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

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

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

4.1.1 Assessment of non-native segmental and suprasegmental perception

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

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

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

### 4.1.2 Perception of tones by native Mandarin speakers

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

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

4.1.3 Perception of Mandarin tones by non-native speakers

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

Learning to use tonal information in lexical access is another crucial issue in L2 tonal acquisition. To our knowledge, no systematic empirical research has been done to investigate tone processing in lexical access by L2 learners. Recently, learning to use tonal information in a lexically contrastive way has been investigated in several perceptual training studies by testing naive non-native speakers of Mandarin. Using the sound-to-word learning paradigm, which trains participants to associate members of minimal tone pairs with different meanings, these studies examined the contribution of individual variability in cue weighting in tone learning (Chandrasekaran, Sampath, & Wong, 2010), the effect of individual musical experience (Wong & Perrachione, 2007), as well as the influence of tonal context in tone learning (Chang & Bowles, 2015). Some studies also found a training-induced change in the participants’ neural system (Wong, Chandrasekaran, Garibaldi, & Wong, 2011; Wong, Perrachione, & Parrish, 2007). While the focus varies across these studies, the results lead to the convergent finding that naive non-native speakers of Mandarin can be trained to use pitch information lexically.

Based on this finding, it can be assumed that L2 learners of Mandarin can also use tonal information effectively in word processing. However, to what extent can they achieve native-like pattern? Are different tonal contrasts equally difficult for them in lexical access? These questions remain less understood. Therefore, the current study sets out to investigate the use of tonal information in lexical access by L2 learners of Mandarin in a lexical decision task.

4.1.4 The present study

In this study, the main research questions are:

(1) What is the developmental trajectory of the Dutch learners’ phonological processing of tonal contrasts?

(2) Can lexical tones be used properly in lexical activation by Dutch learners of Mandarin?

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

In this task, participants were asked to learn to associate two members of di-syllabic minimal tone pairs with the keys “a” and “b” in a training phase with feedback. In the test phase, a sequence of non-words was presented, and the task for the parti-
4.2 Experiment 1: Sequence recall task

4.2.1 Participants

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

4.2.2 Materials and design

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

---

5 Dutch has no phonemic contrast between a lax and tense velar stop, so that the marked member [g] is a new sound for Dutch learners, but they should know the sound from English, German and French, as well as from loan words that Dutch borrowed from these languages.
Table 4.1. Non-word stimuli used in the sequence-recall task.

<table>
<thead>
<tr>
<th>Experimental condition</th>
<th>Associated keys</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>T1-T2</td>
<td>ba1ti ba2ti</td>
<td></td>
</tr>
<tr>
<td>T1-T3</td>
<td>di3ka di1ka</td>
<td></td>
</tr>
<tr>
<td>T1-T4</td>
<td>gu1pa gu4pa</td>
<td></td>
</tr>
<tr>
<td>T2-T3</td>
<td>gu3pa gu2pa</td>
<td></td>
</tr>
<tr>
<td>T2-T4</td>
<td>di2ka di4ka</td>
<td></td>
</tr>
<tr>
<td>T3-T4</td>
<td>ba4ti ba3ti</td>
<td></td>
</tr>
<tr>
<td>Segmental control</td>
<td>fu1da fu1ga</td>
<td>Segmental control</td>
</tr>
<tr>
<td></td>
<td>su1bi su1di</td>
<td></td>
</tr>
</tbody>
</table>

