Tone-word recognition in Mandarin Chinese
Yue, Jinxing

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 1

Introduction

In the long history of biological evolution, the faculty of language has become a unique trait of the human brain which distinguishes our species from an enormous amount of others (Hauser, Chomsky, & Fitch, 2002). Speech, the meaningful vocalisations used to convey language by humans (Fitch, 2000), is the primary tool of communication in most cultures. The physical form of speech is acoustic signals composed of various co-occurring frequencies (Fitch, Miller, & Tallal, 1997). Spoken words are recognised by analysing acoustically complex cues at remarkable speed. There has been wide consensus that spoken word recognition is realised by a specialised system (or processor) continuously mapping acoustic input onto multiple levels of representations (see Lieberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Samuel, 2001). These representations can be roughly divided into lexical and sublexical levels (Dahan & Magnuson, 2006). At the sublexical level, the recognition system encodes smaller units of spoken words, such as syllables, phonemes, and phonetic features; at the lexical level, the representations are lexicalised combinations of sublexical units. Apart from the phonological components, lexical representations can also include semantic components (Swingley & Aslin, 2000). The current thesis
focuses on the phonological component of lexical representations, which are also referred to as ‘whole-word phonological representations’.

1.1. The roles of lexical-level representations in word recognition

The means of access to various levels of representations during spoken word recognition has been a matter of debate, but multiple studies have shown that lexical-level representations can be accessed as a spoken word unfolds, and may even influence the perception of sublexical elements. Supportive evidence has been reported in studies with behavioural tasks, revealing that the perception of sublexical features is facilitated by feedback from lexical-level representations (Samuel, 1981; Connine & Clifton, 1987; Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Newman, Sawusch, & Luce, 1997; Samuel, 1996; Samuel, 2001; Magnuson, McMurry, Tanenhaus, & Aslin, 2003; Norris, McQueen, & Cutler, 2003; Mirman, McClelland, & Holt, 2005). For example, by using a phoneme identification task, Samuel (2001) asked native speakers of English to judge whether words ended with [s] (as in tremendous) or [ʃ] (as in diminish). The results showed that participants reported more items with [s] endings when the identification task was preceded by an adaptation with words ending with [s], whereas participants reported more items with [ʃ] endings when the adaptors ended with [ʃ]. Their finding was straightforward evidence of top-down effects from lexical-level representations in processing via sublexical representations.

These data could be explained by the TRACE model, a speech perception model which allows both bottom-up and top-down processing via a three-layer structure of word representation (McClelland & Elman, 1986). The TRACE model proposed that the bottom layer was a detector of acoustic features; the middle layer stored the more abstract representations of phonemes; and the topmost layer was the lexical level of word representations. Interactions within a layer were competition via
Tone-word Recognition in Mandarin Chinese

2.0 focuses on the phonological component of lexical representations, which are also referred to as ‘whole-word phonological representations’.

1.1. The roles of lexical-level representations in word recognition

The means of access to various levels of representations during spoken word recognition has been a matter of debate, but multiple studies have shown that lexical-level representations can be accessed as a spoken word unfolds, and may even influence the perception of sublexical elements. Supportive evidence has been reported in studies with behavioural tasks, revealing that the perception of sublexical features is facilitated by feedback from lexical-level representations (Samuel, 1981; Connine & Clifton, 1987; Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Newman, Sawusch, & Luce, 1997; Samuel, 1996; Samuel, 2001; Magnuson, McMurry, Tanenhaus, & Aslin, 2003; Norris, McQueen, & Cutler, 2003; Mirman, McClelland, & Holt, 2005). For example, by using a phoneme identification task, Samuel (2001) asked native speakers of English to judge whether words ended with [s] (as in tremendous) or [ʃ] (as in diminish). The results showed that participants reported more items with [s] endings when the identification task was preceded by an adaptation with words ending with [s], whereas participants reported more items with [ʃ] endings when the adaptors ended with [ʃ]. Their finding was straightforward evidence of top-down effects from lexical-level representations in processing via sublexical representations.

These data could be explained by the TRACE model, a speech perception model which allows both bottom-up and top-down processing via a three-layer structure of word representations (McClelland & Elman, 1986). The TRACE model proposed that the bottom layer was a detector of acoustic features; the middle layer stored the more abstract representations of phonemes; and the topmost layer was the lexical level of word representations. Interactions within a layer were competition via inhibitory connections, whereas between-layer connections were excitatory in nature and permitted facilitations.

