



Universiteit
Leiden
The Netherlands

Production and perception of tones by Dutch learners of Mandarin
Zou, T.; Zou T.

Citation

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

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/52977>

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

Cover Page



Universiteit Leiden



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

Chapter five

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

5.1 Introduction

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

5.1.1 Tone processing by native Mandarin speakers

In Mandarin Chinese, tones are used to distinguish lexical meaning, and the role of tones in spoken word recognition by native speakers has received much attention in previous studies. Some studies suggest that tone might be a weaker cue in word recognition compared to segmental information. Repp and Lin (1990) found that in a speeded classification task, tone was accessed later than segment information, as Mandarin speakers took longer to respond to tonal distinctions than segmental distinctions. Cutler and Chen (1997) supports this finding by showing that in a lexical decision task, Cantonese listeners more often accepted a non-word as a real word when the only difference between the non-word and the word was in tone. Ye and Connine's study (1999) shows a similar result. In vowel and tone monitoring tasks without linguistic context, Mandarin listeners showed faster monitoring times for vowel mismatch stimuli than for tonal mismatches, which supported a perceptual advantage of vowel information. Studies by Sereno and Lee (2014) and Wiener and Turnbull (2016) also report similar results using direct priming and word reconstruction tasks, respectively. It might

be the case that the difference between tones and vowels are simply due to the difference in temporal availability of the cues. Prosody develops more slowly over time than segmental information, but every cue will be used in word recognition as soon as it is reliably perceived. The above-mentioned studies mainly recorded the end-state responses, which are not able to capture the perception of tones in real time.

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

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

5.1.2 Tone processing by non-tone language speakers

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

More recently, the learning of new tonal categories by speakers without prior tone language learning experience has been tested in several phonetic training studies. Adopting different training paradigms (perception-only training, perception-plus-pro-

duction training and sound-to-word training), these studies examined the role of different modalities in training (Lu, Wayland, & Kaan, 2015), the distinction between reflective and reflexive learning (Chandrasekaran, Yi, & Maddox, 2014), the contribution of individual variability in cue weighting in tone learning (Chandrasekaran, Sampath, & Wong, 2010), the effect of individual musical experience (Wong & Perrachione, 2007), as well as the influence of tonal context in tone learning (Chang & Bowles, 2015). Although the research questions varied across these studies, their results led to a convergent finding that naive non-native speakers of Mandarin can gain significant improvement in tonal identification and discrimination with a proper amount of training, and can learn to use tones in a lexically contrastive way. Some studies also found a training-induced change in participants' neural system (Lu et al., 2015; Wong, Chandrasekaran, Garibaldi, & Wong, 2011). Since most of the studies mentioned above focused on the learning of lexical tones by naive non-native Mandarin speakers and beginning learners of Mandarin, the performance of advanced L2 learners and the developmental trajectory in the time course of tone acquisition have not been studied systematically. Moreover, L2 processing of tonal information has not been investigated using on-line methods. Therefore, the current study sets out to examine the role of tones and segments in auditory spoken word recognition using the Visual World Paradigm by monitoring the eye movements of both beginners and advanced Dutch learners of Mandarin.

5.1.3 The present study

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

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

In the VWP task, participants are presented with a display of four pictures and an auditory stimulus corresponding to one of these pictures, and they are asked to identify the word (i.e., the target) they heard with their eye movements being tracked during the whole process. In studies using VWP, the target word is always presented with a competitor which is phonologically similar to the target, and two phonologically unrelated distractors. For example, in Allopena, Magnuson, and Tanenhaus (1998),

