in Mandarin, in line with the phonological mode of processing predicted by ASP (Strange, 2011). More specifically, they have learned to shift their attention to the previously ignored tonal dimension (given their native language experience). This dimension was therefore "stretched" (Nosofsky, 1986) and the difference of tonal categories along this dimension became more salient to them (as compared to the beginners). The enhanced sensitivity to tonal information actually slowed them down in the classification task both in the segment-and-tone condition and the segment-or-tone condition. This result also suggests that one's perceptual space remains plastic and dynamic and can be further shaped by learning experience with a second language throughout adulthood. The sensitivity to pitch information is flexible and the process of establishing new tonal categories in L2 learning indeed involves the redistribution of attention along perceptual dimensions.

The development of the integrality of the segmental and the tonal dimensions (the third research question) was revealed by comparing the reaction times in the forced-segment and forced-tone conditions with the segment-and-tone condition within each participant group. For native Mandarin listeners, these two dimensions were processed in an integral manner. They were less able to divert their attention from tonal variations when classifying the target along the segmental dimension, and *vice versa*. In addition, the integrality of these two dimensions was asymmetrical, in that the segmental dimension interfered more with the tonal dimension, while the interference from tonal variation was smaller when classification was required along the segmental dimension. This finding is consistent with the results of Tong et al. (2008). They tested the interactions between the segmental and the suprasegmental dimensions of Mandarin by asking participants to attend to one dimension while ignoring the other one. Their results suggested that variations in the segmental dimension interfered more with tone classification than *vice versa*.

For native Dutch listeners, our results suggest that the two dimensions were processed separately. The variation in the tonal dimension did not affect the processing of segments. They could direct their attention to the segmental dimension and ignore the other dimension in the forced-segment condition. The beginning learners demonstrated a pattern like that of the native Dutch listeners. That is, they adopted a similar strategy used in the processing of their native language in the processing of tonal information. The advanced learners, in contrast, behaved more like the Mandarin native speakers. They also showed a similar asymmetry with more segmental interference for the tonal dimension than the reverse. One may note that the RTs were longer for advanced learners than for native Mandarin listeners in both the forced-segment and forced-tone conditions. The slower performance by advanced learners is probably due to the fact that their L2 selective processing routines were still in development and were not as automatic as those of native Mandarin listeners. Alternatively, it may be that even though they have acquired the lexical tones phonologically, their processing of the non-native contrasts still requires more attention, as would be predicted by the ASP model.

Lin and Francis (2014) employed the Garner test to examine the differences in attention to consonants and tones by Mandarin learners of English and native English listeners. The experiment was done in both English and Mandarin modes. Results showed that in both the Chinese and English contexts, Mandarin learners of English processed consonant and tone in an integral manner, while English listeners processed these two dimensions in a separate manner. It is worth noting that Mandarin listeners did not give up the segment-tone-integration strategy in an English context although they were proficient L2 English learners. That is, they maintained the processing strategy in their native tone language when processing words in a non-tone second language. This is in contrast to the advanced Dutch learners of Mandarin in this study who had not only successfully acquired the distinctions of tonal categories, but also had developed a strategy similar to that of native Mandarin listeners in terms of segmental-tonal integrality.

In conclusion, a developmental path in phonological tone processing was observed for Dutch learners of Mandarin in the current study. Our results suggest that learners' sensitivity to pitch information is flexible and the acquisition of new tonal categories in L2 can indeed involve a gradual change in the distribution of attention along perceptual dimensions and the development of segment-tone integrated processing.

Chapter four

The representation and accessing of lexical tones by Dutch Learners of Mandarin Chinese

4.1 Introduction

Considerable evidence indicates that Mandarin tone presents a great challenge for adult non-tone language speakers learning Mandarin as a second language (L2) (e.g., Hao, 2012; Shen, 1989; Wang, Jongman, & Sereno, 1999). We know that L2 Mandarin learners can be trained to improve their tone identification in monosyllabic words but typically to a suboptimal plateau (Wang et al., 1999). What could have hampered the ultimate attainment of native-like tonal processing? This study investigated two potential levels of processing costs that may explain the suboptimal tonal identification accuracy by non-tonal second language learners: phonological tonal processing and lexical accessing. We did so by examining the developmental trajectory of phonological/lexical tonal processing by beginning and advanced Dutch learners of Mandarin in a sequence recall task and a lexical decision task, respectively.

4.1.1 Assessment of non-native segmental and suprasegmental perception

Adult learners are often confronted with difficulties in non-native segmental and suprasegmental perception when learning an L2, especially when they have to learn novel phonemic contrasts in the target language. Such difficulties have been demonstrated by many studies using perception tests ranging from basic phonetic discrimination and identification to tasks testing more abstract phonological representations and lexical activation (e.g., Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Pallier, Bosch, & Sebastián-Gallés, 1997; Sebastián-Gallés, Echeverria, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999; Strange & Dittmann, 1984). The automatic selective perception (ASP) model of first and second language speech processing has been proposed to account for such difficulties (Strange, 2011). It emphasizes the role of attention and differentiates between phonological and phonetic modes of perception. The phonological mode is employed by native listeners, in which automatic selective perception routines are used to detect phonologically contrastive information for identifying word forms. This automatic processing is shaped by language experience, costing little cognitive effort. The phonetic mode is employed by native speakers to detect fine-grained allophonic details, and requires more cognitive effort. It is hypothesized that at the beginning stage of L2 learning, the phonetic mode of perception has to be used when processing novel contrasts, and the L2 learning process involves the development of new selective perception routines that optimize the attunement to the information reliable for word-form recognition. The role of the task is also emphasized by this model: in tasks with a high memory load and phonetic variability, L2 listeners are less likely to detect fine-grained phonetic details, and therefore have to use the phonological mode of processing; in less demanding tasks with simple stimuli, the phonetic mode can be used.

The problem of Japanese listeners' discrimination of the English /r/-/l/ contrast (Strange & Dittmann, 1984) is a good example of acquisition difficulty that can be accounted for by the ASP model. The L2 listeners performed well in basic identification and discrimination tasks, in which the phonetic mode of processing could be used. In a more demanding task with complex stimuli asking for the phonological mode of processing, their performance was poor, since the selective perception routines of English had not been established yet. Likewise, for perception of a non-native /e/-/ɛ/ contrast the level of difficulty is a function of task and stimulus factors, as predicted by the ASP model (Pallier et al., 1997; Sebastián-Gallés, Echeverria, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999).

Although most research on non-native sound perception was directed at the segmental aspect, research on non-native suprasegmental perception has yielded some similar results, especially in non-native lexical stress perception. French native listeners have been reported to have difficulties in perceiving word stress contrasts, which are absent in French. Dupoux, et al. (1997) showed that French participants did not have detectable problems with stress contrasts in a basic AX discrimination task. In a later study, a sequence-recall task, with tokens from multiple speakers, was used to test stress processing of French and Spanish native speakers. It was found that French speakers' performance was significantly below that of native Spanish listeners, showing great difficulty in processing stress contrasts (Dupoux, et al., 2001). This suggests that French speakers can use acoustic stress cues in a discrimination task, but cannot encode stress in their phonological representation of words. These findings also suggest the important role of demanding tasks and stimulus complexity in the assessment of participants' processing ability of non-native segmental and suprasegmental contrasts.

4.1.2 Perception of tones by native Mandarin speakers

Mandarin is a language with a rather complex tone system. There are four full tones and a neutral tone. Tone 1 (T1) is a high tone; Tone 2 (T2) is a rising tone, Tone 3 (T3) is a low tone and Tone 4 (T4) is a falling tone (Chen & Gussenhoven, 2008; Duanmu, 2000). When produced in a prepausal position or in isolation, Tone 3 (T3) is realized as a dipping tone. This tone also has two variants in connected speech: it becomes a low falling tone preceding T1, T2, T4 and neutral tone, and it is realized with a rising contour similar to T2 preceding another T3. The neutral tone always comes at the end of a word or phrase, and is associated with a weak syllable. It has a static and mid target, but the target is realized with more pitch variability than lexical full tones: the pitch of a syllable with neutral tone is substantially influenced by the tone in the preceding syllable (Chen & Xu, 2006). The use of lexical tones effectively reduces the "rampant" segmental homophony (Liu & Samuel, 2007). Some studies demonstrated that the tones have a functional load comparable to that of vowels (Oh, Pellegrino, Coupé, & Marsico, 2013; Surendran & Levow, 2004).

As for the role of tonal information in lexical access and selection, some studies suggested that tone might be a weaker cue in word recognition compared to segmental information, using tasks of speeded classification (Repp & Lin, 1990), vowel and tone monitoring (Ye & Connine, 1999) and priming (Sereno & Lee, 2014). However, some recent studies using online measures such as eye-tracking and event-related potentials (ERPs) showed parallel processing of segments and tones in word recognition, arguing that the role of tonal information is comparable to that of segmental information (Malins & Joanisse, 2010; Malins & Joanisse, 2012; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Zhao, Guo, Zhou, & Shu, 2011). Taken together, tonal information plays an important role in lexical access for native speakers, which can be captured and reflected more effectively in on line measurements. Moreover, the relative importance of tone (and word prosody in general) versus segmental information is variable and depends mainly on the speech quality. Prosody is more resistant to noise, and will assume greater importance in poor-quality speech (see, e.g., Cutler & McQueen, 2014; Wang, Zhou, & Xu, 2011).

4.1.3 Perception of Mandarin tones by non-native speakers

The important role in the language system makes tone acquisition a crucial aspect of Mandarin learning. The performance of L2 learners in tone perception has been investigated in previous research. Hao (2012) found that English and Cantonese learners of Mandarin performed better in a Mandarin tone mimicry task, involving low-level auditory perception and articulation only, than in tone identification and reading tasks, which require a more abstract representation of tones. This suggests that the main difficulty is the establishment of robust associations between pitch contours and tone categories. The results also showed that distinguishing T2 and T3 was most difficult, followed by distinguishing T1 and T2. Since T2-T3 confusion was the major problem for both native English speakers and native Cantonese participants, the confusion could be L1 independent, and probably stems from the acoustic similarity of these two tones, in that they have similar pitch contours and there is overlap in their pitch ranges (Moore & Jongman, 1997). This confusion also has been found for adult native Mandarin speakers (Bent, 2005; Zhang, Samuel, & Liu, 2012), and in first language acquisition (Li & Thompson, 1977). Wang et al. (1999) showed that English learners of Mandarin improved their tone identification accuracy in monosyllabic words from 69% to 90% after a two-week training. The training-induced improvement also generalized to new words and speakers. In both pre-test and post-test, the most difficult tone pair was {T2 T3}. The pair {T1 T4} was most resistant to improvement, and became the second most difficult one in the post-test. Zou, Chen and Caspers (2016) tested the phonological processing of Mandarin T2 and T4 by beginners and advanced Dutch learners of Mandarin in an ABX task. Compared to beginners, the advanced learners had improved significantly in tonal discrimination and showed a more native-like pattern in redistributing attention between segmental and tonal information, as well as integrated processing of these two types of information. However, the above mentioned studies mostly used basic identification and discrimination tasks in which phonetic mode of processing can be employed by participants to focus on phonetic details. In cognitively demanding tasks, however, listeners are not likely to detect fine-grained phonetic details and have to use phonological mode of processing. The performance of advanced L2 learners in cognitively demanding task, therefore, becomes an interesting issue to investigate. Can they develop a new "selective perception routine" and show a native-like pattern in phonological tonal processing? The current study therefore sets out to test the processing of all tonal contrasts by beginning and advanced Dutch learners of Mandarin in a cognitive demanding task, trying to reveal the L2 developmental trajectory and shed some light on the issue of ultimate attainment in L2 tonal processing.

Learning to use tonal information in lexical access is another crucial issue in L2 tonal acquisition. To our knowledge, no systematic empirical research has been done to investigate tone processing in lexical access by L2 learners. Recently, learning to use tonal information in a lexically contrastive way has been investigated in several perceptual training studies by testing naive non-native speakers of Mandarin. Using the sound-to-word learning paradigm, which trains participants to associate members of minimal tone pairs with different meanings, these studies examined the contribution of individual variability in cue weighting in tone learning (Chandrasekaran, Sampath, & Wong, 2010), the effect of individual musical experience (Wong & Perrachione, 2007), as well as the influence of tonal context in tone learning (Chang & Bowles, 2015). Some studies also found a training-induced change in the participants' neural system (Wong, Chandrasekaran, Garibaldi, & Wong, 2011; Wong, Perrachione, & Parrish, 2007). While the focus varies across these studies, the results lead to the convergent finding that naive non-native speakers of Mandarin can be trained to use pitch information lexically. Based on this finding, it can be assumed that L2 learners of Mandarin can also use tonal information effectively in word processing. However, to what extent can they achieve native-like pattern? Are different tonal contrasts equally difficult for them in lexical access? These questions remain less understood. Therefore, the current study sets out to investigate the use of tonal information in lexical access by L2 learners of Mandarin in a lexical decision task.

4.1.4 The present study

In this study, the main research questions are:

(1) What is the developmental trajectory of the Dutch learners' phonological processing of tonal contrasts?
(2) Can lexical tones be used properly in lexical activation by Dutch learners of Mandarin?

As reviewed above, Dupoux et al. (2001, 2008) found that sequence recall task provides a robust paradigm for testing the phonological processing of novel L2 phonemic contrasts. So, in the present research, a sequence recall task with non-words and high phonetic variability and memory load has been applied to answer the first research question.

In this task, participants were asked to learn to associate two members of disyllabic minimal tone pairs with the keys "a" and "b" in a training phase with feedback. In the test phase, a sequence of non-words was presented, and the task for the participants was to transcribe the sequence in the correct order by typing a series of "a" and "b". To induce a high memory load, sequences of four non-words were used. The phonetic variability was introduced in that the four non-words in a sequence were always produced by four different voices in random order. In addition, for each voice, four different tokens of the same non-word were used in different trials.

To answer the second research question, the use of tone in lexical access by Dutch learners of Mandarin was tested in an auditory lexical decision task. The stimuli were disyllabic Mandarin real words and non-words (i.e., with a wrong tone on the first syllable). Disyllabic word-non-word pairs differing only in a consonant were used as a control condition. The participants were asked to decide whether the presented item was a real word or not by pressing a key as soon as possible. Both response correctness and reaction time (measured from stimulus onset) were recorded.

4.2 Experiment 1: Sequence recall task

4.2.1 Participants

Twenty-six Dutch learners of Mandarin and 15 Mandarin controls participated in the experiment. All Dutch learners of Mandarin received formal Chinese training from the Chinese Studies program at Leiden University. The beginning group consisted of 6 males and 8 females (mean age = 20.8, SD = 2.8). Their Mandarin learning and speaking experience varied between 0.5 and 2 years (M = 1.2, SD = 0.5), and they had never lived in China. The other 14 participants (8 males and 6 females; mean age = 24.8, SD = 3.6) were advanced Mandarin learners, who had Mandarin experience between 3 and 14 years (M = 5.4, SD = 3.3), and had spent at least one year in China. The native Mandarin control group had 3 males and 12 females (mean age = 26.9, SD = 2.8). All were from the Northern part of China and could speak standard Mandarin.

4.2.2 Materials and design

All six tone pairs (T1-T2, T1-T3, T1-T4, T2-T3, T2-T4, and T3-T4) were tested in the experiment. In the experimental condition, three similar CVCV non-words were used with the target tone on the initial syllable and a neutral tone on the final syllable. The vowel set of the non-words consisted of [a], [i], and [u]. In the consonant set, there were three pairs of stops (labial: [p]-[p^h]; alveolar: [t]-[t^h]; velar: [k]-[k^h]). The three non-words were combined with different tone pairs in a counterbalanced way (see Table 4.1). Two minimal pairs differing only in a consonant were used as the segmental control condition (*fuda-fuga*; *subi-sudi*).[5] They were produced with T1 on the initial syllable and a neutral tone on the second. The segmental control condition was introduced as a baseline. The difficulty in tone processing will be revealed by comparison between segmental control condition and the experimental condition.

---

[5] Dutch has no phonemic contrast between a lax and tense velar stop, so that the marked member [g] is a new sound for Dutch learners, but they should know the sound from English, German and French, as well as from loan words that Dutch borrowed from these languages.

Table 4.1. *Non-word stimuli used in the sequence-recall task.*

Experimental condition	Associated keys	
	A	b
T1-T2	ba1ti	ba2ti
T1-T3	di3ka	di1ka
T1-T4	gu1pa	gu4pa
T2-T3	gu3pa	gu2pa
T2-T4	di2ka	di4ka
T3-T4	ba4ti	ba3ti
Segmental control condition	fu1da	fu1ga
	su1bi	su1di

The stimuli were recorded four times by four native Mandarin speakers (two females and two males) from northern China. In addition, the word "OK" was recorded by a female speaker. All items were recorded with a Sennheiser MKH416T microphone in the Leiden University Phonetics Lab (44.1 kHz, 16 bit).

There are 16 possible combinations for sequences of four non-words. To select the most difficult combinations, two Dutch learners of Mandarin and two native Mandarin listeners participated in a pilot with 192 stimuli (16 sequences × 2 repetitions × 6 tonal pairs). More errors were found for sequences with more variation in combinations. That is, the sequence abba with one transition from a to b and another transition from b to a was more difficult than the sequence of aabb with only one transition from a to b. So, out of all 16 possible sequences, the eight sequences with two and three transitions were selected (aaba, abaa, abba, baab, babb, bbab, abab, baba). In every sequence, the four non-words were produced by four different voices. The order of these four voices was counterbalanced over sequences. Each non-word was recorded four times by the four speakers. So, for each tone/segmental pair, all of these tokens were used in the eight sequences. That is, for T1T2, 16 tokens (4 voices × 4 tokens) of the non-word *ba1ti* were used. The design of this experiment is: 8 tonal/segmental pairs × 8 sequences × 2 repetitions, yielding 128 experimental trials.

4.2.3 Procedure

Each participant was tested individually in the Leiden University phonetics lab with all 128 trials with the auditory stimuli presented through Beyerdynamic DT-770 Pro headphones. The three groups of participants received instructions (in their native language) that they were going to learn some words in a foreign language. The six tonal pairs were tested separately in six experimental blocks. Each block embraces a word learning phase, a training phase and a main experimental phase. In the learning phase, participants were first asked to press "a" on the keyboard to hear the first foreign word. A sound token of one non-word from a tonal pair produced by a female speaker was played at this time (e.g., *ba1ti*). Then they were asked to press "b" on the keyboard, upon which the other sound token produced by the same female speaker was played (e.g., *ba2ti*). After that, the participants were presented with "a" or "b" on the screen.

Pressing the letter on the screen lead to the playing of one token of the corresponding non-word. In this way, participants heard all 16 tokens of each non-word within a tonal pair in random order by pressing the associated key. Subsequently in the training phase, the non-word-key association and distinctions between non-words were further trained in an identification task. After hearing a non-word, the participants were asked to press the associated key. They got feedback on their choice with the message of "Correct" or "Incorrect" on the screen. There were 16 training trials with stimuli produced by different speakers, presented in random order. An accuracy of 80% was defined as the success criterion. The participants moved on to the main experiment when they had reached this criterion. Two beginning learners could not satisfy this criterion and did not continue with the experiment. The remaining twelve beginners moved on to the main experiment. All advanced learners and native Mandarin listeners satisfied the criterion.

In the main experimental phase, there were two warm-up trials and 16 experimental trials. In each trial the participants heard a sequence of four non-word tokens produced by four speakers and a following "OK" produced by a female voice. In order to reduce the possibility of the participants translating the non-words into the associated letters immediately when listening to the stimuli, the inter stimulus interval among the four non-words was kept very short (50 ms) (Dupoux et al., 2008). The "OK" following the non-word sequence was adopted to prevent the participant from using echoic memory (Dupoux, et al., 2008; Morton, Crowder, & Prussin, 1971). The task for participants was to reproduce the order of the sequence by typing the associated keys as quickly and accurately as possible after hearing the word "OK". After the response, the next trial came after a 1500-ms pause.

The order of the six tonal blocks was random among participants. In each block, the participants repeated the training and sequence-recall procedures. The control condition with two blocks of segmental minimal pairs was tested after the six tonal blocks. In total there were eight blocks. Each block took about 5 minutes to complete, and there was a one-minute break between blocks.

4.2.4 Results

Analysis of transcription result (i.e., correct or incorrect transcription of the non-word sequence) was performed with a mixed effects logistic regression model using R and the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). For all trials, a model was constructed with Participant Group (i.e., native Mandarin listeners, beginning Dutch learners, and advanced Dutch learners), Tone Pair (i.e., six tone pairs and one segmental control condition) and their interaction as fixed effects. Intercepts for participants and items, as well as by-participant slopes for the effect of Participant Group were added as random effects. Post-hoc comparisons of differences between different levels within each effect were conducted with Bonferroni adjustment using the Multcomp package in R (Hothorn, Bretz, & Westgall, 2008).

The statistical results for the model of response accuracy are presented in Table 4.2. The χ^2 and corresponding p values for fixed and random effects were obtained from likelihood ratio tests. There were significant main effects of Participant Group, Tone Pair as well as a significant interaction between Participant Group and Tone Pair.

In the following, we will present a more detailed post-hoc analysis of the interaction of Participant Group and Tone Pair.

Table 4 2. *Summary of a mixed effects logistic model for response accuracy. (The fixed effect Tone Pair comprises six tone pairs and the segmental control pair.)*

Fixed effects	Accuracy		
	df	χ^2	*p*
Participant Group	2	30.65	< 0.001
Tone Pair	6	140.06	< 0.001
Participant Group × Tone Pair	12	115.82	< 0.001
Random effects			
1+ Participant Group | Participant	6	290.99	< 0.001
Item	1	38.75	< 0.001

The sequence recall accuracy of the six tone pairs and the segmental control condition for three groups is presented in Figure 4.1. In the control condition, the overall accuracy was high across all three participant groups with no statistical difference among groups (BL = 90.6%, AL = 85.7%, NM = 90.6%; all *p* values > 0.05). This indicates that all three groups can process segmental contrasts with little difficulty.

