國立政治大學語言學研究所碩士論文

National Chengchi University
Graduate Institute of Linguistics
Master Thesis

指導教授：蔡介立
Advisor: Jie-Li Tsai

聲調在中文口語字彙觸接的時序處理:眼動研究之證據
TEMPORAL PROCESSING OF LEXICAL TONE IN LEXICAL ACCESS OF CHINESE SPOKEN CHARACTERS: AN EYETRACKING STUDY

研究生：許媛媜　撰
Student : Yuan-Jhen Syu
中華民國　一百零一年七月
July, 2012

TEMPORAL PROCESSING OF LEXICAL TONE IN LEXICAL ACCESS OF CHINESE SPOKEN CHARACTERS: AN EYETRACKING STUDY

BY

Yuan-Jhen Syu

A Thesis Submitted to the

Graduate Institute of Linguistics

in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

July 2012

The members of the Committee approve the thesis of Syu-Yuan Jhen
Defended on July 30th, 2012.

Jie-Li Tsai (蔡介立)
Advisor

Yuwen Lai (賴郁雯)
Committee Member

Yow-Yu Lin (林祐瑜)
Committee Member

Approved: Chiung-chih Huang (Acting)

Kawai Chui(徐嘉慧), Director, Graduate Institute of Linguistics

Acknowledgements

一直以來，我都喜歡走跟別人不一樣的路，這種路很崎嶇難走，充滿荊棘，但是完成目標時，成就感難以言喻，就像完成這本論文的過程一樣，一路上若是沒有貴人摯友和家人們的相挺，我絕對會走到一半就半途而廢。

每個人的生命中都會有好幾個貴人，我人生中重大的貴人之一絕對是我的論文指導教授蔡介立老師，在學術上，他給了我最大的安全感，每每有任何學問上疑惑，只要找老師討論就可以得到最佳的解答，除了學術以外，他還告訴我快樂的重要，要找到快樂的因子，做任何事才會有熱情能支撐著你，人生才有意義，或許在未來，學術上的事情都會還給老師了，但我會把快樂放在心上，謝謝老師帶給我如此正面的訊息，感謝老師願意指導這麼粗枝大葉，和愚昧的我，不辭辛勞的字字句句的幫我修改論文，包容我的毛病，幫我處理任何奇怪的事情，例如莫名其妙的推薦信之類的，能遇到如此天神般的老師，真的覺得我好幸運！

回想碩一面對完全未知的語言學世界，語言所上的優秀教授們帶給了我驚奇不斷的學術知識，感謝黃瓊之老師、蕭宇超老師、何萬順老師、萬依萍老師、徐嘉慧老師教授我的一切，還有，特別感謝惠鈴助教學姊每一次對我熱切的關心，叮嚀我任何一件大小瑣事。

在 EMR baby lab 中待了兩年，確實隨時隨地都有 **E**xcel, **M**atlab, **R** 來搞得我日日做惡夢，好險還有 EMR baby 們，總是在我整夜沒睡的隔天，滿面愁容地看著我說媛媜你還好吧，準備回家的時候跑到我位置關心我的情況，叮嚀我要吃飯和休息，每當有問題時，總能放下手邊火燒屁股的進度，充滿耐心地幫我解決問題，你們像天使般溫暖我的心情，這兩年的回憶對我來說絕對永生難忘，感謝家興學長拒絕打球邀約願意聽我報告；聰慧機靈的婉雲 baby 幫我解決各式各樣的怪問題，關心我的身體和生活；很有內容的熊 baby 教我實驗設計的開端還有 Matlab，他總是幫我建立自信心；佩如 baby 總能帶給大家歡樂，很不可思議的幫我帶個小精靈給我；善良第一的怡璇 baby 讓我感到最大的溫暖，夜晚一起奮鬥，互相勉勵；帶給了我繼續完成的動力！天使轉世的宛柔 baby 在總能在我頭腦渾沌時，幫我解決任何問題，體貼的給我關懷；最美的瑪莉 baby 總是陪我在夜晚走回家，讓我開心，經歷過好多事情，一起大哭一起大笑，任何事跟你分享總能得到最大的共鳴！可愛的柏亨 baby 總是在夜晚陪伴怕鬼又怕狗的我，讓我不再害怕，還有已經離開實驗室的 baby，翠屏、雅嵐、和致潔，每一次見面的關懷總能帶給我力量！讓我能在挫折中爬起來，擦乾眼淚繼續往前走。

還有最親愛的語言所同學們，心綸、琬婷、淑禎、曉貞、晉瑋、書豪、美杏、侃彧、柏溫、婉君，好懷念跟你們一起吃飯的時光，大家開讀書會的情形，我喜歡大家的聰明，還有幽默，即使研究了不同的領域每每見面總能很有話題，笑聲不斷，能認識你們我覺得好榮幸喔！

最後在這裡要感謝從小到大陪伴在我身旁的垃圾桶，老騷，小瓜，余胖，巫季珍，14 姊妹，四千金，若沒有你們，我會是個一事無成，寫不出論文的人，謝謝你們總能在夜晚接受我的騷擾電話，陪伴我度過所有低潮，散發正面能量給我，每個都是我的情人，你們讓我覺得我好幸福！最後要感謝我媽和我最美麗的兩個姊姊，尤其是我老媽，他最不關心我的論文，因為只擔心我的身體和生活起居，深怕我身體有任何閃失，還有那個聽說在草地上匍匐前進時，也心繫著我論文進度的人，一直在我旁邊給我最大的支持，包容我的神經，最後我要把這本論文送給我在天上的老爸，我知道你一直為我感到驕傲！爸～我終於寫完啦！

TABLE OF CONTENTS

List of Tables

List of Figures

國　立　政　治　大　學　研　究　所　碩　士　論　文　提　要

研究所別：語言學研究所

論文名稱：聲調在中文口語字彙觸接的時序處理:眼動研究之證據

指導教授：蔡介立

研究生：許媛媜

論文提要內容：(共一冊，22,251 字，分 5 章 20 節，並扼要說明內容)

本文主要探討中文聲調在口語字彙觸接過程中所扮演的角色。實驗一藉由眼動實驗中的 Visual World Paradigm 作業，觀察中文聲調影響口語字彙辨識的時序歷程。受試者在聽到指導語和目標字之後，用滑鼠在螢幕上點選聽到的目標字，例如，螢幕上出現的字包含一個目標字：「摸」、一個競爭字(與目標字只有聲調相同：「挖」或是與目標字只有音段相同：「抹」)，以及兩個聲調與音段和目標字完全不同的無關字：「怒」、「菊」。為了觀察目標字、競爭字及無關字在口語字彙處理時的競爭，我們會計算各個字彙的凝視比例。實驗一中由於聲調與目標字相同的競爭字與目標字的第一個音段就開始產生差異，因此未觀察到聲調早期介入的影響。實驗二透過與實驗一相同的實驗程序及方法，操弄目標字和競爭字中聲調和前兩個音段(Cohort)的異同以探測更早期的聲調影響。螢幕呈現包含一個目標字「湯」、一個競爭字(前兩個音段和聲調皆與目標字相同：「胎」，或是只有前兩個音段相同但聲調與目標字不同：「泰」)，以及兩個聲調與音段和目標字完全不同的無關字「剖」、「痕」。結果顯示，聲調在語音訊息前兩個音段時就會產生影響，也就是聲調的影響在語音結束前即有作用。再者，本文發現聲調無法單獨且獨立地對於語音辨識產生影響，此看法與聲調表徵需以“toneme” node 獨立地存在於 the modified TRACE model 的看法不盡相同 (Malins & Joanisse, 2010; Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011)。

