MASTER THESIS

L2 perception of lexical tones by beginning and advanced Dutch learners of Mandarin

A thesis submitted in fulfillment of the requirements for the degree of Master of Arts

at the

Research Master Linguistics, Language & Cognition

Faculty of Arts

Author: Judith Kelholt
E-mail: j.kelholt@student.rug.nl
S-number: 3273695
Word count: 17756

Supervisor: Prof. dr. Wander Lowie
Second reader: dr. Anja Schüppert

2 August 2018
Table of contents

1. **Introduction** .......................................................... 5
2. **Background** .......................................................... 8
   2.1. The role of pitch information in Mandarin and Dutch ........ 8
       2.1.1. Mandarin lexical tones .................................. 8
       2.1.2. The Dutch intonation system ............................. 9
       2.1.3. The categorical perception paradigm ..................... 11
2.2. Theories on non-native phonetic perception ..................... 13
       2.2.1. The PAM and the SLM models ............................ 13
       2.2.2. An opposing model on L2 speech perception ............ 15
       2.2.3. The ASP model ........................................... 17
2.3. Native Mandarin tone perception in infancy and adulthood ..... 19
2.4. Non-native Mandarin tone perception in adulthood after short-term training ........................................ 24
2.5. Non-native Mandarin tone perception in adulthood after long-term training ............................................. 26
3. **The present study** .................................................. 31
   3.1. Experiment 1: Identification task ............................. 33
       3.1.1. Participants .............................................. 33
       3.1.2. Materials and design .................................... 36
       3.1.3. Procedure ............................................... 37
       3.1.4. Results .................................................. 38
   3.2. Experiment 2: Discrimination task ............................. 42
       3.2.1. Materials and design .................................... 42
       3.2.2. Procedure ............................................... 43
       3.2.3. Results .................................................. 44
4. **Discussion** ......................................................... 47
5. **Conclusion** .......................................................... 52
References ........................................................................ 55
Appendices ........................................................................ 61
   A1: Language Background Questionnaire Short Version ........ 61
   A2: Language Background Questionnaire Extended Version ...... 65
Acknowledgments

During the course of writing this thesis, I was lucky enough to collaborate with many great people.

First of all, I had the honor to work in Dr. Carolyn Quam’s lab at Portland State University. Our conversations about Mandarin tonality and speech perception planted the seed for the topic of this thesis. Moreover, Dr. Quam was a great supervisor who I could turn to anytime. Thank you, Carolyn!

After returning from Portland, I started working on setting up the experiments and start looking for people with Mandarin as a mother tongue and Dutch people learning Mandarin. For being able to recruit over 70 participants, I owe my thanks to language teachers at the Groningen Confucius Institute, the Language Centre Groningen and Chinese Studies at Leiden University. I am also grateful for everyone who took part in the study and shared their experiences with Chinese language and culture.

Once the data collection process was finished, the writing process started. This not being my strongest side, my supervisor prof. dr. Wander Lowie was a great help in organizing my thoughts and writing them down in an academic (yet understandable) way. Thank you, professor Lowie!

I would also like to thank my family, friends and classmates who were supporting me during the course of the master’s program.

Last, but definitely not least, there is one person who I cannot thank enough. V, no words needed, or maybe just one: Szeretlek.
(NE LÉGY SZELES...)

Ne légy szeles.
Bár a munkádon más keres –
dolgozni csak pontosan, szépen,
ahogy a csillag megy az égen,
ugy érdemes.

[Attila József, 1935-1937]
1 Introduction

In recent years, there has been an increase in the number of Dutch educational institutions offering courses in Chinese language and culture (Funnekotter, 2013). For instance, a pilot at the pre-university education has started in 2010 where initially 9 schools offered a final exam in Chinese. In 2018, the number of schools increased to 70, and by 2018, Chinese became an official final exam (Visbeen, 2018). With the increasing interest in Chinese culture and language in the Netherlands, the difficulties of the learning Mandarin, one of the most dominant Chinese dialect groups (Thurgood & LaPolla, 2017) among Dutch speakers became more apparent and therefore of high interest to study.

The biggest challenge in learning Mandarin as a second language is generally attributed to the acquisition of the lexical tone system. Difficulties arise especially when the learner has no prior experience with tonal languages (Burnham & Jones, 2002). Lexical tones are used to differentiate meaning in spoken and written Mandarin. That is, if the same Mandarin word is pronounced or written in different tones, it will carry four different meanings. Therefore, the correct use and production of tones in Mandarin is essential for meaningful communication (Yip, 2002).

Tones can be phonetically described as different pitch contours with fundamental frequency (F0) as most relevant varying acoustic property. In non-tonal languages like Dutch, pitch is not used to assign meaning to words, except for a dozen exceptional word pairs where pitch signals word stress and along with other linguistic cues, helps to distinguish meanings of identical word forms (e.g. VOORkomen ‘to occur’ vs. voorKOMEN ‘to prevent’) (Cutler & Van Donselaar, 2001). Apart from these rare exceptions, pitch in Dutch is generally used to mark sentence types (Van Heuven & Haan, 2002), or express paralinguistic meanings like emotion or attitude of the speaker (Shattuck-Hufnagel & Turk, 1996).

This fundamental difference in the role of pitch in Mandarin and Dutch is manifested in listeners’ speech perception as early as infancy. Starting from 6 months
of age, speakers of a tonal language show fine-tuning towards native speech categories like lexical tones and become less alert to acoustic differences in non-native speech (Mattock & Burnham, 2006). Later, in adulthood, tonal speakers are shown to categorize continuous tonal input into distinct categories and be less sensitive to acoustic differences within those categories. This phenomenon is usually examined with the categorical perception paradigm (CP) (Warren, 2012), where listeners are presented with acoustically manipulated tones that make up a continuum and are asked to categorize and discriminate the sounds they heard. When speakers of a non-tonal L1 are tested on tones in a CP paradigm, they do not show categorical perception of tones. However, they tend to be highly accurate in discriminating the components of the tonal continuum (X. Wang, 2013). Thus, whereas tonal L1 speakers tend to categorize tonal input into distinct categories and suppress slight variation in the speech input, non-tonal L1 speakers fail to form categories and are more attentive to the differences in the tonal continuum.

Based on the observations of tonal and non-tonal L1 speakers, the question arises how lexical tones are perceived by beginner and advanced learners of a tonal language with a non-tonal L1. In order to achieve native-like tone perception, L2 learners should become less sensitive to variation in the speech input and detect phonological contrasts to identify lexical tones. Thus, the central question of this study is how exposure to non-native tones affects the perception of L2 tones.

To address this question, a group of adult Dutch beginner and advanced learners of Mandarin Chinese took part in a categorical perception study consisting of an identification and discrimination task. In the identification task, participants listened to Mandarin fragments manipulated in their fundamental frequency and were asked to categorize them in the corresponding tonal category. In the discrimination task, learners heard three Mandarin fragments where two were identical and one was a neighbor in the tonal continuum. Participants were asked to pay close attention to the pitch movement of the fragments and decide which fragment was the odd one out.
By testing two groups of learners differing in their length of training and exposure to Mandarin, the effect of tonal exposure on L2 tone perception were examined. Furthermore, other confounding factors that influence L2 speech perception were controlled for, such as experience with non-linguistic tones and dialectal tones. The examination of two learning groups aims to elucidate whether increasing L2 exposure changes the sensitivity to tone contrast changes and facilitates more native-like perception.
2 Background

2.1. The role of pitch information in Mandarin and Dutch

2.1.1. Mandarin lexical tones

Mandarin is a tone language. The linguistic term tone refers to a distinctive phonemic category that distinguishes two utterances in a language. In phonetic terms, Mandarin tones are described by components such as duration, amplitude and fundamental frequency (F0). In order to find out what the primary cue is that is used for identifying tones in auditory input, previous studies tested the contributions of these components towards the ability of L1 Mandarin speakers’ ability to discriminate tonal fragments. In a study by Abramson (1979), the acoustic properties of tones were manipulated so that all acoustic cues were removed except for the F0. In this case, native Mandarin speakers were still able to discriminate between the tonal stimuli they heard. However, when F0 was removed and other cues like duration and amplitude were left intact, tonal discrimination decreased among the Mandarin native speakers. In this latter case, amplitude seemed to have become the most important cue for discriminating tones. Interestingly, Mandarin amplitude and fundamental frequency are closely related so that over the course of a falling tone, their contour will be falling at the same rate. These contours in the speech signal are also referred to as pitch movements. In sum, whereas both F0 and amplitude contribute to tone discrimination as their pitch movements are in some tones mirrored, F0 does indeed seem the primary cue for distinguishing tones (Yip, 2002).

In Mandarin Chinese, tonal information is an integral part of the speech input, and the pitch contour in which a segment is uttered defines its lexical meaning. Thus, if the same utterance, for example the syllable /ma/ is pronounced in four different tones (henceforth abbreviated to T1, T2, T3 and T4), it will have four different meanings: ‘mother’ in /maT1/, ‘hemp’ in /maT2/, ‘horse’ in /maT3/ and ‘curse’ in
L2 perception of lexical tones by beginning and advanced Dutch learners of Mandarin

J. Kelholt

/maT4/. As tones are imposed onto segments and occur together, they are also referred to as suprasegmentals. The four main tones in Mandarin are illustrated in Figure 1.

![Figure 1. Pitch contours of the four Mandarin tones (Nixon, Chen, & Schiller, 2015).](image)

Expressed in terms of pitch contours, T1 is a high-level tone; T2 is a mid-rising tone, T3 is a low-level tone; T4 is a high-falling tone. The fifth, neutral tone always comes at the end of a word or a phrase, and is associated with a weak syllable (Chen & Xu, 2006). The current study will focus on the T1-T4 contrast. As Figure 1 shows, both tones have approximately the same starting point both in time and pitch height. While T1 is pronounced with a constant high pitch, T4 gradually decreases in pitch over time. Due to this difference in the course of pitch movement between T1 and T4, this contrast is judged by native listeners as one of the most acoustically salient in Mandarin (Huang & Johnson, 2010).