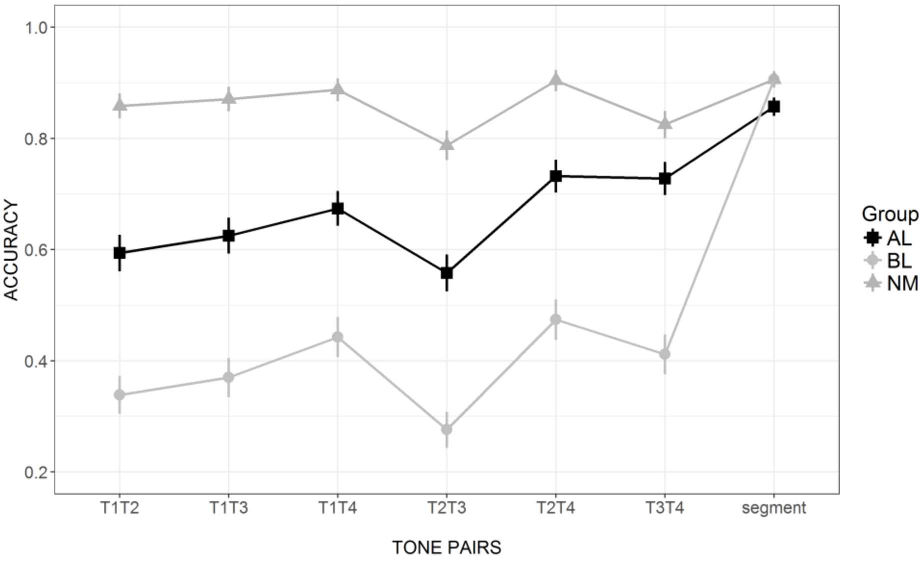

Figure 4.1. *Mean sequence-recall accuracy of three groups of participants in six tone pairs and the segmental control condition. BL: beginning Dutch learners of Mandarin; AL: advanced Dutch learners of Mandarin; NM: native Mandarin listeners. Error bars are ±1SE.*

In the tonal conditions, pairwise comparison demonstrated that the accuracy was significantly different between each two groups of participants for all tone pairs (all p values < 0.05) except T3T4. In all tone pairs, the accuracy of beginning learners was at the bottom but still well above chance performance level (chance level for 4-word sequence equals $1/2^4$, which is 7%). Compared to beginners, the advanced learners had improved significantly in tone processing, but their performance was still below that of the native Mandarin listeners. The only insignificance was between the performance of advanced learners and native Mandarin listeners in T3T4 (AL = 72.8%, NM = 82.5%; Est. = 0.77, $z = 2.10$, $p > 0.05$). The performance of advanced learners in this pair was comparable to that of native Mandarin listeners.

For the native Mandarin speakers, the performance in tonal and segmental conditions was comparable in general. Pairwise comparison showed that the accuracy of segmental pairs was only significantly higher than the accuracy of T2T3 ($z = -4.15$, $p < 0.001$). Among tone pairs, T2T3 (78.8%) was the one with the lowest accuracy, significantly lower than T2T4 (90.4%) which was the most accurate pair (Est. = 1.09, $z = 3.40$, $p < 0.05$). The accuracy was comparable among all other tone pairs.

For beginning learners, the accuracy in the segmental condition was significantly higher than that of all tone pairs (all p values < 0.001). Among tone pairs, the most difficult was T2T3 (27.0%), followed by T1T2 (33.9%), T1T3 (37.0%), T3T4 (41.1%), T1T4 (44.3%) and T2T4 (47.4%). Post-hoc analyses reveal that the accuracy of T2T3 was significantly lower than that of T1T4 (Est. = −0.83, $z = -3.05$, $p < 0.05$) and T2T4 (Est. = 0.93, $z = 3.58$, $p < 0.01$). The accuracies of other tone pairs were not significantly different from each other.

Similar to beginning learners, advanced learners were significantly more accurate in the segmental condition than in all tone pairs (all $p < 0.02$). Within the tonal conditions, T2T3 (55.8%) was again the most difficult pair, followed by T1T2 (59.4%), T1T3 (62.5%), T1T4 (67.4%), T3T4 (72.8%) and T2T4 (73.2%). The accuracy of T2T3 was significantly lower than T3T4 (Est. = 0.82, $z = -3.23$, $p < 0.05$) and T2T4 (Est. = 0.85, $z = 3.32$, $p < 0.05$).

4.3 Experiment 2: Lexical decision task

4.3.1 Materials and design

Ten disyllabic word-non-word pairs were chosen for each tone pair. For Dutch listeners, the stimulus ends with T1, T2 and T3 can be potentially interpreted as non-final (H%), which signals either continuation or question. T4 sounds like a final fall (H*LL%). So, to avoid the influence of different boundary tones from listeners' L1, we kept the tone on the second syllable constant, only using real words with T1 on the second syllable as stimuli. The non-words were constructed by only changing the tone on the first syllable of the real words. Tone pairs were tested bi-directionally, which means that there were 12 pairs in total (T1→T2, T2→T1, T1→T3, T3→T1, T1→T4, T4→T1, T2→T3, T3→T2, T2→T4, T4→T2, T3→T4 and T4→T3). For example, for tone pair T1→T2, 10 real words with T1 on the first syllable were selected, and the corresponding non-words were constructed by changing T1 to T2 on the first syllable. E.g., the corresponding non-word for the real word 春天 [tʂʰuən1tʰiɛn1] 'spring' was [tʂʰuən2tʰiɛn1]. As a control condition, another 40 disyllabic word-non-word pairs which differed only in the initial consonant of the first syllable were chosen. There were ten words with T1 on the initial syllable, and ten words each with T2, T3 and T4. The non-words were constructed by changing the manner of articulation of the initial consonant in the initial syllable. For instance, the corresponding non-word for the real word 公园 [kuŋ1yɛn2] 'park' was [kʰuŋ1yɛn2]. To make sure the learners were familiar with all words, the real words were selected from the first year text books of the Chinese studies program at Leiden University (see Appendix A4 for a list of the stimuli).

It is reported that the non-word type can influence the wordlikeness judgment in Mandarin. It has been shown that non-words with phonotactic violations (e.g., with consonant clusters) can be easily and correctly identified, whereas non-words which do not violate phonotactics but form a segment-tone combination gap (e.g., *dai2*) could not be easily and quickly ruled out by native speakers (Wiener & Turnbull, 2015). To maintain a similar wordlikeness level across non-words, only phonotactically legal syllables were used when constructing non-words in this experiment. It is further known that word frequency and lexical neighborhood density can affect the RT of lexical decision. Neighborhood density of a word (or non-word) is defined as the number of words that exist in the lexicon that differ from the target item in the addition, deletion or substitution of precisely one segment according to Neighborhood Activation Model (NAM) (Luce & Pisoni, 1998). Past research found that high-density words could elicit longer RTs and high-frequency words could elicit shorter RTs. The frequency effect was more salient for low-neighborhood density (Goh, et al., 2009; Luce & Pisoni, 1998). So, these two factors were carefully controlled in this experiment. The overall word frequency, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010), did not differ significantly across tone pairs and the segmental control condition [$F(12, 147) = 0.04, p > 0.99$]. The neighborhood density of the first and second syllable of the disyllabic words, computed as the number of homophones according to the Modern Chinese Dictionary, was also comparable across tone pairs and the segmental condition [first syllable: $F(12, 147) = 0.57, p = 0.87$; second syllable: $F(12, 147) = 0.80, p = 0.65$].

All stimuli were recorded by a female native Mandarin speaker from northern China speaking standard Mandarin. The recording was conducted with a Sennheiser MKH416T microphone in the Leiden University Phonetics Lab using Adobe Audition (44.1 kHz, 16 bit). The design of this experiment is 16 (12 tone pairs) × 10 word-non-word pairs + 4 segmental conditions × 2 word type (real word-non-word), yielding 320 experimental trials.

4.3.2 Procedure

The same groups of participants as in Experiment 1 were tested. All three groups received instructions in their native language. They were asked to decide whether the word they heard was a real word in Mandarin or not as quickly as possible by pressing the button "1" or "2" on the keyboard. The participants were informed that the non-words were very similar to real words but with a difference in tone or initial consonant on the first syllable. The order of the 320 stimuli was randomized for each participant. Two pairs of word-non-words differing in the initial consonant of the first syllable were presented before the main experiment as warming-up stimuli (the warming-up phase was used to let the participants get familiar with the associated buttons (1 vs. 2), so only two were presented). None of these words was used in the main experiment. During the warming-up phase, the participants received a "Correct" or "Incorrect" message on the screen as feedback. The main experiment consisted of four blocks of 80 trials. Each trial began with the presentation of a fixation cross on the screen for 500 ms. The stimulus was presented 500 ms after the disappearance of the cross. After the participant's response, the next trial came after a 1500 ms pause

4.3.3 Results

The response to each trial was classified as a hit (H) (correctly recognizing a real word), a false alarm (F) (mistakenly classifying a non-word as real word), a miss (failing to recognize a real word), or a correct rejection (correctly rejecting a non-word). For each participant, an A' ("A prime") score was calculated for each tone pair across items with formula (4.1) (see Stanislaw & Todorov (2009)):

$$A' = 0.5 + \left[\frac{sign(H-F)\left((H-F)^2 + |H-F| \right)}{4\max(H,F) - 4HF} \right] \tag{4.1}$$

A' is a bias-free estimate of sensitivity to word-non-word classification, which takes account of both hit rate and false-alarm rate. The range of an A' score is from 0.5, which indicates real words cannot be distinguished from non-words, to 1, which corresponds to perfect performance in word-non-word classification (Macmillan & Creelman, 2004; Stanislaw & Todorov, 2009).

Analyses of A' scores were performed with a linear mixed-effects model using R and the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). For all trials, a model was constructed with Participant Group (i.e., native Mandarin listeners, beginning Dutch learners, and advanced Dutch learners), Word Pair (i.e., 12 tone pairs and the segmental condition) and their interaction as fixed effects. Intercepts for Participant was used as random effect.

For reaction time, the raw RT data was converted to natural-logarithmic RT to achieve better normalcy. The analysis of log RT was also performed with a linear mixed effect model using R and the lme4 package (Bates et al., 2014). A model was constructed with Participant Group, Word Pair and their interaction as fixed effects. Intercepts for participants and items were entered as random effects. For both models of accuracy and reaction time, post-hoc comparisons of differences between different levels within each effect were conducted using the Multcomp package in R with Bonferroni adjustment (Hothorn, Bretz, & Westgall, 2008). The mean A' scores for each Participant Group are shown in Figure 4.2 broken down by the twelve tone pairs and the segmental condition. The log-transformed RTs for the three groups in different conditions are similarly presented in Figure 4.3.

For A' scores (see Table 4.3), there was a significant main effect of Participant Group and Word Pair. The interaction between Participant Group and Word Pair was also significant. For RT (see Table 4.3), there was a significant main effect of Participant Group and of the interaction between Participant Group and Word Pair.

Table 4.3. *Summary of mixed effects models for A' score and* RT.

Fixed effects	A' score			RT (log)		
	df	χ^2	*P*	df	χ^2	*P*
Participant Group	2	76.12	< 0.001	2	54.72	< 0.001
Word Pair	12	102.78	< 0.001	12	12.64	n.s.
Participant Group × Word Pair	24	174.4	< 0.001	24	101.58	< 0.001
Random effects						
Participant	6	507.38	< 0.001	6	2381	< 0.001
Item				1	847.64	< 0.001

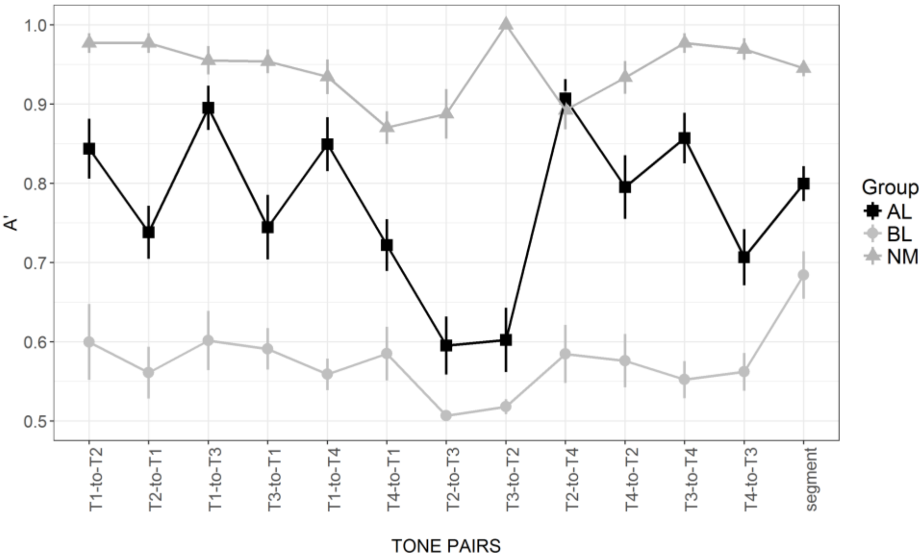

Figure 4.2. *Mean A' score of three groups of participants for 12 tone pairs and the segmental condition. BL: beginning Dutch learners of Mandarin; AL: advanced Dutch learners of Mandarin; NM: native Mandarin listeners. Error bars are ±1SE.*

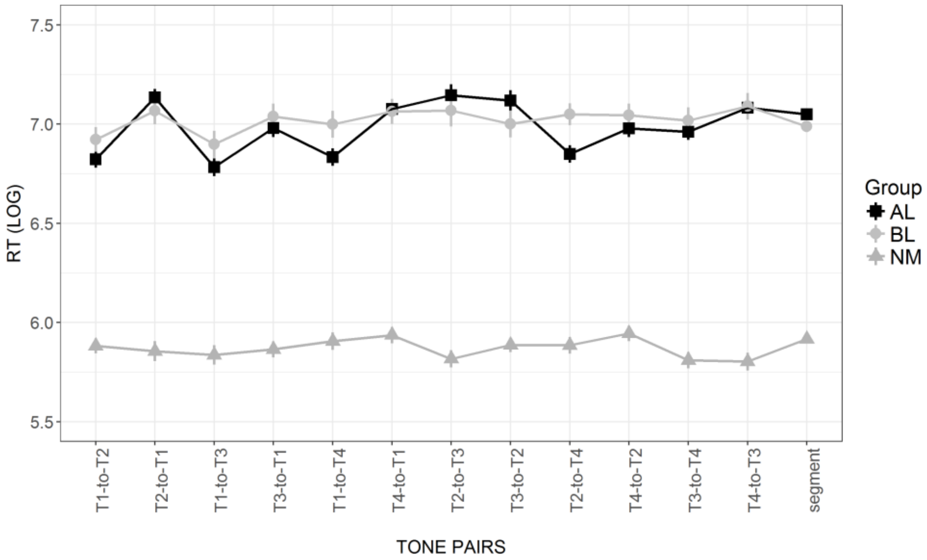

Figure 4.3. *Mean natural log-transformed reaction time of three groups of participants to 12 tone pairs and the segmental condition. BL: beginning Dutch learners of Mandarin; AL: advanced Dutch learners of Mandarin; NM: native Mandarin listeners. Error bars are ±1SE.*

In the segmental control condition, the A' scores of all three groups significantly differed from each other (all p values < 0.05). That is, the advanced learners showed a significant improvement compared with the beginning learners, but still do not perform like native Mandarins. The mean RTs of beginners and advanced learners did not differ from each other, but both learner groups responded significantly slower than native Mandarin listeners (BL vs. NM: Est. = −1.08, $z = -8.03$, $p < 0.001$; AL vs. NM: Est. = −1.15, $z = -8.94$, $p < 0.001$). In the tonal condition, the three groups demonstrated a similar pattern in A' scores compared to that in the segmental control condition. The advanced learners showed a significant improvement in most tone pairs compared to the beginning learners (all p values < 0.05), except for the pairs of T2→T3 and T3→T2, indicating that the sensitivity to T2 and T3 was still very low and resistant to improvement. The A' scores of the native Mandarins are significantly higher than those of the advanced learners in most tone pairs (all p values < 0.001), except for the pairs T2→T4 (Est. = −0.01, $z = -0.39$, $p = 1$), T1→T3 (Est. = 0.06, $z = 1.58$, $p = 0.343$) and T1→T4 (Est. = 0.09, $z = 2.24$, $p = 0.075$). For RT, both learner groups responded significantly slower than the native Mandarins in all tone pairs (all p values < 0.001). The RTs of the two learner groups were comparable (all p values > 0.05). Native Mandarin listeners showed high sensitivity in both the segmental control condition and the tonal condition, and there was no significant difference between these two conditions for both A' scores and RTs. Within the tonal condition, the overall A' score was high across tone pairs. Among tone pairs, T4→T1 was the one with the lowest sensitivity score, significantly lower than T3→T2, T1→T2, T2→T1, T3→T4 and T4→T3 (all p values < 0.05). T3→T2 was the one with the highest sensitivity score, significantly higher than that of

T4→T1 (Est. = −0.13, $z = -4.39$, $p < 0.001$), T2→T3 (Est. = 0.11, $z = 3.80$, $p < 0.5$) and T2→T4 (Est. = 0.11, $z = 3.64$, $p < 0.05$). For each two tones, only T2 and T3 showed a directional difference in sensitivity, with an A’ score for T3→T2 significantly higher than T2→T3. That is, when T3 was produced as T2, native Mandarin listeners were more likely to make a correct response than the other way round. In the initial position, the category of T3 was more well-established than T2. For native Mandarin listeners, the RT was not significantly different across tone pairs.

For beginning learners, their A’ score in the segmental control condition was on average higher than in the tone condition. Pairwise comparison showed that the A’ score in the segmental condition was significantly higher than for tone pairs T2→T3, T3→T2, T1→T4, T2→T1, T3→T4 and T4→T3 (all p values < 0.05). In the tonal condition, both A’ scores and RT were comparable among tone pairs.

For advanced learners, the A’ score of the segmental control condition was comparable with the scores in the tone condition, only significantly higher than that of T2→T3 (Est. = −0.20, $z = -6.68$, $p < 0.001$), T3→T2 (Est. = −0.20, $z = -6.45$, $p < 0.01$) and T2→T4 (Est. = 0.11, $z = 3.51$, $p < 0.05$). The RT of the segmental condition was shorter than that of tone pair T1→T3 (Est. = −0.27, $z = -3.66$, $p < 0.05$). Among tone pairs, T2→T4 was the one with the highest A’ score (0.91), followed by T1→T3 (0.90), T3→T4 (0.86), T1→T4 (0.85), T1→T2 (0.84), T4→T2 (0.80), T3→T1 (0.74), T2→T1 (0.74), T4→T1 (0.72), T4→T3 (0.71), T3→T2 (0.60) and T2→T3 (0.60). Post-hoc tests demonstrated that the A’ scores of T2→T3 and T3→T2 were significantly lower than of other tone pairs (all p values < 0.01). The RT of T2→T3 was significantly longer than T1→T2 (Est. = 0.34, $z = 3.58$, $p < 0.05$), T1→T3 (Est. = 0.39, $z = 4.09$, $p < 0.01$), T1→T4 (Est. = 0.33, $z = 3.49$, $p < 0.05$) and T2→T4 (Est. = −0.34, $z = -3.53$, $p < 0.05$). The response to T3→T2 was significantly slower than the response to T1→T2 (Est. = 0.33, $z = 3.45$, $p < 0.05$) and T1→T3 (Est. = 0.38, $z = 3.97$, $p < 0.01$). That is, compared to beginning learners, sensitivity to tone information in lexical access improved for advanced learners, but this improvement was not equally distributed among tone pairs, with the confusion between T2 and T3 most resistant to improvement. The sensitivity score of T2→T3 and T3→T2 was comparable, indicating that these two tones were symmetrically confusable: it was equally difficult to make a correct response when T2 was produced as T3 and *vice versa*.

It is noteworthy that this is not true for other tone pairs. There was a significant difference in A’ score between T1→T2 and T2→T1 (Est. = −0.11, $z = -3.45$, $p < 0.05$), T1→T4 and T4→T1 (Est. = −0.13, $z = -4.17$, $p < 0.01$) as well as T1→T3 and T3→T1 (Est. = −0.15, $z = -4.93$, $p < 0.01$). That is, it was easier for advanced learners to correctly recognize real words with T1, and to correctly reject non-words with T1 replaced by T2, T3 or T4 than *vice versa*. Similarly, there were significant differences between A’ scores of T2→T4 and T4→T2 (Est. = −0.11, $z = -3.66$, $p < 0.05$), T3→T4 and T4→T3 (Est. = −0.15, $z = -4.92$, $p < 0.01$), as well as T1→T4 and T4→T1 (Est. = −0.13, $z = 4.17$, $p < 0.01$). These directional differences show that there is an asymmetry between T4 and the other three tones. It was more difficult for advanced learners to make a correct response when T4 was produced as another tone than the other way round. There was also a trend of shorter RTs in T1→T2, T1→T3, and T1→T4 than to T2→T1, T3→T1 and T4→T1 respectively, although these differences did not reach statistical significance. These asymmetric patterns for T1 vs. other tones and T4 vs. other tones were only found for advanced learners.

4.4 Discussion and conclusion

In the present study, we investigated the phonological processing of tonal contrasts and the role of tone in lexical access by Dutch learners of Mandarin. A clear developmental path was found by comparing the performance of two learner groups in the sequence recall task: the advanced learners with more Mandarin experience showed a significant improvement compared to the beginning learners, approximating the performance of native Mandarin speakers. The improvement found for the advanced learners can be explained by the relatively important role of lexical tones in spoken Mandarin. Earlier research suggested that the perceptual difficulty of some non-native contrasts cannot only be attributed to the interaction of the L1-L2 phonological systems, but also came from the relatively "weak" role of the contrast in the target language. Mora, Keidel and Flege (2010) demonstrated that even early bilinguals failed to present a native-like pattern in categorical identification and discrimination tasks using certain vowel continua. They perceived the high-mid vowel contrasts (/i/-/e/, /u/-/o/), which exist in both Spanish and Catalan, more categorically than the mid-mid vowel contrasts (/e/-/ɛ/, /o/-/ɔ/), which phonemically contrast only in Catalan. It was argued that such a persistent difficulty can be attributed to the low functional load of mid vowel contrasts: the number of minimal pairs involving the contrast /e/-/ɛ/ is limited. Catalan mid vowels are also neutralized to /ə/ in unstressed syllables. Different from the cases of lexical stress and /e/-/ɛ/ learning, the high functional load makes tone acquisition a crucial aspect of Mandarin learning. The important role of tone was explicitly pointed out to the students as part of their training. Moreover, all advanced learners lived at least one year in China, so the large amount of high quality tonal input they received may have facilitated the formation of tonal categories.