In line with these psycholinguistic findings, studies recording neural activities with electroencephalography (EEG) and magnetoencephalography (MEG) have reported evidence that, under experimental contexts, the perception of a specific phoneme can lead to rapid access to lexical representations. For example, by using the mismatch negativity (MMN) as the neural indicator of access to lexical representations in the long-term memory, Pulvermüller et al. (2001) found divergent MMN responses between a real Finnish word stimulus and a pseudo-word stimulus around 150 ms after the recognition point, after which a word form can be recognised given the subsequent phoneme(s); the onsets of the phoneme corresponding with the underlined letter in lakki and pseudo-word *vakki are the recognition points. In a more recent study, MacGregor, Pulvermüller, van Casteren, and Shtyrov (2012) used a passive listening task with multiple items derived from English, finding MEG evidence that lexical access could begin as early as 50-80 ms post the recognition point. More importantly, by carefully matching the acoustic features and setting control conditions, these studies revealed that the different electrical or magnetic responses were induced by lexicality contrasts instead of low-level acoustic and/or phonemic contrasts (for more evidence, see Shtyrov, Nikulin, & Pulvermüller, 2010; Shtyrov, 2011; Shtyrov, Kimppa, Pulvermüller, & Kujala, 2011).

Taking the above-mentioned studies together, the existing behavioural and neurophysiological data have provided two important implications for the role of lexical-level representations. First, word representations at a lexical level are dynamically involved in the analysis of acoustic input, which can even influence the perception of sublexical features in a top-down fashion. Second, access to lexical-level representations can occur readily even before attention sets in.

The materials used in most previous studies distinguished word meanings only based on segmental cues, namely consonants and vowels. For example, *pipe is a meaningful English word but *pite is not (Shtyrov et al., 2010). In contrast, many languages express distinctive meanings with
the same segmental patterns when they are overlaid with different lexical
tones, a suprasegmental cue. These languages are called tone languages
and include Mandarin Chinese, Thai, and Shona; the words in these
languages are called tone words. For a tone word, the meaning is
determined not only by the phonemic sequence but also by the lexical tone.
However, the role of lexical-level representations during recognition of
words carrying lexical tones remains unclear. Thus, the present thesis
follows the two indications of studies in non-tone languages to explore the
role of lexical-level representations in the recognition of spoken tone
words.

1.2. Segmental and suprasegmental cues

A lexical-level representation has been primarily regarded as a
conceptualised combination of segmental cues in studies of European
languages. These studies can shed limited light on words exploiting rich
suprasegmental cues, another type of acoustic device used to distinguish
word meanings between identical segment patterns. Suprasegmental cues
in a broad sense refer to properties of speech such as timber, loudness,
length, and pitch accent, features which cannot be derived from the
phonemes (Nooteboom, 1997). Pitch accent is a feature typically
represented by the fundamental frequency (f0) of a speech sound, which
can be extended over a syllable, a word or even the length of a sentence
segment (e.g., interrogative intonation).

For spoken words in most Indo-European languages, suprasegmental
cues refer to a specific type of pitch accent: lexical stress, whose variation
pattern in the same segment is quite limited. For example, the typical stress
patterns of most English words like abide, abnormal, situate, and
operation are fixed¹. The poverty of suprasegmental variation may have
led to an exclusion of stress patterns in the current psycholinguistic models
of spoken word recognition in Indo-European languages (e.g., TRACE

¹ The stress patterns can vary in diverse contexts. However, these variations are not
recognised as typical stress patterns, nor lead to changes in the lexical semantics.
model, McClelland & Elman, 1986; Neighbourhood Activation Model, Luce & Pisoni, 1998; Cohort Model, Marslen-Wilson, 1987), even though a number of studies have shown that lexical stress indeed plays a role in constraining word recognition in non-tone languages (e.g., van Heuven, 1985; Cutler & Van Donselaar, 2001; Donselaar, Koster, & Cutler, 2005).

The perception of segmental and suprasegmental cues has been found to be related to differential neural networks. The neural network involved in analysing segmental cues is more lateralised to the left hemisphere, whereas the network for processing pitch information in non-tonal languages has been primarily localised in the right hemisphere (e.g., right inferior frontal gyrus), activated together with involvement of the left-hemispheric homologue (Whalen & Liberman, 1987; Liberman & Whalen, 2000; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre & Belin, 2001; Friederici & Alter, 2004; Zatorre & Gandour, 2008). Physically, segmental cues are rapidly changing cues with rich temporal information (e.g., voice onset time), whereas suprasegmental cues are slowly varying cues represented by spectral features (Zatorre & Gandour, 2008). One account for such hemispheric lateralisation is that the left and right hemispheres have biased sensitivities to analysing acoustic cues on different timescales (Boemio, Fromm, Braun, & Poeppel, 2005), and eventually develop specified neural representations to respond to more abstract phonological inputs such as phonemes and stress patterns (Zatorre & Gandour, 2008).