2.1.1. The Dutch intonation system

In Standard Dutch, the most widely spoken variety of Dutch (Taalunie, 2017), pitch movements can also be used on the segmental level. In a few exceptional word pairs like VOORkomen (‘to occur’) and voorKOMEN (‘to prevent’) where meaning is expressed by different pitch movements, pitch however does not generally play a role
in establishing lexical meaning in Dutch. Instead, the most important functions of pitch movements occur on the post-lexical level. This means that in Dutch, pitch mostly conveys additional meaning to an utterance after its lexical meaning has already been derived. The functions of pitch can be either linguistic and paralinguistic. In conveying linguistic meanings, pitch movements can firstly indicate prominence at the word or sentence levels (Yi Xu & Xu, 2005). Thus, applying pitch to a syllable in a word, or a word in a sentence can express its relative prominence in the linguistic context it is embedded in. Secondly, pitch can delimit prosodic constituents, such as intonation phrases or paragraphs in multi-sentence utterances (Cole, 2015). That is, in natural speech, when a sentence or multiple sentences are read aloud with a different pitch than other sentences in the speaker’s speech, it will place the concerning sentence(s) in focus and indicate the relative importance or different status in the utterance for the listener. Lastly, pitch movements can also provide information on the sentence type, such as questions, statements and interrogatives (Van Heuven & Haan, 2002). Pitch movements can also convey paralinguistic meanings, such as emotions (e.g. joy or anger) and attitude (e.g. politeness or irony) of the speaker (Shattuck-Hufnagel & Turk, 1996).

Summarized, both the Mandarin and Dutch use pitch information to convey post-linguistic meanings. However, the functions and levels at which they do this are strikingly different. In Mandarin, pitch differences make up four main tone categories. They are suprasegmentals applicable to words or segments and are used to define lexical meaning. In Dutch, pitch is mostly used on the post-lexical level to convey linguistic and paralinguistic meanings. Due to these fundamental differences between the phonetic and phonemic systems of Dutch and Mandarin, the difficulties regarding the acquisition of lexical tones become more evident. As learners of Standard Dutch have no experience in using pitch to convey lexical meaning in their L1, they encounter difficulties in correctly perceiving and using pitch in Mandarin. In order to find out what the difficulties are in correctly acquiring tones for speakers of a non-tonal language, it is important to understand how pitch movements like tones are perceived.
The following section will describe the categorical perception paradigm, a tool that is widely used to study speech perception.

2.1.3. The categorical perception paradigm

The categorical perception (CP) paradigm is a method based on the observation that auditory input manipulated along a phonetic continuum is perceived as belonging to categories, rather than the continuous steps they are (Warren, 2012). Most CP experiments consist of an identification task and a discrimination task. In the identification task, listeners are presented with samples from the continuum one at the time and are asked to assign them to one of the two categories in question. After plotting the mean responses for each step, an abrupt change in the number of responses at one point on the continuum is visible. This switching point in responses suggests a category boundary along the continuum. When plotting the responses in a line diagram, this is manifested by an S-curve, referring to the shape of the identification curve, illustrated in Figure 2.

![Figure 2. A typical S-curve of identification responses along a phonetic continuum (Warren, 2012).](image)
As Figure 2 shows, instead of a smooth linear line, the mean responses (y-axis) for each step (x-axis) seem to form categories from steps 1 to 4 and steps 4 to 7. In this example, step 4 is the category boundary that divides the two types of responses (X or Y) along the continuum. In the discrimination task, listeners are presented with neighbouring steps in the continuum and are asked whether the pairs are the same or different. According to the CP hypothesis, listeners are less accurate in discriminating the stimuli when they fall within the same category boundary and are more accurate when they occur in two different categories (Warren, 2012; Repp, 1985).

The earliest studies using the CP paradigm primarily investigated the perception of segments (Abramson & Lisker, 1967). In a cross-language study including Spanish, English and Thai, participants were presented with a continuum that was manipulated in its voice onset time (VOT). Supporting the CP paradigm, they found category boundaries in the continuum for all language groups, specific to their L1. Although the CP paradigm has most widely been used for segments, it has been extended to suprasegmentals such as lexical tones as well (Burnham & Jones, 2002; Hallé, Chang, & Best, 2004; Peng et al., 2010; Yisheng Xu, Gandour, & Francis, 2006). In most studies investigating the categorical perception of tones, two tones are selected and manipulated in their F0. In the identification task, listeners are asked to identify each stimulus as one of the existing tones in their native language. In the discrimination task, they are generally presented with pairs or triplets of tones that are one step apart in the tonal continuum. CP studies conducted with tonal speakers reveal clear category boundaries along the tonal continuum, suggesting categorical perception of tones. When the same experiment is conducted with speakers of a non-tonal language, no CP pattern is visible (Hallé et al., 2004; Peng et al., 2010).

In the following sections, theoretical models will be introduced that address the mechanisms of speech perception in L1 and L2. Most importantly, they formulate hypotheses on whether perception can be altered so that non-tonal speakers are able to perceive tones in a categorical fashion. In section 2.3., empirical studies on infant and adult phonetic perception will be discussed. Sections 2.4. and 2.5. will provide an
overview of adult L2 perception studies in after short-term and long-term training in Mandarin lexical tones. Chapter 3 will be devoted to the discussion of the current study, consisting of the forced-choice tone identification task (section 3.1.) and the AXB tone discrimination task (section 3.2.). The remaining chapters are devoted to the discussion of the results (Chapter 4), and the final conclusions (Chapter 5).

2.2. Theories of native and non-native perception

2.2.1. The PAM and the SLM models

Studies conducted in the 1970’s and 1980’s investigated the perception of L1 and L2 phonemes by native speakers and L2 learners of various language pairs. The most influential studies of these decades demonstrated that phonemes – either L1 or L2 – are generally perceived based on the existing L1 phonemic categories (Abramson & Lisker, 1970; Werker & Tees, 1984; Williams, 1977). To explain how L2 sounds are assimilated into L1 categories, theoretical models were proposed, among that of Best (1994) and Flege (1995).

The *Perceptual Assimilation Model* (PAM) by Best (1994) accounts for the role of learners’ L1 phonological systems in the perception of L2 sounds. By comparing the articulatory similarities of an L1 and L2 sounds, the PAM makes predictions about how well L2 sounds can be assimilated to existing L1 sound categories. If an L2 sound contrast is assimilated to separate L1 categories, the perception of the particular contrast should be excellent. However, if the L2 contrast is assimilated to the same L1 category, the learner will not be able to perceive the difference between them and discrimination will therefore be poor. The PAM model was recently extended to explain how suprasegmentals such as lexical tones are assimilated in a non-tonal L1. This model is called the *Perceptual Assimilation Model for Suprasegmentals* (PAM-S) (So & Best, 2014a). The PAM-S refers to the Mandarin tonal system as prosodic categories.
to which non-tonal intonation categories (i-Categories) are assimilated to. For example, the Mandarin T1 (high-level tone) shows similarities to the English pitch movement of a statement, one of the i-Categories. When the four Mandarin tones are assimilated to four different i-Categories, discrimination of the Mandarin tonal contrasts are should to be excellent. Consequently, the acquisition of the Mandarin phonetic system will be successful. However, if the learner fails to assign the Mandarin tones into distinct i-Categories and assimilates them into a single category, discrimination of Mandarin tones is predicted to be poor and the Mandarin phonetic system will be incorrectly acquired. The PAM and PAM-S models were applied in several studies including various L1 and L2 pairs. Studies including Mandarin as L1 and a non-tonal language as L2 found support for assimilation patterns predicted by the model (Hallé et al., 2004; Reid et al., 2015; So & Best, 2014b). Although these findings support the PAM-S, a point of critique can be raised regarding the use of i-Categories. While the assimilation of tones into non-tonal intonation categories align with the model’s predictions, this does not necessarily justify the existence of such i-Categories that a non-tonal listener is supposed to use. Intonation patterns in non-tonal languages like statement, question and exclamation are more likely functions, rather distinct perceptual categories in the L1 language system. Therefore, they cannot be compared with the lexical tones which have been extensively studied using the categorical perception paradigm and were indeed found to be perceived as distinct categories (Abramson & Lisker, 1967; Gandour, 1983; Repp, 1985).

Another model that addresses the acquisition of L2 sounds is the Speech Learning Model (SLM) by Flege (1995). The SLM focuses on how speech perception plays a role in L2 sound acquisition by comparing the acquisition of ‘new’ and ‘similar’ L2 sounds. According to this model, new sounds are not identifiable with any L1 sound, while similar sounds show acoustic similarities to an L1 sound. According to the SLM, L2 acquisition uses the same learning mechanisms and processes that were involved in the L1 and as they remain intact over the life span, they can be used in L2 acquisition for new sounds. On the other hand, acquisition of similar sounds is hindered by the so
called equivalence classification process and as a result, no new phonetic categories can be established for similar sounds. While the SLM discriminates two acquisition processes based on whether the L2 sound is new or, it lacks elaboration on how equivalence classification of similar sounds proceeds. Moreover, this model does not reflect on whether or how the equation of similar sounds is affected by increasing exposure to L2 sounds during the course of learning.

It is evident that the two models take different approaches to explain the assimilation of L2 sounds in the L1. In order to compare a tonal L1 and a non-tonal L2, the PAM-S model claims the existence of i-Categories that non-native tones are assimilated to. Contrary to PAM-S, the SLM makes the distinction between L1 and L2 sounds rather than intonation and tone categories. According to the SLM, an L2 sound can be classified as either ‘new’ or ‘similar’, based on similarities with an existing L1 sound. Although the PAM and SLM models take different approaches in explaining the perception of non-native speech sounds, they both acknowledge the existence of tone categories and the categorical perception of lexical tones by tonal L1 listeners. Furthermore, they both entail that the correct perception of L2 speech sounds is possible for L2 listeners, although this is based on how the L2 sound is assimilated to existing speech sounds or intonation categories. An alternative speech learning model that contradicts this assumption is the L2 phonology acquisition model by Brown (2000), described below.