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

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

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

5.2 Method

5.2.1 Participants

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

5.2.2 Material

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

Table 5.1. *Experimental stimuli.*

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

Each word set includes one critical word e.g., *chuang1* ‘window’ and four competitors similar to the design in Malins & Joannis, (2010). The segmental competitor shared all phonemes with the critical word, but differed in tone e.g., *chuang2* ‘bed’; the cohort competitor shared the initial consonant and tone, but differed in the rest of the syllable e.g., *che1* ‘car’; the rhyme competitor shared the rime and tone, but differed in initial

consonant, e.g., *guang1* ‘light’;⁶ the tonal competitor shared the tone, but differed in all phonemes, e.g., *ji1* ‘chicken’. Since the contribution of tonal information in word recognition is one of the main issue we would like to examine, the stimuli (critical word-segmental competitors) in the current study cover all 12 possible tone contrasts (see Table 5.1 leftmost column) to get rid of the perception variability between tone pairs. The pair T2 versus T3 is more confusable than other tone pairs to discriminate for native Mandarin speakers (Bent, 2005; Shuai & Malins, 2017; Zhang, Samuel, & Liu, 2012). For L2 learners of Mandarin, the pairs of T2 vs. T3 as well as T1 vs. T4 are also more confusable than other tone pairs (Hao, 2012; Wang, Jongman, & Sereno, 1999; Wang, Sereno, Jongman, & Hirsch, 2003).

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

5.2.3 Procedure

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

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

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

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

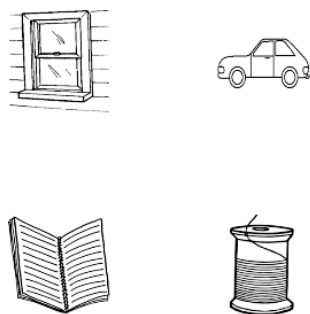


Figure 5.1. Example of the visual stimulus display, including a target *chuang1* ‘window’, a cohort competitor *che1* ‘car’, and two phonological unrelated distractors *ben3* ‘notebook’ and *xian4* ‘thread’.

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

5.2.4 Data analysis

In each trial, the visual display was divided according to the areas of the four items (i.e., target, competitor and distractors), and only looks within the corresponding areas were included in analyses. The proportion of looks to targets was calculated at each time point. The reciprocal trials were excluded from data analysis, leaving only the trials in which target words were heard. Trials in which participants failed to respond or

selected items other than the target were also excluded from the analyses of eye movements (102 out of 9,840 trials). Moreover, the trials in which participants did not initially look at the fixation cross were also left out (78 out of 9,840 trials). Eye movement data was analyzed from 200 ms to 1400 ms after stimulus onset. The starting point was chosen since it takes about 200 ms to launch an eye movement (Altman & Kamide, 2004; Hallet, 1986). The upper limit was chosen to capture the point at which the proportion of looks for the target word reached maximum for the learner groups (Malins & Joannis, 2010). The data recorded at 2-ms intervals was resampled so that the proportion of looks was calculated every 16 ms (62.5 Hz), comparable with the sampling rate used in Malins and Joannis (2010). The plotting of the proportion of looks was also based on the data of 62.5 Hz.

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

5.3 Results

5.3.1 Behavioral results

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

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

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

5.3.2 Fixation analysis for different participant groups

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

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

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

For the native Mandarin speakers, the proportion of looks to target as a function of time gradually ramped up and reached the maximum of approximate 90% around 1100 ms in all conditions, exhibiting a sigmoidal curve. The proportion of looks to target in the segmental condition (Figure 5.2a) showed a slightly delayed raising pattern and reached the maximum later than that in the baseline condition, but this difference was subtle and only the quadratic component reached significance (Est. = 0.10, $z = 3.58$, $p < 0.05$). The proportion of looks to the competitor in the segmental condition exhibited a higher peak compared to the baseline condition, as suggested by significant

difference in intercept (Est. = 0.01; $\chi = 6.85$, $p < 0.001$). The delay in target looks and the high proportion of looks to the competitor in the segmental condition indicated that segmental overlaps competed for lexical activation. Figure 5.2*b* compares the cohort and the baseline conditions. The looks to target in the cohort condition differed significantly from the baseline condition only in the cubic component (Est. = -0.16 , $\chi = -5.65$, $p < 0.001$). The looks to competitor, however, did not differ significantly between cohort and baseline conditions, indicating that the participants did not launch more looks to cohort competitors than to the baseline condition.