It is estimated that 60-70% of the languages in the world (Yip, 2006), including Mandarin, Cantonese, Thai, Vietnamese, and Shona, employ lexical tones to regulate word meanings together with segmental cues. Unlike most Indo-European languages, these tonal languages are comprised of word forms wherein the same segment can have completely different meanings based on contrastive tones. This role of lexical tone in determining word meanings is on a par with the role of phonemes in non-tone languages. Therefore, for tone languages, tones have been described
as tonemes\textsuperscript{2} in some psycholinguistic studies of Mandarin Chinese (Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011). The way spoken words carrying lexical tones are recognised has become an intriguing issue. Not surprisingly, existing studies have mainly investigated the characteristics of lexical-tone perception (Ren, Yang, & Li, 2009; Wong, Parsons, Martinez, & Diehl, 2004; Hsieh, Gandour, Wong, & Hutchins, 2001; Gandour et al., 2000) and/or compared the similarity and disparity between the processing of segmental and tonal cues (Luo et al., 2006; Li et al., 2010; Zhao et al., 2011; Sereno & Lee, 2015). However, relatively little is known about the role of lexical-level representations in spoken tone-word recognition. The present thesis attempts to add to the understanding of how lexical-level representations operate in the recognition of monosyllabic words of Mandarin Chinese.

In the following part of this chapter, we will review some most important concepts of Mandarin Chinese and relevant studies on Mandarin word recognition. A few studies on Cantonese and Thai will also be mentioned as a supplement. The review section is followed by an elaboration of the aims and contents of this thesis.

1.3. Linguistic features of Mandarin Chinese

Mandarin Chinese, called *Putonghua* in Mainland China, is an official language in most Chinese speaking countries and regions. A syllable in Mandarin consists of an onset, a rime and a lexical tone. The onset, equivalent to *shengmu* in Chinese, refers to the initial position in the syllable and allows only one consonant but no consonant clusters (e.g., [sk]). The rime, *yunmu* in Chinese, can be a vowel or a diphthong followed by an optional coda which is restricted to one of two nasal consonants, namely [n] and [ŋ]. When the rime is a vowel, the vowel is the nucleus of the syllable. Diphthong rimes include two types: one type has a glide preceding a central vowel (e.g., /ia/; /ua/), whereas the other type has a

\textsuperscript{2} Toneme is a linguistic concept, and was originally introduced by linguists around 1940s (Jones, 1944; Pike, 1947).
Tone-word Recognition in Mandarin Chinese

As tonemes in some psycholinguistic studies of Mandarin Chinese (Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011). The way spoken words carrying lexical tones are recognised has become an intriguing issue. Not surprisingly, existing studies have mainly investigated the characteristics of lexical-tone perception (Ren, Yang, & Li, 2009; Wong, Parsons, Martinez, & Diehl, 2004; Hsieh, Gandour, Wong, & Hutchins, 2001; Gandour et al., 2000) and/or compared the similarity and disparity between the processing of segmental and tonal cues (Luo et al., 2006; Li et al., 2010; Zhao et al., 2011; Sereno & Lee, 2015). However, relatively little is known about the role of lexical-level representations in spoken tone-word recognition. The present thesis attempts to add to the understanding of how lexical-level representations operate in the recognition of monosyllabic words of Mandarin Chinese.

In the following part of this chapter, we will review some most important concepts of Mandarin Chinese and relevant studies on Mandarin word recognition. A few studies on Cantonese and Thai will also be mentioned as a supplement. The review section is followed by an elaboration of the aims and contents of this thesis.

1.3. Linguistic features of Mandarin Chinese

Mandarin Chinese, called Putonghua in Mainland China, is an official language in most Chinese speaking countries and regions. A syllable in Mandarin consists of an onset, a rime and a lexical tone. The onset, equivalent to shengmu in Chinese, refers to the initial position in the syllable and allows only one consonant but no consonant clusters (e.g., [sk]). The rime, yunmu in Chinese, can be a vowel or a diphthong followed by an optional coda which is restricted to one of two nasal consonants, namely [n] and [ŋ]. When the rime is a vowel, the vowel is the nucleus of the syllable. Diphthong rimes include two types: one type has a glide preceding a central vowel (e.g., /ia/; /ei/), whereas the other type has a glide succeeding a central vowel (e.g., /ai/; /ei/). Mandarin has four lexical tones, denoted as Tone1, Tone2, Tone3, and Tone4. Lexical tones are physically determined by the fundamental frequency (f0) of a syllable. The shapes of the f0 contours of each tone are described as follows: Tone 1 is flat (−), Tone 2 is rising (\), Tone 3 is low dipping (\), and Tone 4 is high falling (\) (Duanmu, 2002).

