CHAPTER 2

Interactions between Lexical Tone and Lexical-level Representations: Evidence from Lexical Decision Tasks

2.1. Introduction

The human brain must process two types of acoustic cues to recognise a spoken word: segmental and suprasegmental cues. Segmental cues refer to information that can be used to identify the boundaries of phonemes and/or phonemic combinations as lexical forms unfold. Therefore, segmental cues are mainly related to phonemes. Suprasegmental cues in spoken words refer to acoustic features that can be extended over a syllable or even over several words. In linguistics, suprasegmental cues include stress patterns, intonation and lexical tones. In words of most Indo-European languages, the use of suprasegmental cues is restricted to stress patterns, and thus, the recognition of a spoken word form mainly relies on mapping the acoustic input onto entries of lexicalised combinations of segmental units (i.e., phonemes), also called segment patterns.
Current linguistic models have focused on how segmental features are represented and accessed in speech perception without taking suprasegmental cues into consideration (TRACE model in McClelland & Elman, 1986; Neighbourhood Activation Model in Luce & Pisoni, 1998; Cohort Model in Marslen-Wilson, 1987; Shortlist Model in Norris, 1994). These models agree that phonological knowledge is stored at two separate levels of mental representations, namely the lexical level and the sublexical level (Dahan & Magnuson, 2006). Representations at the sublexical level encode phonemic knowledge (e.g., consonants and vowels), whereas the lexical level consists of representational units in the form of whole words. For example, the combination of the sublexical features [k] [ʌ] [p] yields a lexical entry, cup, which is a representational unit at the lexical level.

Lexical-level representations and sublexical representations are interactively accessed during spoken word recognition. For example, a series of studies have shown that, given sufficient contexts, the input of a phoneme or a syllable can directly access the lexical-level representation leading to word responses (Pulvermüller et al., 2001; Shtyrov et al., 2010; Shtyrov et al., 2011; MacGregor et al., 2012). Furthermore, a considerable number of studies have shown that the existence of lexical-level representations in turn can affect the perception of phonemes (Samuel, 2001; Magnuson et al., 2003; Norris et al., 2003; Mirman et al., 2005). These studies strongly support the TRACE model (McClelland & Elman, 1986) which suggests that lexical and sublexical representations have interconnections allowing both bottom-up and top-down processing.

Unlike most Indo-European languages, approximately 60-70% of languages in the world, including Chinese and Bantu languages, employ systematic variation of suprasegmental cues to distinguish word meanings (Yip, 2006). Lexical tone is an example of such suprasegmental cues. It is not yet completely clear how spoken words are recognised via their mental representations in tonal languages. This study aimed to investigate this issue, with a special focus on how lexical-level representations interact with sublexical representations of lexical tones, by examining participants’ lexical decisions of Mandarin word forms carrying different lexical tones.
2.1.1. Spoken word recognition in Mandarin Chinese

Mandarin Chinese is an ideal language for the purpose of the present study. Mandarin, referred to as *Putonghua* in mainland China, is the official language in most Chinese speaking countries and regions. A single syllable in Mandarin can be a word or a morpheme, and consists of an onset, a rime and a lexical tone. The onset only allows one consonant, no consonant clusters. The rime position hosts vowels and their nasalised forms with [n] and [ŋ]. Mandarin has four lexical tones, denoted as Tone1, Tone2, Tone3, and Tone4. Each syllable must be overlaid with one type of lexical tone in order to distinguish the different possible meanings of the same segment. For example, given the segment /ni/, /ni1/ (妮) means ‘girl’, /ni2/ (泥) means ‘mud’, /ni3/ (你) means ‘you’, and /ni4/ (腻) means ‘greasy’.

Two models of monosyllabic Mandarin word recognition are based on a division of lexical and sublexical representations, on a par with models for European languages (Ye & Connine, 1999; Zhao et al., 2011). In Mandarin Chinese, the sublexical-level representations are phonemes (onset consonants and rime vowels) and tonemes (Tone1, Tone2, Tone3, and Tone4). Both of the models allow bi-directional processing between the lexical and sublexical levels of representation. That is, in word recognition, the perception of sublexical features can lead to access to lexical-level representations, which in turn can influence processing via representations at the sublexical level. The two models have one major contrast in the structure of the lexical-level representation. Ye and Connine (1999) proposed a word-level representation that includes all lexical information (see Malins, 2013 for a similar proposal), whereas Zhao et al. (2011) held that the lexical-level representation can be further divided into an upper layer of semantic representations, and a lower layer of phonological representations, storing segment-tone patterns (or morphemes). Furthermore, Zhao et al. (2011) argued that the two layers have excitatory interconnections as well.
2.1.2. Lexical tones in Mandarin Chinese