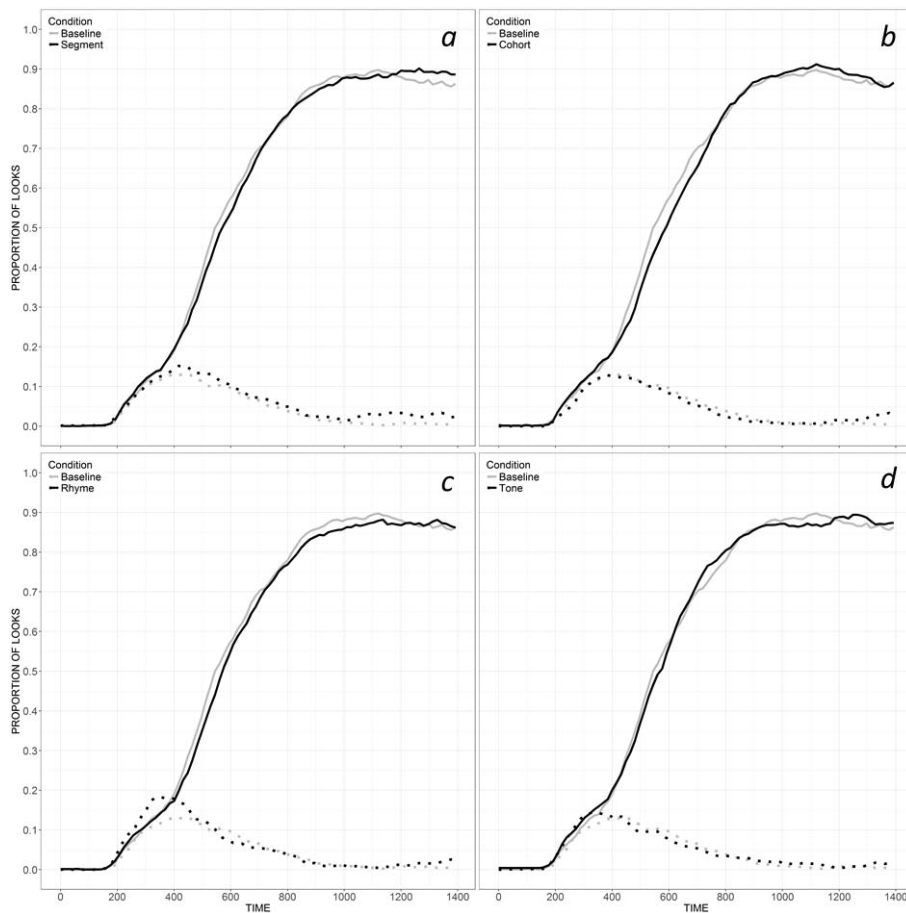


Figure 5.2. Mean proportion of looks to the target (solid line) and the competitor (dotted line) in different conditions averaged across participants and items for native Mandarin speakers. The segmental, cohort, rhyme and tonal conditions were plotted against the baseline condition for comparison.

As shown in Figure 5.2*c*, for native Mandarin speakers, the rhyme condition showed a lower proportion of looks to the target than the baseline condition during the entire

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

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

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

target in the tonal condition showed an early and lower peak compared to the baseline curve (Figure 5.3*d*), as reflected by a significant difference in the linear component (Est. = -0.11 , $\chi = -3.51$, $p < 0.05$), the quadratic component (Est. = -0.11 , $\chi = -3.65$, $p < 0.05$), and the cubic component (Est. = 0.19 , $\chi = 6.18$, $p < 0.001$). The general pattern of proportion of looks to the competitor was similar between the tonal and the baseline conditions, with significant differences only in the linear component (Est. = -0.09 , $\chi = -4.74$, $p < 0.001$) and the cubic component (Est. = -0.07 , $\chi = -3.53$, $p < 0.05$).