The rime of each syllable must be overlaid with one lexical tone to distinguish the meanings of words having the same segment. A well-known example is the combination of /ma/ with the four tones, yielding four words with completely different meanings: /ma1/ ‘mother’, /ma2/ ‘hemp’, /ma3/ ‘horse’, and /ma4/ ‘to curse’. Figure 1.1 illustrates the f0 contours of the four words based on the segment /ma/ produced by a female speaker of Mandarin.

![Figure 1.1](image.png)

**Figure 1.1** The f0 contours of the four tones. The demonstrated words are /ma1/ ‘mother’, /ma2/ ‘hemp’, /ma3/ ‘horse’, and /ma4/ ‘to curse’.

---

3 A spoken Chinese word form is transcribed in Pinyin script, an official Latinised Chinese transcription system, between / /. The tone number is written after the segments making up the syllable.
Syllables with lexical tones are the basic units of the Chinese lexicon. A syllable is a morpheme that is either a word by itself or is part of a polysyllabic word. For example, /chi1/ (吃) means ‘eat’ as an independent word. However, when it is combined with /xiao3/ (小), forming the noun /xiao3chi1/ (小吃), the meaning is ‘snack’ (literally translated as small eat). Monosyllabic words are the most frequently used word type in modern Chinese (Language Teaching Research Centre of Beijing Language Institute, 1986). Therefore, the current thesis targets monosyllabic spoken words for the initial probe into the role of lexical-level representations in spoken tone-word recognition.

1.4. Cognitive and neural bases for tone-word recognition

Lexical tone processing has been the topic of much debate by researchers of tone-word recognition. An overview of previous studies has led to two important indications. First, lexical tones and segmental cues can be processed separately via differential cognitive and neural mechanisms. Second, the processing of lexical tones and segmental cues is subject to top-down regulation of lexical-level representations.

1.4.1. Separate lexical-tone and segment perception

A body of literature has documented evidence that segmental cues (i.e., onset and rime) and lexical tone are recognised via separate cognitive and neural mechanisms, even though lexical tones are physically entangled with rimes. For example, Liang and van Heuven (2004) reported a case study of an aphasic speaker of Chinese with left-hemispheric lesions. They found selective impairment of lexical tone identification with spared production and perception of vowels. This case was strong evidence that the segmental and tonal aspects of Chinese words are processed separately via mechanisms lateralised to the left hemisphere. In an electrophysiological study, Luo et al. (2006) used Mismatch Negativity (MMN) to indicate the brain’s responses to consonants and lexical tones.
Syllables with lexical tones are the basic units of the Chinese lexicon. A syllable is a morpheme that is either a word by itself or is part of a polysyllabic word. For example, /chi1/ (吃) means 'eat' as an independent word. However, when it is combined with /xiao3/ (小), forming the noun /xiao3chi1/ (小吃), the meaning is 'snack' (literally translated as small eat).

Monosyllabic words are the most frequently used word type in modern Chinese (Language Teaching Research Centre of Beijing Language Institute, 1986). Therefore, the current thesis targets monosyllabic spoken words for the initial probe into the role of lexical-level representations in spoken tone-word recognition.

1.4. Cognitive and neural bases for tone-word recognition

Lexical tone processing has been the topic of much debate by researchers of tone-word recognition. An overview of previous studies has led to two important indications. First, lexical tones and segmental cues can be processed separately via differential cognitive and neural mechanisms. Second, the processing of lexical tones and segmental cues is subject to top-down regulation of lexical-level representations.

1.4.1. Separate lexical-tone and segment perception

A body of literature has documented evidence that segmental cues (i.e., onset and rime) and lexical tone are recognised via separate cognitive and neural mechanisms, even though lexical tones are physically entangled with rimes. For example, Liang and van Heuven (2004) reported a case study of an aphasic speaker of Chinese with left-hemispheric lesions. They found selective impairment of lexical tone identification with spared production and perception of vowels. This case was strong evidence that the segmental and tonal aspects of Chinese words are processed separately via mechanisms lateralised to the left hemisphere. In an electrophysiological study, Luo et al. (2006) used Mismatch Negativity (MMN) to indicate the brain's responses to consonants and lexical tones of Mandarin words in the pre-lexical processing stage. The MMN is an event-related brain potential (ERP) component, elicited by infrequent auditory events (‘deviant’ stimuli) presented intermittently among other, more frequently presented auditory events (‘standard’ stimuli). The MMN can reflect the neural activities involved in discriminating the acoustic difference between deviant and standard stimuli (Näätänen, 1995). Luo and colleagues found that MMNs in lexical-tone contrasts were more pronounced at the right-hemispheric recording electrodes relative to MMNs to consonant contrasts, which were stronger over the left-hemispheric recording sites. Moreover, the tone-related MMNs had earlier latencies than consonant-related MMNs. These findings suggested that lexical tones and segmental cues in Mandarin Chinese are processed differently in terms of their time course and the underlying neural anatomy.