Abstract

The present study aims to examine the role of tonal information during Mandarin Chinese spoken character recognition. Two eye-tracking experiments were conducted with the visual world paradigm, which participants heard a Chinese monosyllabic character and used a mouse to click on the corresponding character in a visual array of 4 characters on the screen. Experiment 1 manipulated the relationship between the spoken target characters and written characters on the screen, including a target (e.g., /mɔ1/‘touch’), a tonal competitor (the tone was the same as target except segment: e.g., /wa1/‘dig’) or a segmental competitor (the segmental structure was the same with the target except tone: e.g., /mɔ3/ ‘wipe’), and two unrelated distractors (the segments and tone were different from target: e.g., /nu4/ ‘anger’, and /tɕy2/ ‘chrysanthemum’). The fixation proportions on target, competitors and the unrelated distractors were computed during the unfolding of the auditory target stimuli. The results showed tonal difference was detected before the end of auditory stream. However, no early involvement of tonal information was found, which may due to the tonal competitor and target shared no segment from the first phoneme. In order to examine the earlier tonal processing, Experiment 2 manipulated two types of cohort competitors sharing the initial two segments with the target (e.g., */tʰaŋ1/* “soup”), a cohort-tone competitor,

e.g., */t^{h}aj1/* "fetus" (both tone and initial two segments are the same with target) and a cohort-only competitor e.g., */t^{h}aj4/* "peaceful" (initial two segments is the same with the target but with different tone). Result showed that tone affected spoken character recognition while processing the two initial segments. In addition, tone could not affect spoken character processing independently, which might be inconsistent with the assumption that tone is a separate level of representation, called "toneme" node, in the modified TRACE model (Malins & Joanisse, 2010; Ye & Connine, 1999; Zhao et al., 2011).

Chapter1

Introduction

1.1 General background

How do listeners recognize a word in a streaming of continuing auditory inputs? In order to decode a string of spoken utterance, the information needs to be extracted from the acoustic signal and mapped onto different forms of internal representation in the mental lexicon. Recently, the models of spoken word recognition demonstrated that as the spoken auditory inputs are unfolding, a set of lexical candidates compete for recognition (Carroll, 2008; Frauenfelder & Tyler, 1987; W. D. Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986).

The present study aims to examine the tonal processing during spoken character recognition of Mandarin Chinese. Lexical tone belongs to the prosodic information, which could also be called suprasegmental information because it goes beyond and spans over segments. For Chinese spoken character recognition, the segmental and suprasegmental information are processed from acoustic signals. Segmental elements supporting a cluster of distinctive features(Roca & Johnson, 1999), including vowels and consonants. Supra-segmental elements involve pitch variations to form tones and intonations distinctions. (Jongman, Wang, Moore, & Sereno, 2006).

According to the supra-segmental features, there are non-tone languages and tone languages. For most European languages such as English, French, and Dutch, the meaning of a word do not change irrespective of whether it is said on a rising pitch or a falling pitch. On the contrary, for tone languages like Chinese, Cantonese, lexical tones are pitch variations that serve to provide contrasts in word meaning(Ladefoged, 2005).

In Beijing Mandarin, there are four different tones, each displaying a distinct pitch inflection: high-level (Tone 1), high-rising (Tone 2), low-dipping (Tone 3), and high-falling (Tone 4). The same segmental context can carry different meanings as altering pitch inflection. For example, the segment ma pronounced with a high-level tone means “mother”, while the identical segment pronounced with a high-falling tone means “scold”. In the present study, we used Taiwan Mandarin tone. According to Chang (2010), Taiwan Mandarin tone differed from Beijing Mandarin counterparts in terms of f0 height (tonal registers) and contours. All four tones in Taiwan Mandarin are with lower tonal registers. Tones 2 and 3 tend to respectively become dipping and low-falling. Tone 3 has a low-falling pattern (i.e., half 3rd tone) in isolation. Figure 1 shows an illustration of the F0 contours of the four Taiwan Mandarin tones (吳聲弘, 2012).

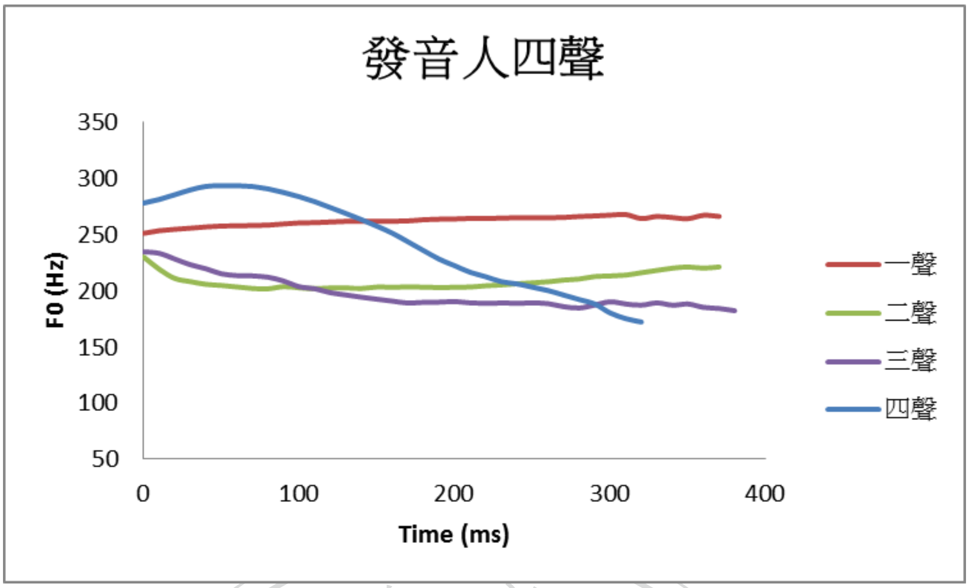


Figure 1. F0 contours for the four Taiwan Mandarin tones, each combined with the syllable *ma* (吳聲弘, 2012).

In the non-tone languages, such as English and Dutch, there is evidence showing how lexical stress influences spoken word recognition (Cutler, 1986; van Donselaar, Koster, & Cutler, 2005). Some studies suggest that Dutch listeners rely on the prosodic information for spoken word recognition (van Donselaar et al., 2005) while English listeners does not (Cutler, 1986).

Early studies exploring the issue on the processing of tone languages, such as Chinese and Cantonese, utilized the tasks of lexical decision and homophone judgment. These researches demonstrated that tone was accessed later than segmental information; that is, tone plays a minor role in spoken word recognition (Cutler & Chen, 1997; Taft & Chen, 1992; Ye & Connine, 1999). However, Lee(2007) found that lexical tone played a role as important as segmental information. Differing from the behavioral task, recent studies applied the experimental techniques such as event related potentials (ERPs) or eye-tracking to examine the on-line auditory processing. These results suggest that the tonal and segmental information are accessed at a similar temporal point. Therefore, the tone and segment information might play a comparable role during spoken word recognition (Malins & Joanisse, 2010; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Tsang, Jia, Huang, & Chen, 2011).