2.2.2. An opposing model on L2 speech perception

Based on the PAM and SLM models that describe the influence of the L1 on L2 sound perception, Brown aims to address why this is the case. To understand adult L2 perception, this model first takes a developmental approach and looks into infant speech perception. Initially, infants as young as one-month old are already sensitive to both native and non-native contrasts in the language input. This capacity of discriminating acoustically different non-native contrasts decreases with increasing L1
exposure until the child is only able to discriminate native contrasts. As the perceptual ability decreases for non-native contrasts, it increases for native contrasts. The improvement of discriminating native contrasts and consequently the acquisition of the L1 phonological system then imposes perceptual boundaries within which phonemic categories are perceived in speech perception development. Assumingly, the L1 phonological system continues to constrain speech perception in adults as well. In adulthood, the phonological system filters out allophonic details in the acoustic signal and funnels them into a single L1 category, a phenomenon also described in the categorical perception paradigm. This process aids processing and comprehension for native speakers who are continuously faced with variable realizations due to coarticulation or inter-speaker variability. By filtering out this irrelevant “noise” in the signal, processing can proceed more quickly. So far, Brown’s model is in line with the hypotheses of the categorical perception paradigm and therefore with the PAM-S and SLM models as well. However, Brown takes a counterview by posing that while categorical perception is a useful method for native speech perception, it will hinder correct perception of L2 contrasts when variation that is filtered out by the L1 is relevant in the L2. This will then enable the perception (including identification and discrimination) of L2 contrasts and will hinder the L2 acquisition in other linguistic domains as well. Brown’s model then makes an even stronger claim by posing that discriminating non-native contrasts will not only be hindered when the contrast is not present in the L1, but for any contrast along an acoustic dimension defined by a given underlying feature, for example the F0 in lexical tones. Thus, since tonal contrasts are not phonemic in Dutch, as opposed to Mandarin, this will lead to a persistent impairment in tonal perception for Dutch learners of Mandarin and should therefore not trigger acquisition. In addition to L1 features, existing L1 phonological categories also hinder the construction of new L2 categories. This again will result in the incorrect acquisition of the Mandarin tone system. While Brown’s model seems to provide a more elaborate view on L2 speech perception than the PAM and SLM do, it does not encounter for the role of L2 exposure in the perception of non-native contrasts. If
Brown’s model is correct, the same L1 perceptual mechanisms are used for L2 acquisition. If the L1 and L2 do not share the same underlying feature, none of the non-native contrasts will be acquired, regardless of the contrast being similar or new. Moreover, the role of L2 exposure in this model is neglected as well, meaning that no matter how much the learner is exposed to the non-native contrast, it will not be correctly acquired. In the next section, a fourth – more plausible – speech perception model will be introduced that does assume sensitivity to non-native contrasts and the possibility of acquiring them by using different modes of perception.

2.2.3. The ASP model

A model that seems the most plausible to explain L1 and L2 speech perception and the processes involved in L2 phonetic acquisition is the Automatic Selective Perception (ASP) model by Strange (2011). This model emphasizes the role of attention in language acquisition and differentiates between a phonological and phonetic mode of perception. In the phonological mode, perception is accomplished via the employment of L1 selective perception routines (SPRs) that draw the attention of the listener to phonologically contrastive information for identifying word forms in the L1. Therefore, fine-grained allophonic details are ignored in this mode of perception. These SPRs in the phonological mode are primarily used by native speakers as a result of intensive exposure to the L1 and become overlearned routines used in word-form detection. Therefore, this mode of processing costs the least cognitive effort. This mode of perception is in essence the manifestation of the categorical perception paradigm and aligns with the other speech learning models’ theories on L1 speech perception. Opposed to the phonological mode, L2 learners employ the so called phonetic mode of processing when processing novel L2 phonological contrasts. As L2 learners have not had sufficient exposure to the L1, they are using their native SPR’s to detect phonologically contrastive information in the L2. As a result, beginning L2 learners’ attention will be drawn to fine-grained allophonic details in the L2 input that might
not necessarily contribute to word form recognition. This mode of processing involved the processing of redundant details in the input and therefore slows down the detection of phonological contrasts that are relevant in the L2. Consequently, this mode of processing requires bigger cognitive effort than the phonological mode.

Although the ASP model attributes the use of the phonological and phonetic modes of processing primarily to native listeners and L2 learners respectively, it predicts that L2 learners are able to shift to the phonological mode and L1 listeners can activate the phonetic mode of processing. For L2 learners, this translates to the ability to develop new SPR’s that optimize attention to phonologically relevant information in the L2 and suppress the detection of allophonic details that do not contribute to word-form recognition. If L2 learners are indeed able to use the phonological mode of perception, they should also be able to attain native-like perception of non-native phonological categories. This view on evolving L2 perception is shared by other speech learning models like PAM-S that also supports that speech perception can evolve and become more native-like as a result of L2 exposure. However, it does not align with the predictions of Brown (2000) who disagrees with the possibility of attaining native-like perception of L2 speech sounds. As for native speakers, the ASP also predicts the ability to switch to the phonetic mode of processing and detect allophonic details in the language input if necessary. This suggests that native speakers do not lose the ability to perceive non-contrastive details. According to the ASP model, the ability to switch from one mode to the other as an L1 or L2 listener depends on the task demands and stimulus complexity. In a task where the memory load is high and the stimulus consists of high phonetic variability, the phonological mode of processing is required. As beginner L2 learners do not yet have the selective perception routines as L1 listeners do, they are predicted to perform poorly on such task. Native listeners, however should perform well as it requires the use of their primary, phonological mode of perception. In a task with low stimulus variability and cognitive load, the phonetic mode of processing is asked for and as beginner learners are hypothesized to primary rely on this mode, their performance should be good. Native speakers should
in this case be able to switch to the phonetic mode and perform well on this task as well.

Thus far, no systematic research has been done to investigate the processing of suprasegmentals in both identification and discrimination tasks by beginning and advanced L2 learners. The current study therefore studies the same group of participants on both an identification task and a discrimination task to test to what extent they are able to identify tones in the phonological mode of processing and discriminate between the allophonic details in the phonetic mode of perception.

To get a better insight in the difficulties around acquiring Mandarin lexical tones as a second language learner, it is important to understand how Mandarin native speakers learn and perceive tones. Findings from L1 tone perception studies can then be applied to short-term intensive laboratory training studies where native non-tonal language speakers are exposed to lexical tones and shed light on how non-native input is perceived by naïve listeners. While short-term training studies provide useful information about L2 tone perception, carrying out studies in a cross-sectional manner are perhaps even more informative. In cross-sectional investigations, the same experimental tasks are used as in short-term studies, but the participants are not naïve non-tonal language speakers. They are already trained in Mandarin tones and can be classified as beginner or advanced learners, mostly based on the years and level of instruction in Mandarin. The following paragraphs aim to provide an overview of the relevant studies carried out in lexical tones perception in tonal and non-tonal infants and adults, as well as short-term and long-term training studies.

2.3. Native Mandarin tone perception in infancy and adulthood

In the first year of life, infants can discriminate a wide range of speech contrasts, both native and non-native. After the first year, however, the discrimination of non-native
contrasts gradually decreases and attention will be more drawn to native contrasts instead. This process is called *perceptual reorganization* and when it is completed, categorical perception of native sounds will be achieved. After the formation of native categories, the perception of non-native contrasts will be more difficult (Werker & Logan, 1985). In a longitudinal study on perceptual reorganization for Mandarin lexical tones, 6- and 9-month old Mandarin and English infants were tested (Mattock & Burnham, 2006). As an additional factor, infants heard speech and non-speech (violin sound) stimuli. It was found that Chinese infants were overall better at discriminating speech and non-speech stimuli at both 6 and 9 months. English infants’ discrimination abilities for speech tones declined between 6 and 9 months of age, while their non-speech discrimination remained constant. This suggests that attunement to native speech contrasts is visible as early as 6- and 9 months and is specific for contrasts in language input (Anderson, Morgan, & White, 2003). These findings are supported with work by Yeung and colleagues (2013) who investigated the schedule of the perception of speech segments. They found that language-specific perception in Mandarin and Cantonese infants first emerges for phonemic contrasts like lexical stress and tone, until ca. 5 months of age.

By the time a Mandarin-learning baby turns 1-year-old, (s)he is also able to discriminate contrasts within the lexical tone categories. Using an acoustically salient tone contrast and other acoustically similar, less salient tone contrasts, Tsao (2008) found that 1 year-old Mandarin babies are better at discriminating salient tone contrasts than similar ones. This finding aligns with predictions of the CP paradigm in discrimination tasks where within-category contrasts are supposed to be harder to discriminate that between-category contrasts (Warren, 2012).

In Mandarin-speaking adults, the perceptual categories for lexical tone are fully developed after childhood and have become well-grounded, thanks to extensive L1 tonal input in adulthood (Yip, 2002). In an investigation by Hallé, Chang and Best (2004), native Taiwanese (Mandarin speaking) and French listeners were tested on their Mandarin tone identification and discrimination abilities. They were presented
with a tonal continuum varying in fundamental frequency and were asked to identify and discriminate the stimuli in the continuum. Results indicated that Taiwanese listeners perceived tones in a categorical fashion, while French listeners were more sensitive to the differences in steps along the continuum. Thus, while Taiwanese listeners neglected acoustic variation that is not informative to identify the Mandarin tones, French listeners were more attentive to psychophysical variation in the stimuli, despite the absence of tonal contrasts in the French phonetic system, suggesting the use of the phonetic mode of perception. These results lend support for the categorical perception paradigm where native tonal listeners are expected to form categories using the ASP model’s phonological mode and neglect acoustic variation that is not contrastive for tones. For non-native listeners, categorical perception does not occur because they can only access the phonetic mode of perception.

In a study by Peng and colleagues (2010) also using the CP paradigm, native German and native Chinese listeners were tested on their Mandarin and Cantonese perception abilities. Interestingly, both the German and the Chinese listeners showed CP in identifying the tones. However, the German listeners’ category boundaries were less sharp than that of the Chinese listeners. In the discrimination tasks, the native German listeners were more sensitive to changes in the tonal continuum, while the Chinese listeners exhibited linguistic boundaries which reflected their native Mandarin tone categories.

A study by Burnham and Jones (2002) extended the CP paradigm by testing native Thai (a tone language that distinguishes five lexical tones) and native Australian English listeners on musical tones, sine waves and an artificial tone system, in addition to a Thai tonal continuum. They found that both groups showed CP, however, Thai speakers perceived the Thai tones and the artificial tones more categorically than the English listeners did. However, for the non-speech continua (music and sine stimuli), the perception of the Thai listeners was equivalent to the English listeners. Based on these findings, the authors concluded that L1 experience with lexical tones facilitate CP of unfamiliar tones. However, there is no carry-over of the CP effect to other non-
linguistic domains. Another important conclusion is that the native English speakers showed CP-like perception for both Thai, artificial and the non-linguistic continua even after a short exposure to F0 manipulated stimuli.

In a related study, Xu, Gandour and Francis (2006) tested Mandarin Chinese and English listeners on the perception of a Mandarin tone contrast continuum and non-speech tones. Contrary to Burnham and Jones’s findings with Australian English listeners, the English listeners in this study did not show any CP pattern in the Mandarin tone condition, only the Mandarin group. These findings suggest that long-term exposure to categorical representations of tonal categories facilitates CP. In the non-speech condition, the Mandarin group performed similarly to the speech condition. The English speakers also showed a CP pattern, however weaker than the Mandarin group. This suggest that CP might be domain-general and therefore accessible for non-tonal speakers as well.