The result of lexical decision task also showed that, compared to beginners, advanced learners performed significantly better in correctly identifying real words and rejecting non-words which were minimally different from real words in tones. That is, effectively using tones in lexical access is also in function of Mandarin experience.

The improvement of advanced learners in both tasks demonstrated that they were shaping new selective perception routines, and their phonological mode of tone processing was in development which is in line with the ASP model (Strange, 2011). However, it should be noted that the RTs were longer for advanced learners than for native Mandarin listeners in all conditions, indicating that their L2 selective processing routines were still in development and were not as automatic as those of native Mandarin listeners.

Despite the general improvement in tone processing between the beginners and the advanced learners, the tone pair of T2 and T3 remains the most difficult one and resists improvement, as reflected in both tasks. In the sequence recall task, T2-T3 was also the most difficult pair with the lowest accuracy for all three groups of participants. The confusion of this tone pair has been attributed to the acoustic similarity of these two tones in previous studies (e.g., Hao, 2012; Wang et al., 1999). However, it should be noted that, naturally produced disyllabic non-words were used as stimuli in the present sequence recall task. T3 in a non-word like *di3ka* was realized as a variant with low falling contour. T2 and T3 share some acoustic similarity in citation form (both have a rising part in the pitch contour), but in connected speech, the low-falling variant of T3 is different from T2. Therefore, it can be the case that for both native speakers and learners, multiple variants can co-exist as representations of T3.

When hearing the low falling T3 variant, the listeners may recover the canonical form of T3 which bears some acoustic similarity with T2, and this restoration of the canonical T3 may result in the confusion of these two tones.

In the lexical decision task the A' scores were very high across tone pairs for native Mandarin listeners. However, differences among tone pairs still exists. For native Mandarin listeners, there was an asymmetry for tone pair T2 and T3. They performed better in recognizing real words with T3 and rejecting non-words in which T3 was produced as T2 than *vice versa*. That is, the category of T3 (a low tone) in word initial position was more well-established compared to T2 (a rising tone). This might be accounted for by the fact that an initial low tone can only be perceived as T3 by native listeners, but an initial rising tone could have two underlying tone forms: T2 or the sandhi form of T3: according to the tone sandhi rule, T3 becomes a rising tone (which sounds like T2) when followed by another T3. This may hinder the participants in making correct responses when T2 is followed by another syllable. However, this hypothesis still needs to be tested in further experiments with more participants and stimuli. For the two learner groups, T2 and T3 are mutually confusable in lexical access. This confusion is in line with the findings of other research (Hao, 2012; Wang et al., 1999).

In the lexical decision task, the confusion for T2 and T3 was bi-directional for both groups of learners. Comparing to beginning learners, only small improvement was found in this tone pair for advanced learners. In contrast, larger improvements were found in other tone pairs. However, for these tone pairs, the confusion of the two tones was not bi-directional. The advanced learners were significantly more accurate in recognizing real words with T1 and rejecting non-words in which T1 was produced as the other three tones than the other way round. That is, the category of T1 had been relatively well established when compared to the other three tones. Contrastively, it was difficult for advanced learners to make a correct response when T4 was produced as T1, T2 or T3, that is, the category of T4 was relatively less well-established when compared to the other three tones in pairs. These results are potentially related to the prosodic features of the learners' native language, since Gandour (1983) showed that compared to tone-language speakers, intonational language speakers are more sensitive to pitch height than to pitch direction. Alternatively, such asymmetric tone perception can be attributed to the influence of intonation patterns in Dutch. The acquisition of tonal categories can be influenced by similar intonation contours in learners' L1. The pitch fall in T4 is similar to the falling pitch accent in Dutch, which is the most neutral form of pitch accent in Dutch (Gussenhoven, 2005). This similarity may make T4 less marked for Dutch learners of Mandarin, and therefore Dutch learners might be less attentive to T4 in the time course of tone acquisition.

In conclusion, the advanced learners showed a significant improvement in tonal processing at phonological level and using lexical tones in lexical access compared to beginning learners. This improvement in tone acquisition can be attributed to the important role of tones in Mandarin. Different tone pairs were not equally difficult to learners in lexical access, and the source of such differences can be attributed to acoustic similarity between particular tones as well as interference from L1 suprasegmental features.

Chapter five

The role of lexical tonal and segmental information in spoken word recognition for Dutch learners of Mandarin

5.1 Introduction

For adult second language learners of Mandarin Chinese, the acquisition of lexical tones is crucial since tones use pitch information to distinguish lexical meaning. However, non-tonal learners of Mandarin often find difficulties in tonal processing, since different pitch movements do not serve a lexically contrastive role in their native languages. Past studies have shown that experienced and novice Mandarin learners can gain significant improvement in tone acquisition after perception training (e.g., Hao, 2012; Lu, Wayland, & Kaan, 2015; Wang, Jongman, & Sereno, 2003). However, these studies largely focus on learners' performance in simple tonal discrimination and identification tasks. The role of tones in the time course of auditory spoken word recognition for non-tonal learners has remained much less understood. Can tones be exploited in spoken word recognition by L2 learners? What is the role of tonal information compared to segmental information? This study was therefore designed to examine the role of tonal information, in comparison to segmental information, in word recognition by Dutch learners of Mandarin using an online eye-tracking experiment.

5.1.1 Tone processing by native Mandarin speakers

In Mandarin Chinese, tones are used to distinguish lexical meaning, and the role of tones in spoken word recognition by native speakers has received much attention in previous studies. Some studies suggest that tone might be a weaker cue in word recognition compared to segmental information. Repp and Lin (1990) found that in a speeded classification task, tone was accessed later than segment information, as Mandarin speakers took longer to respond to tonal distinctions than segmental distinctions. Cutler and Chen (1997) supports this finding by showing that in a lexical decision task, Cantonese listeners more often accepted a non-word as a real word when the only difference between the non-word and the word was in tone. Ye and Connine's study (1999) shows a similar result. In vowel and tone monitoring tasks without linguistic context, Mandarin listeners showed faster monitoring times for vowel mismatch stimuli than for tonal mismatches, which supported a perceptual advantage of vowel information. Studies by Sereno and Lee (2014) and Wiener and Turnbull (2016) also report similar results using direct priming and word reconstruction tasks, respectively. It might be the case that the difference between tones and vowels are simply due to the difference in temporal availability of the cues. Prosody develops more slowly over time than segmental information, but every cue will be used in word recognition as soon as it is reliably perceived. The above-mentioned studies mainly recorded the end-state responses, which are not able to capture the perception of tones in real time.

Recently some studies using online measures such as eye-tracking and event related brain potentials investigated the dynamics in the time course of word recognition. Parallel processing of segments and tones was found in these studies, suggesting the role of tonal information is comparable to that of segmental information. For example, Schirmer, Tang, Penney, Gunter, and Chen (2005) used event related potentials (ERPs) to investigate the role of tone and segmental information in Cantonese word processing. Comparing the ERPs elicited by semantically congruous words and by tonally and segmentally induced semantic violations, they found that both segments and tones were accessed at a similar point in time and elicited an N400-like negativity (although this task probably tap into a different stage of the word recognition process. Since semantics work post-access). The ERP study by Malins and Joanisse (2012) offered further support for the comparable roles of segments and tones, showing that both segmental and tonal information could be accessed and used as soon as they become available during word processing. Zhao, Guo, Zhou, and Shu (2011) also reported that segmental and tonal mismatches equally modulate the amplitudes and time courses of the N400.

Taken together, the existing literature suggests that tonal information is exploited in spoken word recognition. It plays an early constraining role in lexical activation, and word with non-matching tone would not be activated as candidate. This effect can be captured and revealed more effectively in online measurements with tasks more similar to real communication situations. Therefore, the eye-tracking method is employed in the current study to provide a continuous measure of the auditory word recognition process.

5.1.2 Tone processing by non-tone language speakers

There have also been abundant studies that have tested the perception and production in beginning Mandarin L2 learners. For example, Wang et al. (1999, 2003) show that English learners of Mandarin improved their tone identification accuracy in monosyllabic words from 69% to 90% after a two-week training. The training-induced improvement also generalized to new words and speakers. Other than tones in isolated syllables, the perception of longer stimuli has also been tested. Hao (2012) found that both English and Cantonese learners of Mandarin performed better on monosyllabic tonal identification than on disyllabic identification. Both learner groups showed better performance on Mandarin tone mimicry than in tone identification and reading-aloud tasks. Mimicry only involved low-level auditory perception and articulation while the latter tasks required a more abstract representation of tones. This suggests that the main difficulty in tone learning is how to establish robust associations between pitch contours and tone categories.

More recently, the learning of new tonal categories by speakers without prior tone language learning experience has been tested in several phonetic training studies. Adopting different training paradigms (perception-only training, perception-plus-production training and sound-to-word training), these studies examined the role of different modalities in training (Lu, Wayland, & Kaan, 2015), the distinction between reflective and reflexive learning (Chandrasekaran, Yi, & Maddox, 2014), the contribution of individual variability in cue weighting in tone learning (Chandrasekaran, Sampath, & Wong, 2010), the effect of individual musical experience (Wong & Perrachione, 2007), as well as the influence of tonal context in tone learning (Chang & Bowles, 2015). Although the research questions varied across these studies, their results led to a convergent finding that naive non-native speakers of Mandarin can gain significant improvement in tonal identification and discrimination with a proper amount of training, and can learn to use tones in a lexically contrastive way. Some studies also found a training-induced change in participants' neural system (Lu et al., 2015; Wong, Chandrasekaran, Garibaldi, & Wong, 2011). Since most of the studies mentioned above focused on the learning of lexical tones by naive non-native Mandarin speakers and beginning learners of Mandarin, the performance of advanced L2 learners and the developmental trajectory in the time course of tone acquisition have not been studied systematically. Moreover, L2 processing of tonal information has not been investigated using on-line methods. Therefore, the current study sets out to examine the role of tones and segments in auditory spoken word recognition using the Visual World Paradigm by monitoring the eye movements of both beginners and advanced Dutch learners of Mandarin.

5.1.3 The present study

An eye-tracking experiment with Visual World Paradigm (VWP) was employed in the current study to test (1) the relative role of segmental and tonal information in lexical activation and selection by native speakers and Dutch learners of Mandarin, and (2) the developmental trajectory for Dutch learners of Mandarin in using segmental and tonal information effectively in spoken word recognition. Both beginners and advanced Dutch learners of Mandarin were recruited and native Mandarin speakers were tested as a control group.

VWP has typically been used in eye-tracking studies to investigate on-line auditory word recognition (Righi, Blumstein, Mertus, & Worden, 2009; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; also see a review in Heuttig, Rommers, & Meyer, 2011). Many related factors have been tested, such as the effect of frequency (Dahan, Magnuson, & Tanenhaus, 2001) and neighborhood density (Magnuson, Dixon, Tanenhaus, & Aslin, 2007). This paradigm has also been employed to test spoken word recognition by participants with language impairment (e.g., McMurray, Munson, & Tomblin, 2014; McMurray, Samelson, Lee, & Tomblin, 2010; Mirman, Yee, Blumstein, & Magnuson, 2011; Yee, Blumstein, & Sedivy, 2008). A recent study also demonstrated that this paradigm can provide reliable measurement for individual behavior (Farris-Trimble & McMurray, 2013).

In the VWP task, participants are presented with a display of four pictures and an auditory stimulus corresponding to one of these pictures, and they are asked to identify the word (i.e., the target) they heard with their eye movements being tracked during the whole process. In studies using VWP, the target word is always presented with a competitor which is phonologically similar to the target, and two phonologically unrelated distractors. For example, in Allopena, Magnuson, and Tanenhaus (1998), participants were presented with a target word (e.g., *beaker*), a cohort competitor (e.g., *beetle*) (which shares the onset and vowel with the target), and two phonologically unrelated distractors (e.g., *parrot*, *carriage*). The results show that when hearing the instruction *Pick up the beaker*, participants tended to look at both *beaker* and *beetle* initially. As the word was unfolding, the target picture gained a greater proportion of looks. A similar effect was also found for the competitor that shared the rhyme with the target (e.g., *speaker*), indicating that listeners continuously extract segmental information from the acoustic signal and that phonologically similar lexical candidates can be gradually activated. This result supports the linking hypothesis between eye movements and lexical access, showing that fixations are time locked to the details of the speech input. This finding also closely matches the word recognition mechanisms posited by the TRACE model (McClelland & Elman, 1986), which assumes that when a word is heard, phonologically similar lexical candidates can be activated at any point in overlap with the speech input. Speech sounds presented at a feature layer can be mapped onto phoneme and word layer. With between-layer excitation and within-layer competitive inhibition, one word can be recognized among phonologically similar lexical candidates. The model suggests a lateral inhibition among lexical candidates, which can lead to a delayed activation of the correct target word. Studies using VWP have found evidence for this mechanism, in which the delayed activation is reflected as a delay in fixation on the target word (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Tanenhaus, Magnuson, Dahan, & Chambers, 2000).

The VWP has also been used in a recent study to examine the role of tonal information in spoken word recognition (Malins & Joanisse, 2010). In this study, Mandarin listeners were asked to identify the corresponding picture from four pictures while hearing a word. Both cohort competitor (sharing word initial phonemes and tone with the target) and segmental competitor (sharing segmental structure with the target and differing only in tone) caused slower eye movements to correct targets, indicating that tonal and segmental information play comparable roles in constraining lexical activation. Based on this finding and suggestions from previous studies (Malins & Joanisse, 2010; Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011), tones have been incorporated into the TRACE model in a recent simulation of monosyllabic spoken word recognition of Mandarin Chinese (Shuai & Malins, 2017).

The processing of segmental and tonal information in word recognition by Dutch learners of Mandarin is tested in the current study. To test the competition effect of segmental versus tonal cues, the target words are presented with different types of competitors: cohort competitors sharing the initial consonant and tone with the target (e.g., *che1* 'car'; for the target of *chuang1* 'window'); rhyme competitors with rhyme and tone overlap (e.g., *guang1* 'light'); segmental competitors with complete segmental overlap (e.g., *chuang2* 'bed'), and tone competitors with tone overlap (e.g., *ji1* 'chicken'). The probability of fixation to the target and competitors was recorded since it may reflect the activation of the corresponding items.

5.2 Method

5.2.1 Participants

Fifteen Mandarin control participants and 26 Dutch learners of Mandarin participated in the experiment (11 beginners and 15 advanced learners). The native Mandarin control comprised 5 males and 10 females (mean age = 26.9, SD = 2.5). All were from the Northern part of China and spoke Standard Chinese on a daily basis. All Dutch learners of Mandarin received formal Chinese training from the Chinese Studies program at Leiden University. The beginner group consisted of 4 males and 7 females (mean age = 20.1, SD = 2.3). Their Mandarin learning and speaking experience varied between 0.5 and 2 years (mean = 1.2, SD = 0.5), and they had never lived in China. The other 15 participants (4 males and 11 females; mean age = 23.5, SD = 3.0) were advanced Mandarin learners, who had Mandarin experience between 3 and 14 years (mean = 4.9, SD = 2.6), and had spent at least one year in China.

5.2.2 Material

The stimuli in the eye-tracking experiment consisted of 12 sets of monosyllabic Mandarin words (Table 5.1).

Table 5.1. *Experimental stimuli.*

	target		segmental competitor		cohort competitor		rhyme competitor		tonal competitor	
T1-T2	*chuang1*	window	*chuang2*	bed	*che1*	car	*guang1*	light	*ji1*	chicken
T2-T1	*tang2*	sugar	*tang1*	soup	*tou2*	head	*wang2*	king	*chuan2*	boat
T1-T3	*shu1*	book	*shu3*	mouse	*shan1*	mountain	*zhu1*	pig	*deng1*	lamp
T3-T1	*bing3*	pie	*bing1*	ice	*ben3*	notebook	*ying3*	shadow	*san3*	umbrella
T1-T4	*hua1*	flower	*hua4*	painting	*hei1*	black	*gua1*	mellon	*xin1*	heart
T4-T1	*xia4*	summer	*xia1*	shrimp	*xian4*	thread	*jia4*	shelf	*yao4*	medicine
T2-T3	*yu2*	fish	*yu3*	rain	*yun2*	cloud	*ju2*	orange	*pan2*	plate
T3-T2	*bi3*	pen	*bi2*	nose	*ben3*	notebook	*mi3*	rice	*gou3*	dog
T2-T4	*qi2*	flag	*qi4*	gas	*qiao2*	bridge	*li2*	pear	*yang2*	sheep
T4-T2	*mao4*	hat	*mao2*	fur	*mian4*	noodle	*yao4*	medicine	*tu4*	rabbit
T3-T4	*shu3*	mouse	*shu4*	tree	*shou3*	hand	*tu3*	dirt	*yan3*	eye
T4-T3	*dian4*	electricity	*dian3*	point	*dou4*	bean	*mian4*	noodle	*lu4*	road

Each word set includes one critical word e.g., *chuang1* 'window' and four competitors similar to the design in Malins & Joanisse, (2010). The segmental competitor shared all phonemes with the critical word, but differed in tone e.g., *chuang2* 'bed'; the cohort competitor shared the initial consonant and tone, but differed in the rest of the syllable e.g., *che1* 'car'; the rhyme competitor shared the rime and tone, but differed in initial consonant, e.g., *guang1* 'light';[6] the tonal competitor shared the tone, but differed in all phonemes, e.g., *ji1* 'chicken'. Since the contribution of tonal information in word recognition is one of the main issue we would like to examine, the stimuli (critical word-segmental competitors) in the current study cover all 12 possible tone contrasts (see Table 5.1 leftmost column) to get rid of the perception variability between tone pairs. The pair T2 versus T3 is more confusable than other tone pairs to discriminate for native Mandarin speakers (Bent, 2005; Shuai & Malins, 2017; Zhang, Samuel, & Liu, 2012). For L2 learners of Mandarin, the pairs of T2 vs. T3 as well as T1 vs. T4 are also more confusable than other tone pairs (Hao, 2012; Wang, Jongman, & Sereno, 1999; Wang, Sereno, Jongman, & Hirsch, 2003).

The word frequencies, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010), do not differ significantly across critical words and competitors [$t(56) = 0.897$, $p = 0.38$]. The competitor conditions do not differ from each other, either [$F(3,44) = 1.62$, $p = 0.20$]. All stimuli were recorded by a female native Mandarin speaker from Beijing. The recording was made through a Sennheiser MKH416T microphone in the Leiden University Phonetics Lab (44.1 kHz, 16 bit). All the words used as stimuli are unambiguously picturable nouns. A black-white line drawing was selected for each word with the assistance of a native Mandarin speaker and a native Dutch speaker to make sure the pictures can appropriately depict their intended words.

5.2.3 Procedure

Participants were tested individually in a sound-attenuated room in the Leiden University Eye-tracking Lab. Before the eye-tracking recording, there was a familiarization session to ensure that participants were familiar with all the stimuli. In this session, they were first shown the stimulus list with pictures and words in pinyin. Then in a picture-naming task, they were presented with all the pictures in random sequence and were instructed to name each picture aloud. If the responded name was different from the word intended for the experiment, participants were given the intended name and were asked to say the name again.

In the subsequent eye-tracking session, participants performed an auditory-visual picture matching task and the eye movements were recorded using an Eyelink 1000 device with a 16mm lens with a 500-Hz sampling rate. A 24-inch BenQ XL2410T monitor was used to display visual stimuli. The participants were seated before the screen with a chin and forehead rest set. Their eyes were set at a distance of 69cm from the screen. Gaze position was calibrated with a 9-point grid prior to the test, and there was a drift check before each trial. On each trial, the participants were presented with a four-picture display. To ensure the non-overlapping of the parafoveal view in picture looking, the stimulus size and position were calculated and adjusted according to the size of the screen with a resolution of 1920 × 1080 pixels. The size of each picture was 5 × 5 cm, subtending a visual angle of approximately 6 degrees (Li, 2016; Miellet, O'Donnell, & Sereno, 2009). The four pictures in each experimental trial contained a target, a phonological competitor (one of the four competitor conditions: segmental, cohort, rhyme and tone), and two phonological-unrelated distractors (Figure 5.1). The

---

[6] We followed the conventional analysis of the Chinese syllable structure in treating the /u/ in *guang* and *gua* and /i/ in *xian*, *jian*, *mian* as part of the rhyme (e.g., Lee & Zee, 2003).

competitor in the baseline trial shares no phonological similarity with the target. That is, the targets were presented with the other three phonologically unrelated words in this condition. The position of the target and competitor was counterbalanced across trials to ensure the target occurred equally often in each cell position. The relative position of the target and competitor (adjacent or opposite) in each four-picture display was also counterbalanced across trials. At the beginning of each trial, a fixation cross ("+") appeared in the centre of the screen for 500 ms. Then, a four-picture display was presented on the screen. Concurrently, an auditory word was presented over a Beyer DT-770 Pro dynamic headphone at a constant and comfortable hearing level. Participants were instructed to first look at the fixation cross and then to select the word they heard by mouse clicking on the corresponding picture. After the participant's response, the next trial proceeded after a 1000-ms pause.