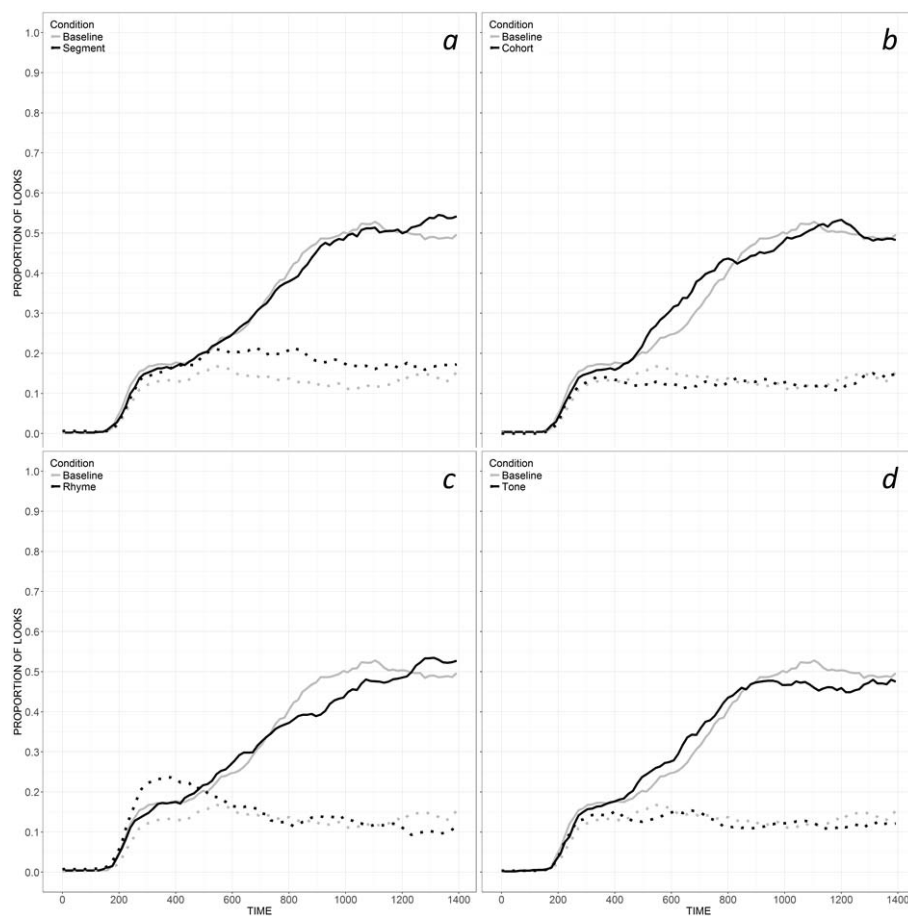


Figure 5.3. Mean proportion of looks to the target (solid line) and the competitor (dotted line) in different conditions averaged across participants and items for beginning Dutch learners of Mandarin. The segmental, cohort, rhyme and tonal conditions were plotted against the baseline condition for comparison.