1.2 Research questions

The present study conducts two eye movement experiments to examine the time course of tone information processing during spoken character recognition. Specific research questions to be addressed are as follows:

(1) When is tonal information processed during spoken character recognition? Is tonal information processed in early phase of spoken character recognition? Or, is it accessed in a relatively late stage of spoken character processing?

(2) In what way does tone affect lexical process with segmental information? Does lexical tone affect spoken processing independently? Or, does tonal information influence lexical process depending on segmental information.

Chapter 2

Literature Review

2.1 Processing the spoken language signal

Processing of speech perception roughly include three levels: the auditory level, the phonetic level, and the phonological level (Carroll, 2008; Frauenfelder & Tyler, 1987; Lass, 1976; Studdert-Kennedy, 1976). At the auditory level, the signal is represented in terms of its frequency, intensity, and temporal attributes, which could be shown on a spectrogram. At the phonetic level, the individual phones are identified by a combination of acoustic cues such as the formant transitions. At the phonological level, the phonetic segment is converted into a phoneme, and phonological rules are applied to the sound sequence. These levels are successively processed by listeners when decoding speech signals (Carroll, 2008). Listeners firstly discriminate auditory signals from other sensory signals and decide whether the auditory stimuli are something they have heard. Then listeners identify the particular properties and qualify it as speech. Lastly, the properties would be recognized as the meaningful speech of a particular language (Carroll, 2008).

2.1.1 Perception of phonetic segments

Concerning the speech perception, many researchers have great interest in how listeners manage to decode speech signals into phonetic units and derive meaningful words. The properties such as vowels and consonants help listeners identify phonetic segments are tightly intertwined and overlapped (Gleason & Ratner, 1998). One of the issues for speech perception is how individual words from the complex speech input are separated and then further identify them appropriately.

Moreover, there is no one-to-one correspondence between the phonemes and their acoustic realization. This problem could be termed as lack of invariance, which results from the phenomenon of context conditioned variation(Carroll, 2008; Frauenfelder & Tyler, 1987; Gleason & Ratner, 1998). The context conditioned variation refers to the production of same phonetic segment varies depending on the environment in which the segment is produced. However, there are also some studies suggest that the speech perceptions are relied on both invariant and context-conditioned cue (Cole & Scott, 1974).

Another issue about the segmental perception is the phenomenon of categorical perception. Categorical perception is typically found on contrasts between many different pairs of consonants. For categorical perception, perceptual systems transform relatively linear sensory signals into absolute or categorical non-linear mental representations. In speech, listeners convert the continuous auditory signals into discretely meaningful words. According to Liberman, Harris, Hoffman, and Griffith (1957), listeners' ultimate task is to identify [p] or [b] which belongs to one or another category of speech sounds. The minimal feature between the [p] and [b] is the voicing. To notice the difference between the voiced [b] and the voiceless [p], the time when the sound is released at the lips and when the vocal cord starts to vibrate is crucial. The vibration of voiced [b] occurs immediately but the vibration of voiceless [p] occurs after a short lag, which is termed as voice onset time (VOT). Some of the categorical perception studies construct synthesized speech syllables to examine whether categorical perception holds for nonspeech such as chirp or only for speech(Jusczyk & Luce, 2002; Liberman et al., 1957). The researchers found that categorical perception was used in speech rather than the nonspeech. However, there is still no firm argument regarding whether there is a special mode of speech perception (Jusczyk & Luce, 2002; Liberman et al., 1957).

Due to the continuous and noncategorical characteristic of vowels, vowel perception is different from consonant perception (Fry, Abramson, Eimas, & Liberman, 1962). Vowel has longer and larger formant but consonants are presented by the formant transitions, which transient cues forces listeners to impose a categorical identity on the stimuli more rapidly than for vowels. Therefore, after the stimuli have been identified, the cues for the consonants are lost, and only the coded stimuli remain. Additionally, because of the relatively longer duration of vowels, the perception course suggests that vowels are processed longer at the auditory level than consonant (Carroll, 2008; Frauenfelder & Tyler, 1987; Garman, 1990).

2.1.2 Lexical access and models

In addition to the issues on discrimination and categorization of phonetic segments, many researchers are interested to expand the inquiry domain to the processes which spoken words are recognized for retrieving meanings. Psycholinguists are eager to understand how listeners use phonological and prosodic knowledge to parse the sensory input during word recognition (Grosjean & Gee, 1987; Lyn, 1987; Uli H & Tyler, 1987).

Models of spoken word recognition generally assume that phonological information is continuously integrated during spoken word recognition. When the speech is unfolding, lexical candidates compete for recognition as a function of phonological similarity with the speech input (Foss & Hakes, 1978; Garman, 1990; Gleason & Ratner, 1998; Myers, Laver, & Anderson, 1981). The models are different in explaining the temporal dynamics of spoken word recognition between the incoming speech stimuli and potential lexical representation.

One of the significant models is Cohort model (W. D. Marslen-Wilson & Welsh, 1978; William D & Marslen-Wilson, 1987). Cohort model proposes that the onset of a word activates a set of lexical candidates competing for recognition. In the first, autonomous stage, when the first phoneme of a word is heard, all of the candidates with the phonological resemblance of the words are activated. For example, if the phoneme /d/ in the word “drive” is heard, then the words beginning with /d/ may activate many candidates such as “dive,” “drink,” “date,” “dunk” and so on. This set of activated words is called the “cohort”. The words in the cohort are not assumed to affect the activation levels of one another, which mean that at this stage, word recognition is a completely data-driven or bottom-up process. In the second stage, once a cohort structure is activated, all possible sources of the auditory information may begin to influence the selection of the target word from the cohort. The additional auditory phonetic information may eliminate some of the cohort words. The coming phonetic information is assumed to work in a strictly left-to-right fashion. However, in this stage, the sources of higher levels information may also help to eliminate the hypothesized word cohorts. For instance, if the phoneme of the /r/ presents following the phoneme /d/, this further acoustic-phonetic information may eliminate the cohort words such as “date” and “dunk.” And then the higher level sources of the information may appear and eliminate other words of the cohort word such as “dive” and "drink," which might be not suitable for the semantic or syntactic available information. The spoken word recognition is finally achieved when a single candidate remains in the cohort. A latter revised cohort model extends to consider other sources of information such as word frequency effect (Frauenfelder & Tyler, 1987; Gleason & Ratner, 1998; Jusczyk & Luce, 2002; W. Marslen-Wilson & Tyler, 1980; William D & Marslen-Wilson, 1987).

The TRACE model is an interactive model (McClelland & Elman, 1986), assuming three levels of primitive processing units: the features, the phonemes, and the words (Figure 2) . These processing units have excitatory connections between levels and inhibitory connections within the levels. These connections can both excite and inhibit the activation levels of the nodes according to the stimulus input and the activity in the system. For example, the stimuli with voicing such as the consonants /b/, /d/, or /g/ will make the voiced feature at the phoneme level of the model become active. The activeness in turn passes its activation to all voiced phonemes at the next level, which in turn activates the words having those phonemes. Furthermore, via lateral inhibition among units in a level, the most activated unit may come to dominate other competing units which are also temporarily concordant with the input. For example, the word unit *cat* at the lexical level will inhibit the similar and competing lexical units (e.g., *pat*). This inhibition helps to make sure that the best candidate word will win the competition in the process (Gleason & Ratner, 1998; Jusczyk & Luce, 2002; McClelland & Elman, 1986).