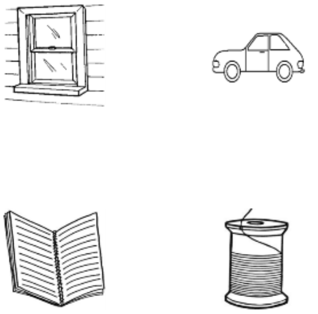

Figure 5.1. *Example of the visual stimulus display, including a target chuang1 'window', a cohort competitor che1 'car', and two phonological unrelated distractors ben3 'notebook' and xian4 'thread'.*

The eye-tracking task consisted of 6 blocks of 72 trials, i.e., 432 trials in total. Among these, 384 were experimental trials (96 trials × 4 repetitions) and 48 (12 trials × 4 repetitions) were baseline trials. The 96 experimental trials were built for each type of competitor. In half of the trials, the relationship between the target and its competitors was as listed in Table 5.1. The others were reciprocal trials, in which the relationship of the targets and competitors were inverted so that the probability of hearing targets or competitors on any given trial was the same. The order of the trials was randomized across participants, and in each block, the same target word did not appear more than three times. Six trials were presented before the main experiment as a warming-up.

5.2.4 Data analysis

In each trial, the visual display was divided according to the areas of the four items (i.e., target, competitor and distractors), and only looks within the corresponding areas were included in analyses. The proportion of looks to targets was calculated at each time point. The reciprocal trials were excluded form data analysis, leaving only the trials in which target words were heard. Trials in which participants failed to respond or selected items other than the target were also excluded from the analyses of eye movements (102 out of 9,840 trials). Moreover, the trials in which participants did not initially look at the fixation cross were also left out (78 out of 9,840 trials). Eye movement data was analyzed from 200 ms to 1400 ms after stimulus onset. The starting point was chosen since it takes about 200 ms to launch an eye movement (Altman & Kamide, 2004; Hallet, 1986). The upper limit was chosen to capture the point at which the proportion of looks for the target word reached maximum for the learner groups (Malins & Joanisse, 2010). The data recorded at 2-ms intervals was resampled so that the proportion of looks was calculated every 16 ms (62.5 Hz), comparable with the sampling rate used in Malins and Joanisse (2010). The plotting of the proportion of looks was also based on the data of 62.5 Hz.

The statistical analysis of the eye movement data curves was conducted with growth curve analysis (GCA) (Mirman, 2014; Mirman, Dixon, & Magnuson, 2008) with linear mixed modeling in R. Using GCA, the changing of the gaze distribution probability over time was captured and fitted using four-order orthogonal polynomials. The intercept term indicates the average overall fixation proportion; the linear term indicates a monotonic change in the general direction of the curve, while the quadratic, cubic and quartic terms tend to reflect the minor details of the steepness of the curve (Mirman, Dixon, & Magnuson, 2008; Malins & Joanisse, 2010). In the base linear mixed model, the only fixed effect was Time (up to the fourth-order polynomial), and the random effect included the by-Subject and by-Item (different tone pairs, level = 12) intercepts. Other fixed effects were added in a stepwise fashion. Only the effects that significantly improved the model fit were kept. Post-hoc comparisons were conducted using the *glht* function with Bonferroni adjustment in the Multcomp package in R (Hothorn, Bretz, & Westgall, 2008).

5.3 Results

5.3.1 Behavioral results

Mean accuracy for mouse clicking and reaction time (measured from the onset of the stimulus) are shown in Table 5.2. For native Mandarin speakers, A repeated measures ANOVA showed that the main effect of condition was not significant for both accuracy and RT (both p values > 0.05). A significant main effect of condition on accuracy was found for beginning learners [$F(4, 40) = 16.18$, $p < 0.001$]. They were significantly less accurate in the segmental condition than in the other conditions (all p values < 0.05 in post-hoc analysis). A significant main effect of condition on RT was also found for beginners [$F(4, 40) = 12.04$, $p < 0.001$]. They were significantly slower in the segmental condition compared to the other conditions (all p values < 0.05). The advanced learners were also less accurate in the segmental condition, which was reflected by a significant main effect of condition [$F(4, 56) = 18.11$, $p < 0.001$]. Post-hoc analysis showed significant differences between the segmental condition and the other conditions (all p values < 0.001). A significant main effect of condition on RT was also found for advanced learners [$F(4, 56) = 30.43$, $p < 0.001$], and the post-hoc comparison showed that the RT in the segmental condition was significantly longer than all the other conditions (all p values < 0.01).

Table 5.2. *Mean accuracy and reaction time for picture identification for three groups of participants. Standard deviations are shown in parentheses.*

Condition	Accuracy (%)			RT (ms)		
	NM	BL	AL	NM	BL	AL
Baseline	100.0 (0.0)	97.9 (4.6)	100.0 (0.0)	1265.7 (107.0)	2189.4 (659.7)	1603.9 (317.3)
Cohort	100.0 (0.0)	95.7 (8.0)	99.7 (0.6)	1292.7 (134.8)	2213.4 (626.2)	1565.5 (241.3)
Rhyme	99.7 (0.8)	96.1 (5.1)	99.2 (0.8)	1367.0 (141.2)	2330.1 (619.2)	1680.0 (296.1)
Segmental	99.6 (0.9)	83.0 (15.2)	94.0 (5.2)	1358.0 (139.4)	2486.6 (676.2)	1874.5 (311.0)
Tonal	99.9 (0.3)	97.9 (4.5)	100.0 (0.0)	1289.5 (114.6)	2088.8 (574.1)	1608.6 (277.2)

5.3.2 Fixation analysis for different participant groups

For the fixation data of three groups of participants in four conditions, linear mixed models were built for both target looks and competitor looks. Both models had Time (up to the fourth-order component), Group (native Mandarin speakers, beginning learners, advanced learners), Competitor Condition (baseline, segmental, cohort, rhyme, tonal) and their interactions as fixed effects. By-subject and by-item intercepts were entered as random effects. The statistical results of the fixed effects are presented in Table 5.3.

Table 5.3. *Summary of fixed effects for the models of Looks to target and Looks to competitor.*

Fixed effects	Looks to target			Looks to competitor		
	df	χ^2	p	df	χ^2	p
Linear component	1	46869.0	<.001	1	122.0	<.001
Quadratic component	1	3974.8	<.001	1	6377.2	<.001
Cubic component	1	3541.8	<.001	1	4201.5	<.001
Quartic component	1	586.2	<.001	1	36.0	<.001
Group	2	51.8	<.001	2	33.7	<.001
Condition	4	272.3	<.001	4	1778.6	<.001
Group: Condition	8	211.8	<.001	8	512.4	<.001
Group: Time components	8	13042.0	<.001	8	2224.8	<.001
Conditon: Time components	16	244.7	<.001	16	1735.5	<.001
Group: Condition: Time components	32	257.6	<.001	32	543.1	<.001

For the native Mandarin speakers, the proportion of looks to target as a function of time gradually ramped up and reached the maximum of approximate 90% around 1100 ms in all conditions, exhibiting a sigmoidal curve. The proportion of looks to target in the segmental condition (Figure 5.2*a*) showed a slightly delayed raising pattern and reached the maximum later than that in the baseline condition, but this difference was subtle and only the quadratic component reached significance (Est. = 0.10, $z = 3.58$, $p < 0.05$). The proportion of looks to the competitor in the segmental condition exhibited a higher peak compared to the baseline condition, as suggested by significant difference in intercept (Est. = 0.01; $z = 6.85$, $p < 0.001$). The delay in target looks and the high proportion of looks to the competitor in the segmental condition indicated that segmental overlaps competed for lexical activation. Figure 5.2*b* compares the cohort and the baseline conditions. The looks to target in the cohort condition differed significantly from the baseline condition only in the cubic component (Est. = −0.16, $z = -5.65$, $p < 0.001$). The looks to competitor, however, did not differ significantly between cohort and baseline conditions, indicating that the participants did not launch more looks to cohort competitors than to the baseline condition.

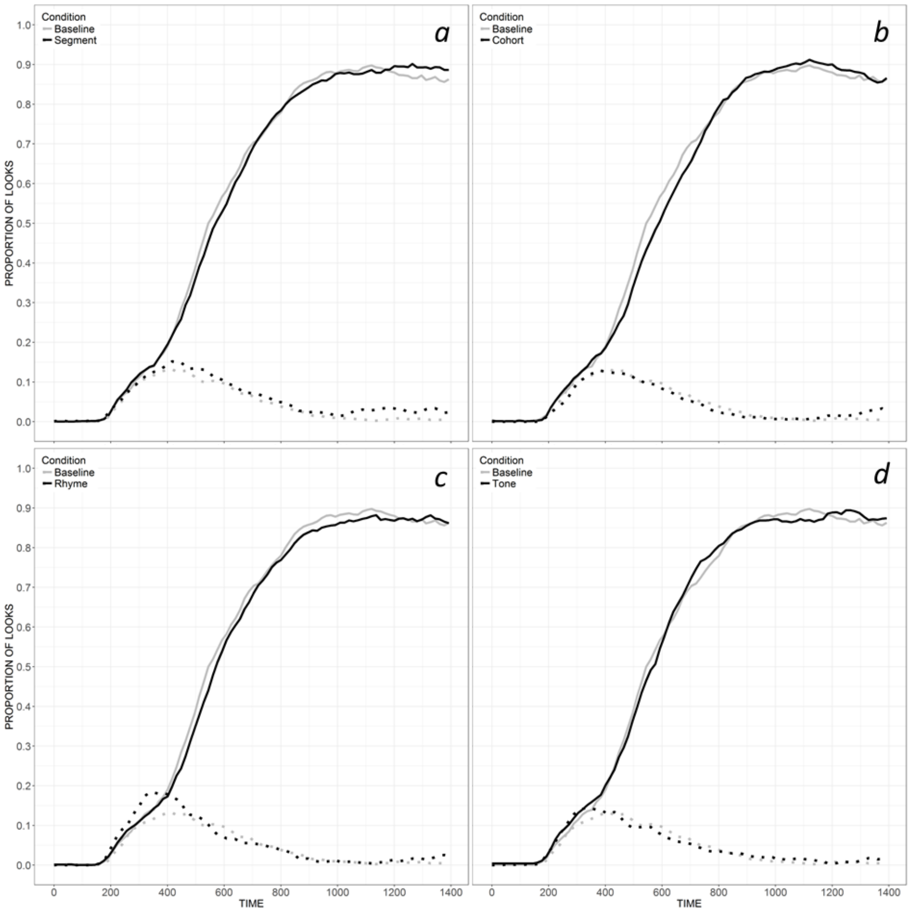

Figure 5.2. *Mean proportion of looks to the target (solid line) and the competitor (dotted line) in different conditions averaged across participants and items for native Mandarin speakers. The segmental, cohort, rhyme and tonal conditions were plotted against the baseline condition for comparison.*

As shown in Figure 5.2*c*, for native Mandarin speakers, the rhyme condition showed a lower proportion of looks to the target than the baseline condition during the entire processing time, as indexed by significant differences in the mean height (Est. = −0.02, $z = -5.22$, $p < 0.001$). The greater looks to rhyme competitor compared to baseline in the initial part also complemented the pattern found in the target looks. There was an early peak at ca. 350 ms for looks to competitor, while the rate of looks to the rhyme competitor was even higher than that to target in the time window between 200 and 500 ms, indicating a strong competition effect. Significant differences between competitor looks in rhyme and baseline conditions were found in overall height (Est. = 0.01, $z = 5.16$, $p < 0.001$), the linear component (Est. = −0.08, $z = -4.59$, $p < 0.001$) and the quadratic component (Est. = 0.12, $z = 7.16$, $p < 0.001$). In the tonal condition, looks to target were not significantly different from baseline in any aspect. Looks to competitor only differed significantly from baseline in the quadratic (Est. = 0.07, $z = 4.03$, $p < 0.01$) component, which indicated that competitors only sharing tonal information cannot be activated effectively to compete with the target during word recognition.

It should be noted that both rhyme and segmental competitors received higher proportion of looks than the baseline condition, indicating a competition effect from both of them. Compared to the rhyme competitor, the fixation curve of the segmental competitor exhibited a significantly lower and delayed peak (linear component: Est. = 0.12, $z = 6.70$, $p < 0.001$; quadratic term: Est. = −0.11, $z = -6.55$, $p < 0.001$), indicating that the mismatch in tone was used to constrain the activation of the incompatible lexical candidate.

Figure 5.3*a* shows that for beginning learners, the proportion of looks to target in the segmental condition diverged from baseline from 700 ms, presenting a lower increasing rate than baseline, which was reflected by a marginally significantly different intercept (Est. = −0.01, $z = -3.18$, $p = 0.07$) between these two conditions. There is also a greater rate of looks to competitor in the segmental condition than the baseline condition, which is confirmed by a significantly different overall height (Est. = 0.04, $z = 20.35$, $p < 0.001$) and quadratic component (Est. = −0.14, $z = -7.49$, $p < 0.001$). The proportion of looks to segmental competitor remained high (about 20%) during the whole trial, which reflected the beginners' difficulty in distinguishing minimal tone pairs. Figure 5.3*b* shows that the proportion of looks to target increased faster than baseline before 800 ms. After 800 ms, the increase became slower and the target looks in the cohort condition reached its maximum later than the baseline condition. A significantly different cubic component (Est. = 0.16, $z = 5.15$, $p < 0.001$) was found between the baseline and the cohort conditions. For the proportion looks to competitor, the beginning learners did not launch more looks to competitor in the cohort condition than in the baseline condition. There were even fewer looks to cohort competitor between 400 and 800 ms compared to the baseline condition. The post-hoc comparison showed no significant difference in competitor looks in any aspect between these two conditions. As shown in Figure 5.3*c*, there is clearly smaller proportion of looks to the target in the rhyme condition than in the baseline condition, which is supported by significant overall height (Est. = −0.01, $z = -4.05$, $p < 0.01$), quadratic component (Est. = 0.11, $z = 3.51$, $p < 0.05$), and cubic component (Est. = 0.24, $z = 7.73$, $p < 0.001$). There was also a greater proportion of looks to competitor in the rhyme condition than baseline, confirmed by a significant overall height (Est. = 0.02, $z = 9.55$, $p < 0.001$), linear component (Est. = −0.34, $z = -17.97$, $p < 0.001$) and quartic component (Est. = −0.11, $z = -5.95$, $p < 0.001$) between these conditions. The rate of looks to the competitor was even higher than to the target between 200 and 500 ms, indicating a strong competition from the rhyme competitor. The proportion of looks to target in the tonal condition showed an early and lower peak compared to the baseline curve (Figure 5.3*d*), as reflected by a significant difference in the linear component (Est. = −0.11, $z = -3.51$, $p < 0.05$), the quadratic component (Est. = −0.11, $z = -3.65$, $p < 0.05$), and the cubic component (Est. = 0.19, $z = 6.18$, $p < 0.001$). The general pattern of proportion of looks to the competitor was similar between the tonal and the baseline conditions, with significant differences only in the linear component (Est. = −0.09, $z = -4.74$, $p < 0.001$) and the cubic component (Est. = −0.07, $z = -3.53$, $p < 0.05$).

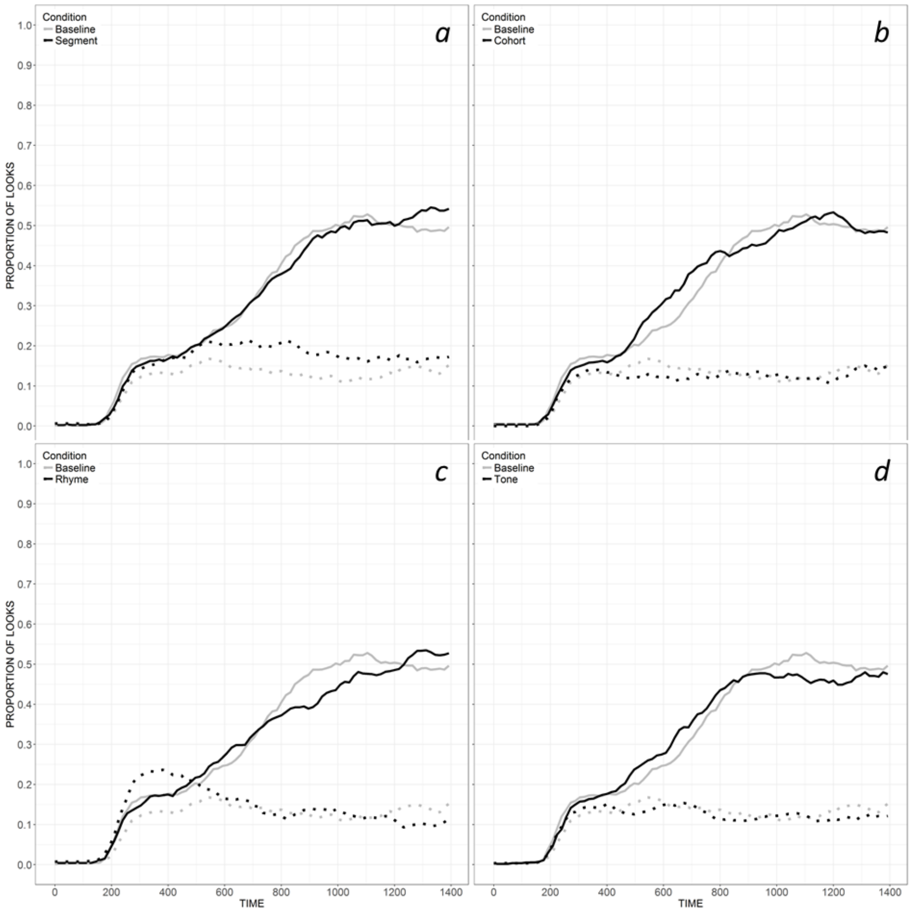

Figure 5.3. *Mean proportion of looks to the target (solid line) and the competitor (dotted line) in different conditions averaged across participants and items for beginning Dutch learners of Mandarin. The segmental, cohort, rhyme and tonal conditions were plotted against the baseline condition for comparison.*

Figure 5.4*a* shows that for advanced learners, there was a less proportion of looks to the target in the segmental condition than baseline for the whole trial, as suggested by significant difference in the overall height (Est. = −0.05, z = −17.93, p < 0.001), the linear component (Est. = −0.15, z = −5.59, p < 0.001) and the quadratic component (Est. = 0.16, z = 6.06, p < 0.001). The fixation curve of the segmental competitor was higher than baseline for the whole trial, complementing the pattern found in the target curve.

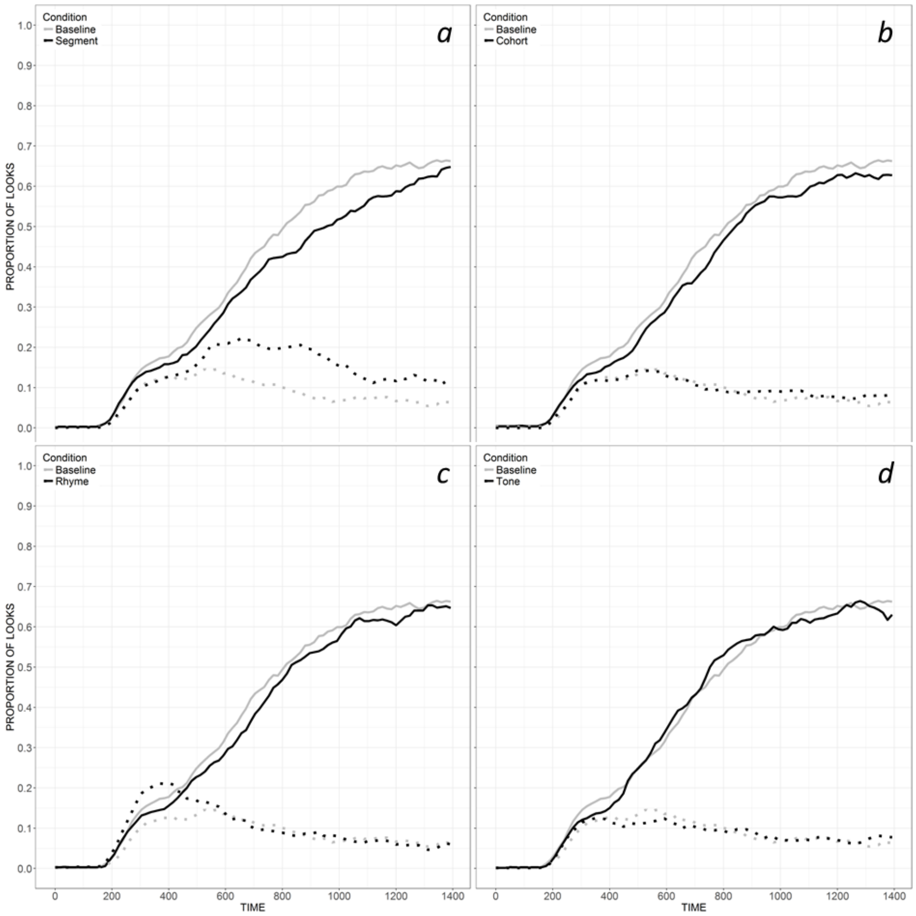

Figure 5.4. *Mean proportion of looks to the target (solid line) and the competitor (dotted line) in different conditions averaged across participants and items for advanced Dutch learners of Mandarin. The segmental, cohort, rhyme and tonal condition was plotted against the baseline condition for comparison.*

Significantly different overall height (Est. = 0.06, $z = 31.59$, $p < 0.001$), linear component (Est. = 0.13, $z = 8.10$, $p < 0.001$), quadratic component (Est. = −0.25, $z = -15.15$, $p < 0.001$) and quartic component (Est. = 0.15, $z = 9.03$, $p < 0.001$) were found in the competitor curves between these two conditions. The lower rate of looks to target and higher looks to competitor combined to indicate a strong competition effect for segmental competitor. Compared to native Mandarin speakers, tonal information cannot be used effectively by advanced learners to inhibit the prosodically incompatible candidate. As shown in Figure 5.4*b*, there was a smaller proportion of looks to target in the cohort condition than baseline during the whole trial, reflected in significantly different overall height (Est. = −0.03, $z = -9.71$, $p < 0.001$). The curve for the cohort competitor, however, did not differ significantly in any aspect compared to baseline. Figure 5.4*c* presents that for the whole trial, the proportion of looks to the target in the rhyme condition was smaller than that in the baseline condition, which was confirmed by significant differences in the overall height (Est. = −0.02, $z = -8.09$, $p < 0.001$) and the quadratic component (Est. = 0.09, $z = 3.58$, $p < 0.05$). The proportion of looks to the competitor in the rhyme condition was greater than baseline, showing a peak between 300 and 400 ms. This was confirmed by a significant difference in the overall height (Est. = 0.02, $z = 8.36$, $p < 0.001$), linear component (Est. = −0.21, $z = -12.96$, $p < 0.001$), quadratic component (Est. = 0.11, $z = 6.56$, $p < 0.001$) and quartic component (Est. = −0.09, $z = -5.33$, $p < 0.001$). The smaller proportion of looks to the target and greater proportion of looks to the competitor indicated a strong competition effect exerted by the rhyme competitor. For the tonal condition (Figure 5.4*d*), the proportion of looks to target showed a faster increase than baseline before 950 ms, which is reflected by a significantly different quadratic component (Est. = −0.10, $z = -3.69$, $p < 0.05$). The curve of the tonal competitor was not significantly different from baseline in any aspect. The greater proportion of looks to target in the tonal condition than baseline and the lack of difference in competitor looks combined to indicate that for advanced learners, the competitor with only tonal overlap cannot be activated and compete directly with the target during word recognition.