The four lexical tones in Mandarin Chinese have different recognition trajectories due to their peculiar characteristics. Lexical tones are physically determined by the fundamental frequency (f0) of speech sounds. The most important f0 parameters of Mandarin tones are pitch height and pitch contour (Howie, 1976). The contour shapes of the four tones are described as follows: Tone1 is flat (−), Tone2 is rising (/), Tone3 is dipping (∨), and Tone4 is falling (\) (Chao, 1968; Duanmu, 2002). Moreover, the typical onset and offset pitch heights of the four tones can be described in a 5-point scale: Tone1 is 55, Tone2 is 35, Tone3 is 24 and Tone4 is 51. Since 5 is the highest and 1 is the lowest on this scale, Tone1 and Tone4 have similar onset pitch heights which are higher than those of Tone2 and Tone3. Accordingly, Tone1 and Tone4 start with a high register, and Tone2 and Tone3 have low register onsets. Moreover, Tone1, Tone2 and Tone3 have similar shapes for the initial part of their pitch contours (see Figure 1.1 in Chapter 1 for an illustration).

The four tones are not isolated (i.e., correctly recognised) in the same way due to discrepancies in their pitch heights and contour shapes. For example, by using a gating paradigm, Wu and Shu (2003) observed the isolation trajectories in spoken words of four tones gated with a 40-ms increment based on the initial 80-ms proportion (i.e., Gate 1 presented 80-ms of a word, Gate 2 presented 120-ms of a word, Gate 3 presented 160-ms of a word, and so on). Participants were instructed to provide a word according to the acoustic information they heard. By analysing the tonal dimension of their answers in each gate, they found that Tone2 required the longest time to be isolated compared to Tone1, Tone3 and Tone4. Moreover, further analyses on tone confusion revealed that in early gates, Tone1 and Tone4 were mutually confusing. This effect was attributed to the similar onset pitch height of the two tones. Moreover, Tone2 and Tone3 words were more likely to be misidentified as Tone1 words in the initial gates. Considering that Tone1, Tone2, and Tone3 have similarly shaped of pitch contours, this kind of mistake likely reflects a default use
of a high register tone with similar contour shapes when the available cue is limited for low register tones.

However, Wu and Shu (2003) did not match the occurrence probability of the segments. That is, some segment patterns occurred more often than the others. Consequently, the results can be confounded because the recurrence of the same segment might induce different processing strategies. To overcome this confounding factor, in another gating study, Lai and Zhang (2008) used eight word quadruplets, each of which consisted of words with the same segment carrying the four tones (i.e., /ma1/, /ma2/, /ma3/, and /ma4/). The researchers found that tones with high onset registers (Tone1 and Tone4) were recognised faster than those with low onset registers (Tone2 and Tone3). Furthermore, Tone1 was recognised faster than Tone4, but Tone2 and Tone3 had similar isolation points. The analyses revealed the same mistake patterns as in Wu and Shu (2003) that Tone1 and Tone4 were mutually confusing, and the onsets of Tone2 and Tone3 were more likely to be misidentified as Tone1, which has a higher onset pitch register but similar initial contour shapes as Tone2 and Tone3.

Although the pitch height and pitch contour have shown to affect lexical tone identification, it remains unclear whether the perception of pitch heights and pitch contours can be influenced by lexical-level representations. To examine this issue, we investigated lexical decision performance in Tone1, Tone2, and Tone4 real words and pseudo-words.

### 2.1.3. Lexical decisions in real words and pseudo-words

An auditory lexical decision task (LDT) entails timed classification of words and pseudo-words according to the lexical status of a spoken word form (Goldinger, 1996). Reaction time (RT) and accuracy are two dependent variables which have been used to infer the processes of spoken word perception. Higher RTs and lower accuracy rates are usually interpreted as indicating greater processing difficulty. The auditory LDT is believed to reflect lexical access because it is sensitive to word frequency (e.g., Slowiaczek & Pisoni, 1986), lexical competition (e.g.,
Gaskell & Dumay, 2003) and lexical status (e.g., Mimura, Verfaellie, & Millberg, 1997). The response latencies of real words are usually shorter than those of pseudo-words, and study participants are more likely to judge real words correctly relative to pseudo-words (Radeau et al., 1989).

Moreover, since pseudo-words do not have lexical entries in the mental lexicon, researchers can explore processing mechanisms at the lexical and sublexical levels of representations by comparing participants’ behavioural and neural responses to real words and pseudo-words (Radeau et al., 1989; Gaskell & Dumay, 2003; Sumner & Samuel, 2007; Shtyrov, Osswald, & Pulvermüller, 2008; McGettigan et al., 2011). To generate pseudo-words, researchers usually substitute one or more phonemes of a real word, resulting in a phonotactically possible phoneme or phonemic sequence that is not a real word. This kind of pseudo-word has no lexical meaning and does not violate sublexical principles. Therefore, such segment-manipulated pseudo-words can reflect sublexical processing without triggering additional processing costs induced by phonotactically impossible combinations of phonemes (e.g., *[stk] is phonotactically impossible in English).