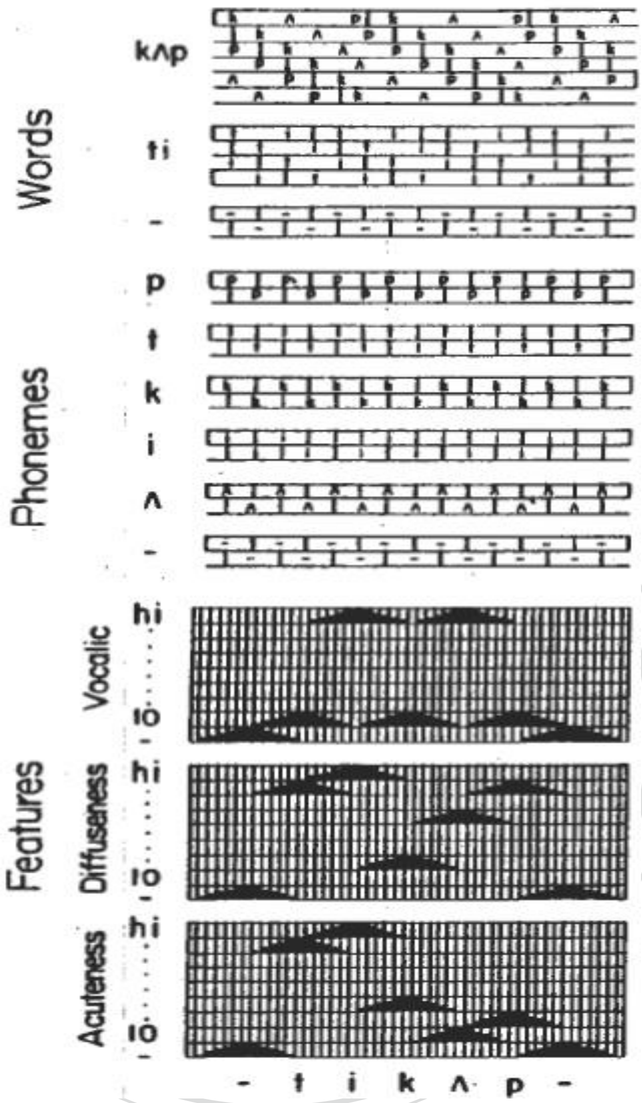


Figure 2. A subset of the units in the TRACE. Each rectangle represents a different unit. The labels indicate the item for which the unit stands, and the horizontal edges of the rectangle indicate the portion of the TRACE spanned by each unit. The input feature specifications for the phrase “tea cup,” preceded and followed by silence, are indicated for the three illustrated dimensions by the blackening of the corresponding feature units (McClelland & Elman, 1986).

There are differences between the Cohort and the TRACE models (Table 1). First, the Cohort model emphasizes on the temporal dynamics of spoken word recognition.

Cohort model suggests the significance of the initial word, which means that spoken words may be identified before their offsets if similar competitors are not active. However, the TRACE theory suggested the duplicative nodes and connections of its system through successive time slices of input. This might be questionable in treating the temporal dynamics in spoken word recognition. The time-slice solution results in an extremely complex structure. Second, although the TRACE model is relatively complex, its highly interactive feature makes it possess the computational specificity, which results in a relatively easy way to conduct a direct test of behavior simulation. Therefore, this feature helps in accounting for phenomena with a broad range. On the contrary, the lack of interactive feature causes the poverty of computational specificity in Cohort model. Last, the Cohort model emphasizes on the exact match between auditory input and lexical representation rather than the sublexical representation. However, the TRACE model has the phonemes level which is between the words level and features level (Jusczyk & Luce, 2002).

Table 1. The features for the Cohort and the TRACE models (Jusczyk & Luce, 2002)

	Cohort	TRACE
Activation	Constrained	Radical
Units and levels (arrows indicate direction of information flow)	Words ↑ Features	Words ↑ ↓ Phonemes ↑ ↓ Features
Lexical competition via lateral inhibition	No	Yes
Sublexical-to-lexical interaction (bottom-up)	Facilitative and inhibitory	Facilitative
Lexical-to-sublexical interaction (top-down)	No	Facilitative
Distinguishing features	1. Focus on time-course of recognition 2. Interactivity	1. Highly interactive, simple processing units 2. Computationally explicit 3. Attempts to account for broad range of phenomena

2.2 Prosody in spoken word recognition

According to Cutler, Dahan, and van Donselaar (1997), prosody is an intrinsic determinant of the spoken form in languages. This intrinsic determinant is realized as an effect on the timing, amplitude, and frequency spectrum of the utterance. Prosody includes intonation, duration, stress, and tone. One of the important features is that it spans over long segments such as syllables, words, and the utterances in speaking style, sentence type and so on. Prosodic cues can convey lexical and nonlexical information; for example, the function of distinguishing lexical meaning in tone, the prominence function in stress, or the emotion expression in the sentence intonation. Any part of the speech has duration, amplitude, and fundamental frequency. Therefore, when listeners recognize the speech, they are processing the variation determined by prosody (Cutler et al., 1997; Leena, 2012).

When and how might prosodic information play a role in the processing? Early findings suggested that prosody plays an organizing role in speech. For example, nonsense syllables are recalled better only if the string of the nonsense syllables presented with sentence prosody (Epstein, 1961). In addition,.Cutler et al. (1997) suggested that processing of speech input is facilitated by coherent prosodic structure appropriate for sentences. Studies of such facilitated effects have established a significant role for temporal patterning. Thus temporal envelops of spoken utterance, preserving amplitude information but virtually without spectral variation, allow listeners to recognize short utterances and even nonsense syllables almost perfectly (Cutler et al., 1997; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Second, listeners use relevant acoustic information as soon as it becomes available. For instance, listeners take coarticulatory information efficiently from one segment to another(Whalen, 1991). Thus, some researchers propose that whenever the prosodic information could constrain initial lexical activation, it is important to see what and how such prosodic is processed by listeners(Cutler et al., 1997).

Because of the varied characteristics of the prosodic information, the prosody information, such as stress and tone, in spoken word recognition has been investigated. Most research on lexical access have been carried out in English, hence, the prosodic structure which have been investigated is stress (Cutler, 1986; van Donselaar et al., 2005). In English, the stress pattern can only be contrasted in multisyllabic domain rather than in monosyllable like in tone languages. Tone languages such as Cantonese or Mandarin are good examples to be illustrated because tone contrasts may be realized in a monosyllable (Cutler & Chen, 1997; Taft & Chen, 1992).

2.3 Stress in lexical processing

Studies of English vocabulary structure suggest that listeners could use the stress-pattern information in word recognition. However, some studies showed that the stress information does not facilitate English listeners in auditory lexical decision or in the grammatical category judgment. In Cutler and Clifton (1984) the participant performed a grammatical category judgment of the bisyllabic syllables with or without the standard stress pattern (for example, initial stress for noun or final stress for verbs). The result showed that the reaction time was not affected by the different stress pattern. Cutler (1986) used a cross-modal priming task to distinguish the contrast pattern of stress such as OBject-obJECT, and FORbear-forBEAR. If the stress information was used by listeners, the prime and the target would not be considered as homophones and no homophonic priming effect would be expected. Subjects were asked to listen to a sentence contained a prime which meaning was related to the target and then performed the lexical decision task. The resulted showed that the pair could prime each other. Subjects considered the stress minimal pairs as homophones, suggesting that the access code did not influenced by the stress prosodic information. Listeners did not discriminate from these two words for msec in the initial access to the lexicon.