Moreover, compared to the rhyme competitor, the looks to the segmental competitor showed a significantly lower and delayed peak intercept (Est. = 0.04, $z = 23.21$, $p < 0.001$), linear component (Est. = 0.35, $z = 21.04$, $p < 0.001$) and quadratic component (Est. = −0.36, $z = -21.68$, $p < 0.001$), which suggested that the mismatch in tonal information slowed down the activation of the segmental competitor.

5.3.3 Comparison of fixation results across participant groups

Figure 5.5 illustrates the developmental path in proportion of looks to target in different conditions.

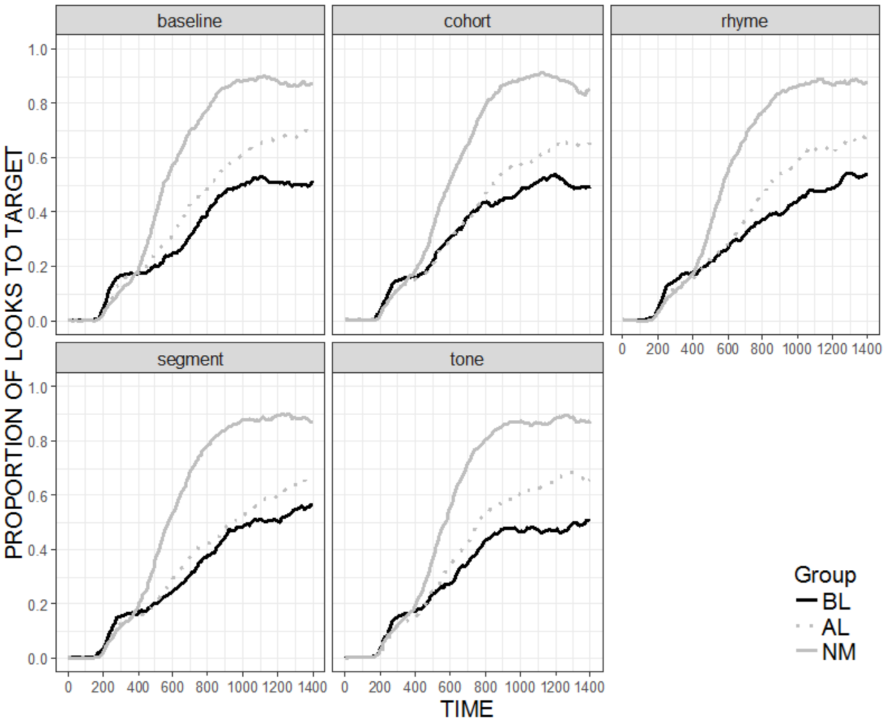

Figure 5.5. *Mean proportion of looks to target in different conditions. The data from three participant groups are plotted together for comparison.*

The proportion of looks to target in the native Mandarin speakers was significantly greater than for the two learner groups in all conditions, as suggested by significant differences in overall height, linear component and quadratic component (all p values < 0.01). Within the two learner groups, the advanced learners obtained a higher proportion of looks to target than the beginners in all conditions. In the baseline condition, the target looks of the advanced learners kept rising during the whole trial, while for the beginners, the proportion of looks to target reached a plateau of 50% at approximately 1100 ms. This discrepancy between the two learner groups was also suggested by significant differences in the intercept, linear, quadratic and cubic components (all p values < 0.01). In the cohort condition, the proportion of looks to target increased more rapidly for the advanced learners than beginning learners, indexed by significant differences in linear component (Est. = −0.53, $z = -18.56$, $p < 0.001$) and cubic component (Est. = 0.09, $z = 3.16$, $p < 0.05$) between the two groups. The advanced learners also showed a faster increase than beginners in the proportion of looks to target in the rhyme condition, as suggested by a significant difference in the linear component (Est. = −0.57, $z = -20.05$, $p < 0.001$), quadratic component (Est. = 0.11, $z = 3.97$, $p < 0.01$) and cubic component (Est. = 0.14, $z = 4.89$, $p < 0.001$). In the segmental condition, the proportion of looks to target only differed significantly between the learner groups in the linear component (Est. = −0.33, $z = -11.63$, $p < 0.001$). For tonal condition, the proportion of looks to target for advanced learners was larger than that of the beginners and showed a continuously rising tendency during the trial. This difference was confirmed by significantly different overall height (Est. = −0.09, $z = -3.19$, $p < 0.05$), linear component (Est. = −0.68, $z = -23.54$, $p < 0.001$) and quadratic component (Est. = 0.09, $z = 3.02$, $p < 0.05$). Taken together, the advanced learners exhibited a larger proportion looks to target than beginning learners in almost all conditions, and approximated the performance of native Mandarin speakers.

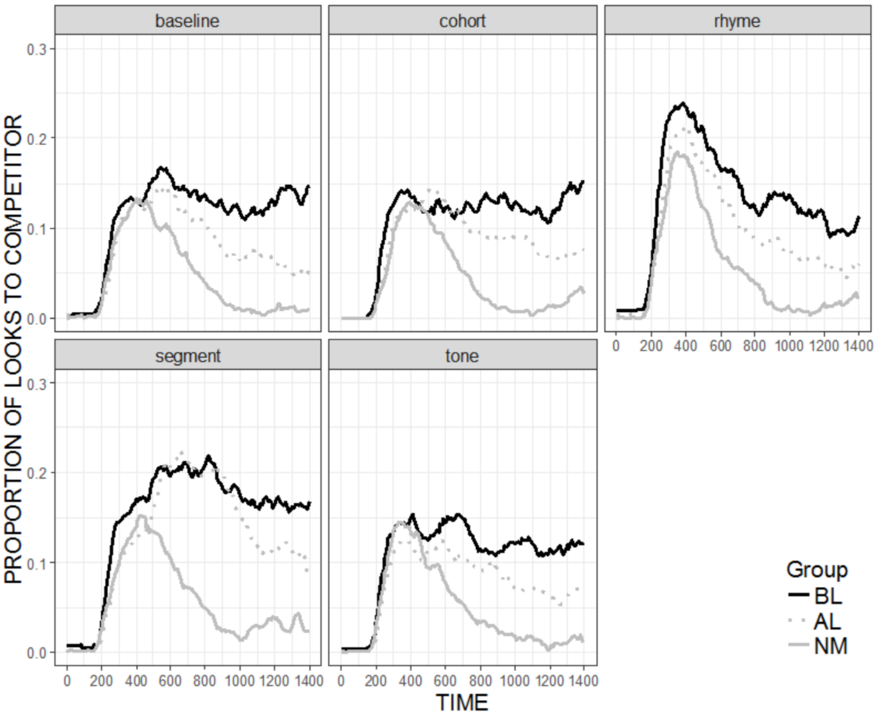

Figure 5.6. *Mean proportion of looks to competitor in different conditions. The data from three participant groups are plotted together for comparison.*

For native Mandarin speakers (Figure 5.6), the competitor look reached its maximum at ca. 400 ms, and then declined rapidly as the auditory stimulus unfolded in all conditions. The higher and early peak in the rhyme condition indicated that the competitor with rhyme and tone overlaps was fully activated to compete for recognition. The segmental competitor shared rhyme with the target but differed in tone. Compared to the rhyme competitor, the curve of the segmental competitor exhibited a lower and later peak. For the two learner groups, there was also an early peak for the proportion of looks to competitor, but the competitor looks did not decline rapidly. The post-hoc comparison showed significant differences between native speakers and the learner groups in the overall height and linear component in all conditions (all p values < 0.05). Within the two learner groups, the proportion of looks to competitor for beginners was greater than the advanced learners in the baseline condition (intercept: Est. = 0.04, $z = 3.08$, $p < 0.05$; linear component: Est. = 0.17, $z = 9.86$, $p < 0.001$; quadratic component: Est. = 0.06, $z = 3.16$, $p < 0.05$). The competitor looks between the two learner groups also differed significantly in the cohort condition (linear component: Est. = 0.15, $z = 8.52$, $p < 0.001$; quadratic component: Est. = 0.07, $z = 4.24$, $p < 0.001$; cubic component: Est. = −0.05, $z = -3.03$, $p < 0.05$) and rhyme condition (intercept: Est. = 0.04, $z = 3.50$, $p < 0.01$). In the segmental condition, the proportion of looks to competitor was high for both learner groups (about 20%), indicating a great difficulty in discriminating minimal tone pairs. Significant differences were found between the learner groups in the linear (Est. = 0.09, $z = 5.38$, $p < 0.001$) and quadratic components (Est. = 0.16, $z = 9.11$, $p < 0.001$). In the tonal condition, the proportion of looks to competitor decreased more rapidly for advanced learners than for the beginners, which was reflected by a significant difference in the linear component (Est. = 0.08, $z = 4.27$, $p < 0.001$).

5.4 Discussion and conclusion

In this study, eye-tracking with the Visual World Paradigm was used to investigate the competition effect of segmental versus tonal information and their role in constraining lexical access by native Mandarin speakers as well as for Dutch learners of Mandarin.

The results show that for native Mandarin speakers, rhyme competitors competed for lexical activation, as was reflected by a lower proportion of looks and slower eye movements to targets, as well as a greater proportion of looks to competitors compared to baseline condition. The cohort competitors, however, did not show strong competition effects. Native Mandarin speakers did not launch more looks to cohort competitors than in the baseline condition, suggesting that the cohort competitor was not activated enough to compete for recognition. This result is not in line with Malins and Joanisse (2010), who found a strong competition effect for cohort competitors. Such divergent result may be caused by different experimental procedures. In Malins and Joanisse's study, the visual stimulus was presented 1500 ms before the auditory stimulus so that participants could preview the four pictures during that time. As a result, the names of the pictures were already encoded during the preview and were ready to be mapped when the word was made audible. In that case, participants can be very sensitive to the onset of the speech input, which explains the strong competition effect of the cohort competitor in their study. As suggested by Tanenhaus et al. (2000), in a task with long preview time, participants may encode the scene and hold it in working memory for later matching with the auditory input and therefore "bypass" normal lexical processing. So, in our experiment, the visual and auditory information was presented concurrently to directly test real-time lexical processing. Since the duration of the initial consonant was short, when processing the pictures, the participants heard the rhyme part in the meantime. Given that the rhyme competitor was matched with the target in both rhyme and tone, it was activated adequately for competition. This finding is also consistent with the prediction of continuous evaluation of the unfolding speech by the TRACE model. The proportion of looks to the competitor in the segmental condition exhibited a higher peak compared to the baseline condition, indicating that the segmental competitor was activated to a small extent. In the tonal condition, native Mandarin speakers did not launch more looks to the competitor than baseline, indicating that the tonal competitors were not sufficiently similar to the target word, and were not activated enough to compete for recognition. Among conditions, the curve of segmental competitor exhibited a lower and delayed peak compared to the curve of rhyme competitor, indicating the non-matching tonal information constrained the lexical activation of segmental competitors. The rhyme competitor was activated adequately in that it shared both rhyme and tonal information with the target. The segmental competitor, which only shared the rhyme part with the target but differed in tone, was less activated. This result indicates that for native Mandarin speakers, tones play an early constraining role. As the input unfolded, only those candidates matching the input both in segmental and tonal information (i.e., the rhyme competitor) were activated. Segmental competitors with non-matching tonal information were not activated in the early stage.

Compared to the native Mandarin speakers, both groups of learners generally showed less target activation and increased competitor activation across all conditions, which was reflected in lower peaks and much slower increasing rates of fixation to the targets, as well as higher peaks in the curves of fixation to the competitors and more fixation to the competitors toward the end of the time course.

Among conditions, the rhyme competitor, which shared both rhyme and tonal information with the target, was activated most to compete with the target for both learner groups. The segmental competitor, which shared the rhyme part with the target but differed in tone, also received a high proportion of looks, but activated less and later than the rhyme competitor. This suggests that, similar to native speakers, tonal information was exploited during word recognition by both beginners and advanced learners of Mandarin. The non-matching tonal information was used effectively to constrain lexical activation in an early stage. For both learner groups, there was a high plateau in the curve of the segmental competitor in the later part of the trial. This indicated although the learners could use tonal information in the early stage for constraining lexical access, they still experienced great difficulty in deciding between minimal tone pairs. They were less confident with their choice between the target and the segmental competitor and launched more looks to the segmental competitor during the whole trial compared to the other conditions. In the cohort and tonal conditions, the eye movements to target were slower compared to the baseline condition, but the proportion of looks to competitor was not significantly greater than baseline, suggesting that the cohort and tonal competitors were not activated enough for lexical competition for either group of learners.

Moreover, a clear developmental path was observed by comparing the performance of the beginning and the advanced learners. The advanced learners generally showed a higher proportion of looks to the target and fewer looks to the competitor in all conditions, suggesting that they could activate the correct targets and suppress the competitors to a greater extent.

In conclusion, eye-tracking result suggests that native Mandarin speakers use tonal information effectively to constrain lexical activation in an early stage, similar to the way they exploit rhyme information. Beginning and advanced Dutch learners of Mandarin generally showed fewer fixations to the target and more looks to the competitor in the baseline, cohort, rhyme, segmental and tonal conditions, indicating the target was not fully activated and the competitor was not fully suppressed compared to native speakers. Similar to native Mandarin speakers, both rhyme and tonal information was used by Dutch learners of Mandarin in an early stage to inhibit the activation of incompatible lexical candidates. Compared to beginning learners, significant improvement has been found for advanced learners into the native direction.

Chapter six

Conclusion

This dissertation presents results of four experiments which investigated various aspects of the production and perception of lexical tones by beginning and advanced Dutch learners of Mandarin as well as the developmental trajectory in their L2 tone acquisition.

6.1 Recapitulation of research questions

In connected speech, tones can be influenced by the preceding and following tones and show coarticulated f0 realizations which are different from the canonical contours when produced in isolation. Such deviant f0 shapes of coarticulated tones also make it a great challenge for adult non-tonal L2 learners to achieve native-like tone production. Previous research on tonal coarticulation has mainly focused on the directionality (carryover or anticipatory), the nature (assimilatory or dissimilatory), and the magnitude of contextual effects on tonal production by native speakers. However, no systematic empirical research has been devoted to the underlying mechanisms of tonal coarticulation and the acquisition of coarticulated patterns by L2 learners of Mandarin. Whether fine-grained phonetic details such as tonal coarticulation patterns can be acquired by L2 learners is still not clear. So, Chapter 2 in the current dissertation examined disyllabic tonal coarticulation in native Mandarin speakers as well as beginning and advanced L2 learners in a reading-aloud task under cognitive load, based on two research questions:

(1) What are the underlying mechanisms of tonal coarticulation for native speakers and L2 learners?
(2) What is the developmental trajectory in the acquisition of fine-grained phonetic details in tonal coarticulation for L2 learners?

In terms of tone perception, previous studies have demonstrated that non-tonal L2 learners can be trained to improve their ability in tonal identification and discrimination, but how they learn to use pitch information in lexical contrastive way and form new tonal categories is still less understood. Therefore, Chapter 3 in this thesis tested beginning and advanced Dutch learners of Mandarin as well as native Mandarin and native Dutch speakers in an ABX matching-to-sample test, aiming to reveal whether L2 learners can redistribute their attention to segments and tones and develop a more integral processing of these two kinds of information like native Mandarin speakers, as well as the developmental trajectory in these processes during the period of Mandarin acquisition.

Previous findings generally show a good performance of L2 learners of Mandarin in basic identification and discrimination tasks, in which phonetic mode of processing can be employed to focus on phonetic details. However, in cognitively demanding tasks more like real communication situations, listeners are not likely to rely on fine-grained phonetic details and have to use "automatic selective perception routines" (Strange, 2011) to detect phonologically contrastive tonal information. The performance of L2 learners in demanding tasks requiring phonological mode of processing has not been tested before. To our knowledge, another crucial issue in L2 tonal acquisition, learning to use tonal information in lexical access, has also not been systematically investigated. Therefore, Chapter 4 employed a cognitively demanding sequence-recall task and a tonal lexical decision task to answer two additional questions:

(3) To what extent can advanced Dutch learners of Mandarin achieve a native-like pattern in phonological processing of tonal contrasts?
(4) Can lexical tones be used properly in lexical activation by Dutch learners of Mandarin?

The developmental trajectory observed in Chapter 4 suggests that for Dutch learners of Mandarin, tonal information can be exploited in lexical access. However, the time course of lexical activation and the relative role of segmental and tonal information in speech recognition in L2 learners remained an interesting issue to explore. Chapter 5 therefore employed an eye-tracking experiment with the Visual World Paradigm to investigate the relative role of segmental and tonal information in lexical activation by native Mandarin speakers and Dutch learners of Mandarin, as well as the developmental trajectory for Dutch learners of Mandarin in using segmental and tonal information in spoken word recognition.

6.2 Results of individual chapters

Chapter 2 examined disyllabic tonal coarticulation in native Mandarin speakers and L2 learners. For native Mandarin speakers, the result suggested that tonal coarticulation is bidirectional. The carryover effect exerted by the offset of the initial tones exhibited an assimilatory nature: this effect was strong and the influence could still be seen at least two-thirds into the vowel. The anticipatory effect exerted by the following tones on the tones in the initial position was dissimilatory in nature and had a smaller magnitude compared to the carryover effect. For T1, T2 and T4, the dissimilatory effect had been found when followed by most pairs of tones with contrastive onsets. It should be noted that, different from Xu's (1997) finding, the raising effect of the following low onsets was not constrained to the maximum f0. The whole contours of T1 and T2 were raised by the following tones with low onsets (T2 and T3) in the current experiment. The underlying coarticulatory mechanism was investigated in the high-cognitive-load condition. The carryover effect was robust and remained intact under high cognitive load, indicating that the carryover effect does not involve advance planning and is likely to be caused by physiological constraints – which is in line with the findings from segmental coarticulation (Whalen, 1990). The anticipatory effect, on the other hand, decreased with high cognitive load, indicating that anticipatory coarticulation involves advance planning. This finding can be potentially accounted for by the model proposed by

Tilsen (2013, 2009), who argued that an inhibitory speech planning mechanism is used for contemporaneously planned articulatory targets to maintain and maximize the contrasts of different phonemes. In the no-cognitive-load condition of the present study, the inhibitory mechanism functioned well and led to a clear dissimilatory anticipatory effect, maximizing the contrast between the adjacent tones. In the high-cognitive-load condition, however, as the inhibitory mechanism was constrained, the dissimilatory anticipatory effect also decreased.

For L2 learners, a clear developmental path was found in both carryover and anticipatory effects. The beginning learners showed a weaker assimilatory carryover effect compared to native speakers, while the performance of advanced learners was more similar to the native patterns, showing substantial carryover effects in T1, T2 and T4. That is, although the carryover effect did not include advance planning and was mainly a result of physiological constraints, its acquisition was still a gradual process.

Different from the carryover effect, the anticipatory coarticulation was strong for Dutch learners of Mandarin. For beginning learners, the dissimilatory anticipatory effect had been found on T1, T2 and T4 in the first syllable. This effect remained robust for T1 and T2 with high cognitive load. The advanced learners showed an anticipatory effect comparable to that of native Mandarin speakers in the normal condition. Different from the native speakers, the anticipatory effect remained robust for all tones under the influence of high cognitive load for advanced learners.

The strong anticipatory effect in L2 tonal coarticulation can be potentially explained with the help of Tilsen's model (Tilsen, 2013). The inhibitory mechanism had been acquired by beginning and advanced learners as an effective way to maintain the contrast and ensure the perceptibility of different tonal categories in connected speech.

Chapter 3 investigated how beginning and advanced Dutch learners of Mandarin process tonal information. An ABX task was adopted to investigate phonological discrimination of Mandarin tones and segment-tone integration in Dutch learners of Mandarin, with both native Mandarin and Dutch speakers (without tonal learning experience) as control groups. Results showed a developmental path in lexical tone processing. The beginning learners could not process tonal contrast adequately at the phonological level, and they processed segmental and tonal information separately, like native Dutch listeners without Mandarin experience. The advanced learners showed a good phonological discrimination of tonal contrasts. They showed a more native-like pattern in distributing their attention between segmental and tonal information, and they processed the two dimensions in an integrated manner, similar to native Mandarin listeners. This suggests that the acquisition of new tonal categories in L2 involves a redistribution of attention along perceptual dimensions and the development of segment-tone integration.