For the current study of Mandarin Chinese tone words, we created tonal pseudo-word forms by combining a legal segmental string (i.e., a syllable) with a tone in a novel way. For example, /se4/ (色) is a lexical combination of /se/ and Tone4 that together means ‘colour’, whereas /se2/, consisting of /se/ + Tone2, is a pronounceable and phonotactically possible but meaningless combination in standard Mandarin Chinese. The most important characteristic of this kind of pseudo-word, named ‘tone-manipulated pseudo-word’, is that it does not involve any illegal segmental and suprasegmental elements, but the combination still lacks lexical representation at the lexical level.

2.1.4. The present study

The primary goal of the present study was to examine how lexical tones interact with lexical-level representations. For this purpose, we recorded the lexical decision performance in real words and pseudo-words carrying
Tone1, Tone2, and Tone4. As previous studies have revealed, the isolation time of the three tones can be ranked as Tone1 < Tone4 < Tone2 (Lai & Zhang, 2008). Moreover, Tone2 is likely to be misidentified as Tone1 in the initial phase of recognition because of their similar onset pitch contours, and Tone1 and Tone4 are mutually confusing due to their comparable onset pitch heights (Lai & Zhang, 2008; Wu & Shu, 2003). By comparing the lexical decision performance between Tone1, Tone2 and Tone4 in real words, we expected that the lexical decisions on real words are sensitive to lexical factors rather than tonal features because of the top-down influence on tone perception. That is, when the lexical factor was controlled, we did not expect differences between lexical decisions in real words with Tone1, Tone2, and Tone4. In contrast, the lexical decision of pseudo-words should be influenced by the distinctive resolution speed of the three tones and the shared onset pitch features.

The secondary purpose of this study was to investigate whether a division of phonological and semantic representation levels is necessary for Mandarin word recognition (Zhao et al., 2011). For this purpose, we developed two tasks of lexical decision making. In one lexical decision task, participants were asked to decide whether an auditory word form was a meaningful word (LDT of semantics); in another version of the experiment, participants were instructed to judge whether a word form was an existent sound pattern (LDT of phonology). In order to complete the LDT of semantics, participants needed to access the semantic component of lexical representations, whereas in the LDT of phonology they only needed to rely on the phonological components of the lexical-level representations (cf. Poldrack et al., 1999). If the lexical segment-tone patterns were represented in a level below semantics, we would expect differential lexical decision patterns between the two tasks, as measured by reaction times (RT).

2.2. Method

This experiment was a 2×2×3 design with the between-subject factor of Task (LDT of semantics vs. phonology) and two within-subject factors,
Lexicality (real word vs. pseudo-word) and Tone (Tone1 vs. Tone2 vs. Tone4). We did not include Tone 3 to avoid the reported mutual confusion between Tone2 and Tone3 based on their acoustic similarities (Wu & Shu, 2003; Lee, Tao, & Bond, 2008).

2.2.1. Participants

Eighty students from Harbin University of Commerce and Harbin Institute of Technology participated in the experiment. They were randomly divided into two equal groups. One group was instructed to perform auditory LDT of semantics (7 males, 33 females; \( M \) age = 19.8 years), whereas the other group was asked to perform auditory LDT of phonology (8 males, 32 females; \( M \) age = 22.3 years). They were all native speakers of Mandarin Chinese, as assessed by a questionnaire issued prior to the experiment. Of the 80 students, 75 were born and grew up in Heilongjiang Province and Jilin Province, known for their use of Standard Mandarin Chinese. Five participants were from the provinces of Tianjin City, Shanxi, Henan, and Hebei, and reported being monolingual speakers of Mandarin Chinese. The participants reported no hearing or language disorders.

2.2.2. Materials

The materials were monosyllabic Chinese spoken word forms, consisting of 27 real words (RW), 27 pseudo-words (PW) and 36 foils (27 real words and 9 pseudo-words). See Appendix A for a complete list of real and pseudo-words. Real words and pseudo-words were matched for high phonotactic probabilities (see Storkel, 2001), designated by the word frequency according to data from the Chinese Internet Word Frequency List of the Lancaster Corpus of Mandarin Chinese (McEnery & Xiao, 2004). The phonotactic probability of a pseudo-word is represented by the most frequently used word sharing the same segmental template but carrying different tones. Words appearing more than 100 times per million words were considered to have high phonotactic probability (cf. Zhang & Damian, 2009).
The real and pseudo-words consisted of equal numbers of stimuli with each tone (i.e., 9 each of Tone1, Tone2, and Tone3). The words of the three tonal categories were matched for word frequency \( F (2, 24) = 1.26, p = .303 \); pseudo-words were matched for phonotactic probability \( F (2, 24) = 0.425, p = .642 \). Real and pseudo-words were equally divided over three blocks, in each of which there was minimal overlap of onsets and rimes between the experimental stimuli and foils, except for one overlap of the initial onset /n/ for two pseudo-words in the third block.