In Dutch, the stress information involved during spoken word recognition (van Donselaar et al., 2005). van Donselaar et al. (2005) also used the cross-modal priming experiments to examine the role of suprasegmental information in processing Dutch. The result in Dutch showed that the inappropriate stressing could prevent lexical activation. The authors also suggested that the constraining of the suprasegmental during the processing was within a single syllable in Dutch, indicating that it began as soon as the relevant acoustic information was available to modulate the activation of potential candidate words. The inconsistent results between English and Dutch are probably due to that the minimal stress pairs are rare in English. Although English is a lexical-stress language, the stress cue might be redundant in lexical processing. The stress information in English can nearly always be derived from the segmental information (Cutler et al., 1997).

2.4 Tonal processing

2.4.1 Tone perception

Acoustic analysis about tone typically focuses on the fundamental frequency (F0), which is a quantification of the rate of vocal fold vibration and usually expressed in Hertz (Hz). According to Jongman et al. (2006), tone is a function of the rate of vocal fold vibration. To characterize Mandarin tones, the F0 height and the F0 contour are the crucial acoustic parameters.

Researchers have explored the contribution of F0 height and F0 contour to tonal perception. Some studies suggested that for Mandarin listeners, both of the F0 height and F0 height are important. However, some studies claimed a more crucial role of F0 contour (Gandour, 1984; Jongman et al., 2006).

Recently, Tsang et al. (2011) used the event-related potentials (ERPs) to examine how pitch contour and pitch height contributed to early tonal processing in an auditory passive oddball paradigm in Cantonese. Classifying six tones in Cantonese by pitch height and contour, the authors manipulated four conditions: height-large difference (Tone 6/ Tone 1), height-small difference (Tone 6/ Tone 3), contour-early difference (Tone 1/ Tone 2), and contour-late difference (Tone 6/ Tone 2). In the experiment, the stimulus (e.g., /ji1/, /ji2/, /j3/ and /ji6/) was presented while the participant was watching a self-chosen silent movie with closed captions. The result indicated that the turning point on the pitch contour could modulate the effect of pitch height, suggesting that Cantonese speakers did not process pitch contour and pitch height as totally unrelated dimensions. In Mandarin Chinese, Lai and Zhang (2008) used a gating paradigm to examine the amount of tonal information needed to correctly identify the four tones of the target. In the experiment, there were eight tone quadruplets, which contained the same segmental structure but different tones. Subjects were asked to identify the tone for each gated stimulus (40msec increments) and provided a confidence rating on a scale of one to seven for their response by pressing the corresponding button. The stimuli were presented in a duration-blocked fashion, in which participants heard the first gate of the stimuli to the last gate, which always contained the entire syllable. The isolation point, which was the size of the segment needed to correctly identify, was examined. The result showed that the isolation point was different among four tones. The earliest isolation point was the Tone 1, followed by Tone 4, and then followed by Tone 2 and Tone 3. To sum up, the acoustic features of four tones in Mandarin affect tone perception.

Whether the acoustic similarity among the four tones affects tonal perception has been examined in both Chinese and Cantonese. There are six tones in Cantonese. Tone 1 is most distinct from other tones, while other tones bear a similar point on the F0 scale. In a lexical decision task, Cutler and Chen (1997) manipulated the mismatch and match of phonological structure between prime and target In this study, Cantonese tones were separated by "easy group" including Tone 1, and "hard group" including the remaining tones. Tone 1 was in the easy group because of its acoustic distinction from other tones in Cantonese. The "hard" group comprises the tones with similar acoustic contour. The result showed that the "easy" group had lower error rate and faster response time than that of the "hard" group. As for Chinese, Ye and Connine (1999) explored the role of tonal similarity in a vowel and tone monitoring task. In their experiment three, the Tone 2 and Tone 3 were grouped as "close" tone while the Tone 2 and Tone 4 were labeled as "far" tone. Results showed that the reaction time to the "far" tones was longer than the "close" tones in both the vowel and tone monitoring tasks.

2.4.2 Tone in lexical processing

2.4.2.1 The role of tone compared to segment

For tonal languages, many studies had investigated on how listeners extract both tonal and segmental information from the acoustic signal to map onto the mental lexicon. Empirical evidence suggested that compared to segmental distinctions, lexical tone distinction could be slow and play a weaker role in lexical processing. Cutler and Chen (1997) examined the processing of lexical tone in Cantonese. In a lexical decision task, subjects were asked to judge real words and non-words by pressing a button. The non-words differed from the real words in the mismatch of tone, vowel, vowel-tone, onset, onset-tone, onset-vowel, and onset-vowel-tone. The result showed that there were more errors on tone mismatch compared to consonant or vowel mismatch. In the second experiment with same-different judgment task, the two words were different in any one of the three dimensions of the syllable (consonant, vowel, and tone) or any combination of these. Result showed that the response was less accurate and slower when the only difference between them was in the tone. According to this finding, Cutler et al. (1997) argued that tonal information often arrived later than does information about the vowel that with the tone. The Experiment 1 of Ye and Connine (1999) showed the similar result. In a phoneme monitoring task, subjects judged whether there was a tone-plus-vowel (e.g., tone 2-/a/) combination for the target. The stimuli consisted of existent (known) and non-existent syllables, which contained the target. Non-target bearing syllables were different from the target in the vowel or by the tone. The result showed that the response time was slower in tone mismatch compared to vowel mismatch. Similar to Cutler and Chen (1997) ,the authors concluded that the perceptual acquisition of tonal information lagged behind that of vowel information. Both studies indicated that tones were realized on vowels. Therefore, tone could not access until the vowel information was available. In addition, they suggested a perceptual advantage for vowel information while a perceptual disadvantage for tone regarding its relatively later availability.

However, some recent studies proposed a relatively strong constrain of tone compared to earlier studies. Lee (2007) used the direct priming and mediated priming paradigms to examine the role of tone in lexical activation of Mandarin Chinese. If the tone constrained the lexical activation, tonal information would allow listeners to kick out tonally incompatible candidates; therefore, there would be no priming facilitation of sharing segmental structure. In direct priming task, there were four kinds of primes sharing different phonological structure with the target: S prime (sharing segmental structure), T prime (sharing tone), ST prime (sharing both segment and tone), UR prime (no overlapping with the target). In mediated priming task, the prime and target are not directly related, but the prime is form-related to a third word that is not actually presented. For both priming tasks, subjects did lexical decision with two interstimulus intervals (ISIs) 50msec and 250msec. According to the results, Lee (2007) suggested that the "failure" of S primes to facilitate target could be interpreted as listeners' active use of tonal information to constrain lexical activation. However, when the ISI is 50msec, there was a facilitatory priming effect in minimal tone pair condition, suggesting that the mismatch in tone did not prevent activation. This study showed that minimal tone pairs were activated in early phase of lexical access because of the segment overlapping, but listeners were capable of using tonal information quickly to eliminate the tonally incompatible candidate. The result of Lee (2007) showed a stronger tonal effect in lexical processing.