Chapter 4 further investigated the phonological processing of all tonal contrasts and the use of tones in lexical access by Dutch learners of Mandarin using a cognitively demanding sequence recall task and a lexical decision task. The results showed a clear developmental path by comparing the performance of two learner groups in the sequence recall task: the advanced learners, with more Mandarin experience, exhibited a significant improvement compared to the beginners, approximating the performance of native Mandarins. The result of the lexical decision task also showed that, compared to beginners, advanced learners performed significantly better in correctly identifying real words and rejecting non-words which were minimally different from real words in tones. The improvement of advanced learners in both tasks demonstrated that they were shaping new selective perception routines, and their phonological mode of tone processing was in development which is in line with the ASP model (Strange, 2011). However, it should be noted that the RTs were longer for advanced learners than for native Mandarin listeners in all conditions, indicating that their L2 selective processing routines were still in development and were not as automatic as those of native Mandarin listeners.

However, in the sequence recall task, T2 and T3 was the most confusable tone pair for both learner groups. In the lexical decision task, T2 and T3 was mutually confusable for learners and was resistant to improve. Such difficulty may have stemmed from the acoustic similarity between these two tones, in that they have similar pitch contours and there is overlap in their pitch ranges (Moore & Jongman, 1997). Moreover, some asymmetric patterns were found for advanced learners in the lexical decision task. The advanced learners were significantly more accurate in recognizing real words with T1 and rejecting non-words in which T1 was produced as the other three tones than the other way round. That is, the category of T1 had been relatively well established when compared to the other three tones. Contrastively, it was difficult for advanced learners to make a correct response when T4 was produced as T1, T2 or T3, that is, the category of T4 was relatively less well-established when compared to the other three tones in pairs. These results are potentially related to the prosodic features of the learners' native language, since Gandour (1983) showed that compared to tone-language speakers, intonational language speakers are more sensitive to pitch height than to pitch direction.

Taken together, the advanced learners showed a significant improvement in tonal processing at phonological level and using lexical tones in lexical access compared to beginning learners. This improvement in tone acquisition can be attributed to the important role of tones in Mandarin. Different tone pairs were not equally difficult to learners in lexical access, and the source of such differences can be attributed to acoustic similarity between particular tones as well as interference from L1 suprasegmental features.

Chapter 5 investigates the time course of lexical activation and the relative role of segmental and tonal information in speech recognition by testing native Mandarin speakers, as well as the beginning and advanced learners of Mandarin in an eye-tracking experiment. Using visual word paradigm, the participants heard an auditory word and were asked to identify the corresponding picture from a display of four pictures that consisted of the target (*chuang1* 'window'), a phonological competitor (segmental: *chuang2* 'bed'; cohort: *che1* 'car'; rhyme: *guang1* 'light'; tonal: *ji1* 'chicken'), and two phonologically unrelated distractors. The probability of fixation to the target and competitor was recorded since it may reflect the activation of the corresponding items. In this task, the auditory and visual stimuli were presented concurrently to participants, so when processing the pictures, the participants heard the rhyme part at the mean time. For native speakers, slower eye movements to the target were found in the rhyme condition, accompanied by high proportion of looks to the rhyme competitor, indicating that the rhyme competitor was activated adequately for competition. A lower and delayed peak in proportion of looks to the competitor in the segmental condition compared to the rhyme condition indicated that for native Mandarin speakers, tones play an early constraining role. As the input unfolded, only those candidates matching the input both in segmental and tonal information (i.e., the rhyme competitor) were activated. Segmental competitors with non-matching tonal information were not activated in the early stage.

Compared to the native Mandarin speakers, both groups of learners generally showed less target activation and increased competitor activation across all conditions. Among conditions, the rhyme competitor was activated most to compete with the target for both learner groups. The segmental competitor also received a high proportion of looks, but activated less and later than the rhyme competitor. This suggested that, similar to native speakers, the non-matching tonal information was used effectively to constrain lexical activation in an early stage by both learner groups. However, more late fixations to the segmental competitor indicated that both learner groups were confronted with great difficulty in decimating tonal minimal pairs. The cohort and tonal competitors, on the other hand, were not activated enough for lexical competition for both groups of learners.

Furthermore, the advanced learners generally showed a higher proportion of looks to the target and less looks to the competitor in all conditions than the beginning learners, suggesting that the advanced learners could activate the correct targets and suppress the competitors to a greater extent, approaching the performance of native Mandarin speakers.

6.3 General conclusion

In sum, with regard to L2 tone production, we found that the fine-grained patterns in carryover and anticipatory tonal coarticulation can be acquired by Dutch learners of Mandarin. Compared to native Mandarin speakers, advanced learners showed a more robust anticipatory effect. This dissimilatory effect was quite probably adopted by them as a strategy to prevent tonal targets from becoming perceptually indistinct. This finding is in line with Tilsen's model (2013): the dissimilatory effect between contemporaneously planned articulatory targets may function as a general mechanism to maintain perceptual differences between sounds.

In terms of tone perception, we found that, first, advanced learners improved significantly compared to beginning learners and approximated the performance of native Mandarin speakers in discriminating tones and redistributing their attention between segments and tones. These results suggest that Dutch learners of Mandarin can learn to distinguish tonal contrasts, which provides general support to the SLM model. That is, the capacities of forming new sound categories remain intact for adult L2 learners. The PAM-L2 model can also be used to account for L2 tonal acquisition. The case of non-native tonal contrasts fits the Uncategorized-uncategorized scenario best, with both sounds in the tonal contrast falling within the learners' L1 phonetic space, but neither fits any single L1 phonological category. The prediction of this scenario is in agreement with our results.

Second, through the cognitively demanding sequence recall task and the lexical decision task, we found that, compared to beginning learners, advanced learners could process tones at the phonological level more accurately and use tonal information in lexical access more effectively. The learners' success in these tasks is generally in line with the suggestions of the ASP model (Strange, 2011): they are developing new selective perception routines attuned to the most reliable information for recognition of different tone categories.

Finally, in the eye-tracking experiment, advanced learners could activate the correct targets and suppress the phonologically similar competitors to a greater extent than beginning learners, exhibiting a developmental trajectory in using segmental and tonal information in spoken word recognition. Advanced learners could use segmental and tonal information in the early stage to constrain lexical activation, which is consistent with the concept of the TRACE model: the speech sounds are perceived in an incremental way. With the unfolding of the signal, segmental and tonal information was used by the learners to activate compatible word candidates and inhibit the incompatible ones.

These findings of this dissertation thus have some implications for language teaching. To make the production more natural and to facilitate the learning of coarticulation patterns, pronunciation of tones should be trained not only in monosyllables, but also in disyllabic words. Furthermore, tests on auditory and phonological processing of tones can be used to assist individual learners in finding out the problematic aspects in their tone learning.

6.4 Future research

Based on the findings of the this dissertation, some suggestions can be made for future research.

As shown in past research and in Chapter 2 in the current dissertation, tonal contours can be distorted by the influence of the preceding and following tones, showing coarticulated forms which deviate substantially form the canonical contours. In tone perception, some studies have found that native Mandarin speakers can compensate for tonal variations due to coarticulation to some extent (Xu, 1994). The results of Chapter 2 demonstrated that the fine-grained phonetic details of tonal coarticulation can be acquired gradually by L2 learners in tone production. Whether coarticulated tones in running speech can be correctly perceived by L2 learners and whether they can develop a perceptual compensation mechanism for deviant tonal contours become interesting issues to investigate. Such an investigation can also shed light on the issue of ultimate attainment in tone acquisition.

Chapter 3 showed that the advanced Dutch learners of Mandarin had not only successfully acquired the distinctions between tonal categories, but had also developed a strategy similar to that of native Mandarin listeners in terms of attention distribution and segmental-tonal integration. Previous studies showed that bilinguals integrate both languages in a common phonetic space and that learning an L2 may bring aboutt systematic phonetic changes in L1 production, which indicates that the L1 sound system is not static, but remains dynamic and adaptable (Aneta, 2000; Antoniou, 2012; Chang, 2012). It would be interesting to test whether a tonal L2 learning experience can further shape listeners' L1 perception. Specifically, it would be interesting to examine whether Dutch learners of Mandarin, when listening to Dutch, would attach more perceptual weight to pitch information and whether they would process segmental and pitch in a more integral manner compared to native Dutch speakers without any tone language learning experience.

Chapter 4 found that tones can be used in lexical access by Dutch learners of Mandarin. Interestingly, asymmetric patterns in the perception of tonal minimal pairs were found in advanced learners. The category of T1 was more well-established and T4 was less well-established compared to the other tones. Previous studies have also found asymmetric patterns in mapping phonetic information of a sound contrast to lexical representations in L2 listening, with one sound in the contrast more dominant than the other (Cutler, Weber, & Otake, 2006; Weber & Cutler, 2004). To further understand the asymmetric mapping from phonetic tonal information to lexical representations in L2 listening, the eye-tracking method can be employed to probe the temporal dynamics of L2 word recognition.

Furthermore, since this study only focused on the acquisition of lexical tones by Dutch learners of Mandarin, it would be worthwhile testing tonal processing by L2 learners with L1s different in prosodic structure. Such research can reveal how L2 tonal processing can be modulated by different L1s.

References

Abramson, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. In: V. P. T. L. Thongkum, P. Kullavanijaya, & M. R. K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology: In honour of Eugenie J. A. Henderson* (pp. 1-16). Bangkok: Chulalongkorn University Press.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419-439.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In: J.M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347- 386). New York: Psychology Press.

Aneta, P. (2000). L2 Influence on L1 in Late Bilingualism. *Issues in Applied Linguistics*, *11*(2), 175-205.

Antoniou, M., Tyler, M. D., Best, C.T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics, 40*(4), 582-594.

Bartoń, K. (2015). *MuMIn: Multi-Model Inference.* R package version 1.15.1. http://CRAN.R-project.org/package=MuMIn.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. http://CRAN.R-project.org/package=lme4.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48.

Baumotte, H., & Dogil, G. (2008). Coarticulation in non-native speakers of English: /əlV/-sequences in non-proficient versus proficient learners. In: A. Botinis (Ed.), *ISCA Tutorial and Research Workshop on Experimental Linguistics (ExLing 2008)* (pp. 37-40). Athens, Greece.

Beddor, P. S., Harnsberger, J. D., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*(4), 591-627.

Bell-Berti, F., Krakow, R. A., Gelfer, C. E., & Boyce, S. E. (1995). Anticipatory and carryover effects: Implications for models of speech production. In: F. Bell-Berti, & L. Raphael (Eds.), *Producing speech: Contemporary Issues for Katherine Safford Harris* (pp. 77-98). New York: AIP Press.

Bent, T. (2005). Perception and production of non-native prosodic categories. Doctoral dissertation, Northwestern University.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97-103.

Best, C. T. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In: J. Goodman, & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167-224). Cambridge, MA: MIT Press.

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

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

Best, C.T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: O. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins.

Birch, S., & Clifton, C. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language*, *47*(4), 571-588.

Boersma, P & Weenink, D. (2016). Praat: doing phonetics by computer.Version 6.0.20. http://www.praat.org/.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Attention, Perception, & Psychophysics*, *61*(5), 977-985.

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, *39*(4), 585-594.

Brengelmann, T. T., Cangemi, F., & Grice, M. (2015). Tonal coarticulation in L2 standard Chinese. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: University of Glasgow. 9.3, Paper 2. 1-4.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *The Quarterly Journal of Experimental Psychology*, *64*(1), 74-95.

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*(1), 219-252.

Brown. C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research, 14*(2), 136-193.

Brown. C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In: J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4-63). Oxford: Blackwell.

Brunelle, M. (2009). Northern and Southern Vietnamese tone coarticulation: A comparative case study. *Journal of Southeast Asian Linguistics*, *1*, 49-62.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, *5*(6), e10729.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*(1), 456-465.

Chandrasekaran, B., Yi, H. G., & Maddox, W. T. (2014). Dual-learning systems during speech category learning. *Psychonomic Bulletin & Review*, *21*(2), 488-495.

Chang, C.B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249-268.

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *Journal of the Acoustical Society of America*, *138*(6), 3703-3716.

Chang, Y.-C., & Hsieh, F.-F. (2012). Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *The Linguistic Review*, *29*(1), 37-73.

Chao, Yuen-Ren. (1930). A System of Tone Letters. *La Maître phonétique, 45*, 24-27.

Chen, A. (2005). *Universal and language-specific perception of paralinguistic intonational meaning*. Utrecht: LOT.

Chen, A., Gussenhoven, C., & Rietveld, A. (2004). Language-specificity in perception of paralinguistic intonational meaning. *Language and Speech, 47*(4), 311-350.

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

Chen, Y. (2012). Tonal variation. In: A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *The Oxford handbook of laboratory phonology* (pp. 103-114). Oxford: Oxford University Press.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics, 36*(4), 724-746.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech – Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica*, *63*(1), 47-75.

Choi, J. D., & Keating, P. (1991). Vowel-to-vowel coarticulation in three Slavic languages. *UCLA working papers in phonetics, 78,* 78-86.

Choi, W., Tong, X., Gu, F., Tong, X., & Wong, L. (2017). On the early neural perceptual integrality of tones and vowels. *Journal of Neurolinguistics*, *41*, 11-23.

Cole, J. (2015). Prosody in context: a review. *Language, Cognition and Neuroscience*, *30*(1-2), 1-31.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Attention, Perception, & Psychophysics*, *59*(2), 165-179.

Cutler, A., & McQueen, J. M. (2014). How prosody is both mandatory and optional. In: J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller, & E. Van Zanten (Eds.), *Above and Beyond the Segments: Experimental linguistics and phonetics* (pp. 71-82). Amsterdam: Benjamins.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech, 40*(2), 141-201.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269-284.

Dahan, D., & Gaskell, M. G. (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, *57*(4), 483-501.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*(4), 317-367.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534.

DiCanio, C. (2014) Triqui tonal coarticulation and contrast preservation in tonal phonology. In: Y. Bennett, R. Dockum, E. Gasser, D. Goldenberg, R. Kasak, & P. Patterson (Eds.), *Proceedings of the workshop on the sound systems of Mexico and Central America*. New Haven, CT: Yale Department of Linguistics.

Ding, H. (2012). Perception and production of Mandarin disyllabic tones by German learners. *Proceedings of Speech Prosody 2012*, 378-381.

Duanmu, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Dufour, S., Nguyen, N., & Frauenfelder, U. H. (2010). Does training on a phonemic contrast absent in the listener's dialect influence word recognition?. *Journal of the Acoustical Society of America*, *128*(1), EL43-EL48.

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

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

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress 'deafness' in simultaneous French-Spanish bilinguals. *Cognition*, *114*(2), 266-275.

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

Farris-Trimble, A., & McMurray, B. (2013). Test-retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1328-1345.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In: W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*(4), 467-491.

Francis, A. L., & Ciocca, V. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychopysics*, *65*(7), 1029-1044.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349-366.

Franich, K. (2015). The effect of cognitive load on tonal coarticulation. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow, UK: the University of Glasgow. D1.5, Paper 1. 1-5.

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

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and language*, *84*(3), 318-336.

Gandour, J., Potisuk, S. & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics, 22*(4)*,* 477-492.

Gandour, J., Potisuk, S., Ponglorpisit, S., Dechongkit, S., Khunadorn, F., & Boongird, P. (1996). Tonal coarticulation in Thai after unilateral brain damage. *Brain and language*, *52*(3), 505-535.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., ... & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage, 23*(1), 344-357.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*(1), 207-222.

Gao, S., Hu, J., Gong, D., Chen, S., Kendrick, K. M., & Yao, D. (2012). Integration of consonant and pitch processing as revealed by the absence of additivity in mismatch negativity. *PloS One*, *7*(5), e38289.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*(1), 98-123.

Garner, W. R. (2014). *The processing of information and structure*. New York: Psychology Press.

Gathercole, S. E., & Baddeley, A. D. (2014). *Working memory and language*. New York: Psychology Press.

Goh, W. D., Suárez, L., Yap, M. J., & Tan, S. H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. *Psychonomic Bulletin & Review*, *16*(5), 882-887.

Goldstone, R. L. (1993). Feature distribution and biased estimation of visual displays. *Journal of Experimental Psychology: Human Perception and Performance, 19*(3), 564-579.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178-200.

Grosvald, M. (2009). Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics*, *37*(2), 173-188.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. 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* (pp. 57-77). Amsterdam: John Benjamins.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*(5), 2711-2724.

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: The phonology of intonation and phrasing* (pp. 118-145). Oxford: Oxford University Press.

Gussenhoven, C., Rietveld, T., Kerkhoff, J., & Terken, J. (2003). *ToDI second edition*. http://todi.let.kun.nl/ToDI/ home.htm.

Hallé, P. A. (1994). Evidence for tone-specific activity of the sternohyoid muscle in modern standard Chinese. *Language and Speech*, *37*(2), 103-123.

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

Hallet, P. E. (1986). Eye movements. In: K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1-10.122). New York: Wiley.

Han, M. S., & Kim, K. O. (1974). Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics*, *2*(3), 223-232.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269-279.

Hardcastle, W. J., & Hewlett, N. (2006). *Coarticulation: Theory, data and techniques*. Cambridge: Cambridge University Press.

Hardison, D. M., & Saigo, M. M. (2010). Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics*, *31*(1), 81-99.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5-33.

He, X. (2012). *Mandarin-accented Dutch prosody*. Utrecht: LOT.

He, Y., & Wayland, R. (2010). The production of Mandarin coarticulated tones by inexperienced and experienced English speakers of Mandarin. In *Speech Prosody 2010 Conference Proceedings* (pp. 1-4).

Heeren, W. F. L., & Schouten, M. E. H. (2008). Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories. *Journal of the Acoustical Society of America*, *124*(4), 2291-2302.

Heeren, W. F. L., & Schouten, M. E. H. (2010). Perceptual development of the Finnish/t-t/distinction in Dutch 12-year-old children: A training study. *Journal of Phonetics*, *38*(4), 594-603.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*(3), 346-363.

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

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151-171.

Keating, P. A. (1990). The window model of coarticulation: Articulatory evidence. In: J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 451-470). Cambridge: Cambridge University Press.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161-168.

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339.

Lee, C. Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, *50*(1), 101-123.

Lee, R. R. W., Hsu, C. H., Lin, S. K., Wu, D. H., & Tzeng, O. J. L. (2017). Learning transforms functional organization for Mandarin lexical tone discrimination in the brain: Evidence from a MEG experiment on second language learning. *Journal of Neurolinguistics*, *42*, 124-139.

Lee, W. S., & Zee, E. (2003). Standard Chinese (Beijing). *International Phonetic Association. Journal of the International Phonetic Association*, *33*(1), 109-112.

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

Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, *4*(2), 185-199.

Li, Q. (2016). *The production and perception of tonal variation: evidence from Tianjin Mandarin*. Utrecht: LOT.

Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics*, *54*, 123-150.

Lin, M., & Francis, A. L. (2014). Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese. *Journal of the Acoustical Society of America*, *136*(5), 2827-2838.

Liu, S., & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes*, *22*(4), 566-594.

Lu, S., Wayland, R., & Kaan, E. (2015). Effects of production training and perception training on lexical tone perception – A behavioral and ERP study. *Brain Research*, *1624*, 28-44.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, *19*(1), 1-36.

Luthy, M. J. (1983). Nonnative speakers' perceptions of English "nonlexical" intonation signals. *Language Learning*, *33*(1), 19-36.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. New York: Psychology Press.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133-156.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407-420.

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*(8), 2032-2043.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America*, *88*(3), 1286-1298.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *Journal of the Acoustical Society of America*, *69*(2), 559-567.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145-160.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229-258.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, *18*(1), 1-86.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064-1071.

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1344-1362.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, *60*(1), 1-39.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of phonetics*, *32*(4), 543-563.

Meyer, D. E., & Gordon, P. C. (1985). Speech production: Motor programming of phonetic features. *Journal of Memory and Language*, *24*(1), 3-26.

Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2009). Parafoveal magnification: Visual acuity does not modulate the perceptual span in reading. *Psychological Science*, *20*(6), 721-728.

Mirman, D. (2014). *Growth curve analysis and visualization using* R. Boca Raton, FL: CRC Press.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475-494.

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, *117*(2), 53-68.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, *102*(3), 1864-1877.

Mora J. C., Keidel J. L., Flege J. E. (2011). Why are the Catalan contrasts between /e/-/ɛ/ and /o/-/ɔ/ so difficult for even early Spanish–Catalan bilinguals to perceive? In: M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New sounds 2010* (Vol. II, pp. 183–193). Frankfurt am Main: Peter Lang.

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

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189-234.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357-395.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39-57.

Oh, E. (2008). Coarticulation in non-native speakers of English and French: An acoustic study. *Journal of Phonetics*, *36*(2), 361-384.

Oh, Y. M., Pellegrino, F., Coupé, C., & Marsico, E. (2013). Cross-language comparison of functional load for vowels, consonants, and tones. *Proceedings of Interspeech 2013*, 3032-3036.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9-B17.

Pierrehumbert, J. B., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, *46*(4), 181-196.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In: H. C. Nusbaum, & J. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT Press.

Potisuk, S., Gandour, J., & Harper, M. P. (1997). Contextual variations in trisyllabic sequences of Thai tones. *Phonetica*, *54*(1), 22-42.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. *Journal of the Acoustical Society of America*, *76*(6), 1624-1635.

Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics, 18*(4)*,* 481-495.

Righi, G., Blumstein, S. E., Mertus, J., & Worden, M. S. (2010). Neural systems underlying lexical competition: An eye tracking and fMRI study. *Journal of Cognitive Neuroscience*, *22*(2), 213-224.

Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, *17*(1), 1-12.

Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, *72*(2), 111-123.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*(2), 240-255.

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, *58*(2), 131-151.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*(2), 193-247.

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *Journal of the Acoustical Society of America*, *133*(5), 3016-3029.

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association*, *24*(3), 27-47.

Shuai, L., & Malins, J. G. (2017). Encoding lexical tones in jTRACE: a simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods*, *49*(1), 230–241.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103-130.

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149.

Strange, W. (2011). Automatic selective perception ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456-466.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131-145.

Sun, S. H. (1998). *The development of a lexical tone phonology in American adult learners of Standard Mandarin Chinese.* Honolulu, HI: University of Hawaii Press.

Surendran, D., & Levow, G. A. (2004). The functional load of tone in Mandarin is as high as that of vowels. *Proceedings of Speech Prosody 2004*, 99-102.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, *29*(6), 557-580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632.

