Production and perception of tones by Dutch learners of Mandarin

Zou, T.; Zou T.

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

Chapter three

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

3.1 Introduction[4]

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

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

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

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

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

3.1.1 Phonetic and phonological processing of non-native contrasts

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

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

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

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

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

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

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

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

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

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

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

3.1.3 The present study

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

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

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

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

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

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

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

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

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

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

3.2 Methods

3.2.1 Participants

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

3.2.2 Stimuli

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

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

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

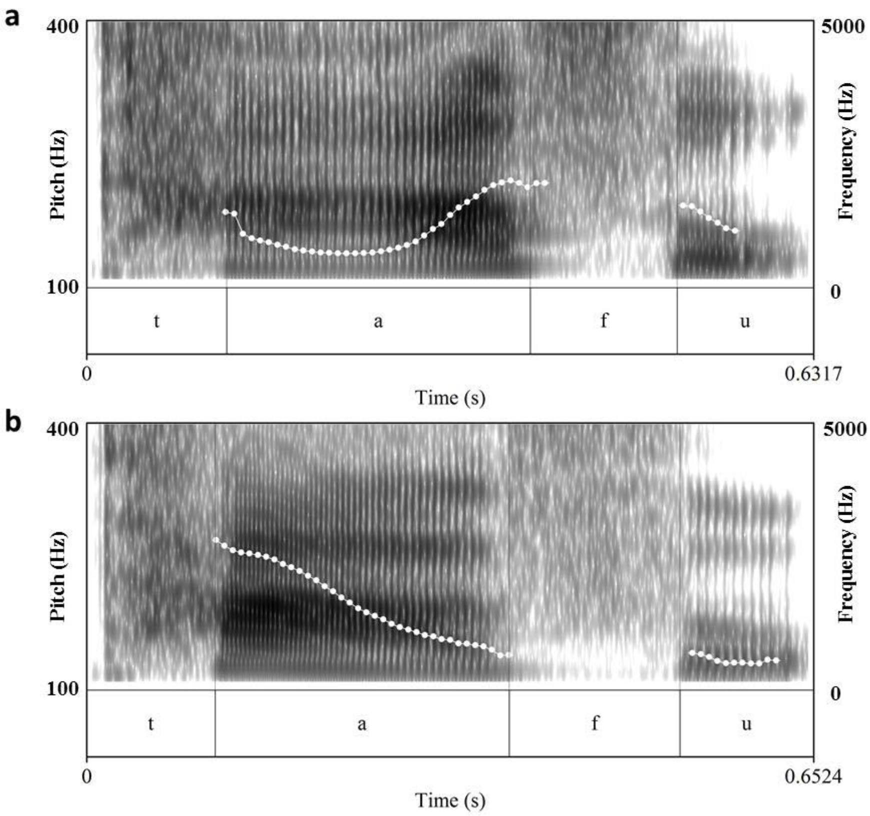

Figure 3.1. *Examples of the tonal sequences produced by a male speaker. Fig. 1a shows the pitch contour and the spectrogram of Tone 2 followed by a neutral tone. Fig. 1b shows the pitch contour and the spectrogram of Tone 4 followed by a neutral tone.*