2.4.2.2 Time course of tonal processing

In order to examine the tonal processing, most of previous studies adopted the "off-line" paradigm, such as the lexical decision task, vowel and tone judgment, same or different judgment and so on. The behavioral measures involving explicit responses by listeners might not access the on-line processing. In addition, the facilitation of form priming task might reflect a task-specific strategy by subjects. For example, in the priming paradigm, listeners could learn that there might be the same segments between initial prime and target and then were more ready to response to the target word (Lee, 2007). It was worth suspecting whether the behavioral studies were sensitive to the on-line processing. Some studies have explored how tonal processing unfolds in time. Most of these findings suggested tone and segment play a comparable role during spoken processing. Schirmer et al. (2005) used event related potentials (ERPs) to examine the role of tone and segmental information for word processing of spoken Cantonese. In this experiment, participants listened to sentences that were semantically correct (e.g., *beng6*; "illness") or included a semantically incorrect word and then judged whether the sentence was congruent or not. The semantically incorrect words differed from the most expected sentence by tone (e.g., *beng2*; "bisquit"), segmental structure (e.g., *bou6*; "step"), or by both tone and segmental structure (e.g., *gwai3*; "season"). The result showed that the amplitude of the N400-like negativity were comparable between tone and segmental violation. The authors suggested that tonal and segmental information were accessed at a similar point in time and played comparable roles during word recognition in Cantonese. Zhao et al. (2011) used also ERP technique to examine the time course of monosyllabic spoken word recognition. The result was similar to Schirmer et al. (2005). In the study, subjects performed a picture/spoken-word/picture task, in which they listened to spoken words and judged whether a previously presented picture and a subsequently presented picture belonged to the same semantic category. Each target picture was paired with five different pictures, which were either in the same or a different semantic category as the target picture. In addition, each target also corresponded to five different manipulated spoken monosyllables. Spoken words were manipulated for the relation to the name of the target picture: identical pronunciation (e.g., /bi2/-/bi2/), onset mismatched (e.g., /bi2/-/li2/), rime mismatched (e.g., /bi2/-/bo2/), tone mismatched (e.g., /bi2/-/bi3/), and syllable mismatched (e.g., /bi2/-/ge1/). The result indicated that during 400-500msec, both the rime and tone mismatch conditions had a tendency to elicit larger negativities than the identical condition. They suggested that the competitive effect of tone and rime during lexical activation were relatively the same and tonal information might play a comparable role as rime. In addition, this study suggested a modification of the TRACE model to incorporate detectors for lexical tone. Zhao et al. (2011) suggested that Chinese monosyllabic spoken word computation is based on global similarity across the whole syllable rather than phonemic-based segments. Therefore, they added “toneme” nodes with phoneme node and an additional level with syllabic morpheme nodes (Ye & Connine, 1999). According to Ye and Connine (1999) and Zhao et al. (2011), the toneme nodes were at the equivalent level of phoneme nodes, but they were separated from phoneme nodes. These toneme nodes also had inhibitory connections within the other nodes. Zhao et al. (2011) also added a syllabic morpheme node to the modified TRACE model. Thus, they suggested that there were four layers in the modified TRACE model (Figure 3): feature, phoneme/toneme, morpheme, and words (from the lowest to the highest).

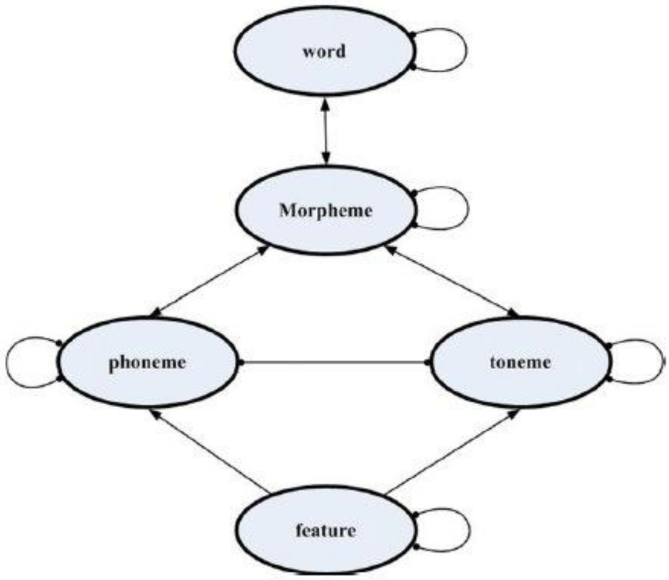


Figure 3. A four-layer modified version of the TRACE model with feature, phoneme/toneme, morpheme, and word node. Bidirectional arrows between levels represent interactive feedforward and feedback excitatory connections (Zhao et al., 2011).

2.5 Visual world paradigm

In the study Cooper (1974), the author asked the participants to listen to short narratives when looking at displays showing common object which were referred to the listening text. He found that listeners' gaze was drawn to the objects which were mentioned or related to auditory stimuli. Additionally, he also found that the eye movements of listeners were closely time-locked to the text. More than 90% of the fixations were on the triggered critical items when the corresponding word was spoken or within 200msec after the offset of the auditory word. He contended that it is a new paradigm for real-time investigation of cognitive process in particular for the detailed study of speech perception.

The paradigm pioneered by Allopenna, Magnuson, and Tanenhaus (1998) is now known as "visual world paradigm." On each experimental trial, participants were presented with an array of four pictures on a computer screen and subsequently heard an auditory stimulus corresponding to one of these items and then clicked on the target picture with a mouse. Usually, the visual display includes a target object or picture that the name is uttered in the auditory instruction, one or two competitors sharing some features with the target, and distractors that is unrelated to the target. The competition effects would stem from the manipulation, which the name of the competitor picture had a phonological relationship to the spoken target. The eye movements of the participants are recorded for later analysis. The proportion of looks to target and manipulated competitors were calculated during the unfolding of the auditory stimuli to see the competition effect. The activation of the name of the picture determines the probability that the participants will shift attention to that picture and thus generate a saccadic eye movement to fixate it. When listeners detected the differences between the target and the competitors could be reflected on the divergent time point of the trajectory of target and competitors.

There are some reasons why visual world paradigm becomes so widely used. Firstly, visual world paradigm can be used in an implicit and simple natural task, which makes it possible to investigate real-time language comprehension in a non-disruptive situation. Second, visual World Paradigm has provided researchers an effective tool to estimate lexical activation over time in studying the time course of spoken word recognition (M. K. Tanenhaus, 2007). This paradigm is highly sensitive on lexical activation and mapping well on the simulation results from the computational models like the TRACE model (Allopenna et al., 1998). The following sections review some studies of spoken word recognition with the visual world paradigm.

2.5.1 Visual world paradigm and spoken word recognition

Allopenna et al. (1998) examined whether competitors effects would be presented for objects with names that rhyme with the target. They showed participants an array of pictures containing a target picture (e.g., *beaker*), a cohort competitor picture (e.g., *beetle*) which shared an onset with the target word, a rhyme competitor (e.g., *speaker*) which shared the rhyme structure with the target, and a picture of distractor item that was phonologically unrelated to the target or competitor. Subjects performed the task by following the spoken instruction such as “Pick up the beaker; now put it below the diamond”. Their eye movements to the pictures of four objects were recorded. The researchers found that the likelihood of fixations to the “beaker” and the “beetle” increased when the participants heard the target “beaker.” When the auditory target “beaker” began to mismatch phonologically with “beetle,” the probability of looks to “beetle” decreased when the probability of looks to the “beaker” continued to rise. The proportion of looks to “speaker” started to increase when the end of the word “beaker” unfolded. The result showed that the onset competitors of the target had a competition earlier than rhyme competitors. The result could suggest that the acoustic information at the initial of the spoken words was more important than the later acoustic information. As predicted by the TRACE model, the onset phonological overlap constrained lexical selection.