All materials were produced by a female native Mandarin speaker in a sound proof studio with a U87 microphone, digitized via a Fireface 800 RME sound card, and recorded with Nuendo 6 (Steinburg Media Technology, Germany) at a sampling rate of 44.1 kHz. Each item was articulated 15 times. The initial and last four articulations were excluded from selection. Of the remaining items, one was selected based on two criteria: 1) the item was clearly articulated and easily recognised aurally, and 2) the f0 contour was clear and smooth based on visual inspection. All items were normalised for the same average intensity (65 dB) and duration (650 ms) using the acoustic software programme PRAAT (Boersma & Weenink, 2013). The f0 contours were not subject to normalisation in order to keep the items free from formant distortion. Figure 2.1 illustrates the f0 contours of all real and pseudo-words.

### 2.2.3. Procedure and Apparatus

Participants were tested individually in a sound attenuated room. The experimental stimuli presentation was programmed with DMDX V4.0.4.6 (Forster & Forster, 2003), and presented with an HP 540 laptop over a Cosonic CD-778MV headphone at 75 percent of the full volume of the laptop. Each participant was randomly assigned to either the LDT of semantics group or LDT of phonology group, and was asked to press ‘O’ for an existent word and ‘P’ for a non-existent word form on the laptop keyboard as quickly and accurately as they could. A practice session with nine trials was administered before the experimental session. A participant was not allowed to proceed to the first experimental block until the
feedback showed no more than three erroneous judgments out of nine trials.

![Figure 2.1](image)

**Figure 2.1** Schematic demonstrations of f0 contours of the 54 experimental items. Tone1 words are in black; Tone2 words are in red; Tone4 words are in blue. Note the similar onset pitch heights between Tone1 and Tone4 word forms, and the similar initial pitch contours between Tone1 and Tone2 word forms.

The experiment consisted of 30 trials in each block (9 real words and 9 pseudo-words). At the beginning of each trial, a fixation mark was presented for 1000 ms in the centre of the screen, followed by the aural presentation of a stimulus. After a 50 ms interval from the offset of the stimulus, a question mark was shown in the screen centre, prompting the participant for their judgment. In each block, the presentation of items was pseudo-randomised on an individual basis. The same experimental item type was not presented more than three times in a row. A 1-minute break was given between each block to avoid potential inference from the phonemic overlaps across blocks.
2.2.4. Data analysis

Data were pre-processed separately for the LDT of semantics and LDT of phonology responses. For the LDT of semantics, RTs of erroneous responses, and responses shorter than 50 ms (Luce, 1986) or longer than 3 $SD$ (1527 ms) were omitted. The 314 omitted data points (195 erroneous, 21 shorter than 50 ms, 198 longer than 3 $SD$) accounted for 14% of the total observations. The rate was lower than the 15% level advised by Ratcliff (1993) for lexical decision studies. For the LDT of phonology, 183 erroneous data responses, 32 trials shorter than 50 ms, and 99 trials longer than 3 $SD$ of 1445 ms were excluded, accounting for 14% of the total trials. The accuracy rates were obtained from the original data without data trimming.

Three-way repeated measures ANOVAs were first performed with the average RTs and accuracy rates by participant and by item separately. The by-item analysis was performed to rule out the possibility that any significant effects in by-subject analysis were caused by some specific items. If interactions between Task, Lexicality and Tone, and between Task and either Lexicality or Tone were identified, further analyses unpacked from the interactions were performed within the two tasks separately. If not, the data of the two tasks were combined for other analyses. Planned comparisons were conducted to identify which factors led to differential lexical decisions among word forms with the three tones. Lastly, we conducted multiple pairwise comparisons (paired $t$-tests) between every two-tone group (Tone1 vs. Tone2, Tone1 vs. Tone4, and Tone1 vs. Tone2) separately for real and pseudo-words with Bonferroni’s correction to directly reveal the responses patterns. Greenhouse-Geisser correction was performed when the sphericity assumption in ANOVA was violated. The alpha level was set to $p < .05$. All $p$ values were corrected when Bonferroni’s correction or Greenhouse-Geisser correction were applied. The statistical analyses were conducted with SPSS V22 (IBM, Armonk, NY, USA).
2.3. Results

The average RTs and accuracy rates in real words and pseudo-words carrying three types of lexical tones in the two tasks can be seen in Table 2.1. Descriptive statistics showed that real words elicited longer RTs and higher accuracy rates than pseudo-words. Moreover, the LDT of semantics were generally longer than those of phonology.

2.3.1. Reaction times

Three-way ANOVAs identified Task as a marginally significant between-subject effect when the RTs were calculated by participant (\(F(1, 78) = 3.91, p = .052\)), and a significant main effect by item (\(F(1, 16) = 36.42, p < .001\)). The results suggest that the RTs in the LDT of phonology are shorter than those in the LDT of semantics. However, there is no interaction between Task and the other two within-subject factors, Lexicality and Tone (by subject \(F(2, 156) = 1.00, p = .370\); by item \(F(2, 32) = .26, p = .775\)), nor interaction between Task and Tone (by subject \(F(2, 156) = 1.37, p = .257\); by item \(F(2, 32) = 0.323, p = .726\)). These results show that the general patterns of RTs are the same in the two tasks. The only effect induced by the tasks is an overall slower response in the LDT of semantics than in the LDT of phonology.