Whether only experience with linguistic tones facilitates CP of non-native linguistic tonal contrasts has been studied by comparing musicians and non-musicians on their tone perception abilities. Alexander and colleagues (2005) for example asked English-speaking non-musicians and musicians to identify and discriminate the four Mandarin tones. They found that both in identification and discrimination, musicians significantly outperformed the non-musicians (see also Lee & Hung, 2008 and Wong, Skoe, Russo, Dees, & Kraus, 2007). In a study by Cooper and Wang (2012), both experience with musical tones and linguistic tones were tested. In a Cantonese tone identification task, native Thai and English listeners were tested, each subdivided into musician and non-musician groups. The findings showed that both language and musical experience facilitated L2 tone identification, with musical experience being more advantageous than tone language background. However, the combination of the two (Thai speakers with musical experience) didn’t result in additional advantage. Thus, studies on non-native tonal perception in musicians and non-musicians show that musical background facilitates lexical tonal perception. According to the authors,
this might be due to similar processing mechanisms involved in lexical and musical tone perception (Schellenberg & Peretz, 2008).

In a recent investigation, the effect of lexical tones in a Dutch dialect, Limburgian were studied by Ramachers (2018). This is an exceptional case of L1 tonal experience as the dialect speakers who were studies also spoke Standard Dutch, a dialect of Dutch without lexical tones. The Limburgian dialect spoken in Roermond distinguishes two tones that can be used to differentiate between lexical meanings, similar to Mandarin or Thai. In a discrimination task, Limburgian speakers and Standard Dutch speakers were presented with F0 modified Limburgian lexical tones. Results showed while dialect speakers performed significantly better than the Standard Dutch speaking group. This suggests that experience with lexical tones in a dialect can facilitate tone perception. A highly interesting follow-up study would be to examine whether dialect speakers of Limburgian would benefit from their tone knowledge in learning a foreign tonal language like Mandarin or Thai.

The results of the investigations using the CP paradigm with native tonal and non-tonal language groups and linguistic and non-linguistic tonal continua do not align in all aspects. While all studies were able to show CP for the native tonal listeners, it is still not clear whether this group is able to extend the CP to an unfamiliar linguistic tonal continua or non-linguistic stimuli. For the non-tonal groups, the findings are mixed as well. While some studies suggest that they are able to perceive linguistic or non-linguistic tonal continua categorically – although to a lesser extent than native tonal listeners do – other studies indicate that CP is not accessible for non-tonal listeners at all.
2.4. Non-native Mandarin tone perception in adulthood after short-term training

Based on the assumptions of the ASP model by Strange (2011) and the studies reviewed above, it can be concluded that speakers of a tonal and non-tonal L1 perceive lexical tones differently. While tone-language speakers assumingly use a phonological mode of processing in the first place, non-tonal language speakers only have access to the phonetic mode of processing. Consequently, native Mandarin listeners initially neglect acoustic variation that does not describe a tone category, while native Dutch listeners are sensitive to acoustic details in acoustically modified Mandarin input. An empirical question based on the ASP model is how learners of a non-tonal language perceive lexical tones and whether they can establish native-like perception routines to do so. To answer this question, several short-term laboratory training studies were conducted where native tonal and non-tonal listeners were trained on novel tonal contrasts. The most relevant studies contributing to the discussion of L2 tone perception and lending supporting the ASP model are described below.

One of the first training studies on lexical tone perception in Dutch learners are the experiments by Leather (1990, 1997). In series of studies, native Dutch adult speakers were trained to either produce or to perceive Mandarin lexical tones. The production group received a one-time production training and were then tested on their perception abilities. Conversely, the perception group received perception training and were tested on their tone production abilities. The distinction between production and perception training in the experimental setup aimed to control for the interrelation between the accurate production and perception of newly learned tone contrasts. The results of the post-test showed that both training procedures resulted in the improvement in perception and production of Mandarin tones by native Dutch speakers. An alternated version of Leather’s training procedure was applied in the investigation of Wang (X. Wang, 2013) who tested native Hmong (a tone language that
distinguishes seven lexical tones), Japanese and English listeners with a perception only or a perception and production training method for Mandarin tones. In a 6-hour intensive laboratory training within 3-4 weeks, participants were trained on how to perceive (or produce) the four Mandarin tones. Then, participants were asked to complete an open-choice identification task. Similar to Leather’s findings, both training methods for all participant groups resulted in improved accuracy in the identification of the L2 tones. Interestingly, the Hmong group’s perceptual accuracy was lower than that of the non-tonal English and Japanese groups. This suggests that L1 tonal experience in this study has not been shown to facilitate the perception of novel tone contrasts.

In an another study, Wang and colleagues (Y. Wang & Spence, 1999) auditorily trained native American listeners to perceive Mandarin lexical tones. In addition to improved identification accuracy, they also found that speakers were able to generalize their knowledge to new stimuli, and this effect was still visible during the retention test six months after training. These results can also be attributed to the fact that participants in this study received 8 one-hour sessions in lexical tones throughout 2 weeks, which is a more intensive and long-term training scheme compared to other training studies.

While the studies described above found improvement after short training in Mandarin tones lending support for sensitiveness of novel L2 sounds in speech perception. These findings thus lend support to models like the ASP who predict sensitivity to novel non-native contrasts and the possibility to use different modes of perception in order to perceive them. However, there have also been conflicting results published, providing support for Brown’s model. For instance, Wayland and Guion (2004) tested native American English and Mandarin speaking groups’ ability to identify and discriminate Thai tone contrasts after auditory training. They found that the Mandarin speaking group was outperforming the English group before and after training. Moreover, the Mandarin group improved significantly in the experimental tasks while the English group did not.
In a study with a similar experimental design, Reid and colleagues (2015) found that training native Australian English and native Mandarin groups on Thai tones with auditory, visual and audio-visual methods only improves the native Mandarin groups’ tone identification accuracy or discrimination abilities, not that of the native English group. Both studies contribute to Browns’ model that predicts facilitation of Mandarin speakers’ L1 system to perceive novel tone contrasts in Thai. As the English groups’ L1 lacks this linguistic feature, they are not able to acquire tonal contrasts and fail to benefit from auditory training.

2.5. Non-native Mandarin tone perception in adulthood after long-term training

This section will provide an overview of studies that investigate the effect of long-term exposure to non-native tones in non-tonal L1 learners. These studies are cross-sectional, meaning that participants are classified in learner groups, based on the length of exposure to tonal languages in the form of formal instruction and the time spent in a Mandarin-speaking country.

In the first experiment in Guion & Pederson (2007), four participant groups were tested on their non-native tone discrimination abilities: native English, native Mandarin, native Japanese speakers and advanced L2 English learners of Mandarin. The learner group was studying Mandarin for at least 4 years and had been visiting China for at least 6 months consecutively. In the AX discrimination task, all groups were presented with a synthesized 9-step tone continuum, varying along two acoustic parameters; average F0 height (low, medium, high) and F0 slope (gradually decreasing in 9 steps). Participants were asked to rate how different the two fragments were from 1 to 9, where 1 is no difference at all and 9 is extremely different. The results revealed that the advanced L2 learning group performed similarly to the native Mandarin
group, while the other groups were performing significantly worse on discriminating the tones. Moreover, it was found that the L2 group used the F0 slope parameter similarly as the native Mandarin group, which is assumed to have contributed to their better performance compared to the L1 groups. Based on these findings, the authors suggest that as a result of intensive L2 learning experience, acoustic cues like F0 slope – a feature that is not used in English to discriminate linguistic stimuli – can be reweighted or brought into active use. This lends support to the ASP model’s that predicts increasing perception abilities as a result of L2 exposure, using the phonological mode of perception. However, this study did not include beginning learners to compare their mode of perception to the advanced and native Mandarin groups. Therefore, the ASP model is only partially supported by the results.

Another long-term training study by Shen & Froud (2016) studied both the tone identification and discrimination abilities of the participants. They tested L1 English, L1 Mandarin and advanced L2 English learners of Mandarin. The criterion for the learners to participate was the completion of at least 3 college semesters of Mandarin courses. Using the classic categorical perception paradigm, participants were asked to complete a forced choice identification task and an AXB discrimination task. The experimental tonal stimuli consisted of the F0 modified tonal continuums ranging between T1-T4 and T2-T3. In the identification task, participants heard a stimulus and had to decide whether it belonged to one of the available tone categories (T1 or T4, T2 or T3). In the discrimination task, three fragments 1 step apart in the continuum were presented. The task here was to decide whether the second fragment (X) was more like the first (A) or the third (B) fragment. In the identification task, the native Mandarin and advanced learner group revealed CP of tones. This was explained as the result of extensive and detailed instruction on Mandarin tone prototypes during the course of learning which resulted phonological mode of perception. As predicted, this was not the case for the L1 English group who showed significantly broader category boundaries in their responses on this task. The results of the discrimination task of this study revealed some unexpected outcomes. According to the ASP model, the L1
English group should be able to discriminate tonal contrasts better than the native Mandarin group and the advanced learners. On the contrary, the L1 English group performed near-chance on this task. This contradicts the ASP model that expects listeners without any prior experience with lexical tone categories to rely on the phonetic mode and gain high accuracy in discriminating tonal contrasts. For the advanced learner group, the mean accuracy rates in this task exceeded that of the native Mandarin group which was another unexpected result. According to the ASP model, the advanced learners could here switch from using the phonological mode in the identification task to using the phonetic mode which is called for in this task. As the native Mandarin group performed above chance-level as well suggests their ability to switch to the phonetic mode when necessary as well.

In another longitudinal study by Hao (2017) the tone discrimination abilities of beginning and advanced learners of Mandarin were compared to that of native English listeners’. The beginning learners were learning Mandarin for 1.47 years on average and the advanced learners for 5.59 years. In an AXB discrimination task, all groups were presented with the T1-T4 and T2-T3 contrasts, where the pitch difference in the former is assumed to be more salient than the latter and therefore easier to perceive. In both contrasts, the beginning and advanced groups showed higher discrimination accuracy than the native English group did. However, the native group performed above chance in both contrasts, suggesting that they are sensitive to tonal contrasts to some extent. Comparing the easy T1-T4 and more difficult T2-T3 contrast, all groups showed a significant difference in accuracy rates where the T1-T4 yielded higher accuracy responses than the T2-T3 contrast. The results of this study are therefore similar to what Shen and Froud (2016) found in the discrimination task performed by the native Mandarin, native English and advanced Mandarin learners. An unexpected result of this study was that the two learner groups did not differ in their performance. This was explained by the possibility of the beginner group being already too advanced in their course of Mandarin learning. This suggest that Hao’s study in reality only compared advanced learners’ and native English speakers’ performance.
Moreover, this study only used one task from the CP paradigm and lacks data from a tone identification task. As a result, the tone identification abilities of the native and advanced learner groups were not compared.