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

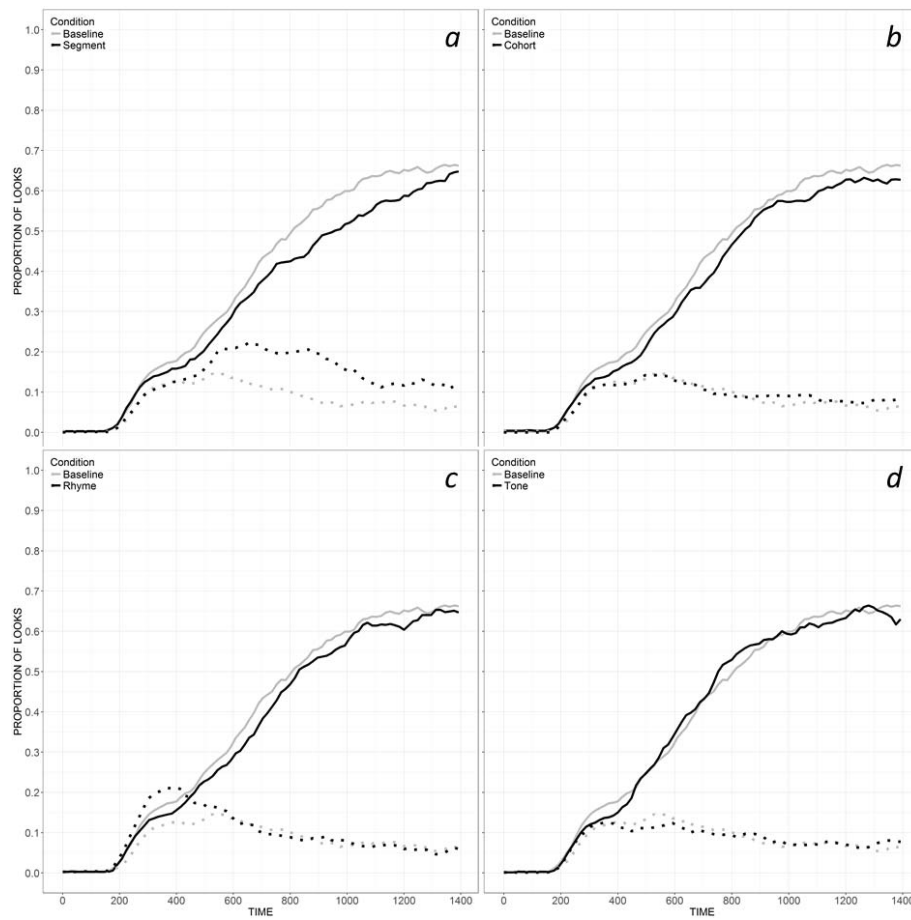


Figure 5.4. Mean proportion of looks to the target (solid line) and the competitor (dotted line) in different conditions averaged across participants and items for advanced Dutch learners of Mandarin. The segmental, cohort, rhyme and tonal condition was plotted against the baseline condition for comparison.

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

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

5.3.3 Comparison of fixation results across participant groups

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

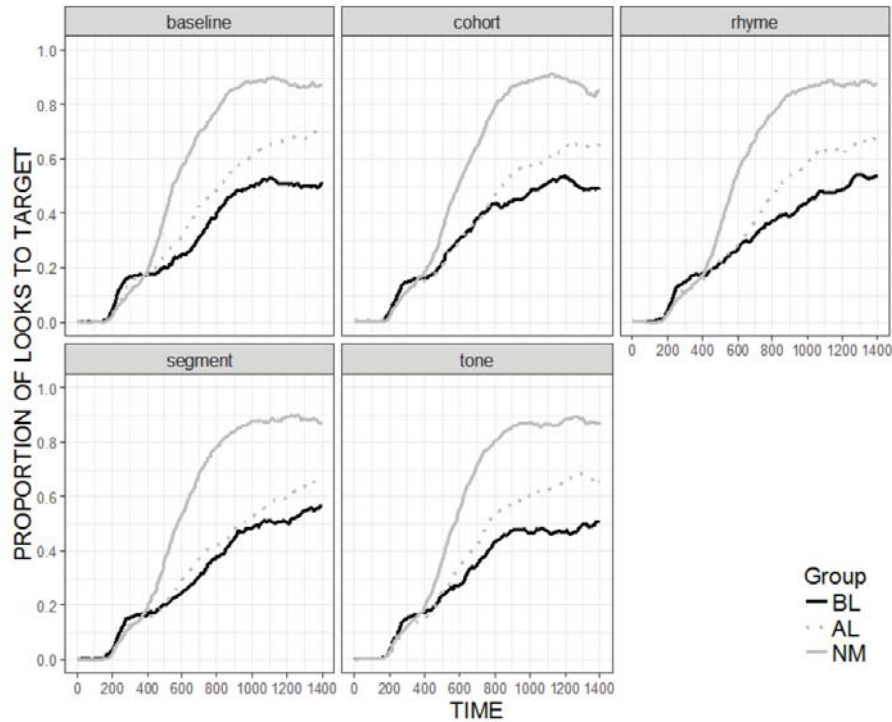


Figure 5.5. Mean proportion of looks to target in different conditions. The data from three participant groups are plotted together for comparison.