Moreover, there was a main effect of Lexicality (by subject \(F(1, 78) = 39.49, p < .001\); by item \(F(1, 16) = 35.85, p < .001\)). This result confirms that real words induce shorter RTs than pseudo-words. There was also a main effect of Tone (by subject \(F(2, 156) = 5.82, p = .004\); by item \(F(2, 32) = 4.39, p = .021\)). However, the interaction between Lexicality and Tone was not significant (by subject \(F(2, 156) = 2.34, p = .100\); by item \(F(2, 32) = 0.994, p = .381\)). This implies that the RT pattern in real words with Tone1, Tone2, and Tone4 does not differ from the pattern in pseudo-words with the three tones.
2.3. Results

The average RTs and accuracy rates in real words and pseudo-words carrying three types of lexical tones in the two tasks can be seen in Table 2.1. Descriptive statistics showed that real words elicited longer RTs and higher accuracy rates than pseudo-words. Moreover, the LDT of semantics were generally longer than those of phonology.

2.3.1. Reaction times

Three-way ANOVAs identified Task as a marginally significant between-subject effect when the RTs were calculated by participant \(F(1, 78) = 3.91, p = .052\), and a significant main effect by item \(F(1, 16) = 36.42, p < .001\). The results suggest that the RTs in the LDT of phonology are shorter than those in the LDT of semantics. However, there is no interaction between Task and the other two within-subject factors, Lexicality and Tone (by subject \(F(2, 156) = 1.00, p = .370\); by item \(F(2, 32) = .26, p = .775\)), nor interaction between Task and Tone (by subject \(F(2, 156) = 1.37, p = .257\); by item \(F(2, 32) = 0.323, p = .726\)). These results show that the general patterns of RTs are the same in the two tasks. The only effect induced by the tasks is an overall slower response in the LDT of semantics than in the LDT of phonology.

Moreover, there was a main effect of Lexicality (by subject \(F(1, 78) = 39.49, p < .001\); by item \(F(1, 16) = 35.85, p < .001\)). This result confirms that real words induce shorter RTs than pseudo-words. There was also a main effect of Tone (by subject \(F(2, 156) = 5.82, p = .004\); by item \(F(2, 32) = 4.39, p = .021\)). However, the interaction between Lexicality and Tone was not significant (by subject \(F(2, 156) = 2.34, p = .100\); by item \(F(2, 32) = 0.994, p = .381\)). This implies that the RT pattern in real words with Tone1, Tone2, and Tone4 does not differ from the pattern in pseudo-words with the three tones.

**Table 2.1** Mean RTs and accuracy rates as functions of the lexicality of stimuli in the LDTs of semantics and phonology separately. (T1=Tone1, T2=Tone2, T3=Tone3, RW = real word, PW = pseudo-word)

<table>
<thead>
<tr>
<th></th>
<th>Reaction Time (ms)</th>
<th>Accuracy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>(SD)</td>
</tr>
<tr>
<td>LDT of semantics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1RW</td>
<td>497.9</td>
<td>219.1</td>
</tr>
<tr>
<td>T2RW</td>
<td>486.8</td>
<td>208.4</td>
</tr>
<tr>
<td>T4RW</td>
<td>507.8</td>
<td>198.5</td>
</tr>
<tr>
<td>T1PW</td>
<td>569.5</td>
<td>251.7</td>
</tr>
<tr>
<td>T2PW</td>
<td>612.4</td>
<td>289.8</td>
</tr>
<tr>
<td>T4PW</td>
<td>646.3</td>
<td>263.2</td>
</tr>
<tr>
<td>LDT of phonology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1RW</td>
<td>425.6</td>
<td>160.6</td>
</tr>
<tr>
<td>T2RW</td>
<td>411.6</td>
<td>141.3</td>
</tr>
<tr>
<td>T4RW</td>
<td>431.2</td>
<td>142.9</td>
</tr>
<tr>
<td>T1PW</td>
<td>513.5</td>
<td>225.3</td>
</tr>
<tr>
<td>T2PW</td>
<td>498.7</td>
<td>224.0</td>
</tr>
<tr>
<td>T4PW</td>
<td>541.0</td>
<td>205.6</td>
</tr>
</tbody>
</table>
Nonetheless, Bonferroni corrected pairwise comparisons between the RTs in every two tone categories in real and pseudo-words still showed that Tone1 pseudo-words were recognised significantly more slowly than Tone4 pseudo-words by subject (t(79) = -3.18, p = .013), and with marginal significance by item (t(17) = -2.94, p = .054). The shorter RTs in Tone2 pseudo-words than Tone4 pseudo-words were a marginally significant effect in the comparison by subject (t(79) = -2.63, p = .06), and a significant effect in the comparison by item (t(17) = -3.35, p = .024). No significant effects were identified in other comparisons. The results are displayed in Figure 2.2.