In line with this right-hemispheric dominance in tone-related MMNs, one study with functional magnetic resonance imaging (fMRI) in an auditory immediate recognition task also revealed more prominent tone-related activation over the right hemisphere (Li et al., 2010). The researchers found that, although monitoring lexical-tone contrasts and segmental contrasts (i.e., onset / rime) in a word sequence led to activations in largely overlapped cortical regions in both hemispheres, stronger activation in the frontoparietal areas of the right hemisphere was identified in direct comparisons between tone-contrast monitoring and consonant-contrast monitoring, and between tone-contrast monitoring and rime-contrast monitoring. These findings indicated that lexical-tone perception is supported by cortical regions distributed over the left and right hemispheres, but it is the right-hemispheric neural response that distinguishes lexical-tone perception from segmental cue perception.

However, a few studies have shown that the right-hemispheric response to lexical tone is due to the perception of acoustic features rather than the activation of tonal representations with linguistic specifications. For example, the functional imaging data reported by Gandour et al. (2000) in a study with Positron Emission Tomography (PET), showed that in judging lexical tones of Thai words, the cortical activation in native speakers was lateralised to the left hemisphere, whereas the activation in
both native Chinese (tone language) and English (non-tone language) speakers was dominant in the right hemisphere (for similar results, see Hsieh et al., 2001; Wong et al., 2004). These results suggest that lexical specification of tones in the left hemisphere hinges on linguistic experience, but the right hemisphere may be responsible for capturing acoustic cues which are slowly varying (Zatorre & Gandour, 2008).

This functional division of the two hemispheres in processing lexical tone is also supported by a PET study (Wong et al., 2004). The researchers found that the left insula was most active when native Chinese-speaking participants performed Chinese tone discrimination. However, when the stimuli were English words superimposed with Chinese tones, the activation of the right insula was similar in native English-speaking participants and native Chinese speakers. Moreover, native English speakers exhibited the same response pattern in the right hemisphere when listening to Chinese tone words. These findings suggest that the way a lexical tone is perceived is highly influenced by higher order auditory representations (Griffiths & Warren, 2002).

1.4.2. The role of lexical tones in word recognition

Whether lexical tones are processed in the same fashion as segmental cues is highly controversial. A number of studies with behavioural measurements have repeatedly shown that the processing of segmental cues has a primacy over the processing of lexical tone. Lexical tones are processed more slowly and less accurately than segmental cues (Repp & Lin, 1990; Experiments 2 and 3 in Cutler & Chen, 1997; Experiments 1 and 2 in Ye & Connine, 1999; Yip, 2001; Mattys, White, & Melhorn, 2005). For example, in a same-different judgment task with monosyllabic spoken words in Cantonese, Cutler and Chen (1997) found that both native Cantonese speakers and speakers without Cantonese knowledge (native Dutch speakers) were prone to more errors and longer response times when two successively presented words contrasted in tones relative to onset and rime contrasts. More recently, by using the ERP technique with high temporal resolution, several studies have found neural evidence that
Tonal information is processed later than segmental cues (Hu, Gao, Ma, & Yao, 2012 with Chinese idioms; Li, Wang, & Yang, 2014 with Chinese poems). For example, Li et al. (2014) asked participants to judge whether the last word in one line of a poem matched the rime and tone of the last word in the previous line as strictly required by the rhyming rules in ancient Chinese poetry. They found that compared to the congruent condition, rime violations, regardless of tone matching, elicited stronger negative-going ERPs time-locked to 300-500 ms post stimulus onset, but tone violations alone only triggered larger positive-going ERPs between 600-1000 ms.