The proportion of looks to target in the native Mandarin speakers was significantly greater than for the two learner groups in all conditions, as suggested by significant differences in overall height, linear component and quadratic component (all p values < 0.01). Within the two learner groups, the advanced learners obtained a higher proportion of looks to target than the beginners in all conditions. In the baseline condition, the target looks of the advanced learners kept rising during the whole trial, while for the beginners, the proportion of looks to target reached a plateau of 50% at approximately 1100 ms. This discrepancy between the two learner groups was also suggested by significant differences in the intercept, linear, quadratic and cubic components (all p values < 0.01). In the cohort condition, the proportion of looks to target increased more rapidly for the advanced learners than beginning learners, indexed by significant differences in linear component (Est. = -0.53 , $\chi^2 = -18.56$, $p < 0.001$) and cubic component (Est. = 0.09 , $\chi^2 = 3.16$, $p < 0.05$) between the two groups. The advanced learners also showed a faster increase than beginners in the proportion of looks to target in the rhyme condition, as suggested by a significant difference in the linear component (Est. = -0.57 , $\chi^2 = -20.05$, $p < 0.001$), quadratic component (Est. = 0.11 , $\chi^2 = 3.97$, $p < 0.01$) and cubic component (Est. = 0.14 , $\chi^2 = 4.89$, $p < 0.001$). In the segmental condition, the proportion of looks to target only differed significantly between the learner groups in the linear

component (Est. = -0.33 , $\chi = -11.63$, $p < 0.001$). For tonal condition, the proportion of looks to target for advanced learners was larger than that of the beginners and showed a continuously rising tendency during the trial. This difference was confirmed by significantly different overall height (Est. = -0.09 , $\chi = -3.19$, $p < 0.05$), linear component (Est. = -0.68 , $\chi = -23.54$, $p < 0.001$) and quadratic component (Est. = 0.09 , $\chi = 3.02$, $p < 0.05$). Taken together, the advanced learners exhibited a larger proportion looks to target than beginning learners in almost all conditions, and approximated the performance of native Mandarin speakers.

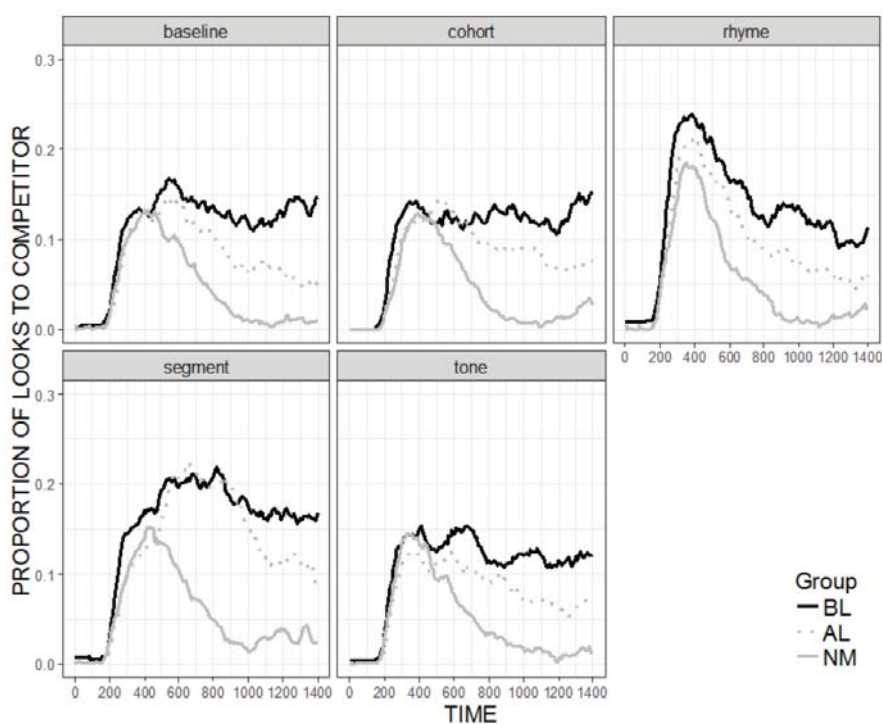


Figure 5.6. Mean proportion of looks to competitor in different conditions. The data from three participant groups are plotted together for comparison.