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics, 37*(3), 276-296.

Tilsen, S. (2013). Inhibitory mechanisms in speech planning maintain and maximize. In: Yu, A. C. (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 112-117). Oxford University Press.

Tong, X., McBride, C., Lee, C. Y., Zhang, J., Shuai, L., Maurer, U., & Chung, K. K. (2014). Segmental and suprasegmental features in speech perception in Cantonese-speaking second graders: An ERP study. *Psychophysiology*, *51*(11), 1158-1168.

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, *23*(5), 689-708.

Van Heuven, V. J. (2017). Prosody and sentence type in Dutch. *Nederlandse Taalkunde*, *22*(1), 3-29, 43-46.

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

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

Wang, S., Chen, H. C., & Yang and Lei Mo, J. (2008). Immediacy of integration in discourse comprehension: Evidence from Chinese readers' eye movements. *Language and Cognitive Processes*, *23*(2), 241-257.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*(2), 1033-1043.

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (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, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*(6), 3649-3658.

Wang, W., Zhou, N., & Xu, L. (2011). Musical pitch and lexical tone perception with cochlear implants. *International Journal of Audiology*, *50*(4), 270-278.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1-25.

Welby, P. (2003). Effects of pitch accent type and status on focus projection. *Language and Speech*, *46*(1), 53-81.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, *18*(1), 3-35.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*(1), 25-47.

Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and speech*, *59*(1), 59-82.

Wong, F. C., Chandrasekaran, B., Garibaldi, K., & Wong, P. C. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *Journal of Neuroscience*, *31*(24), 8780-8785.

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565-585.

Wong, P., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*, *28*(10), 995-1006.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*(4), 2240-2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61-83.

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics, Monograph Series, 17*, 1-31.

Xu, Y. (2009). Timing and coordination in tone and intonation—An articulatory-functional perspective. *Lingua*, *119*(6), 906-927.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319-337.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159-197.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., & Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, *27*(2), 173-183.

Yang, C. (2011). The acquisition of Mandarin prosody by American learners of Chinese as a foreign language (CFL). Doctoral dissertation, The Ohio State University.

Yang, C. (2016). *The Acquisition of L2 Mandarin Prosody: From experimental studies to pedagogical practice*. Amsterdam: John Benjamins.

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics*, *33*, 29-49.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*(5-6), 609-630.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1-14.

Yee, E., Blumstein, S. E., & Sedivy, J. C. (2008). Lexical-semantic activation in Broca's and Wernicke's aphasia: Evidence from eye movements. *Journal of Cognitive Neuroscience*, *20*(4), 592-612.

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

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1493), 1087-1104.

Zhang, X., Samuel, A. G., & Liu, S. (2012). The perception and representation of segmental and prosodic Mandarin contrasts in native speakers of Cantonese. *Journal of Memory and Language*, *66*(3), 438-457.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761-1770.

Zou, T., Chen, Y., & Caspers, J. (2016). The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones. *Bilingualism: Language and Cognition*,1-13. DOI 10.1017/S1366728916000791

Summary

This thesis examines the production and perception of lexical tones by beginning and advanced Dutch learners of Mandarin as well as the developmental trajectory in second-language (L2) tone acquisition.

In Mandarin Chinese, a lexical tone language, pitch configurations such as a high, low and rising tone (cued primarily through fundamental frequency) are used to differentiate between word forms. As documented in previous studies, lexical tone always presents great difficulty for adult L2 Mandarin learners whose native language is non-tonal. In terms of tone production, past research mainly focused on the accuracy of L2 tone pronunciation in isolated words. The L2 acquisition of tonal coarticulation, which leads to deviant tonal contours differing from the canonical forms in natural connected speech, has not been studied systematically. The L2 perception of tone in isolated words has been examined in identification and discrimination tasks in previous works. Yet, how non-natives learn to use pitch information in a lexically contrastive manner and how such learners process tone at the phonological level remains unclear. Moreover, the online processing of tonal information in word recognition by L2 learners is also an interesting issue to investigate. Based on these research questions, this thesis provides a systematic study of the production and perception of lexical tones by beginning and advanced Dutch learners of Mandarin with four well-controlled experiments.

Chapter 1 is a general introduction. After a brief introduction to the tone system in Mandarin, an overview of different aspects of L2 tone acquisition is organized around the basic research questions.

Chapter 2 investigates the tonal coarticulation in two-syllable words by Dutch learners of Mandarin and native Mandarin speakers. All the 16 tonal combinations in disyllables were tested using the non-word /mama/ with each syllable bearing one of the four Mandarin tones. The underlying coarticulatory mechanism was investigated in a high cognitive load condition. This chapter shows that for native Mandarin speakers, the substantial carryover coarticulation is assimilatory and not planned, contrary to the dissimilatory anticipatory coarticulation, which shows a smaller magnitude and involves advance planning. For L2 learners, a developmental trajectory toward the native norm has been found for carryover coarticulation, with the advanced learners showing stronger assimilatory coarticulatory effects than beginning learners. Although the carry-over effect is mainly a result of physiological constraints and does not involve planning, its acquisition by L2 learners is still a gradual process. As for the anticipatory coarticulation, the advanced learners show a strong dissimilatory effect, which is more robust than that found for the native speakers. This may indicate that the learners effectively employ an inhibitory mechanism to maintain the contrast and ensure the perceptibility of different tonal categories in running speech.

Chapter 3 sets out to examine the phonological discrimination of Mandarin tones and segment-tone integration in Dutch learners of Mandarin, with both native Mandarin and Dutch speakers without tonal learning experience as control groups. An ABX task with four conditions is used to test how participants' attention is distributed

between segments and tones. This chapter demonstrates a developmental path toward the native norm in tone processing for L2 learners. The beginning learners do not attend to tonal information and process segmental and tonal information separately, like native Dutch listeners without any Mandarin experience. The advanced learners show a more native-like pattern in distributing their attention between segmental and tonal information. Moreover, they process the two dimensions in an integrated manner, similar to native Mandarin listeners. This chapter suggests that the acquisition of new tonal categories in L2 involves a redistribution of attention between segmental and tonal dimensions as well as the development of segment-tone integration.

Chapter 4 further explores the processing of all tonal contrasts at the phonological level and the use of tones in lexical access by Dutch learners of Mandarin using a cognitively demanding sequence recall task and a lexical decision task. The results of the sequence recall task show a clear developmental path in phonological processing of tonal information by L2 learners. The results of the lexical decision task also indicate that, compared to beginners, advanced learners performed significantly better in correctly identifying real words and rejecting non-words that are minimally different from real words in terms of their tone structure. The improvement of advanced learners in both tasks demonstrates that they are shaping new selective perception routines, and that their phonological mode of tone processing is in development.

Furthermore, this chapter bears out that in the lexical decision task, Tone 2 and Tone 3 are mutually confusable for learners and are difficult to learn. Such difficulty may stem from the acoustic similarity between these two tones. Asymmetric patterns are also found for advanced learners in the lexical decision task. For these learners, the category of Tone 1 has been relatively well established when compared to the other three tones. In contrast to this, the category of Tone 4 is less well-established when compared to the other three tones in pairs. These results are potentially related to the prosodic features of the learners' native language, since previous research showed that, compared to tone-language speakers, intonation-language speakers are more sensitive to pitch height than to pitch direction.

Chapter 5 investigates the time course of lexical activation and the relative contribution of segmental and tonal information to speech recognition by testing native Mandarin speakers, as well as beginning and advanced learners of Mandarin in an eye-tracking experiment using the visual world paradigm. The participants heard a spoken word and were asked to identify the corresponding picture from a display of four pictures that consisted of the target, a phonological competitor, and two phonologically unrelated distractors. The probability of fixation on the target and competitor was recorded since it may reflect the activation of the corresponding items. This chapter demonstrates that native Mandarin speakers use tonal information effectively to constrain lexical activation in an early stage of word recognition, in much the same way as they exploit segmental information. Similar to native Mandarin speakers, tonal information is also used by Dutch learners of Mandarin in an early stage to inhibit the activation of incompatible lexical candidates, although they still experience difficulty in the discrimination of some tonal minimal pairs. Compared to beginning learners, significant improvement has been found for advanced learners toward the native norm.

Finally, **Chapter 6** recapitulates the research questions and summarizes the main findings of this thesis. This chapter also provides suggestions for future research.

Samenvatting

Dit proefschrift onderzoekt de productie en perceptie van lexicale tonen door beginnende en gevorderde Nederlandse leerders van het Mandarijn, als ook het ontwikkelingstraject dat deze leerders afleggen tijdens de verwerving van tonen in een tweede taal (T2).

Mandarijn, of noordelijk Chinees, is de meest gesproken eerste taal (T1) ter wereld, met naar schatting een miljard moedertaalsprekers. Mandarijn is een lexicale toontaal, waarin toonhoogteconfiguraties worden gebruikt zoals hoge, lage en stijgende toon (primair aangegeven door de grondtoon in het spraakgeluid) om verschil aan te brengen tussen woordvormen. Zo betekent dezelfde klankopeenvolging /ma/ met een vlakke hoge toon 'moeder' maar met een stijgende toon 'hennep'. Uit eerder onderzoek is al gebleken dat lexicale tonen (of: woordtonen) altijd een probleem vormen voor volwassen T2-leerders van het Mandarijn als hun moedertaal geen woordtonen kent. Wat spraakproductie betreft heeft eerder onderzoek zich voornamelijk toegespitst op de uitspraak van de T2-tonen in losse woorden. De T2-verwerving van tonale coarticulatie, een proces van wederzijdse beïnvloeding van opeenvolgende tonen waarbij de tooncontouren in natuurlijke, verbonden spraak fors gaan afwijken van hun canonieke vormen, is niet eerder systematisch onderzocht. De waarneming van T2-tonen is in eerder werk in losse woorden alleen bestudeerd aan de hand van identificatie- en discriminatietaken. Hoe niet-moedertaalsprekers toonhoogte-informatie leren gebruiken om woorden uit elkaar te houden en hoe zulke T2-leerders de woordtonen op fonologisch niveau verwerken, is vooralsnog onbekend. Tevens is het van groot belang te weten hoe T2-leerders tooninformatie gebruiken tijdens de online herkenning van gesproken woorden. Om antwoord te geven op deze vragen is in dit proefschrift een systematische studie verricht naar de productie en waarneming van woordtonen door beginnende en gevorderde Nederlandse leerders van het Mandarijn, waarbij vier experimenten zijn uitgevoerd.

Hoofdstuk 1 is een algemene inleiding. Na een korte uitleg van het toonsysteem van het Mandarijn volgt een overzicht van de diverse aspecten van de verwerving van woordtonen in een T2, in overeenstemming met de vier onderzoeksvragen.

Hoofdstuk 2 onderzoekt de tooncoarticulatie in tweelettergrepige woorden door Nederlandse leerders van het Mandarijn en door moedertaalsprekers van die taal. Alle 16 logisch mogelijke tooncombinaties in opeenvolgingen van twee lettergrepen zijn onderzocht aan de hand van het in het Mandarijn niet bestaande woord /mama/, waarin op elke van de twee lettergrepen een van de vier tonen van het Mandarijn is gesproken. Het onderliggend coarticulatiemechanisme is onderzocht in een situatie waarin de spreker cognitief zwaar belast werd. De resultaten wijzen uit dat er bij de moedertaalsprekers van het Mandarijn sprake is van substantiële *carry-over* coarticulatie (d.w.z. dat de toon van de voorafgaande lettergreep de vorm van de toon op de volgende lettergreep beïnvloedt, en niet andersom). De tonen op de opeenvolgende lettergrepen gaan meer op elkaar lijken (assimilatie); deze aanpassing vereist geen planning van de kant van de spreker, in tegenstelling tot anticiperende coarticulatie (de volgende toon beïnvloedt de vorm van de voorafgaande toon), waarbij de doorgaans kleinere aanpassing dissimilerend van aard is (de opeenvolgende tonen gaan juist minder op elkaar lijken), en planning vergt. Voor de T2-leerders vinden we een ontwikkelingstraject in de richting van de T1-norm voor zo ver het *carry-over* coarticulatie betreft, waarbij de toonassimilatie bij de gevorderde leerders sterker is dan bij de beginners. Hoewel het *carry-over* effect voornamelijk wordt veroorzaakt door fysiologische beperkingen en geen planning inhoudt, is de verwerving ervan door T2-leerders toch een geleidelijk proces. Met betrekking tot de anticiperende coarticulatie vertonen de gevorderde leerders een sterke dissimilatie, die bij hen zelfs robuuster optreedt dan bij de moedertaalsprekers. Dit zou erop kunnen wijzen dat de leerders effectief een inhibitiemechanisme toepassen om het contrast tussen opeenvolgende tonen op te scherpen en zo de waarneembaarheid van de verschillende tooncategorieën in lopende spraak te waarborgen.

Hoofdstuk 3 heeft als doel de fonologische discriminatie van de tonen in het Mandarijn te onderzoeken alsmede de integratie van segmentele en tonale informatie, bij Nederlandse leerders van het Mandarijn, met zowel moedertaalsprekers van het Mandarijn als Nederlandse sprekers zonder enige ervaring met het Mandarijn als controlegroepen. We hebben daarbij een ABX vergelijkingstaak gebruikt met vier condities om te testen hoe de deelnemers hun aandacht verdelen tussen de segmenten (klinkers en medeklinkers) en de tonen. Dit hoofdstuk laat zien dat de verwerking van de gesproken input door de T2-leerders zich ontwikkelt in de richting van de moedertaalnorm. De beginnende leerders geven nauwelijks aandacht aan de tooninformatie en verwerken de segmentele informatie apart, net zoals Nederlanders zonder enige ervaring met het Mandarijn dat doen. De gevorderde leerders echter verdelen hun aandacht meer op de manier waarop moedertaalsprekers van het Mandarijn dat doen, namelijk met gelijke(r) aandachtsverdeling tussen de segmentele en tonale informatie. Bovendien verwerken zij de twee informatiestromen geïntegreerd, zoals de T1-luisteraars dat ook doen. Een en ander suggereert dat verwerving van nieuwe tooncategorieën neerkomt op een herverdeling van aandacht tussen segmentele en tonale informatie als ook ontwikkeling van de segment-toonintegratie.

Hoofdstuk 4 verkent nader hoe Nederlandse leerders van het Mandarijn alle tooncontrasten op fonologisch niveau verwerken en welke rol deze tonen spelen bij de lexicale access (toegang tot de mentale woordenschat) aan de hand van een cognitief veeleisende taak waarbij luisteraars een gehoorde reeks woorden in de juiste volgorde uit hun geheugen moeten oproepen, alsook aan de hand van een lexicale decisietaak ('is een gehoorde klankreeks wel of niet een bestaand woord?'). De resultaten van de geheugentaak wijzen op een duidelijke ontwikkeling in de verwerking van tonale informatie door T2-leerders. De resultaten van de lexicale decisietaak geven daarenboven aan dat de gevorderde leerders, in vergelijking met beginners, bestaande woorden significant beter herkenden en non-woorden die in hun toonstructuur slechts minimaal verschilden van bestaande woorden terecht afwezen. De betere prestaties van de gevorderde groep in beide taken laat zien dat deze leerders zich nieuwe selectieve waarnemingsroutines eigen maken en dat hun fonologische toonverwerking zich ontwikkelt.

Dit hoofdstuk laat bovendien zien dat Toon 2 en Toon 3 in de lexicale decisietaak met elkaar verward worden en daarom moeilijk te leren zijn. Deze moeilijkheid wordt mogelijk veroorzaakt doordat deze twee tonen akoestisch sterk op elkaar lijken. In de lexicale decisietaak vinden we voor de gevorderde leerders ook asymmetrische patronen. Toon 1 is bij deze groep leerders vrij goed gedefinieerd in vergelijking met de andere drie tonen, in tegenstelling tot Toon 4. Deze verschillen hebben mogelijk te maken met prosodische eigenaardigheden in de moedertaal van de leerders omdat uit eerder onderzoek is gebleken dat sprekers van een intonatietaal in vergelijking met sprekers van een toontaal gevoeliger zijn voor verschillen in (gemiddelde) toonhoogte dan voor de richting van de toonhoogteverandering (toonhoogtestijging versus -daling).

Hoofdstuk 5 onderzoekt de tijdsontwikkeling van de lexicale activatie en de relatieve bijdrage van segmentele en tonale informatie aan de spraakherkenning. Hiertoe zijn bij moedertaalsprekers van het Mandarijn alsook beginnende en gevorderde leerders van het Mandarijn de oogbewegingen gemeten in een zgn. visuele-wereld-experiment. De deelnemers hoorden een gesproken woord en moesten een daarmee corresponderend plaatje aanwijzen op een scherm met vier plaatjes: een van het doelwoord, een van een klankvormelijk nagenoeg hetzelfde woord en nog eens twee klankvormelijk ongerelateerde afleiders. De relatieve frequentie van een oogfixatie op het doelwoord of op de minimaal verschillende concurrent werd bepaald als indicatie van de activatiegraad van de corresponderende woorden. Dit hoofdstuk laat zien dat de moedertaalsprekers van het Mandarijn de informatie in de tonen effectief gebruiken om lexicale activatie in een vroeg stadium van de woordherkenning in te perken. En evenals de moedertaalsprekers van het Mandarijn gebruiken Nederlandse leerders van het Mandarijn tooninformatie om de activatie van incompatibele woordkandidaten te inhiberen, ook al ervaren deze proefpersonen nog steeds problemen als zij bepaalde minimale toonparen moeten onderscheiden. Vergeleken met de beginnende leerders vertoont de gevorderde groep een significante vooruitgang in de richting van de Chinese moedertaalnorm.

Hoofdstuk 6, ten slotte, recapituleert de onderzoeksvragen en vat de belangrijkste bevindingen van het onderzoek samen. Dit hoofdstuk doet ook suggesties voor verder onderzoek.

摘要

本论文考察了以荷兰语为母语的汉语学习者汉语声调的发音和感知。同时，通过对比初级和高级水平汉语学习者的表现，揭示了第二语言习得中声调的发展轨迹。

作为一种声调语言，汉语普通话利用不同的音高形式区分词义（基频是声调的主要声学线索）。以往的研究表明，对于母语是非声调语言的汉语学习者，声调是汉语习得的主要难点之一。在声调发音方面，过往研究主要关注二语学习者在单字词发音中声调的正确率。在连续语流中，受到协同发音作用的影响，声调往往呈现出与单字词发音时不同的音高曲线。可是，二语学习者声调协同发音的习得状况还缺乏系统性的研究。在感知方面，过往研究主要利用辨别和区分任务，考查汉语学习者单字词的声调感知。然而，学习者声调习得的内在机制以及他们在音系层面对声调的加工还需要进一步研究。同时，二语学习者在词汇认知过程中对声调信息的在线加工也是一个需要探索的课题。因此，本论文包含四个语音学实验，旨在系统性地考察上述研究问题。

第一章为总引言。本章在简要介绍了汉语普通话声调系统后，对本文所关注的第二语言声调习得的不同研究领域进行了综述。

第二章研究了以荷兰语为母语的汉语学习者以及汉语母语者在双字词中的声调协同发音。本章的实验以/mama/作为实验刺激，考查了双字词中 16 种声调组合，并在高度认知负荷 (high-cognitive-load) 的条件下，考查了声调协同发音的内在机制。结果表明，对于汉语母语者，顺向协同发音作用很强，体现出同化的特性，无需语言规划；逆向协同发音在影响程度上较弱，体现出异化的特性，需要语言规划。二语学习者在声调顺向协同发音的习得上，呈现出了明显的发展轨迹，高水平学习者更加接近汉语母语者的发音模式。值得注意的是，声调的顺向协同发音主要是发音生理局限的产物，因此并不需要语言规划。二语学习者对声调顺向协同发音的学习仍然是一个循序渐进的过程。我们还发现高水平学习者的逆向协同发音作用很强，而且稳定性高于母语者。这或许表明二语学习者可以有效地利用抑制机制来保持不同声调间的音高对比，从而保证连续语流中不同声调范畴的辨析度。

第三章研究以荷兰语为母语的汉语学习者在音系层面对声调的区分以及他们对音段和声调信息的整合。实验还包括汉语母语者以及没有学习过声调语言的荷兰语母语者作为对照组。实验使用了 ABX 范式，并设计了四种情况, 以检测被试的注意力在音段和声调这两个层面的分配状况。本章的结果也呈现出学习者的习得发展轨迹。初级水平学习者与没有声调语言背景的荷兰语母语者呈现出相似的感知模式：他们不会关注声调信息，并且对音段和声调信息进行分别加工。高级水平学习者在音段、声调这两个层面的注意力分配上更加接近汉语母语者。同时，与汉语母语者相似的是，他们可以综合加工这两个层面的

信息。这意味着，在二语习得过程中，习得新的声调范畴需要学习者重新调整注意力在音段和声调层面的分配，同时学习者需要学习综合加工这两个层面的信息。

第四章进一步探索了学习者在音系层面对声调信息的加工以及在词汇通达中对声调信息的使用。本章采用了比 ABX 更复杂的实验范式：序列记忆任务以及词汇判断任务。序列记忆实验的结果显示，相较于初级水平学习者，高水平学习者序列记忆的正确率有显著提升。词汇判断实验的结果也表明，相比初级水平学习者，高水平学习者可以更准确地辨别真词，也可以更高效地拒绝只在声调上与真词有差别的假词。高水平学习者在这两个实验中的进步都表明，他们正在形成新的选择感知路径 (selective perception routines), 他们对于声调的音系加工模式也在持续发展。

此外，本章的实验结果也表明，在学习者的声调感知中，二声和三声最容易相互混淆。易混原因可能是这两个声调有相似的声学特点。同时，在词汇判断实验中，高水平学习者表现出不对称的声调感知模式。相比其他声调，一声的声调范畴更加稳定，而四声的声调范畴稳定性较差。要解释这种现象，可以从学习者母语的韵律系统中寻找线索。基于以往研究，相比声调语言的母语者，语调语言的母语者对音高高度更加敏感，而对音高的变化方向敏感度较低。

第五章采用眼动实验，通过视觉情境范式 (visual world paradigm) 考察了被试在词汇激活的时间进程中，音段和声调信息的作用。被试为汉语母语者以及初级和高级水平汉语学习者。实验中，被试会听到一个汉语词，同时他需要在四幅图片中选出与听到的词相对应的图片。这四幅图片的内容包括目标词，一个语音重叠词，以及两个干扰词。实验记录了被试对目标词和语言重叠词的注视概率，以此来测量相应词汇的激活程度。结果显示，汉语母语者在词汇认知的早期就可以有效利用声调信息来约束词汇激活。在这方面，声调起到了与音段信息大致相当的作用。与汉语母语者类似，学习者也可以在词汇认知的早期阶段利用声调信息抑制不适合的词汇候选项，虽然他们在区分某些声调最小对立对的时候还存在困难。相较初级水平学习者，高水平学习者更加接近汉语母语者的认知模式。

最后，**第六章**回顾了研究问题并总结了本论文的主要发现。这一章也对今后的相关研究提出了建议。

Appendices

A1. *Summary of mixed effects models for f0 contours of each tone in the first or second syllable.*

T1 in first syllable

Fixed effects	Df	χ^2	p
Linear and Quadratic terms	2	13.36	<.01
Tonal Context	3	46.74	<.001
Cognive Load	1	4.22	<.05
Participant Group	2	1.02	n.s.
Linear and Quadratic terms : Tonal Context	6	21.52	<.01
Linear and Quadratic terms : Cognitive Load	2	0.48	n.s.
Linear and Quadratic terms : Participant Group	4	7.75	n.s.
Tonal Context : Cognitive Load	3	165.62	<.001
Tonal Context : Participant Group	6	8.42	n.s.
Cognive Load : Participant Group	2	0.10	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	3.41	n.s.
Linear and Quadratic terms : Tonal Context : Participant Group	12	8.14	n.s.
Linear and Quadratic terms : Cognitive Load : Participant Group	4	4.70	n.s.
Tonal Context : Cognitive Load : Participant Group	6	101.19	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	10.53	n.s.
Repetition	3	7.50	n.s.
Repetition: Linear and Quadratic Terms	6	6.11	n.s.