Contradicting the above results, however, a few recent studies reported that during recognition of spoken tone-words, the time course of lexical tone processing did not deviate from that of segmental cue processing (Brown-Schmidt & Canseco-Gonzalez, 2004; Schirmer et al., 2005; Lee, 2007; Malins & Joanisse, 2010; Zhao et al., 2011; Malins & Joanisse, 2012). Zhao et al. (2011), for example, measured the ERPs time-locked to the onset of monosyllabic spoken words in Mandarin. These words were preceded by the presentation of a picture tagged with a corresponding written word (e.g., /bi2/ ‘nose’ and the picture of ‘nose’ tagged with the Chinese character 鼻 ‘nose’ in the congruent condition). By manipulating the phonological mismatches between the spoken word and the Chinese word represented by the picture, they found no differential N400 effects, a neural indicator of lexical access difficulty (Kutas & Hillyard, 1984; Kutas & Federmeier, 2011), amongst the tone, rime, and onset mismatch conditions in terms of amplitude and latency. Moreover, they identified earlier and stronger N400 responses in the whole-syllable mismatch condition (e.g., /bi2/ vs. /ge1/) relative to the other three partial mismatch conditions (e.g., tone mismatch /bi2/ vs. /bi3/, rime mismatch /bi2/ vs. /bo2/, and onset mismatch /bi2/ vs. /li2/). Their findings suggested that lexical tones can constrain tone-word recognition in a similar way as segmental cues. Thus, segmental cues do not have a processing primacy over tonal cues. More interestingly, this study indicated that in the access to lexical-level representations, a syllable is subject to holistic processing, in which the sublexical features are no longer recognised separately but
integrated into one unit which can be mapped onto the lexical entries at the lexical level. These results implied top-down influence of lexical-level representations in the perception of sublexical features in tonal words.

The varying processing weights of lexical tones in different studies have indicated that lexical tones are processed differently under various contexts (Ye & Connine, 1999; Liu & Samuel, 2007). Studies reporting a segment processing primacy usually employed tasks requiring participants to focus on the phonological aspects of stimuli (e.g., same-different judgment task in Cutler & Chen, 1997; rhyme judgment task in Li et al., 2014). In contrast, those studies reporting comparable roles of segments and tones often adopted tasks requiring analyses at a semantic level (e.g., sentence congruence judgment task in Schirmer et al., 2005; cross-modal word matching task in Zhao et al., 2011). Focus on the phonological aspects may elicit processing via sublexical representations and therefore are more likely to reflect the bottom-up processing mechanism (Repp & Lin, 1990; Shuai & Gong, 2014). However, tasks requiring semantic-level analyses may give rise to the activation of lexical representations, leading to a prominent top-down influence on lexical-tone perception (e.g., Schirmer et al., 2005; Zhao et al., 2011).

Furthermore, it should be reiterated that the MMN responses to mismatched lexical tones occurred earlier than the MMNs to mismatched onset consonants as reported in Luo et al. (2006). This result is different from those studies that reported either earlier processing of segmental cues or similar speeds of the processing of segmental and tonal cues (e.g., Hu et al., 2012; Schirmer et al., 2005). The contradiction may be induced by differences in the experimental settings. First, in Luo et al. (2006), the participants did not need to perform any overt tasks that were related to the auditory stimuli, whereas in most previous studies, the participants had to perform some judgment either on the semantic or the phonological aspects. That is, in studies like the Luo et al. (2006), the brain may respond differently to a tone word compared to those studies with an overt task. Second, Luo and colleagues focused on the neural responses in an early time window centring around 200 ms post stimulus onset, whereas the studies with explicit tasks focused on later time windows between 400 ms
and 600 ms after stimulus onset. It may be concluded that the disparities in those studies imply varying processing primacies of different aspects of tone words as a tone word unfolds.

1.5. Models of tone-word recognition

Attempts to model the recognition of monosyllabic Mandarin words have emerged in the last decade and a half. To date, the two well-documented models proposed in Ye & Connine (1999) and Zhao et al. (2011) are both based on the TRACE model (McClelland & Elman, 1986). Both models agreed that a spoken tone word was recognised via representations organised in a high-and-low structure. At the low level, representations were sublexical in nature. The bottom representational layer was a system identifying acoustic features of speech sounds (i.e., the rapidly changing temporal information and the slowly varying spectral information). Above the bottom layer was the representation of more abstract phonemes and tonemes. At the high representational level, lexical representations were composed of existing segment-tone patterns used as the forms of words.

Differing from Ye & Connine’s (1999) model, Zhao et al. (2011) proposed that the lexical representation could be further divided into two layers: a morphemic layer storing the phonological patterns of segment-tone combinations, and a semantic layer encoding the conceptual components of words. Moreover, on a par with the original TRACE model, the two models describe excitatory interconnections between the lexical layer and the phonemic-tonemic layer, and mutually inhibitory connections within the same layer. The between-layer excitatory interconnections allowed bi-directional information flow, and thus made top-down influence of lexical-level representations on sublexical feature perception possible. Moreover, such structure also permitted direct access to lexical-level representations by an input of sublexical features.

Whilst the above models described the representational structure of tone words in Mandarin, two studies have also tapped into the neural mechanism of tone-word recognition. Luo et al. (2011) proposed a two-stage model, operating in a serial fashion. In the initial pre-attentive stage,
the brain processed the speech signal of a spoken tone word as general acoustic cues; in the subsequent attentive stage, the processing occurred via a left-lateralised neural network based on linguistic experience. Luo et al.’s model provided an account for their data that in the pre-attentive processing stage, tone-related brain responses (MMN) were greater in the right hemisphere, but consonant-related responses were more pronounced in the left hemisphere. That is, the right hemisphere was sensitive to slowly varying acoustic cues (i.e., lexical tones), and the left hemisphere was responsible for capturing rapidly changing cues (i.e., segments) (Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002). However, this model did not predict any interaction between general acoustic processing and phonological processing.