2.3.2. Accuracy

Three-way repeated measures ANOVAs with the accuracy data showed that the between-subject factor Task was not a main effect (by subject $F(1, 78) = 0.14$, $p = .706$; by item $F(1, 16) = 0.04$, $p = .844$). This result indicates that accuracy rates of the two versions of LDT do not differ from each other. Moreover, there were no interactions between Task, Lexicality, and Tone (by subject $F(2, 156) = 1.40$, $p = .250$; by item $F(2, 32) = 0.36$, $p = .70$).

With respect to the within-subject factors, Lexicality was a main effect (by subject $F(1, 78) = 32.71$, $p < .001$; by item $F(1, 16) = 7.93$, $p = .012$). This result implies that real words have higher accuracy rates compared to pseudo-words. Moreover, a main effect of Tone and an interaction between Lexicality and Tone were identified (Tone: by subject $F(2, 156) = 11.50$, $p < .001$; by item $F(2, 32) = 3.39$, $p = .046$. Lexicality $\times$ Tone: by subject $F(2, 156) = 29.77$, $p < .001$; by item $F(2, 32) = 7.65$, $p = .002$). Planned comparisons revealed a main effect of Tone in both real words (by subject $F(2, 156) = 11.37$, $p < .001$; by item $F(2, 32) = 3.77$, $p = .033$) and pseudo-words (by subject $F(2, 156) = 24.08$, $p < .001$; by item $F(2, 32) = 6.70$, $p = .004$). These results suggest that accuracy rates do not vary with the task instructions, but the accuracy rates in real words
with the three tones have different patterns from the accuracy rates in the three tonal categories of pseudo-words.

Multiple comparisons revealed significant differences between Tone1 real words and Tone4 real words (by subject $t(79) = 4.17, p < .001$; by item $t(17) = 3.24, p = .03$) and between Tone1 pseudo-words and Tone2 pseudo-words (by subject $t(79) = -6.31, p < .001$; by item $t(17) = -3.70, p = .012$). These results suggest that Tone1 words are more likely to be correctly judged than Tone4 words (see Figure 2.3). However, the LDTs in pseudo-words exhibit a different pattern that show Tone1 pseudo-words are prone to more erroneous judgment than Tone2 pseudo-words.

**Figure 2.1** Mean RTs of lexical decisions on real words and pseudo-words with Tone1, Tone2, and Tone4. Error bars are standard error means. Note the RT of Tone4 pseudo-words is longer than the RTs of Tone1 and Tone2 pseudo-words. # $p = .06$, * $p < .05$ (significance marked according to analyses by subject).
Figure 2.2 Mean accuracy rates of lexical decisions in real words and pseudo-words with Tone1, Tone2, and Tone4. Error bars are standard error means. *** $p < .001$ (significance marked according to analyses by subject).

2.4. Discussion

The present study used two versions of an auditory LDT to examine the interaction between lexical-level representations and lexical tone representations. We compared the RTs and accuracy rates of lexical decisions in three tonal categories of real words and in the same tonal categories of pseudo-words in Mandarin Chinese. Both the RT and accuracy data show that the pattern of decisions between real words with the three tones is distinct from the decision pattern in pseudo-words. Thus, these results suggest that lexical-level representations influence the role of lexical tones in Mandarin word recognition.

2.4.1. The influence of lexical-level representations on tone-word recognition reflected by RTs

The RT data show different response speeds for the decisions on pseudo-words with Tone1, Tone2, and Tone4, whereas the lexical decisions on
real words do not exhibit differential RTs amongst the three tonal categories. These results imply that recognising real words is not affected by lexical-tone perception, whereas the speed of recognising a pseudo tone word may be related to the isolation of its lexical tone.

The lexical decision task has been used to reflect underlying processes in lexical access (Goldinger, 1996). Since the lexical frequencies between words of the three tonal categories are matched, comparable RTs are expected in Tone1, Tone2 and, Tone4 real words. This result would suggest that lexical tones in real words can be instantly integrated into segments and processed by accessing whole-word representations. Our interpretation is in agreement with the study by Zhao et al. (2011), which found that the N400 elicited in a lexical tone mismatch condition had similar latency and amplitude as the N400s in onset mismatch and rime mismatch conditions. The N400 is a neural indicator of difficulties in lexical access, usually observed around 400 ms post stimulus onset, as proposed by Kutas and Hillyard (1984). Their finding indicated that in lexical access, all segmental cues and tonal (i.e., suprasegmental) cues are utilised simultaneously.