In a recent study, Zou (2017) tested the phonological mode of perception beginning and advanced Dutch learners of Mandarin, as well as native Dutch and native Mandarin listeners for Mandarin tones. She presented participants with tonal stimuli using the T2-T4 contrast in high phonetic variability non-words using an ABX discrimination task, asking for the phonological mode of processing. Based on the assumptions of the ASP model, it was hypothesized that native Dutch listeners and beginning learners will use the phonetic mode of perception and therefore perform poor on the task. Native Mandarin listeners and advanced learners, on the other hand were expected to use the phonological mode of perception and perform well. Indeed, results showed that advanced learners were able to perceive tonal contrasts in the phonological mode, and achieve accuracy rates, similar to native Mandarin listeners. The native Dutch group and the beginning learners’ performance was significantly above chance, indicating some sensitivity to the T2-T4 phonological contrast. These findings are in line with those of Shen and Froud (2016) and Hao (2017). While the results of the discrimination task align with previous studies, this investigation – similarly to Hao (2017) – lacks information about the tone identification abilities of the language groups. Without investigating both tone identification and discrimination performance, the CP of the different learner groups cannot be fully understood.

In sum, the most relevant cross-sectional studies with non-tonal learner groups learning Mandarin have used different tasks to provide insights in the L2 perception of Mandarin tones. It is crucial to investigate both discrimination and identification abilities of learners and native listeners as discrimination tasks only provide insights in whether the phonetic mode of processing is used rather than the phonological mode. By performing an additional identification task, the use of the phonological mode of processing can be studied as well. Furthermore, data from both tasks of the same listener may show switching between the two modes and possible discrepancies in
performance depending on the task demands. Another problematic issue of these studies is their different interpretation of ‘beginner’ and ‘advanced’ learners. These groups are mostly made by the experimenter, based on the length of learning and/or years of residence in a Mandarin speaking country. By not measuring the language proficiency (either by self-assessment or a language task) in Mandarin, learner groups might not be distinct groups and as a result, their group performance might not reflect effects from difference in the level of the L2.

The current study addresses these issues by studying both identification and discrimination performance of native Mandarin, native Dutch and beginning and advanced learners of Mandarin. The learner groups will be created using information on the learners’ Mandarin exposure as well as self-assessed proficiency in 5 components in Mandarin. The aim of the study is to investigate whether learners’ speech perception remains sensitive when learning and L2. By using an identification and discrimination task, this study provides better insight in the modes of perception that beginning and learner groups, as well as native Dutch and Mandarin groups can use to perceive Mandarin tone contrasts.
3 The present study

In this study, the main research questions are:

(1) Do beginner and advanced learners of Mandarin show categorical perception of Mandarin lexical tones in a tone identification task?

Previous studies on lexical tone perception indicate that native tonal speakers perceive tones categorically while speakers of a non-tonal language perceive tones in a continuous manner (Hallé et al., 2004; Peng et al., 2010; Yisheng Xu et al., 2006). Whether learners of Mandarin without any L1 tonal experience can perceive tones categorically has not been studied extensively. According to the ASP model, beginning learners should not be able perceive tones categorically as they do not (yet) have access to the phonological mode of perception. Advanced learners however can use this mode of perception and perform well on a tone identification task (Strange, 2011).

(2) Do learners’ perceptual sensitivity change in discriminating tone contrasts as a result of emerging use of the phonological mode of perception?

Based on the predictions of the ASP, beginning learners can only rely on the phonetic mode of perception when they are presented with tone contrasts. In a discrimination task with a highly varying tonal continuum, this mode of perception should therefore facilitate the perception of the continuum steps. Previous studies indeed show a good performance of beginning learners in tone discrimination tasks (Hao, 2017; Zou, 2017). As for the advanced group, listeners are assumed to have developed categorical perception of tones and can therefore rely on the phonological mode of perception. This mode might therefore hinder the perception of tonal contrasts in a high-variability discrimination task.

To address these questions, beginning and advanced Dutch learners of Mandarin were studied in their perception of lexical tones. As control groups, native Mandarin and Dutch listeners were recruited as well. As reviewed above, the categorical perception
paradigm has been shown to be a reliable tool for testing the phonetic perception of non-native contrasts. Therefore, the present study will apply the identification task to answer the first research question. In the identification task, participants were asked to listen to a Mandarin fragment from a frequency-manipulated continuum ranging from T1 and T4. Their task was to identify the fragment as T1 or T4 by clicking on one of the corresponding buttons on the computer screen. The phonetic variability was introduced in that the continuums were recorded four times, thereby creating interspeaker variability. To induce the difficulty of the task, participants were asked to respond as fast and as accurate as they could.

To answer the second research question, the perception of phonetic contrasts by Dutch learners of Mandarin was tested in an AXB discrimination task. Using the same tonal continuum as in the identification task, participants heard a sequence of three Mandarin fragments where fragment X corresponded to either fragment A or fragment B. The target fragment X and the corresponding A or B fragment were one step apart in the tonal continuum. Sequences four steps or further apart in the continuum were included in the control condition. The participants were asked to decide whether X sounded more like A or B by clicking on the corresponding button on the computer screen. Similar to the identification task, participants were asked to respond as fast and as accurate as they could.

In identification and discrimination tasks where the primary acoustic cue for tones, the F0 is manipulated resulting in a high-variability stimulus task, native Dutch listeners and beginning L2 learners are more likely to perceive the variation in the acoustic signal using the L1 phonetic mode of perception. This is hypothesized to result in high discrimination accuracy in the AXB discrimination task and no clear CP pattern in the identification task. In advanced learners and native Mandarin listeners, the phonological mode of processing is expected to emerge, which results in increased perceptual attention to phonologically contrastive information in the stimulus. As a result, advanced learners and native Mandarin speakers are hypothesized to show a CP pattern in the identification task and respond less accurately in the discrimination
task than the native Dutch and beginning learner group. In terms of the categorical perception paradigm, the native Mandarin group are expected to show a CP pattern, in a form of an S-curve in the overall identification task responses. For the native Dutch group, it is expected that they will perceive the tonal continuum in a continuous manner and not show any CP pattern. For the learner groups, a CP-like pattern is expected for the advanced group but not for the beginner group. As for the discrimination task, it is predicted that beginner learners will be more sensitive to the contrasts in the tonal continuum than the advanced learners and native Mandarin group as they are not supposed to have developed CP for tones yet. Consequently, beginner learners are expected to outperform advanced learners and native Mandarin listeners in the tone discrimination task.

3.1. Experiment 1: Identification task

3.1.1. Participants

Twenty Dutch control participants, Twenty-five Mandarin control participants and twenty-six Dutch learners of Mandarin participated in the experiment. The native Dutch control participants (10 males and 10 females) had no knowledge of Mandarin or any other tone languages. The average age of the native Dutch group was 23.89 and that of the native Mandarin group (7 males and 18 females) 22.08. The Mandarin participants were studying or working at the University of Groningen when the study was conducted. They listed a regional Chinese dialect and Standard Mandarin as either their first or second language. Some of them had taken Dutch language courses and 7 of them listed Dutch as their third and 11 of them as their fourth language. From the Dutch learner group, 10 beginner learners and 16 advanced learners were identified, based on their Mandarin learning experience in the Netherlands and/or in a Mandarin speaking environment. Participants in the beginner group (5 males and 5 females) were recruited from the first-year Chinese classes at Leiden University and
HSK1 and HSK2 courses from the Groningen Confucius Institute. The advanced learners (10 males and 6 females) were recruited from the third-year bachelor and first-year master Chinese courses Leiden University and the HSK3, HSK4 and HSK5 courses from the Groningen Confucius Institute. None of the learners were heritage speakers of Mandarin or a Limburgian dialect. Their age, starting age of learning Mandarin, length of learning, the time spent in Mandarin-speaking countries, musical experience and daily musical practice are summarized in Table 1. Independent t-tests were conducted to compare these characteristics of the two learner groups, and the results are also listed in the rightmost column of Table 1. Significant differences at $p \leq 0.05$ level are marked by “*”.

<table>
<thead>
<tr>
<th>Table 1. Comparison of the Mandarin language learning background of the two learner groups.</th>
<th>Beginner group</th>
<th>Advanced group</th>
<th>t-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>20.83 years</td>
<td>24.90 years</td>
<td>$t(24) = 1.174$</td>
</tr>
<tr>
<td>Starting age of learning Mandarin</td>
<td>19.3 years</td>
<td>19.68 years</td>
<td>$t(24) = 0.162$</td>
</tr>
<tr>
<td>Length of learning</td>
<td>1.53 years</td>
<td>5.22 years</td>
<td>$t(24) = 2.410^*$</td>
</tr>
<tr>
<td>Time spent in Mandarin-speaking countries</td>
<td>0.6 months</td>
<td>5.12 months</td>
<td>$t(24) = 3.066^*$</td>
</tr>
<tr>
<td>Musical experience (years)</td>
<td>4.9 years</td>
<td>4.87 years</td>
<td>$t(24) = -0.011$</td>
</tr>
<tr>
<td>Daily musical practice (hours)</td>
<td>0.28 hours</td>
<td>0.26 hours</td>
<td>$t(24) = 0.112$</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, the two learning groups were not significantly different in their age and starting age of learning Mandarin, which was around 19 years. For the length of learning, the advanced group has been learning significantly longer, by approximately four years. Moreover, the advanced group spent significantly more time in a Mandarin-speaking country than the beginner group did. To make sure that none of the learner groups would benefit from experience with musical tones in the study, music-related questions were recorded as well. By gathering data from both learner groups on how many years of musical experience they have and how many hours a day they spend on musical practice, it was hoped that the learner groups are similar and differences in lexical tones perception were not influenced by their musical training. Indeed, the learner groups did not significantly differ in these two musical measures. These statistical analyses support that the two groups indeed differed in
Mandarin experience and were comparable in their musical background are therefore correctly labelled as beginner and advanced learners.

The language learners’ self-estimated Mandarin proficiency in speaking comprehension, speaking production, reading, writing and pronunciation were recorded at the end of the experimental session using the Language Background Questionnaire (LBQ) (Sabourin et al., 2016), and are summarized in Table 2. Participants were asked to indicate their language proficiency on a 6-point scale, consisting of the following steps: very low, low, intermediate, advanced, near-native and native. For the statistical analyses, these scale points were transformed into numerical values so that 1 corresponds to very low and 6 to native. Mann-Whitney U-tests for ordinal data were conducted to compare the learner groups in the language proficiency components.

Table 2. Comparison of the Mandarin self-estimated language proficiency of the two learner groups.