In a more recent study, Shuai and Gong (2014) proposed a three-stage model of lexical-tone perception. A special focus of this model was how phonological knowledge affected the perception of acoustic cues at the acoustic-feature level. The three stages were marked by three auditory ERP responses, namely N1, P2, and N400. N1 and P2 are two auditory evoked brain potentials. The N1 is the first prominent negative deflection, usually centring around 100 ms post stimulus onset. Succeeding the N1, the P2 is a prominent positive-going component, peaking around 200 ms post stimulus onset. In the first stage, the processing of the initial phoneme begins about 100 ms after the onset of a tone word. Lexical tones were far from being recognised in such a short time window. Nonetheless, the limited input could be used to predict general syllable patterns, and lead to top-down influence which could be strengthened by contextual information. In the second stage (100-300 ms), the top-down influence was further enhanced with the accumulation of input cues, whilst the bottom-up processing of lexical tones and segmental cues was going on, as reflected by the P2. In the third stage (300-500 ms), top-down processing due to the access to semantic information became the main stream, along with some bottom-up processing, as indexed by the N400. Although this model elegantly accounts for the time course of tone-word processing in accordance with the time windows of the neurophysiological indices, the authors only considered two sublexical layers of
representations, namely the acoustic analysis and the phonological representation, without any clear consideration of representations at the lexical level.

1.6. The present thesis

To advance the investigation of tone-word recognition, the present thesis aims to underpin the role of lexical-level representations by looking for answers to two general questions:

1) How do lexical-level representations interact with sublexical representations of lexical tones and segments?

2) What is the neural mechanism behind the activation of lexical-level representations?

The basic rationale is to investigate the way lexical-level representations affect tone-word recognition by comparing the responses to real words and phonotactically possible pseudo tone word forms. In the current thesis, all tonal pseudo-words are derived from legal segment patterns and tones, but the combinations of the segments and the lexical tones are meaningless in Mandarin. For example, the combination of /na/ and Tone1, */na1/ is a non-existent segment-tone pattern in Mandarin Chinese, even though /na/ is a legal segment pattern in the combinations with Tone2 (/na2/(拿) ‘hold’), Tone3 (/na3/(哪) ‘where’) and Tone4 (/na4/(哪) ‘there’). We label this kind of pseudo-word a tone-manipulated pseudo-word. Compared with pseudo-words derived from non-existent combinations of onsets and rimes like */tua1/ (in Shuai & Gong, 2014), tone-manipulated pseudo-words can induce no segmental processing difficulties at the sublexical level, but still lack lexical-level representations. Therefore, such pseudo-words are an ideal tool for revealing top-down effects yielded by lexical-level representations.

Four studies are carried out to address the two general research questions, and will be presented in Chapter 2 - 5. The two studies reported in Chapters 2 and 3 are conducted by using auditory lexical decision tasks (LDTs) and a form priming paradigm to measure the behavioural
responses in real and pseudo tonal word forms. In Chapter 2, with auditory LDTs, we examine whether the absence of lexical-level representations in pseudo-words leads to differential responses to pseudo-words in three tonal categories (Tone1, Tone2, and Tone4), and whether this response pattern is distinct from the one in real words of the three tones. Reaction time and accuracy data will be used to indicate the underlying processes (Goldinger, 1996). Two versions of the auditory LDT will be used to detect whether the semantic and phonological components of lexical-level representations influence top-down processing in the same way. Since the perception of the three lexical tones are different in nature (Wu & Shu, 2003; Lai & Zhang, 2008), we expect distinctive response patterns to pseudo-words in the three tonal categories rather than to real words because lexical decisions of real words should be sensitive to lexical factors instead of sublexical features. Moreover, if both semantic and phonological representations are stored at the same lexical level, we do not expect distinctive response patterns due to the different task instructions.

In Chapter 3, with a form priming paradigm, we systematically check the priming effects in lexical decisions of tonal word forms with ten types of prime-target pairs. In form-priming experiments, researchers usually manipulate the lexical decision performance in target words preceded by various types of prime words (Zwitserlood, 1996). The priming effects are characterised by facilitation and inhibition. Facilitatory priming is usually shown by faster responses, and inhibitory priming is reflected by slower reaction times, compared to proper baselines. The primary goal of this study is to reveal the priming effects in paired word forms sharing the same segment but with distinctive tones (word : word, pseudo-word : word, word : pseudo-word, pseudo-word : pseudo-word, in which the initial word form is a prime, followed by a target). These materials allow us to detect the interaction between lexical and sublexical access in tone-word recognition as reflected by facilitatory priming and inhibitory priming effects.