Some studies have investigated the processing of phonetic detail. Bob McMurray, Tanenhaus, and Aslin (2002) used visual world paradigm to examine whether small within-category differences in voice onset time (VOT) affected lexical access. Participants were presented by a display of four pictures which were named by spoken stimuli that varied along 0-40msec VOT continuum. The result showed that spoken word recognition had graded sensitivity to with-category voice onset time. Therefore, the fine-grained phonetic differences were preserved in patterns of lexical activation in the competition among lexical candidates and could be used to maximize the efficiency of on-line word recognition. B. McMurray, Clayards, Tanenhaus, and Aslin (2008) examined the time course of phonetic cue integration by visual world paradigm and found that the phonetic cue was use for lexical access as soon as available. During test, the pictures which names were with the voicing (e.g., /b/ vs. /p/) or manner (e.g., /b/ vs. /w/) relationship were shown to participants. The result showed that after the onset of the target, the probability of eye fixations on target and competitor pictures diverged at different points in time.

2.5.2 Visual world paradigm and tonal processing

The evidence from an eye-tracking study also showed a consistent pattern with the ERP evidence(Schirmer et al., 2005; Zhao et al., 2011). Malins and Joanisse (2010) examined the tonal and segmental role in the Mandarin spoken word recognition with the visual world paradigm. In the experiment, participants were presented with an array of four pictures on a computer screen and subsequently heard an auditory stimulus corresponding to one of these items and then clicked on the target picture with a mouse. The competition effects would stem from the manipulation, which the name of the competitor picture had a phonological relationship with that of the target picture. There were three types of competitor, take the target *chung2* “bed” for example: segmental competitor (e.g., *chung1*, “window”), which shared the segmental structure but differed in tone with target, cohort competitor (e.g., *chuan2*, “ship”) which shared word initial phonemes and tone with target, rhyme competitor (e.g., *huang2*, “yellow”) which shared word-final phonemes and tone with target, and tonal competitor (e.g., *niu2*, “cow”) which shared only tone but differed in segmental structure with target. In order to show the time course of accessing tonal and segmental information, the authors examined the time of fixation proportion curves for cohort or segment competitors diverging from targets and found that the divergence was at a similar time (Figure 4). The authors suggested that the similar time course of resolution of targets from the two types of competitors indicates that segmental and tonal information were accessed concurrently and play a comparable role in constraining spoken word recognition.

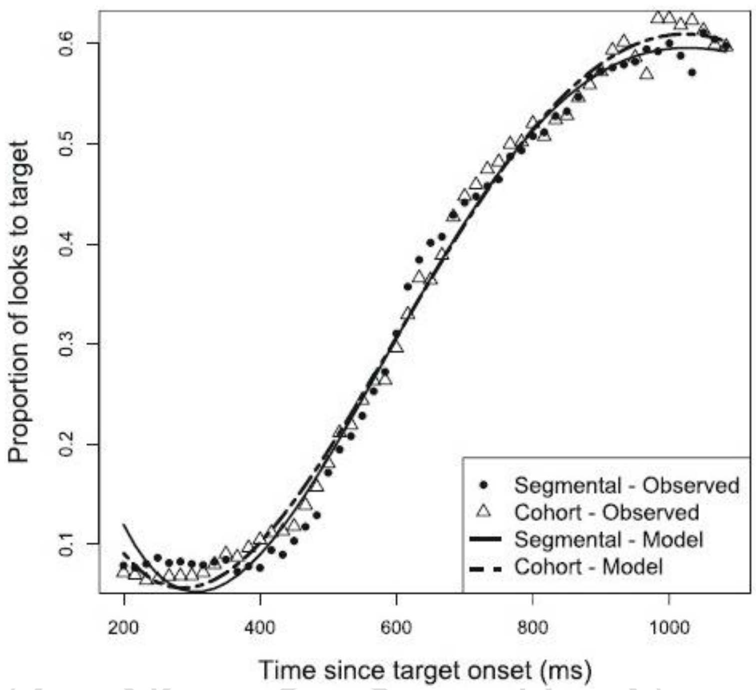


Figure 4. Observed data (symbols) and model fits (lines) of fixation proportions to target for the segmental and cohort conditions (Malins & Joanisse, 2010).

However, it might be awkward to examine the tonal and segmental information under the segmental condition and cohort condition. The fixation proportions of the target and competitor could also be discussed in tonal condition. If the tonal information had the same effect as segmental competitor, why the target in tonal condition had more fixation proportion at the early phase of spoken word recognition? The fixation proportions of target in tonal condition should be lower in early phase of processing if tonal information had comparable effect as segment on the lexical processing. Secondly, there were only seven stimuli set in their material. The number of the experimental stimuli might be too small to show the effects.

2.6 Interim Summary

The issue about the spoken word recognition of Chinese has been investigated by different experimental paradigms. However, the issue of the tonal information during lexical processing is still controversial. The tonal information spans over the syllable; therefore, the off-line experimental paradigms such as lexical decision or tone and segment judgment might not reveal the on-line tonal processing during lexical activation.

The aim of the study is to examine the time course of tonal processing during listening to Chinese spoken character. By using the visual world paradigm with printed characters, we manipulate the phonological structure mismatch between the target and the competitor in two experiments. According to Huettig and McQueen (2007), the printed word version might be more sensitive to phonological manipulation than the traditional picture version. The divergence between the target and the competitor would infer the time point of accessing tone during spoken character processing.

Chapter 3

Experiment One

The present experiment used the visual world paradigm and eye movement recording to examine the tone processing during Chinese spoken character recognition. Two phonological competitors were manipulated, the tonal competitor (TC) sharing only the tone with the target and the segmental competitor (SC) sharing the segmental structure with the target.

In order to investigate when the tonal information begins to affect the Chinese spoken processing, when the fixation proportion curves of targets diverge from two competitors and the unrelated distractors was examined. The time point of the divergence infers when participants could distinguish the tonal or segmental differences.

According to the phonological overlap between the targets and competitors, the first set of analysis examines whether and when the competitors obtain higher fixation proportion than the unrelated distractors. If tonal competitors are activated due to sharing the same tone with target, the fixation proportion on tonal competitor will be higher than the unrelated distractors. In addition, the patterns of four tones may be different because the time needed to identify different tones is varied (Lai & Zhang, 2008). As for the segmental competitor sharing the same segmental structure with the target, it is expected that the fixation proportions to segmental competitor would be higher than that of unrelated distractors.

The second set of analysis examines when the fixation proportion curves of target and the competitors diverge. If tone has an early influence, the time point of divergence between target and the segmental competitor may be early, similar to the divergence of the target and unrelated distractors. However, if tone influences lexical processing after the offset of the target, the divergence between target and the segmental competitor may occur after the offset of the target. As for tonal competitor, if the segmental incongruity can be detected early, the time point of divergence between the target and the tonal competitor should be early.