For native Mandarin speakers (Figure 5.6), the competitor look reached its maximum at ca. 400 ms, and then declined rapidly as the auditory stimulus unfolded in all conditions. The higher and early peak in the rhyme condition indicated that the competitor with rhyme and tone overlaps was fully activated to compete for recognition. The segmental competitor shared rhyme with the target but differed in tone. Compared to the rhyme competitor, the curve of the segmental competitor exhibited a lower and later peak. For the two learner groups, there was also an early peak for the proportion of looks to competitor, but the competitor looks did not decline rapidly. The post-hoc comparison

showed significant differences between native speakers and the learner groups in the overall height and linear component in all conditions (all p values < 0.05). Within the two learner groups, the proportion of looks to competitor for beginners was greater than the advanced learners in the baseline condition (intercept: Est. = 0.04, $\chi = 3.08$, $p < 0.05$; linear component: Est. = 0.17, $\chi = 9.86$, $p < 0.001$; quadratic component: Est. = 0.06, $\chi = 3.16$, $p < 0.05$). The competitor looks between the two learner groups also differed significantly in the cohort condition (linear component: Est. = 0.15, $\chi = 8.52$, $p < 0.001$; quadratic component: Est. = 0.07, $\chi = 4.24$, $p < 0.001$; cubic component: Est. = -0.05, $\chi = -3.03$, $p < 0.05$) and rhyme condition (intercept: Est. = 0.04, $\chi = 3.50$, $p < 0.01$). In the segmental condition, the proportion of looks to competitor was high for both learner groups (about 20%), indicating a great difficulty in discriminating minimal tone pairs. Significant differences were found between the learner groups in the linear (Est. = 0.09, $\chi = 5.38$, $p < 0.001$) and quadratic components (Est. = 0.16, $\chi = 9.11$, $p < 0.001$). In the tonal condition, the proportion of looks to competitor decreased more rapidly for advanced learners than for the beginners, which was reflected by a significant difference in the linear component (Est. = 0.08, $\chi = 4.27$, $p < 0.001$).

5.4 Discussion and conclusion

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

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

the segmental competitor was activated to a small extent. In the tonal condition, native Mandarin speakers did not launch more looks to the competitor than baseline, indicating that the tonal competitors were not sufficiently similar to the target word, and were not activated enough to compete for recognition. Among conditions, the curve of segmental competitor exhibited a lower and delayed peak compared to the curve of rhyme competitor, indicating the non-matching tonal information constrained the lexical activation of segmental competitors. The rhyme competitor was activated adequately in that it shared both rhyme and tonal information with the target. The segmental competitor, which only shared the rhyme part with the target but differed in tone, was less activated. This result indicates that for native Mandarin speakers, tones play an early constraining role. As the input unfolded, only those candidates matching the input both in segmental and tonal information (i.e., the rhyme competitor) were activated. Segmental competitors with non-matching tonal information were not activated in the early stage.

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

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

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

In conclusion, eye-tracking result suggests that native Mandarin speakers use tonal information effectively to constrain lexical activation in an early stage, similar to the way they exploit rhyme information. Beginning and advanced Dutch learners of Mandarin generally showed fewer fixations to the target and more looks to the competitor in the baseline, cohort, rhyme, segmental and tonal conditions, indicating the target was not fully activated and the competitor was not fully suppressed compared to

native speakers. Similar to native Mandarin speakers, both rhyme and tonal information was used by Dutch learners of Mandarin in an early stage to inhibit the activation of incompatible lexical candidates. Compared to beginning learners, significant improvement has been found for advanced learners into the native direction.