<table>
<thead>
<tr>
<th></th>
<th>Beginner group</th>
<th>Advanced group</th>
<th>Mann-Whitney U-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking comprehension</td>
<td>1.7</td>
<td>3.375</td>
<td>W(24) = 147*</td>
</tr>
<tr>
<td>Speaking production</td>
<td>2</td>
<td>3.438</td>
<td>W(24) = 149*</td>
</tr>
<tr>
<td>Reading</td>
<td>2</td>
<td>3.313</td>
<td>W(24) = 142*</td>
</tr>
<tr>
<td>Writing</td>
<td>1.9</td>
<td>3.313</td>
<td>W(24) = 145.5*</td>
</tr>
<tr>
<td>Pronunciation</td>
<td>2.2</td>
<td>2.688</td>
<td>W(24) = 101.5</td>
</tr>
</tbody>
</table>

As can be seen from Table 2, the advanced group estimated their Mandarin proficiency around scale point 3 which corresponds to ‘intermediate’ level. The beginner group estimated their Mandarin proficiency at the high end of very low and low (scale points 1 and 2). The self-estimated scores for each component were statistically higher for the advanced group, except for the pronunciation component where both groups estimated their proficiency as ‘low’. Based on the self-assessment scores of the two language groups, they seem significantly different in their self-assessed proficiency in Mandarin as well.
3.1.2. Materials and design

The T1-T4 tone contrast was selected with the tone-bearing syllable /ta/, identical to the stimuli in (Liu, Chen, & Kager, 2017). The first step in creating a tonal continuum ranging between T1 and T4 was to record the syllables /taT1/ and /taT4/. These syllables are legit words in Mandarin, meaning 'build' when pronounced in T1 and 'big' when pronounced with T4. The syllables were recorded by a female Mandarin-speaking phonetician, using Audacity® recording and editing software. The recordings took place in a sound-proof booth of the Utrecht University phonetics lab, using a Genelec 1029A active speaker. Both syllables were recorded four times to create inter-speaker variation in the stimulus, which resonates a more natural language input (Nygaard, Sommers, & Pisoni, 1995). Using Praat (Boersma & Weenink, 2018), the recordings were marked at four points in time in each stimulus (0%, 33%, 67% and 100%). Then, these points were temporally aligned in seven equal steps, by modifying their fundamental frequency (Hz). After interconnecting de points within the stimuli, new pitch contours were made, and in total 8 stimuli were created (see Figure 3.). This procedure was followed for all four realizations of /taT1/ and /taT4/ pairs, creating 4 continuums, with 32 stimuli in total. The main duration of the stimulus sounds is 412ms, with the average sound intensity level normalized at 70dB.

A pilot study using these stimuli revealed that the first step in each continuum elicited significantly more correct responses and shorted reaction times than other steps in the continua. This may have been the result of the entirely flat pitch contour that has become a robot-like tone, and thus was easy to tell apart from other realizations of the target sound. Therefore, the first step of each continua was lowered with a half semitone, resulting in a slightly more decreasing pitch contour. By this, it was hoped to make this realization more speech-like and less of an outlier in the tonal continuum.

---

1 Audacity® software is copyright © 1999-2018 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.
3.1.3. Procedure

The participants were individually tested in a sound-attenuated booth in the phonetics labs of the Center for Language Cognition Groningen (CLCG) or the Leiden University Centre for Linguistics (LUCL). Before the experiment started, they received oral instructions from the experimenter and written feedback in Dutch for Dutch participants, and English for Mandarin participants. The native Mandarin speakers and the learner groups were asked to listen to the Mandarin fragment they heard and decide whether this was pronounced with T1 or T4. The native Dutch speakers were asked to decide between the categories flat (“vlak”) or falling (“dalend”). All participants were explicitly asked to respond as fast and as accurate possible by clicking on one of the buttons presented on the computer screen (Lenovo ThinkPad T420). Before the test phase started, participants completed a training phase where they heard a clear T1 (flat) and T4 (falling) fragment and were asked to respond accordingly. During this phase, the correct answer remained on the screen, hereby providing the participants with feedback. After the training phase, the test phase

![Figure 3. An example of an 8-step continuum, ranging from T1 to T4. The in-between stimuli are equally distributed in pitch (Hz) (Liu et al., 2017).](image)
started where all 32 stimulus sounds were played twice (64 trials in total) in a fully random order for each participant. During the test phase, participants did not receive any feedback. The intertrial interval (ITI) was 750ms during which a fixation cross was presented in the middle of the screen for 250 ms. Figure 4 below illustrates the structure of a trial in the identification task.

![Diagram](image.png)

*Figure 4. The structure and time course of a trial in the identification task.*

### 3.1.4. Results

In order to visualize the mean tendency of categorization of the continuous stimuli, the mean proportion of T4 responses were calculated per step per participant and depicted in Figures 4 and 5 below. In Figure 5, these responses were then split into the native Dutch and Mandarin groups and in Figure 6 for the beginner and advanced learners. By depicting the learner groups’ responses separately from the native groups’ helps to see the difference between learner and native groups and helps to contrast the learner groups in more detail.
A repeated measures ANOVA with a Greenhouse-Geisser correction for all four language groups determined that the mean responses per step per language group were differ statistically significantly (F(1,358) = 10.382, p < 0.005). The interaction of mean response per step across the two language group was significant as well (F(1,358) = 11.665, p < 0.005). Looking at the native Mandarin and Dutch identification patterns, Figure 5 shows that the native Mandarin group labelled the steps in the tonal continuum in a categorical fashion while the native Dutch group did not. This tendency is the most visible in the right-end of the continuum where steps 4 and 5 were labelled as T4 between 80% and 95% and for steps 6, 7 and 8 almost 100% of the time. For the left-end of the continuum, Mandarin speakers labelled steps 1 and 2 barely as T4. For step 3 however, the mean T4 responses were more varying. Overall, the identification curve of the Mandarin speakers reveals a categorical perception pattern where the far-end steps in the tonal continuum are clearly labelled as non-T4.
and T4 and variability on the tonal boundary around step 3 and 4. The sudden switch in identification response is around step 4 which is halfway through the continuum. A test for within-subject contrasts per language group per step showed statistically different responses for each step for the Mandarin group (F(1,7) = 169.737, p < 0.005. For the Dutch group, this was not significant (F(1,7) = 0.033, p = .857), indicating that the mean T4 responses are not different across the steps and therefore do not show a categorical perception pattern. The mean responses from step 2 onward are around 50% and 60% which is around chance-level. It seems that the Dutch group did not label the tonal continuum categorically and labeled most steps in the continuum with either the label ‘flat’ or ‘falling’. The mean responses of the beginning and advanced learners are shown in Figure 6 below.

Looking at the identification curves of the two learner groups, the general identification patterns seem highly similar. Both groups correctly labelled steps 1 and
2 in the continuum as non-T4, however they reveal moderate variability in their responses. Between steps 3 and 6, the difference between beginning and advanced learners become more apparent. Although both groups gradually identified the tonal continuum as T4, the variability in responses is higher for the beginning learners than for the advanced group, indicating uncertainty in the identification of these steps. The mean responses overlap in steps 5 and 6, however the variability in responses seems higher for beginning learners, especially in step 5. Looking at the general identification curve, beginning learners also labelled steps 3 and 4 more often as T4 than the advanced learners did. As for the right-end of the continuum, the high variability in responses for the beginning groups remains and that of the advanced group decreases. Furthermore, the advanced group identified step 7 and 8 as T4 around 90% of the time, opposed to 80% for the beginning group. A test for within-subject contrasts per language group per step showed statistically different responses for each step for the beginning learners group (F(1,80) = 4.353, p = .040) and the advanced learners group (F(1,112) = 6.312, p = .013). However, the difference between the mean responses for the learner groups across all steps was not statistically significant (F(1,206) = 0.013 p = .908).

Comparing the identification curves across Figure 5 and 6, there seems to be a different identification pattern for all four groups. The native Mandarin speakers show a categorical perception pattern with low variability responses for steps at the second at the far ends of the continuum and more variability at the category boundary. The native Dutch speakers show no identification pattern with mean response values around chance level for all steps, except for step 1. The two learner groups do not show a categorical perception pattern as the Mandarin group did but they certainly show a tendency to label the left-end of the tonal continuum more often as non-T4 and the right-end as T4. In comparison to the native Mandarin group, there is a gradual change in the labels used along the continuum and there is no clear switch point where the steps are clustered in a single label. The identification patterns of the beginning group
show higher variability than the advanced groups’, indicating uncertainty to label the
tonal continuum steps.

3.2. Experiment 2: Discrimination task

3.2.1. Materials and design

For the AXB discrimination task, the same stimulus materials were used as in the
identification task (see section 3.1.2.). A pilot study of this experiment with 7 native
Dutch speakers revealed that presenting trials and asking for same-different
judgements between stimuli that were two perceptual steps apart resulted in ceiling
effects. That is, participants discriminated the fragments with little or no errors.
Therefore, experimental participants all completed a one-step AXB discrimination
task, resulting in 7 possible step combinations (1-2, 2-3, 3-4, 4-5, 5-6, 6-7 and 7-8). The
data of the participants who took part in the pilot study were excluded from the
discrimination task analysis.

The experimental trials consisted of experimental and filler trials (104 each). The
AXB experimental trials were three Mandarin fragments played after each other,
where the target sound (X) was either identical to the first fragment (A) or the third
fragment (B). The AXB trials were combined in four possible ways; AAB, ABB, BAA
and BBA, where the same AA and BB represent the position of the identical fragment
in the trial. The experimental trials were counterbalanced for the four tone continua
(A-D), four possible combinations (AAB, ABB, BAA, BBA), and answer type (fragment
1 or fragment 3), resulting in 104 trials in total. The filler trials were constructed the
same way as the experimental trials, except for the perceptual step between A-X and
X-B. In these trials, the perceptual step between the different fragments within the
trials was 4 steps or more, which made them acoustically salient, compared to the
experimental trials. In total, 208 trials were presented in two blocks in the experiment.
3.2.2. Procedure

After completing the identification task, the same group of participants were tested in the discrimination task where they heard two competitor stimuli A and B and target stimulus X played in an AXB order. Their task was to listen to all three fragments and then decide whether the first (A) or the third (B) stimulus sounded the same as the second (X) by clicking on one of the two buttons presented on the screen as accurately and fast as possible. Between the A-X and X-B sounds, a 500 ms inter-stimulus interval (ISI) was used. The inter-trial interval was 750 ms during which the fixation cross was presented on the screen for 250 ms. Before the experiment started, participants heard two examples where the A-X and X-B fragments were 8 steps apart, thus the target and the distractor fragments were examples of the natural lexical tones 1 and 4. During the training phase, the correct answer remained on the screen. During the test phase, participants did not receive any feedback. Figure 7 below illustrates the structure of a trial in the AXB discrimination task.