Although behavioural measurements such as lexical decisions give valuable insight into the organisational structure of the mental lexicon of
Tone-word Recognition in Mandarin Chinese

In Chapter 2, with auditory LDTs, we examine whether the absence of lexical-level representations in pseudo-words leads to differential responses to pseudo-words in three tonal categories (Tone1, Tone2, and Tone4), and whether this response pattern is distinct from the one in real words of the three tones. Reaction time and accuracy data will be used to indicate the underlying processes (Goldinger, 1996). Two versions of the auditory LDT will be used to detect whether the semantic and phonological components of lexical-level representations influence top-down processing in the same way. Since the perception of the three lexical tones are different in nature (Wu & Shu, 2003; Lai & Zhang, 2008), we expect distinctive response patterns to pseudo-words in the three tonal categories rather than to real words because lexical decisions of real words should be sensitive to lexical factors instead of sublexical features. Moreover, if both semantic and phonological representations are stored at the same lexical level, we do not expect distinctive response patterns due to the different task instructions.

In Chapter 3, with a form priming paradigm, we systematically check the priming effects in lexical decisions of tonal word forms with ten types of prime-target pairs. In form-priming experiments, researchers usually manipulate the lexical decision performance in target words preceded by various types of prime words (Zwitserlood, 1996). The priming effects are characterised by facilitation and inhibition. Facilitatory priming is usually shown by faster responses, and inhibitory priming is reflected by slower reaction times, compared to proper baselines. The primary goal of this study is to reveal the priming effects in paired word forms sharing the same segment but with distinctive tones (word: word, pseudo-word: word, word: pseudo-word, pseudo-word: pseudo-word, in which the initial word form is a prime, followed by a target). These materials allow us to detect the interaction between lexical and sublexical access in tone-word recognition as reflected by facilitatory priming and inhibitory priming effects.

Although behavioural measurements such as lexical decisions give valuable insight into the organisational structure of the mental lexicon of tone words, such methods cannot faithfully reveal online processes because of the involvement of post-lexical processing and response planning during task completion (Balota & Chumbley, 1984). Therefore, in Chapters 4 and 5, we employ ERPs, a neurophysiological technique indexing post-synaptic activities in large bundles of neurons aligned in a parallel orientation, related with specific stimulus events (Coles & Rugg, 1995). This technique provides high temporal resolution at millisecond-level accuracy and allows researchers to investigate the hemispheric lateralisation of brain functions in response to auditory stimuli (Alho et al., 1998; Rinne et al., 1999; Shtyrov et al., 2010).

In Chapter 4, we investigate the time course of access to lexical-level representations in tone-word recognition by referring to the MMNs elicited in an oddball experiment. The speech-related MMN, usually identified between 150-250 ms post stimulus onset, has been used as an ERP indicator of auditory discrimination and automatic access to the long-term memory traces of spoken words (for reviews, see Pulvermüller & Shtyrov, 2006; Shtyrov & Pulvermüller, 2007). Moreover, the MMN over the left-hemispheric region has been linked with specified neural networks for language processing (Alho et al., 1998; Rinne et al., 1999; Näätänen et al., 1999; Koyama et al., 2000). Following a protocol of rapid word learning (Shtyrov et al., 2010), we compare the MMN responses elicited in real words and pseudo-words, repeatedly presented during perceptual training. If lexical-level representations can modulate MMN responses, the MMNs in the final phase of perceptual training should be enhanced relative to the MMNs in the early training period. Moreover, we expect the MMNs to the trained pseudo-word to resemble the MMNs of the real word. In addition, the expected learning effects should be over the left-hemispheric recording sites.

In Chapter 5, we explore whether lexical representations can be accessed in the pre-MMN time windows, namely the N1 and P2 time windows. The N1 and P2 are known to be highly sensitive to the physical features of stimuli, and thus, it is difficult to explore top-down processing in these time windows by referring to the two ERP components. To overcome this issue, we adopt a habituation paradigm which has been used
to identify disparities between the N1 responses to auditory stimuli with high and low levels of representations, namely speech and non-speech sounds (Woods & Elmasian, 1986; Teismann et al., 2004). Based on these studies, we expect to observe differential auditory habituation patterns in real words and pseudo-words.

The following four chapters (Chapters 2 to 5) will report our studies in detail. The ‘Discussion and Conclusion’ in Chapter 6 will summarise and discuss the general findings of the thesis.