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

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

### 4.2.3 Procedure

Each participant was tested individually in the Leiden University phonetics lab with all 128 trials with the auditory stimuli presented through Beyerdynamic DT-770 Pro headphones. The three groups of participants received instructions (in their native language) that they were going to learn some words in a foreign language. The six tonal pairs were tested separately in six experimental blocks. Each block embraces a word learning phase, a training phase and a main experimental phase. In the learning phase, participants were first asked to press “a” on the keyboard to hear the first foreign word. A sound token of one non-word from a tonal pair produced by a female speaker was played at this time (e.g., ba1ti). Then they were asked to press “b” on the keyboard, upon which the other sound token produced by the same female speaker was played (e.g., ba2ti). After that, the participants were presented with “a” or “b” on the screen.
Pressing the letter on the screen lead to the playing of one token of the corresponding non-word. In this way, participants heard all 16 tokens of each non-word within a tonal pair in random order by pressing the associated key. Subsequently in the training phase, the non-word-key association and distinctions between non-words were further trained in an identification task. After hearing a non-word, the participants were asked to press the associated key. They got feedback on their choice with the message of “Correct” or “Incorrect” on the screen. There were 16 training trials with stimuli produced by different speakers, presented in random order. An accuracy of 80% was defined as the success criterion. The participants moved on to the main experiment when they had reached this criterion. Two beginning learners could not satisfy this criterion and did not continue with the experiment. The remaining twelve beginners moved on to the main experiment. All advanced learners and native Mandarin listeners satisfied the criterion.

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

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

4.2.4 Results

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

The statistical results for the model of response accuracy are presented in Table 4.2. The $\chi^2$ and corresponding $p$ values for fixed and random effects were obtained from likelihood ratio tests. There were significant main effects of Participant Group, Tone Pair as well as a significant interaction between Participant Group and Tone Pair.
In the following, we will present a more detailed post-hoc analysis of the interaction of Participant Group and Tone Pair.

Table 4.2. Summary of a mixed effects logistic model for response accuracy. (The fixed effect Tone Pair comprises six tone pairs and the segmental control pair.)

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>df</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Group</td>
<td>2</td>
<td>30.65</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Tone Pair</td>
<td>6</td>
<td>140.06</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Participant Group × Tone Pair</td>
<td>12</td>
<td>115.82</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

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

Figure 4.1. Mean sequence-recall accuracy of three groups of participants in six tone pairs and the segmental control condition. BL: beginning Dutch learners of Mandarin; AL: advanced Dutch learners of Mandarin; NM: native Mandarin listeners. Error bars are ±1SE.
In the tonal conditions, pairwise comparison demonstrated that the accuracy was significantly different between each two groups of participants for all tone pairs (all p-values < 0.05) except T3T4. In all tone pairs, the accuracy of beginning learners was at the bottom but still well above chance performance level (chance level for 4-word sequence equals 1/24, which is 7%). Compared to beginners, the advanced learners had improved significantly in tone processing, but their performance was still below that of the native Mandarin listeners. The only insignificance was between the performance of advanced learners and native Mandarin listeners in T3T4 (AL = 72.8%, NM = 82.5%; Est. = 0.77, \(z = 2.10, p > 0.05\)). The performance of advanced learners in this pair was comparable to that of native Mandarin listeners.

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

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

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

4.3 Experiment 2: Lexical decision task

4.3.1 Materials and design

Ten disyllabic word-non-word pairs were chosen for each tone pair. For Dutch listeners, the stimulus ends with T1, T2 and T3 can be potentially interpreted as non-final (H%), which signals either continuation or question. T4 sounds like a final fall (H*LL%). So, to avoid the influence of different boundary tones from listeners’ L1, we kept the tone on the second syllable constant, only using real words with T1 on the second syllable as stimuli. The non-words were constructed by only changing the tone on the first syllable of the real words. Tone pairs were tested bi-directionally, which means that there were 12 pairs in total (T1→T2, T2→T1, T1→T3, T3→T1, T1→T4, T4→T1, T2→T3, T3→T2, T2→T4, T4→T2, T3→T4 and T4→T3). For example, for tone pair T1→T2, 10 real words with T1 on the first syllable were selected, and the corresponding non-words were constructed by changing T1 to T2 on the first syllable. E.g., the corresponding non-word for the real word 春天 [tʰun1tʰɪn1] ‘spring’ was
As a control condition, another 40 disyllabic word-non-word pairs which differed only in the initial consonant of the first syllable were chosen. There were ten words with T1 on the initial syllable, and ten words each with T2, T3 and T4. The non-words were constructed by changing the manner of articulation of the initial consonant in the initial syllable. For instance, the corresponding non-word for the real word 公园 [kʊŋ1yɛn2] ‘park’ was [kʰʊŋ1yɛn2]. To make sure the learners were familiar with all words, the real words were selected from the first year text books of the Chinese studies program at Leiden University (see Appendix A4 for a list of the stimuli).