![Diagram of AXB discrimination task](image)

*Figure 7. The structure of a trial in the AXB discrimination task.*
3.2.3. Results

The mean accuracy of the responses per language group are plotted in Figure 5 below, also indicating the significant differences between groups (marked with ‘*’). The mean accuracy here means the correct discrimination across all pairs. Thus, an accurate response is when the participant responded ‘first’ when A = X ≠ B and ‘second’ when A ≠ X = B. As Figure 8 shows, the mean accuracy rates across all pairs was the highest for the native Dutch group with 73.58% mean discrimination accuracy. For the beginning learners, this was 69.34% and the advanced learners 67.02%. The native Mandarin speakers has the lowest accuracy rates overall, with 61.86%.

Figure 8. A boxplot of the mean accuracy rates in the AXB discrimination task per language group. The double arrows with ‘*’ indicate statistically significant difference between the groups.
In a RM-ANOVA, multiple comparisons for the mean accuracy rates per language group were performed with Tukey’s HSD post-hoc test. The difference in mean accuracy rates for the native Dutch group was significantly higher than that of the native Mandarin group and the advanced Dutch learners (native Dutch vs. native Mandarin: $p = .000$; native Dutch vs. advanced learner: $p = .049$). The difference in mean accuracy rates was also significant between the native Mandarin group and the beginning learners ($p = .009$). Although the native Mandarin groups’ performance was low, a one-tailed t-test showed their performance was significantly above chance level (50%) (data aggregated by items: $t(167) = 89.864$, $p < 0.001$). Between the beginning and advanced learner groups, there was no significant difference in mean accuracy, but there was a slight trend of the beginning learners performing better within each pair than the advanced learners ($p > 0.05$). This is illustrated in Figure 9 where the mean accuracy per pair for the two learner groups is illustrated.

![Figure 9. The mean responses per pair in the AXB discrimination task for the learner groups.](image)
As Figure 9 illustrates, the highest accuracy for both learner groups was achieved in the 1-2 and 2-3 pair. In the following three pairs, the accuracy rates for the groups is similar and lays around 60% and 70%. In pair 6-7, the difference in accuracy between the groups becomes bigger, showing better accuracy for the advanced group. In the 7-8 pair, the beginner group outperforms the advanced group by almost 20% higher accuracy and the advanced group shows chance-level discrimination of this pair. Overall, the beginning group shows a tendency of discriminating all pairs more accurately than the advanced learners does, except for pairs 5-6 and 6-7. Again, the overall difference between the mean responses of these groups was not statistically different.
4 Discussion

This study investigated whether Dutch learners of Mandarin show categorical perception of Mandarin lexical tones and whether they can switch to a native-like phonological mode of perception when discriminating acoustically highly varying tone contrasts. To see whether increasing experience with Mandarin changes the perceptual sensitivity of learners, beginner and advanced groups were compared in a tone identification and discrimination task. In addition, native Mandarin and native Dutch listeners were included as control groups to see whether the learner groups’ performance becomes more native-like by deviating from the native Dutch groups’ performance and becomes more identical to that of the native Mandarin group. The results clearly demonstrated that the experience with Mandarin has an effect on the both Dutch learner groups’ perception of Mandarin tones. As the learner groups showed improved performance compared to the native Dutch group in both identifying and discriminating tones. This counters Brown’s model (2000) which would predict a persistent impairment in tonal perception for the learner groups regardless of their proficiency in Mandarin. The ASP model, on the other hand, does predict the emergence of a new (phonological) mode of processing which is primarily used by native speakers.

The results of the identification task show that native Mandarin speakers perceived the T1-T4 continuum categorically, with a sharp switch-point halfway-through the continuum (step 4). This CP effect was clearly demonstrated by the S-curve in the identification plot as well (see Figure 5). This supports the predictions of the CP paradigm that predicts categorical perception for L1 speakers (Warren, 2012), and suggests the use of the phonological mode, proposed in the ASP model (Strange, 2011). According to the ASP, when native speakers are presented with high-variability stimuli and are asked to identify them in certain native categories, the phonological mode of perception is activated. By using this mode, the variation in the continuum will not be perceived and distinct tone categories will be formed, even though the
stimuli being a continuum (Strange, 2011). Results for the native Dutch group without knowledge of Mandarin did not reveal CP effects in this task. This suggests that non-tonal listeners with no prior experience did not perceive the tonal continuum categorically and seem to have relied on their phonetic mode of perception. Even though their performance was significantly above chance for each step, the mean proportion of either labels (‘flat’ or ‘falling’) was not significantly different when the steps were compared. This is not surprising since tonal information in Dutch is not used contrastively on the lexical level and therefore it is not perceived contrastively in native Dutch speakers. Analogous findings have been reported in Hallé, Chang and Best (2004) and Xu, Gandour and Francis (2006), who demonstrated CP effects for native tonal speakers (Taiwanese and Mandarin) and no CP for non-tonal speakers (English and French) using a high-variability tone continuum. Nonetheless, the observation that the native Dutch group performed above chance in identifying the tones suggests that they are not ‘tone deaf’, as suggested by Brown (2000). Compared to the native Dutch group, the learner groups show difference in their responses across the continuum steps. However, they do not show a clear CP pattern as the native Mandarin speakers did. Instead, their mean T4 responses seem to gradually increase as the stimulus becomes more ‘T4-like’. This suggests that they are sensitive to the F0 patterns of the continuum steps and therefore used the phonetic perception mode of perception, proposed in the ASP model (Strange, 2011). According to the ASP model, beginning learners have not yet developed L2 SPR’s and therefore can only use the phonetic mode of perception, which is supported by the findings of this study. However, the model predicts a more native-like pattern for the advanced group who assumingly have developed new SPR’s and are therefore able to switch to the phonological mode of perception in a high-variability task. The use of the phonological mode is not reflected in the advanced learners’ identification pattern found in this study. However, looking at the identification accuracy for each step per learner group, there is a tendency for the advanced learners to correctly label steps 1-4 more often as T1 and steps 5-8 as T4 than the beginner group. This is rather a slight tendency as the
difference in the responses for each step per language group were not statistically significant. In a related study by Hao (2017) that neither showed a difference between the beginning and advanced learners’ identification performance, this was attributed to the way the learner groups were created. According to Best and Tyler (2007), the length of learning for beginning learners should be shorter, for example between 6 and 12 months. This is based on the fact that lexical tones are fundamental parts of learning Mandarin and are therefore intensively taught already in the first months of learning Mandarin. In the course of learning, Mandarin exposure may therefore not have a consistent effect on learners’ perceptual abilities. In the current study, the mean length of learning for the beginning group was 1.53 years which exceeds the 12-month cut-off point suggested by Best and Tyler (2007). Even though the advanced group has been exposed to Mandarin significantly longer by learning Mandarin for over 5 years and have resided in a Mandarin-speaking country for 5 months on average, their identification performance was still not significantly more accurate than that of the beginning learners. The fact that there was no difference between the learner groups contradicts the ASP model’s predictions about advanced learners’ ability to use the phonological mode of perception in a native-like manner. However, this might reflect incorrect classification of the learners based on their length of learning in the experimental design.

The results of the discrimination task with high-variability stimuli showed that the native Dutch group was able to perceive the differences between the stimuli the best with an 73.58% average accuracy. The native Mandarin group, on the other hand performed with the lowest accuracy across all groups (61.68% mean accuracy) which however significantly exceeds chance level. The findings for the native Dutch group are in line with the predictions of Strange’s ASP model (2011) for high-variability tasks where the phonetic mode of perception should be activated. As the native Dutch listeners only have access to the phonetic mode, they are predicted to perform well on this task. With regards to the native Mandarin group, the ASP model would suggest that they could switch from their default phonological mode to the phonetic mode and
therefore should be able to perform well on this task as well. However, this is not supported by the findings on the Mandarin group in this study. Rather, it seems that the Mandarin group was not able to perceive the differences between the stimuli, hereby supporting Brown’s model that poses that the Mandarin tone categories hinder perception of variation in the stimuli by filtering out variation that is irrelevant to identify Mandarin tones. The two learner groups’ performance on the discrimination task followed different patterns from the native Dutch and native Mandarin group. By achieving higher discrimination accuracy than the native Mandarin group, they show higher sensitivity to variation in the tonal input. This suggests that the learner groups did not use the same modes of perception than the native Mandarin group did. As they were more accurate in discriminating the phonetically varying stimuli, it looks like they relied more on the phonetic mode of perception than the native Mandarin group did. When comparing the beginning and advanced learners’ performance to that of the native Dutch group, an interesting pattern emerges: While the beginning group seems to perform similarly to the native Dutch group showing relatively high accuracy rates, the advanced group showed significantly lower accuracy rates than the native Dutch group did. This suggests that the discrimination abilities of beginning learners are more comparable to the native Dutch groups’ while the advanced group deviates from the native discrimination pattern. These findings align with that of Shen and Froud (2016) where the English advanced learners outperformed the native Mandarin group in the discrimination task in both tonal continua (T1-T4 and T2-T3). The findings of the discrimination task suggest that while native Mandarin speakers are performing poorly when discriminating tone contrasts, the fact that they perform above-chance indicates that they – at least to some degree – are able to perceive the contrasts even though they are not informative for identifying the tones. In terms of the ASP model, they seem to use the phonological mode of perception which hinders the correct discrimination of the high-variability stimulus. Looking at the learners’ performance in contrast with the native Dutch groups’, it seems that both learner groups’ perception deviates from the native Dutch perception. As the native Dutch
group can only use the phonetic mode which is asked for in this task and are not hindered by established tone categories, they discriminate tone contrast with the highest accuracy. The learner groups’ accuracy rates fall between the native Mandarin and native Dutch accuracy. Note, however that there was no significant difference between the two learner groups in this task. The lack of group difference may be due to the same amount of training in tone identification in the first 6-12 months of learning Mandarin which both learner groups already received when they were tested. This could have led to the identical performance in both tone identification and discrimination tasks.
5 Conclusion