T2 in first syllable

Fixed effects	Df	χ^2	p
Linear and Quadratic terms	2	81.41	<.001
Tonal Context	3	32.29	<.001
Cognive Load	1	5.89	<.05
Participant Group	2	32.47	<.001
Linear and Quadratic terms : Tonal Context	6	71.26	<.001
Linear and Quadratic terms : Cognitive Load	2	15.04	<.001
Linear and Quadratic terms : Participant Group	4	3.42	n.s.
Tonal Context : Cognitive Load	3	29.19	<.001
Tonal Context : Participant Group	6	13.79	<.05
Cognive Load : Participant Group	2	1.39	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	8.10	n.s.
Linear and Quadratic terms : Tonal Context : Participant Group	12	27.17	<.01
Linear and Quadratic terms : Cognitive Load : Participant Group	4	3.02	n.s.
Tonal Context : Cognitive Load : Participant Group	6	112.03	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	16.07	n.s.
Repetition	3	42.08	<.001
Repetition: Linear and Quadratic Terms	6	7.04	n.s.

T3 in first syllable

Fixed effects	Df	χ²	p
Linear and Quadratic terms	2	59.99	<.001
Tonal Context	3	33.26	<.001
Cognive Load	1	2.49	<.05
Participant Group	2	10.64	<.01
Linear and Quadratic terms : Tonal Context	6	86.96	<.001
Linear and Quadratic terms : Cognitive Load	2	5.28	<.05
Linear and Quadratic terms : Participant Group	4	24.92	<.001
Tonal Context : Cognitive Load	3	31.39	<.001
Tonal Context : Participant Group	6	6.09	n.s.
Cognive Load : Participant Group	2	0.24	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	6.14	n.s.
Linear and Quadratic terms : Tonal Context : Participant Group	12	21.52	<.05
Linear and Quadratic terms : Cognitive Load : Participant Group	4	6.12	n.s.
Tonal Context : Cognitive Load : Participant Group	6	41.39	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	32.82	<.001
Repetition	3	4.67	n.s.
Repetition: Linear and Quadratic Terms	6	11.70	n.s.

T4 in first syllable

Fixed effects	Df	χ²	p
Linear and Quadratic terms	2	91.96	<.001
Tonal Context	3	21.07	<.001
Cognive Load	1	4.80	<.05
Participant Group	2	0.15	n.s.
Linear and Quadratic terms : Tonal Context	6	66.72	<.001
Linear and Quadratic terms : Cognitive Load	2	12.41	<.01
Linear and Quadratic terms : Participant Group	4	16.25	<.01
Tonal Context : Cognitive Load	3	55.66	<.001
Tonal Context : Participant Group	6	7.44	n.s.
Cognive Load : Participant Group	2	4.23	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	34.87	<.001
Linear and Quadratic terms : Tonal Context : Participant Group	12	24.10	<.05
Linear and Quadratic terms : Cognitive Load : Participant Group	4	1.17	n.s.
Tonal Context : Cognitive Load : Participant Group	6	38.11	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	68.71	<.001
Repetition	3	27.29	<.001
Repetition: Linear and Quadratic Terms	6	52.24	<.001

T1 in second syllable

Fixed effects	Df	χ^2	p
Linear and Quadratic terms	2	34.16	<.001
Tonal Context	3	21.89	<.001
Cognive Load	1	3.73	n.s.
Participant Group	2	1.92	n.s.
Linear and Quadratic terms : Tonal Context	6	90.10	<.001
Linear and Quadratic terms : Cognitive Load	2	0.64	n.s.
Linear and Quadratic terms : Participant Group	4	12.86	<.05
Tonal Context : Cognitive Load	3	82.51	<.001
Tonal Context : Participant Group	6	13.95	<.05
Cognive Load : Participant Group	2	0.04	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	6.80	n.s.
Linear and Quadratic terms : Tonal Context : Participant Group	12	36.93	<.001
Linear and Quadratic terms : Cognitive Load : Participant Group	4	1.10	n.s.
Tonal Context : Cognitive Load : Participant Group	6	160.43	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	19.44	n.s.
Repetition	3	78.74	<.001
Repetition: Linear and Quadratic Terms	6	3.33	n.s.

T2 in second syllable

Fixed effects	Df	χ^2	p
Linear and Quadratic terms	2	91.24	<.001
Tonal Context	3	33.16	<.001
Cognive Load	1	0.19	n.s.
Participant Group	2	16.55	<.001
Linear and Quadratic terms : Tonal Context	6	89.86	<.001
Linear and Quadratic terms : Cognitive Load	2	7.07	<.05
Linear and Quadratic terms : Participant Group	4	12.21	n.s.
Tonal Context : Cognitive Load	3	83.29	<.001
Tonal Context : Participant Group	6	11.48	n.s.
Cognive Load : Participant Group	2	4.59	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	17.43	<.01
Linear and Quadratic terms : Tonal Context : Participant Group	12	15.03	<.01
Linear and Quadratic terms : Cognitive Load : Participant Group	4	0.60	n.s.
Tonal Context : Cognitive Load : Participant Group	6	47.62	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	54.72	<.001
Repetition	3	6.16	n.s.
Repetition: Linear and Quadratic Terms	6	16.38	<.05

T3 in second syllable

Fixed effects	Df	χ^2	p
Linear and Quadratic terms	2	56.16	<.001
Tonal Context	3	33.26	<.001
Cognive Load	1	2.54	n.s.
Participant Group	2	15.08	<.001
Linear and Quadratic terms : Tonal Context	6	61.76	<.001
Linear and Quadratic terms : Cognitive Load	2	0.58	n.s.
Linear and Quadratic terms : Participant Group	4	28.02	<.001
Tonal Context : Cognitive Load	3	18.52	<.001
Tonal Context : Participant Group	6	18.02	<.01
Cognive Load : Participant Group	2	2.27	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	30.37	<.001
Linear and Quadratic terms : Tonal Context : Participant Group	12	11.46	<.01
Linear and Quadratic terms : Cognitive Load : Participant Group	4	3.42	n.s.
Tonal Context : Cognitive Load : Participant Group	6	20.79	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	37.02	<.001
Repetition	3	9.98	<.05
Repetition: Linear and Quadratic Terms	6	41.44	<.001

T4 in second syllable

Fixed effects	Df	χ^2	p
Linear and Quadratic terms	2	66.56	<.001
Tonal Context	3	21.35	<.001
Cognive Load	1	7.42	<.01
Participant Group	2	2.51	n.s.
Linear and Quadratic terms : Tonal Context	6	82.75	<.001
Linear and Quadratic terms : Cognitive Load	2	5.02	n.s.
Linear and Quadratic terms : Participant Group	4	8.64	<.05
Tonal Context : Cognitive Load	3	11.69	<.01
Tonal Context : Participant Group	6	12.75	<.05
Cognive Load : Participant Group	2	5.81	n.s.
Linear and Quadratic terms : Tonal Context : Cognitive Load	6	19.37	<.01
Linear and Quadratic terms : Tonal Context : Participant Group	12	41.71	<.001
Linear and Quadratic terms : Cognitive Load : Participant Group	4	7.22	n.s.
Tonal Context : Cognitive Load : Participant Group	6	35.71	<.001
Linear and Quadratic terms : Tonal Context : Cognitive Load : Participant Group	12	64.63	<.001
Repetition	3	42.35	<.001
Repetition: Linear and Quadratic Terms	6	11.24	n.s.

A2. *Summary of mixed effects models for duration of target tones in the first and second syllable.*

Duration of the vowel in the first syllable

Fixed effects	Df	χ^2	p
Participant Group	2	18.39	<.001
Tone of 1st syllable	3	313.11	<.001
Tone of 2nd syllable	3	37.39	<.001
Cognitive Load	1	40.81	<.001
Participant Group : Tone of 1st syllable	6	249.96	<.001
Participant Group : Tone of 2nd syllable	6	27.05	<.001
Tone of 1st syllable: Cognitive Load	3	3.77	n.s.
Tone of 2nd syllable: Cognitive Load	3	2.79	n.s.
Participant Group : Tone of 1st syllable: Cognitive Load	8	31.97	<.001
Participant Group : Tone of 2nd syllable: Cognitive Load	6	3.83	n.s.
Repetition	3	0.59	n.s.

Duration of the vowel in the second syllable

Fixed effects	Df	χ^2	p
Participant Group	2	16.30	<.001
Tone of 1st syllable	3	8.15	<.05
Tone of 2nd syllable	3	2619.60	<.001
Cognitive Load	1	62.51	<.001
Participant Group : Tone of 1st syllable	6	11.60	n.s.
Participant Group : Tone of 2nd syllable	6	423.95	<.001
Tone of 1st syllable: Cognitive Load	3	3.35	n.s.
Tone of 2nd syllable: Cognitive Load	3	2.88	n.s.
Participant Group : Tone of 1st syllable: Cognitive Load	8	7.27	n.s.
Participant Group : Tone of 2nd syllable: Cognitive Load	6	5.91	n.s.
Repetition	3	0.97	n.s.

A3. *Pairs of non-words used in the ABX task.*

kasu-tafu
difo-biso
guta-duka
podi-kobi
bapi-gati
tigu-pidu
fuko-supo
soba-foga
muba-nuda

A4. *Stimuli used in the lexical decision task.*

T1→T2		T2→T1	
Words	Non-words	Words	Non-words
应该 *ying1gai1* should	**ying2gai1*	时间 *shi2jian1* time	**shi1jian1*
听说 *ting1shuo1* heard about	**ting2shuo1*	房间 *fang2jian1* room	**fang1jian1*
出发 *chu1fa1* set out	**chu2fa1*	十分 *shi2fen1* very	**shi1fen1*
餐厅 *can1ting1* restaurant	**can2ting1*	爬山 *pa2shan1* mountain climbing	**pa1shan1*
春天 *chun1tian1* spring	**chun2tian1*	黄瓜 *huang2gua1* cucumber	**huang1gua1*
声音 *sheng1yin1* sound	**sheng2yin1*	阳光 *yang2guang1* sunshine	**yang1guang1*
发音 *fa1yin1* pronunciation	**fa2yin1*	成功 *cheng2gong1* success	**cheng1gong1*
出租 *chu1zu1* rent	**chu2zu1*	房租 *fang2zu1* rent	**fang1zu1*
秋天 *qiu1tian1* autumn	**qiu2tian1*	维修 *wei2xiu1* maintain	**wei1xiu1*
西瓜 *xi1gua1* water mellon	**xi2gua1*	服装 *fu2zhuang1* clothing	**fu1zhuang1*

T1→T3		T3→T1	
Words	Non-words	Words	Non-words
医生 *yi1sheng1* doctor	**yi3sheng1*	已经 *yi3jing1* already	**yi1jing1*
今天 *jin1tian1* today	**jin3tian1*	打开 *da3kai1* open	**da1kai1*
关心 *guan1xin1* concern	**guan3xin1*	手机 *shou3ji1* cell phone	**shou1ji1*
中心 *zhong1xin1* center	**zhong3xin1*	饼干 *bing3gan1* cookie	**bing1gan1*
香蕉 *xiang1jiao1* banana	**xiang3jiao1*	小偷 *xiao3tou1* thief	**xiao1tou1*
欧洲 *ou1zhou1* Europe	**ou3zhou1*	首都 *shou3du1* captital	**shou1du1*
西装 *xi1zhuang1* suit	**xi3zhuang1*	北京 *bei3jing1* Beijing	**bei1jing1*
冬天 *dong1tian1* winter	**dong3tian1*	打针 *da3zhen1* injection	**da1zhen1*
身高 *shen1gao1* height	**shen3gao1*	果汁 *guo3zhi1* juice	**guo1zhi1*
书桌 *shu1zhuo1* desk	**shu3zhuo1*	取消 *qu3xiao1* cancel	**qu1xiao1*

T1→T4		T4→T1	
Words	Non-words	Words	Non-words
参加 *can1jia1* participate	**can4jia1*	大家 *da4jia1* everyone	**da1jia1*
糟糕 *zao1gao1* bad	**zao4gao1*	上班 *shang4ban1* work	**shang1ban1*
星期 *xing1qi1* week	**xing4qi1*	再说 *zai4shuo1* what's more	**zai1shuo1*
沙发 *sha1fa1* sofa	**sha4fa1*	蛋糕 *dan4gao1* cake	**dan1gao1*
中间 *zhong1jian1* middle	**zhong4jian1*	客厅 *ke4ting1* living room	**ke1ting1*
冰箱 *bing1xiang1* fridge	**bing4xiang1*	亚洲 *ya4zhou1* Asia	**ya1zhou1*
交通 *jiao1tong1* traffic	**jiao4tong1*	菜单 *cai4dan1* menu	**cai1dan1*
西方 *xi1fang1* west	**xi4fang1*	电车 *dian4che1* tram	**dian1che1*
公斤 *gong1jin1* kilogram	**gong4jin1*	互相 *hu4xiang1* mutual	**hu1xiang1*
搬家 *ban1jia1* move	**ban4jia1*	信息 *xin4xi1* information	**xin1xi1*

T2→T3		T3→T2	
Words	Non-words	Words	Non-words
离开 *li2kai1* leave	**li3kai1*	小心 *xiao3xin1* be careful	**xiao2xin1*
回家 *hui2jia1* go home	**hui3jia1*	纽约 *niu3yue1* New York	**niu2yue1*
直接 *zhi2jie1* direct	**zhi3jie1*	火车 *huo3che1* train	**huo2che1*
昨天 *zuo2tian1* yesterday	**zuo3tian1*	好吃 *hao3chi1* delicious	**hao2chi1*
聊天 *liao2tian1* chat	**liao3tian1*	海边 *hai3bian1* beach	**hai2bian1*
前天 *qian2tian1* day before yesterday	**qian3tian1*	打工 *da3gong1* work	**da2gong1*
时差 *shi2cha1* jet lag	**shi3cha1*	耳机 *er3ji1* earphone	**er2ji1*
其中 *qi2zhong1* among	**qi3zhong1*	有关 *you3guan1* related	**you2guan1*
职工 *zhi2gong1* staff	**zhi3gong1*	打车 *da3che1* take a taxi	**da2che1*
骑车 *qi2che1* cycling	**qi3che1*	普通 *pu3tong1* ordinary	**pu2tong1*

T2→T4		T4→T2	
Words	Non-words	Words	Non-words
明天 *ming2tian1* tomorrow	**ming4tian1*	认真 *ren4zhen*1 serious	**ren2zhen1*
结婚 *jie2hun1* get married	**jie4hun1*	第三 *di4san1* third	**di2san1*
年轻 *nian2qing1* young	**nian4qing1*	现金 *xian4jin1* cash	**xian2jin1*
提高 *ti2gao1* improve	**ti4gao1*	大哥 *da4ge1* big brother	**da2ge1*
读书 *du2shu1* reading	**du4shu1*	衬衫 *chen4shan1* shirt	**chen2shan1*
十八 *shi2ba1* eighteen	**shi4ba1*	录音 *lu4yin1* sound recording	**lu2yin1*
旁边 *pang2bian1* next to	**pang4bian1*	夏天 *xia4tian1* summer	**xia2tian1*
离婚 *li2hun1* divorce	**li4hun1*	念书 *nian4shu1* study	**nian2shu1*
黄金 *huang2jin1* gold	**huang4jin1*	下车 *xia4che1* get off	**xia2che1*
南京 *nan2jing1* Nanjing	**nan4jing1*	大约 *da4yue1* about	**da2yue1*

T3→T4		T4→T3	
Words	Non-words	Words	Non-words
简单 _jian3dan1_ simple	_*jian4dan1_	律师 _lv4shi1_ lawyer	_*lv3shi1_
紧张 _jin3zhang1_ nervous	_*jin4zhang1_	汽车 _qi4che1_ car	_*qi3che1_
老师 _lao3shi1_ teather	_*lao4shi1_	健康 _jian4kang1_ health	_*jian3kang1_
小说 _xiao3shuo1_ novel	_*xiao4shuo1_	唱歌 _chang4ge1_ sing	_*chang3ge1_
起飞 _qi3fei1_ take off	_*qi4fei1_	面包 _mian4bao1_ bread	_*mian3bao1_
好听 _hao3ting1_ pleasant to hear	_*hao4ting1_	大衣 _da4yi1_ coat	_*da3yi1_
海关 _hai3guan1_ customs	_*hai4guan1_	半天 _ban4tian1_ quite a while	_*ban3tian1_
转机 _zhuan3ji1_ transfer	_*zhuan4ji1_	用功 _yong4gong1_ hardworking	_*yong3gong1_
许多 _xu3duo1_ many	_*xu4duo1_	战争 _zhan4zheng1_ war	_*zhan3zheng1_
酒吧 _jiu3ba1_ bar	_*jiu4ba1_	更加 _geng4jia1_ more	_*geng3jia1_

Segment control pairs			
Words	Non-words	Words	Non-words
生活 *sheng1huo2* life	**cheng1huo2*	成为 *cheng2wei2* become	**sheng2wei2*
超级 *chao1ji2* super	**shao1ji2*	回答 *hui2da2* answer	**kui2da2*
刚才 *gang1cai2* just now	**kang1cai2*	学习 *xue2xi2* study	**jue2xi2*
公园 *gong1yuan2* park	**hong1yuan2*	厨房 *chu2fang2* kichen	**shu2fang2*
中国 *zhong1guo2* China	**chong1guo2*	前年 *qian2nian2* year before last	**xian2nian2*
开门 *kai1men2* open the door	**gai1men2*	着急 *zhao2ji2* worry	**shao2ji2*
中文 *zhong1wen2* Chinese	**chong1wen2*	足球 *zu2qiu2* football	**su2qiu2*
商人 *shang1ren2* businessman	**zhang1ren2*	实习 *shi2xi2* internship	**chi2xi2*
忽然 *hu1ran2* suddenly	**ku1ran2*	荷兰 *he2lan2* Netherlands	**ke2lan2*
高楼 *gao1lou2* high-rise	**hao1lou2*	零食 *ling2shi2* snack	**ning2shi2*

Segment control pairs			
Words	Non-words	Words	Non-words
小时 *xiao3shi2* hour	**jiao3shi2*	认为 *ren4wei2* think	**zhen4wei2*
警察 *jing3cha2* policeman	**xing3cha2*	进来 *jin4lai2* come in	**xin4lai2*
解决 *jie3jue2* solve	**xie3jue2*	大学 *da4xue2* university	**ta4xue2*
检查 *jian3cha2* check	**xian3cha2*	去年 *qu4nian2* last year	**ju4nian2*
旅行 *lv3xing2* trip	**nv3xing2*	适合 *shi4he2* suitable	**chi4he2*
打球 *da3qiu2* play ball games	**ta3qiu2*	上楼 *shang4lou2* go upstairs	**chang4lou2*
美元 *mei3yuan2* dollar	**bei3yuan2*	客人 *ke4ren2* guest	**he4ren2*
打折 *da3zhe2* give a discount	**ta3zhe2*	地图 *di4tu2* map	**ti4tu2*
奶油 *nai3you2* cream	**dai3you2*	气球 *qi4qiu2* balloon	**ji4qiu2*
海牙 *hai3ya2* The Hague	**kai3ya2*	课文 *ke4wen2* text	**ge4wen2*

Curriculum vitae

Ting Zou was born on February 21 in 1986 in Tianjin, China. She attended Nankai University from 2004 until 2008 and obtained bachelor degrees in English and in Chinese language and literature. In 2009 she started her master programme at Beijing Language and Culture University and obtained her master's degree in Linguistics and Applied linguistics in 2012. From 2009 to 2012, she received academic training in Experimental phonetics and conducted experimental research on the acquisition of Mandarin by Japanese learners. During these three years, she also worked as a part-time language instructor, teaching Mandarin Chinese to students with various language backgrounds. She began her PhD research on tone acquisition by Dutch learners of Mandarin at Leiden University Centre for Linguistics in 2012. The present thesis is the result of that research.