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

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

4.3.2 Procedure

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

4.3.3 Results

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

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

(4.1)

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

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

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

For \( A' \) scores (see Table 4.3), there was a significant main effect of Participant Group and Word Pair. The interaction between Participant Group and Word Pair was also significant. For RT (see Table 4.3), there was a significant main effect of Participant Group and of the interaction between Participant Group and Word Pair.
Table 4.3. Summary of mixed effects models for A’ score and RT.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>P</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Group</td>
<td>2</td>
<td>76.12</td>
<td>&lt; 0.001</td>
<td>2</td>
<td>54.72</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Word Pair</td>
<td>12</td>
<td>102.78</td>
<td>&lt; 0.001</td>
<td>12</td>
<td>12.64</td>
<td>n.s.</td>
</tr>
<tr>
<td>Participant Group × Word Pair</td>
<td>24</td>
<td>174.4</td>
<td>&lt; 0.001</td>
<td>24</td>
<td>101.58</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant</td>
<td>6</td>
<td>507.38</td>
<td>&lt; 0.001</td>
<td>6</td>
<td>2381</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Item</td>
<td>1</td>
<td>847.64</td>
<td>&lt; 0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.2. Mean A’ score of three groups of participants for 12 tone pairs and the segmental condition. BL: beginning Dutch learners of Mandarin; AL: advanced Dutch learners of Mandarin; NM: native Mandarin listeners. Error bars are ±1SE.
In the segmental control condition, the A’ scores of all three groups significantly differed from each other (all p values < 0.05). That is, the advanced learners showed a significant improvement compared with the beginning learners, but still do not perform like native Mandarins. The mean RTs of beginners and advanced learners did not differ from each other, but both learner groups responded significantly slower than native Mandarin listeners (BL vs. NM: Est. = −1.08, z = −8.03, p < 0.001; AL vs. NM: Est. = −1.15, z = −8.94, p < 0.001). In the tonal condition, the three groups demonstrated a similar pattern in A’ scores compared to that in the segmental control condition. The advanced learners showed a significant improvement in most tone pairs compared to the beginning learners (all p values < 0.05), except for the pairs of T2→T3 and T3→T2, indicating that the sensitivity to T2 and T3 was still very low and resistant to improvement. The A’ scores of the native Mandarins are significantly higher than those of the advanced learners in most tone pairs (all p values < 0.001), except for the pairs T2→T4 (Est. = −0.01, z = −0.39, p = 1), T1→T3 (Est. = 0.06, z = 1.58, p = 0.343) and T1→T4 (Est. = 0.09, z = 2.24, p = 0.075). For RT, both learner groups responded significantly slower than the native Mandarins in all tone pairs (all p values < 0.001). The RTs of the two learner groups were comparable (all p values > 0.05). Native Mandarin listeners showed high sensitivity in both the segmental control condition and the tonal condition, and there was no significant difference between these two conditions for both A’ scores and RTs. Within the tonal condition, the overall A’ score was high across tone pairs. Among tone pairs, T4→T1 was the one with the lowest sensitivity score, significantly lower than T3→T2, T1→T2, T2→T1, T3→T4 and T4→T3 (all p values < 0.05). T3→T2 was the one with the highest sensitivity score, significantly higher than that of
T1→T1 (Est. = −0.13, z = −4.39, p < 0.001), T2→T3 (Est. = 0.11, z = 3.80, p < 0.5) and T2→T4 (Est. = 0.11, z = 3.64, p < 0.05). For each two tones, only T2 and T3 showed a directional difference in sensitivity, with an A’ score for T3→T2 significantly higher than T2→T3. That is, when T3 was produced as T2, native Mandarin listeners were more likely to make a correct response than the other way round. In the initial position, the category of T3 was more well-established than T2. For native Mandarin listeners, the RT was not significantly different across tone pairs.

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

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

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

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

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

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

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

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

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

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