This study investigates the perception of L2 tones by learners with a non-tonal L1, in view of the inconsistent findings of previous investigations. In addition, the current study compares two learner groups – a beginning and an advanced group – with native Dutch and native Mandarin in order to gain insights in the effect of L2 tone experience (or in the case of the native Dutch group, the absence of it) on L2 lexical tone perception. The results of a tone identification task demonstrate that without any experience with lexical tones, native Dutch speakers are not able to perceive Mandarin tones categorically while the native Mandarin group show categorization of the tonal continuum in their existing tonal categories. These findings are in line with previous studies on native tonal and non-tonal perception of lexical tones. Furthermore, they align with the predictions of the CP paradigm that predicts categorical perception for native tonal speakers and continuous perception for native non-tonal speakers. When comparing the beginning and advanced learner groups, the results indicate that they are able to categorize Mandarin tones more categorically than the native Dutch group. This suggests that Dutch learners can perceive tonal contrast in Mandarin as a result of exposure to lexical tones even though tones do not convey lexical meaning in Dutch. Therefore, the results of the native Dutch and native Mandarin group support the ASP model that predicts the learning of new tones and counter Brown’s model that predicts no improvement in tone perception for Dutch learners. In the tone discrimination task, the native Dutch group outperforms both the learner groups and the native Mandarin group suggesting that discrimination of tonal contrasts is easier without experience with tones in the L1. This finding suggests that without prior knowledge of lexical tones, the phonetic mode of perception is activated and therefore tonal contrasts are better perceived than in the groups with knowledge of lexical tones. This could also explain how native Mandarin speakers performed with the lowest accuracy on the discrimination task. As they perceive lexical tones categorically by using the phonological mode of perception, they are not able to discriminate tonal contrasts as
well as native Dutch listeners do. However, as the native Mandarin group still performed above chance on the discrimination task, it seems that they are able to perceive tonal contrast to some extent. As for the beginning and advanced learner groups, their discrimination accuracy is higher than the native Dutch groups’ but not identical to the native Mandarin groups’. This suggests that L2 experience with lexical tones changes sensitivity to discriminate tone contrast so that redundant variation is perceived less than in native Dutch speakers, however not to the same extent as in native Mandarin listeners. This is also predicted by the ASP model which predicts the development of a new, phonological mode of perception in advanced learners as a result of exposure to Mandarin tones. In sum, Dutch learners’ perception of tones seems to be altered by increased exposure to Mandarin tones. This is supported by the identification pattern of both beginning and learner groups that is more categorical than that of the native Dutch group. This suggest that the learner groups have become less sensitive to variation in the tonal input and perceive tones driven by the phonological mode. This explains the decreased accuracy in discriminating tonal contrast in the tonal continuum presented in the discrimination task.

Whereas the results in this study overall show increased perceptual sensitivity in Dutch learners to perceive as a result of exposure to Mandarin lexical tones, it failed to show a difference between the beginning and advanced groups’ tone perception. This could be the result of the study’s limitation in the number of participants in the learner groups. By increasing the number of both beginning and advanced groups, the statistical power of the study could be increased which could lead to group differences that could not be shown in the current study. Another possible why this study failed to show different patterns for beginning and advanced learners in the identification and discrimination tasks might be due to the criteria for beginning learners’ length of training in Mandarin tones. As suggested by Best and Tyler (2007), training in perceiving the four main tones in Mandarin is one of the first aspects that are taught in Mandarin courses. Therefore, the beginning learners in this study with approximately 1,5 years on average training in Mandarin could not rightfully be
considered as beginning learners. Future studies might need to investigate learners in the first months into their training when they have not reached the level of an advanced learner yet. While this investigation focused on the effect of linguistic tone exposure on perception, it would be highly interesting to see how additional exposure to non-linguistic tones affects the perception of tones in non-tonal speakers and learners. For instance, future investigations could perform the CP tasks with Dutch learners of Mandarin with and without significant exposure musical tones in order to provide more insights in the nature of L2 tone perception. Another highly interesting subject of investigation is the effect of exposure to linguistic tones on other tone languages and dialects. More cross-linguistic and interdisciplinary studies would provide meaningful insights in the mechanisms behind lexical tone perception and contribute to our understanding in the acquisition of Mandarin as a second language.
References


Appendices

A1: Language Background Questionnaire Short Version

LANGUAGE BACKGROUND QUESTIONNAIRE – SHORT VERSION

Participant information – To be filled out by the researcher

Date: _______________    Participant code: ___________   Participant group: ______________

The information you share by filling out this questionnaire will be stored anonymously and is strictly confidential, used for research purposes only.

1. Basic language information

For an example, see the attachment. Please ask if you have any questions.

A.

INSTRUCTIONS:

• In the first row, please list all spoken or signed languages of which you have ANY current or previous knowledge in order of DOMINANCE, that is, how comfortable you are using them. Please include your mother language(s) and any regional dialects (e.g. Limburgian) as well!

• In the second row, please write the age at which you were first exposed to this language, that is, when it first became present in your environment.

• In the third row, please indicate whether you were ever IMMERSED in this language, that is, if you received significant exposure to this language in either a community/home setting or a school. If so, please also indicate the age and school year at which the immersion began.

<table>
<thead>
<tr>
<th>Dominance</th>
<th>1. __________</th>
<th>2. __________</th>
<th>3. __________</th>
<th>4. __________</th>
<th>5. __________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start 1st exposure:</td>
<td>Age:</td>
<td>Age:</td>
<td>Age:</td>
<td>Age:</td>
<td>Age:</td>
</tr>
<tr>
<td></td>
<td>School year:</td>
<td>School year:</td>
<td>School year:</td>
<td>School year:</td>
<td>School year:</td>
</tr>
<tr>
<td>Immersed:</td>
<td>YES / NO</td>
<td>YES / NO</td>
<td>YES / NO</td>
<td>YES / NO</td>
<td>YES / NO</td>
</tr>
<tr>
<td>Age/school year:</td>
<td>__________</td>
<td>__________</td>
<td>__________</td>
<td>__________</td>
<td>__________</td>
</tr>
</tbody>
</table>
### B.

**INSTRUCTIONS:**
For each of the above listed languages, please provide your current and highest ever attained level of proficiency. This consists of your *general* proficiency. If you feel that you are between categories, simply indicate where you belong along the scale you feel your proficiency is best represented.

<table>
<thead>
<tr>
<th>Language 1:</th>
<th>Current level of general proficiency:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
<tr>
<td></td>
<td>Highest level of proficiency ever attained:</td>
</tr>
<tr>
<td></td>
<td>I------------------I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language 2:</th>
<th>Current level of general proficiency:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
<tr>
<td></td>
<td>Highest level of proficiency ever attained:</td>
</tr>
<tr>
<td></td>
<td>I------------------I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language 3:</th>
<th>Current level of general proficiency:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I------------------I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
<tr>
<td></td>
<td>Highest level of proficiency ever attained:</td>
</tr>
<tr>
<td></td>
<td>I------------------I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language 4:</th>
<th>Current level of general proficiency:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
<tr>
<td></td>
<td>Highest level of proficiency ever attained:</td>
</tr>
<tr>
<td></td>
<td>I------------------I------------------I------------------I</td>
</tr>
<tr>
<td></td>
<td>very low low intermediate advanced near-native native</td>
</tr>
</tbody>
</table>
2. Musical background

Do you have any musical experience (e.g. choir, playing an instrument, making music as a DJ)? If yes, how many years/months?

______________________________________________________________________________

In what context are you engaged in this musical activity (e.g. recitals, band practice)?

______________________________________________________________________________

What type of musical instructions do/did you receive?

______________________________________________________________________________

How many hours of daily practice do you devote to this activity on average?

______________________________________________________________________________

3. About the tasks

How did the tasks go? What did you have the most difficulty with?

______________________________________________________________________________

Did anything sound like words you already knew? Do you have any idea what these tasks were about?

______________________________________________________________________________

Were you using any strategies for the tasks? If yes, describe it!
4. Biographical information

Month and year of birth: _________        Current age: ______        Gender: (M/F)

Where were you born (country, province/region, city)?: ____________________________

Did you ever move? ______

If yes, fill out the table below. Please list all the moves, including moves to another city or country. Please also include internships or exchange/immersion programs!

<table>
<thead>
<tr>
<th>Where did you move?</th>
<th>At what age?</th>
<th>How long did you stay?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Are you left- or right handed?  LEFT / RIGHT

Do you have any known hearing impairments? ________________________________

Have you ever had a serious head injury? ________________________________

What is your current profession? ________________________________

Thank you for filling out this questionnaire!
A2: Language Background Questionnaire Extended Version

LANGUAGE BACKGROUND QUESTIONNAIRE – EXTENDED VERSION

Participant information – To be filled out by the researcher

Date: ________________  Participant code: __________  Participant group: ______________

The information you share by filling out this questionnaire will be stored anonymously and is strictly confidential, used for research purposes only.

1. Languages in your environment during infancy (0 to 24 months of age inclusively)

INSTRUCTIONS:
Please list all the languages that you were exposed to during your infancy and, for each language, the approximate percentage of the time that you heard it on a weekly basis.
Note: This should add up to 100%.

<table>
<thead>
<tr>
<th>Language:</th>
<th>% of exposure during infancy:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Current Language Proficiency

Please evaluate your current level of proficiency for all languages that you have ever been exposed to:
L2 perception of lexical tones by beginning and advanced Dutch learners of Mandarin

<table>
<thead>
<tr>
<th>Language 1: _________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking comprehension:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Speaking production:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Writing proficiency:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Reading proficiency:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Pronunciation:</td>
</tr>
<tr>
<td>very accented</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language 2: _________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking comprehension:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Speaking production:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Writing proficiency:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Reading proficiency:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Pronunciation:</td>
</tr>
<tr>
<td>very accented</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Language 3: _________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking comprehension:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Speaking production:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Writing proficiency:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Reading proficiency:</td>
</tr>
<tr>
<td>very low</td>
</tr>
<tr>
<td>Pronunciation:</td>
</tr>
<tr>
<td>very accented</td>
</tr>
</tbody>
</table>
If you have been exposed to more than the above 5 languages, please list the others here and comment on your general proficiency for each:

__________________________________________________________________________________________
__________________________________________________________________________________________
__________________________________________________________________________________________
3. Evolution of Language Use

**INSTRUCTIONS:**
- In the **first row**, fill in the ages, grades, or calendar years corresponding to the education level specified on top of each column.
- In each cell, use **percentages** to indicate your usage of/exposure to Dutch, Mandarin, and other languages (combined) for the corresponding context and age.
- **Note:** If your language use changed within these age groups, or if the age group is inaccurate with respect to education level, please specify by writing it inside the box or explain in the comments sections below.

If you have any questions about filling out this table, feel free to ask the researcher for help!

<table>
<thead>
<tr>
<th>Ages/school years/calendar years:</th>
<th>High school</th>
<th>Bachelor’s</th>
<th>Master’s</th>
<th>Other (exchange, internship):</th>
</tr>
</thead>
</table>
| School (=language of instruction) | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
| At home: interactions with family, significant other, roommates | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
| Friends: interactions with friends and significant other (if you did not live with them) | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
| Media use: social media, leisurely reading, television, cinema, radio, internet, music, etc. | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
| Extracurricular activities: sports, hobbies, work (if less than 20h/week), etc. | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
| Daily activities in the community: grocery store, shopping mall, restaurants, gas station, etc. | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
| Work (if over 20h/week): | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
| Other: | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % | Dutch: ___ %  
Mandarin: ___ %  
Other: ___ % |
Thank you for filling out this questionnaire!