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

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

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

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

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

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

3.2.3 Procedure

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

3.2.4 Statistical analyses

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

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

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

3.3 Results

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

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

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

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

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

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

3.3.1. Phonological processing of tonal contrasts

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

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

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

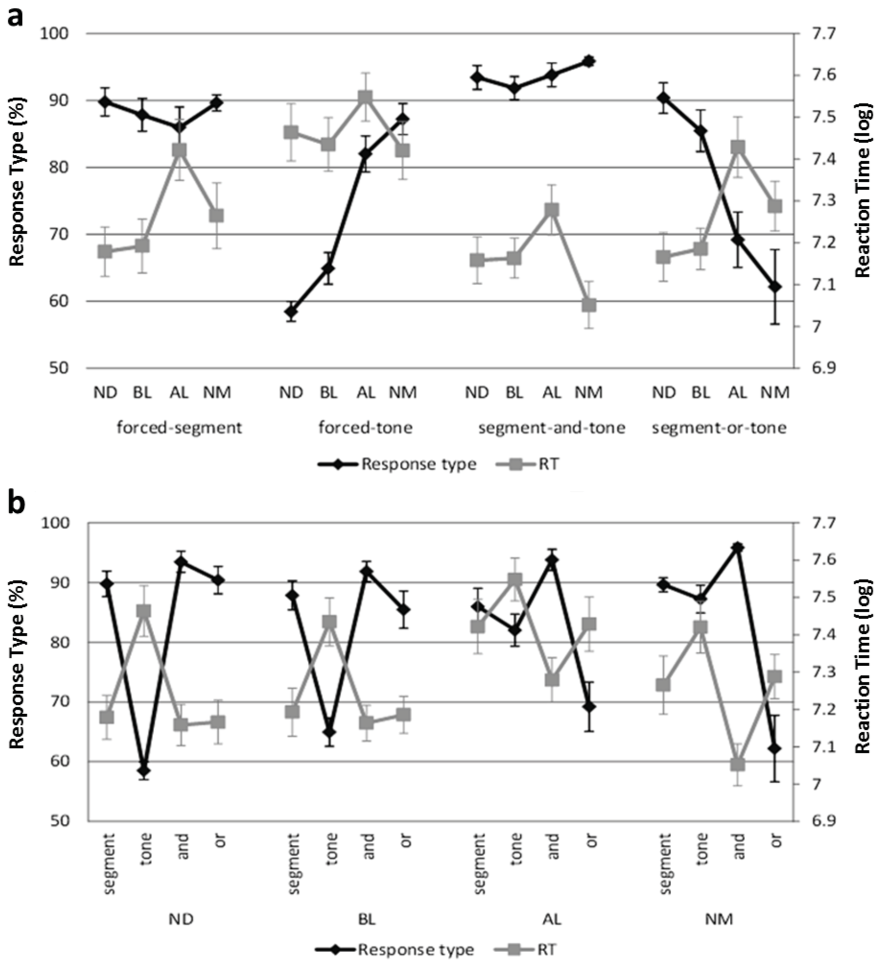

Figure 3.2. *Mean percentage of response types and RTs across participants with standard errors for four participant groups in four conditions. The black line shows percentage of segment-based, tone-based, segment-and-tone-based, and segment-based classification in forced-segment, forced-tone, segment-and-tone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

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

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

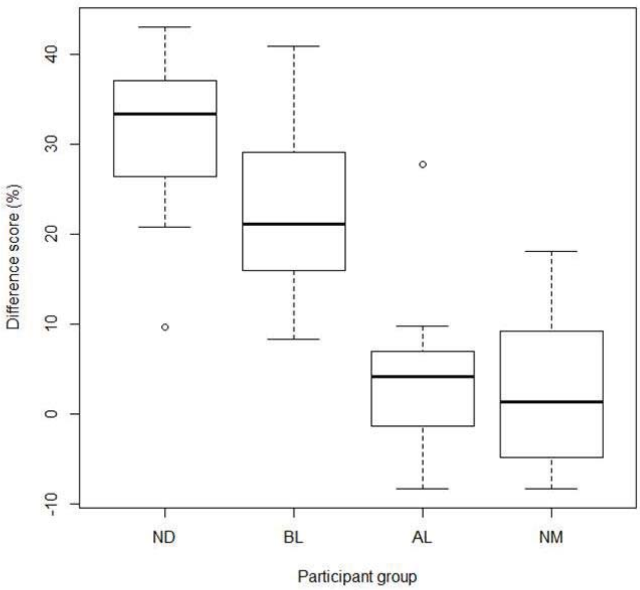

Figure 3.3. *Difference scores of the four participant groups. ND: native Dutch listeners without Mandarin experience; BL: beginning Dutch learners of Mandarin; AL: advanced Dutch learners of Mandarin, NM: native Mandarin listeners. The difference scores were calculated as the percentage of correct segmental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

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

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

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

In the segment-or-tone condition, native Mandarin listeners (62.2%) and advanced learners (69.2%) classified the stimuli along the segmental dimension significantly less often than beginning learners (85.5%) and native Dutch listeners (90.4%) (NM vs. BL: $z = -6.00$, $p < 0.001$; NM vs. ND: $z = -7.73$, $p < 0.001$; AL vs. BL: $z = 4.42$, $p < 0.001$; AL vs. ND: $z = 6.18$, $p < 0.001$). (Note that in Figure 3.2, the black line for this condition refers to the percentage of the segment-based classification.) Within each subgroup ({ND BL} versus {AL NM}), there was no significant difference, but there was a slight trend for native Mandarin listeners to be more attentive to the tonal dimension than advanced listeners, as well as for beginning learners to beg more attentive to the tonal dimension than native Dutch listeners (all p values < 0.1). For RT in this condition, advanced learners responded significantly slower than native Dutch listeners and beginning learners (AL vs. ND: $z = 3.23$, $p = 0.007$; AL vs. BL: $z = 3.00$, $p = 0.015$). Native Mandarin listeners were also slower than native Dutch listeners and beginning learners, but the differences were not statistically significant (all p values > 0.05).

3.3.3 Integrality of segmental and tonal information

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

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

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

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

3.4 Discussion and conclusion

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

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

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

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

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

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

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

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

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