In contrast with the similar RTs in reals words of the three tonal categories, Tone4 pseudo-words were judged more slowly than Tone1 and Tone2 pseudo-words in the present study. Due to the lack of lexical-level representations in pseudo-words, our results suggest that such temporal differences are a reflection of recognising pseudo-words via sublexical representations. Thus, we attribute the RT contrasts between Tone1 and Tone4 pseudo-words, and between Tone2 and Tone4 pseudo-words, to the differential time courses for isolating the three tones. The slower responses to Tone4 pseudo-words than to Tone1 pseudo-words are in line with one previous study showing that the isolation point of Tone1 is earlier than that of Tone4 (Lai & Zhang, 2008). However, contradictory to our data, existing studies have reported that Tone2 has a later isolation point than Tone1 and Tone4 (Wu & Shu, 2003; Lai & Zhang, 2008). In our study, there are no differences between the response latencies of Tone1 and Tone2 pseudo-words, and Tone2 pseudo-words show a tendency to be judged faster than Tone4 pseudo-words. One possible explanation is
that it takes more time to distinguish Tone4 from Tone3. This is because Tone4 has a falling pitch contour (\), whereas Tone3 has a falling-dipping contour shape (\\). Therefore, the brain may recognise Tone4 only after confirmation that a falling tone is not accompanied by a dipping tone. These cognitive processes do not apply to the flat Tone1 (–) and rising Tone2 (/). However, since the current experiment does not involve Tone3 in its design, an alternative explanation is also needed.

We tentatively propose another account that may lie in the normalisation of stimulus duration. In the current study, the average duration of Tone1 pseudo-word recordings is 570 ms, Tone2 is 691 ms, and Tone4 is 654 ms. In order to normalise the duration into 650 ms, Tone1 pseudo-words needed to be stretched, while Tone2 pseudo-words needed to be compressed. Tone4 pseudo-words had to undergo very small temporal changes. Consequently, the isolation points might occur later for Tone1 and earlier for Tone2, while remaining more or less the same for Tone4. Therefore, we observe the tendency of slowest responses in Tone4 pseudo-words, and no difference in responses between Tone1 and Tone2 pseudo-words. However, further studies need to be conducted to provide greater evidence of this explanation.

2.4.2. Accuracy and perception of tonal features

Accuracy data show that the lexical decisions of Tone4 real words are more prone to errors than those of Tone1 real words, but Tone1 pseudo-words are more likely to be misjudged than Tone2 pseudo-words. Since Tone1 and Tone4 have similar onset pitch heights, and Tone1 and Tone2 share similar initial pitch contours, the accuracy data may reflect how lexical-level representations interact with the sublexical features of lexical tones.

It is striking that the accuracy data show quite a different picture from the one reflected by the RT patterns. Our data reveal that lexical tones not only lead to contrastive judgment between Tone1 and Tone2 pseudo-words, but also in real words carrying Tone1 and Tone4. Specifically, in lexical decisions on real words, Tone4 words are more likely than Tone1
words to be misjudged as pseudo-words; in decisions on pseudo-words, participants have greater difficulty in deciding the lexical status of Tone1 pseudo-words than Tone2 pseudo-words.

To explain these results, two questions must be answered. First, why are errors committed contrastively between two tones in real words (i.e., Tone1 vs. Tone4) and pseudo-words (i.e., Tone1 vs. Tone2)? Second, why are the error patterns different between real words and pseudo-words? We answer the two questions by considering the acoustic confusion between Tone1 and Tone2, and between Tone1 and Tone4.

Acoustically, Tone1 shares common points with both Tone4 and Tone2 (Guo, 1993). For example, as previously mentioned, Tone1 and Tone4 have the similar onset pitch heights, and the initial proportions of the pitch contours of Tone1 and Tone2 have similar flat shapes. However, the rising Tone2 and falling Tone4 are highly contrastive. In line with these linguistic facts, studies have found that when the tonal input is limited, Tone1 and Tone4 are mutually confusing, and Tone2 is more likely to be misidentified as Tone1 (Wu & Shu, 2003; Lai & Zhang, 2008). We argue that in our experiment, the processing of lexical tones may undergo the same tonal interference between Tone1 and Tone2, and between Tone1 and Tone4. Therefore, the errors become contrastive between Tone1 and Tone2 pseudo-words, and between Tone1 and Tone4 real words.

If our account holds true, how do we explain the distinctive error patterns between real words and pseudo-words? Our suggestion is that this lexicality selectivity in errors is induced by the different ways to perceive lexical tones in word forms with and without lexical-level representations. That is, when lexical-level information is available, there is interference between the perception of tone words with similar onset pitch heights (Tone1 vs. Tone4), implying that participants can perceive the onset pitch height but cannot use the pitch contour cue to differentiate the two tones. Contrastively, when there is no lexical-level representation, the lexical decision on pseudo-words is more error-prone in distinguishing between tones with similar pitch contours (Tone1 vs. Tone2), indicating sensitivity
to pitch contour and hindrance to the perception of pitch height (Tone1 vs. Tone4).

It can be further inferred that real-word recognition may be sensitive to pitch height which is mediated by the top-down influence of lexical-level representations, whereas the better resolution of pitch contours in pseudo-word recognition suggests dependence on sublexical level processing, given a lack of top-down influence. Our inference is possible because top-down influence can facilitate lexical-tone perception in the very beginning stages of word recognition, whereas lexical-tone perception in pseudo-words does not have such facilitation, and naturally becomes more sensitive to slow varying cues like pitch contour.

One additional issue that must be noted is that interference between two contrastive tones has specific orientations. In our study, although Tone4 real words have lower accuracy rates, most Tone1 words are correctly judged. In contrast with this pattern, Tone1 pseudo-words are more likely to be misjudged, but decisions on Tone2 pseudo-words have high accuracy. This means that the proposed sublexical interference only affects the lexical decision of Tone4 real words and Tone1 pseudo-words. A parsimonious explanation is that Tone4 real words sound more like pseudo-words, and Tone1 pseudo-words sound more like real words. Furthermore, the lexical decision performance just so happens to be contrastive between Tone1 and Tone4 real words, and between Tone1 and Tone2 pseudo-words.

However, the factorial design of this study allows us to perform a rigorous examination of lexical performance, which reveals highly systemic variations contributed by an interaction between word forms’ lexical status and the tones they carry. Therefore, we refute the coincidental view, and hold that the orientations of tonal interference are related to specific tonal perception modulated by lexical-level representations. Since Tone1 and Tone4 are mutually confusing in the initial phase of processing (Lai & Zhang, 2008; Wu & Shu, 2003), Tone1 and Tone4 may compete with each other in the early phase of processing for similar onset pitch heights. At the sublexical level, Tone1 can be isolated earlier than Tone4, making Tone4 word perception subject to
more interference from the competition with Tone1. Such competition may induce lower quality of feedforward activation of the lexical-level representation, and eventually lead to less accurate lexical decisions in Tone4 words. For decisions in pseudo-words, due to the lack of top-down facilitation, lexical-tone perception occurs at a relatively slow pace via sublexical representations. Therefore, in the current study, Tone1 could be prone to competition not only with Tone2 (similar pitch contours) but also with Tone4 (similar onset pitch height), whereas Tone2 only needs to compete with Tone1. Consequently, the more complex competition interferes with lexical decision accuracy in Tone1 pseudo-words.

To summarise, the contrastive accuracy patterns between lexical decision performance in real words and pseudo-words carrying different tones imply divergent sensitivities to onset pitch height when lexical-level representation is available, and to pitch contour shape when lexical-tone processing is via sublexical representations. These results suggest that in the current experiment, accuracy and RT reflect different cognitive processes in tone-word perception. Future studies should be conducted using electrophysiological techniques to look for more online evidence for our explanations.

2.4.3. Semantic and phonological contents at the lexical level

The secondary purpose of the current study was to test whether there is a division between the semantic and phonological aspects of lexical-level representations (Zhao et al., 2011). By using the LDT of semantics and the LDT of phonology, our study achieved results that do not support this view. The task requiring participants to judge lexicality according to the semantic dimension (i.e., LDT of semantics) only led to generally prolonged RTs than in the task requiring participants to judge the existence of a sound pattern (i.e., LDT of phonology). This finding suggests that access to phonological representations is faster than that of semantic representations, as previous studies have shown (Friederici, 2002; Rodriguez-Fornells et al., 2002; Schmitt, Kutas, & Münte, 2002). Moreover, the overall faster responses in the LDT of phonology can be
explained by both of the views that lexical phonology and semantics are at the same level (Ye & Connine, 1999; Malins, 2003) or by the view of separate layers (Zhao et al., 2011). Such evidence can be explained as both successful inhibition of semantic units within the same representational level or by top-down facilitation from the above semantic level.

Apart from the RT findings, there is no differentiation in accuracy between the two versions of LDT. This result suggests that the semantic and phonological representations at the lexical-level do not have substantial differences in terms of exerting top-down influence on the perception of sublexical features. Moreover, the two tasks did not influence the perception of word forms of the three tonal categories differently. This result indicates that task instructions that are biased on the word’s semantics or phonology do not change the way lexical tones and lexical-level representations interact with each other. Therefore, our study suggests that the lexical-level representation can be a multi-dimensional structure hosting both semantic and phonological components. Moreover, it is not necessary to place the semantic representation and the phonological representation in two separate layers as suggested by Zhao et al. (2011).

### 2.5. Conclusion

In this study, we investigated how lexical tones and lexical-level representations affect lexical decision behaviours. Using two versions of an auditory LDT, we found evidence that lexical tones modulated lexical decisions differently in word forms with and without lexical-level representations. Moreover, different dependent variables in the LDT can indicate distinctive aspects of lexical-tone perception in spoken word recognition. In addition, the semantic and phonological components of the lexical-level representation modulate lexical-tone perception similarly.