3.1Method

3.1.1 Participants

Thirty-two participants, including 21 females and 11 males were recruited through on-line sign-up sheets and paid to participate in the experiment. Their mean age was

23.5 years old, ranging from 19 to 30 years old. All participants had normal or correct-to-normal vision and were native speakers of Mandarin Chinese.

3.1.2 Material

3.1.2.1Stimuli

There were 140 monosyllabic Chinese characters in the experimental stimuli, including 28 target characters, 28 tonal competitor characters, 28 segmental competitor characters, and 56 unrelated characters. The segmental structure of these stimuli comprised 60 CV, 60 CVC, and 20 CGVC. The stimuli was consisted of 36 characters with Tone 1, 32 characters with Tone 2, 29 characters with Tone 3, and 33 characters with Tone 4. The features of target and competitors were controlled as follows. First of all, the frequency of these characters were controlled in a range between 7~200, as computed from the CKIP Electronic Dictionary (The CKIP Electronic Dictionary is an electronic lexicon for Mandarin Chinese containing 88,000 entries).There was no significant difference in character frequency across the target and the segmental and tonal competitors ($F (2, 81) = 1.931$, $p=.152$). Secondly, the average stroke of these characters were controlled in a range between 5~20. After deleting tone segmental competitor of which stroke was 20, there was no significant difference of the stroke across the target and two types of competitors was balanced. (F (2, 78=2.514, p=.088).Lastly, the average number of homophone was under 10. However, the difference was significant across the target and the competitors. (F (2, 81) = 9.608, p=.00).

Table 2. Means and SDs of character frequency, strokes, and homophone number for target, tonal competitor and segmental competitor

	Frequency		Stroke		Homophone	
	Mean	SD	Mean	SD	Mean	SD
TAR	52.88	25.42	11.25	2.35	1.36	0.56
TC	73.20	51.89	10.29	2.97	2.61	1.20
SC	55.82	43.68	12.32	2.99	3.11	2.31
UR	63.14	44.66	11.41	2.95	4.20	2.10

Note. TAR: target; TC: tonal competitor; SC: segmental competitor; UR: unrelated distractor

3.1.2.2 Recording

The target and the competitors were recorded by a 25-year-old female Chinese native speaker through the Audio-technica MB 4k/c cardioid condenser microphone. The recording data was digitalized at a sampling rate of 44100 Hz, 16 bits through the software CoolEdit Pro 2.0.The mean durations of target, tonal competitor, and segmental competitor were 713.2 msec, 685.2 msec, and 710.1 msec, respectively.

3.1.2.3 Tonal recognition pretest

To ensure that the tone of the auditory stimuli was clear for the subject to recognize, a survey was conducted. Five female whose average age was 24.2 years old took part in the survey. The participants listened to all of the auditory stimuli and then identified the tone (Tone 1 to Tone 4) they considered to be. The result showed that average of the accuracy of each experimental stimulus was 98%.

3.1.3 Design

Two conditions were manipulated in this experiment. The tonal competitors shared only tone with the target. The segmental competitors shared only the segmental structure with the target. Hence, for example, a stimuli set included a target */mɔ1/* ‘touch’, a tonal competitor */wa1/* ‘dig’, a segmental competitor */mɔ3/* ‘wipe’, and two unrelated distractors (the segmental and tone were different from target: */nu4/* ‘anger’, and */tɕy2/* ‘chrysanthemum’). An experimental trial comprised one target, one competitor, which was either tonal competitor or segmental competitor, and two unrelated distractors. The entire experiment consisted of 62 trials, including 56 experimental trials, 4 filler trials and 2 practices. The filler trials and practice trials were not included for analysis. The experimental trials were mixed and randomly distributed into four lists. In each of the lists, the number of each condition was equal and the conditions were counterbalanced across subjects. There were two blocks of 32 trials, of which the first two trials were fillers. The relationship between the target and the competitor in the first block was exchanged in the second block.

3.1.4 Layout of visual stimuli

Stimuli were presented on a computer monitor (1024×768 pixels). The size of the printed characters presented on the screen was 64×64 pixels in 標楷體 font, which were about 2 centimeters on the monitor (One centimeter on the screen corresponded approximately to 0.0133° of visual arc). Four black printed characters were presented in a diamond array on a grey background. The distance of the four characters between the characters to the fixation point was 104 pixels.

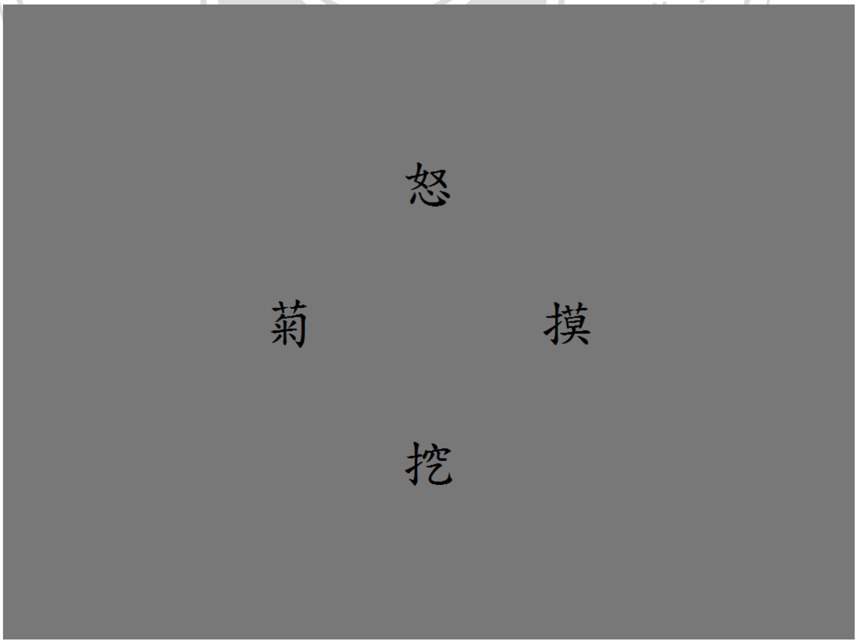


Figure 5. A sample display containing pictures of a target item (*/mɔ1/* ‘touch’), a tonal competitor (*/wa1/* ‘dig’), and two unrelated distractors (*/nu4/* ‘anger’, and */tɕy2/* ‘chrysanthemum’)

3.1.5 Apparatus

Participants' eye movements are recorded using Eyelink 1000 Desktop Mount eye-tracker, manufactured by SR Research. Viewing was binocular, and eye movements were recorded from their dominant eye. The eye tracker sampled gaze position every millisecond. Participants were seated 70 cm away from the chin rest to the screen. Stimuli are presented on a computer monitor (1024x768 pixels).

3.1.6 Procedure

Before the experiment, participants were given a consent form and tested for their dominant eye. After the test, they were seated in front of the monitor with their heads in a forehead and chin rest to eliminate head movement during testing. The instruction was given at the beginning of the experiment. Participants were instructed how to perform the experiment. The five-point calibration and validation were performed in initial trial of each block (two blocks in total). After checking the calibration, an experimental trial started and the participants were asked to fixate a cross, which was in the middle of the monitor. The cross were not vanished until the end of the spoken instruction 「請用滑鼠點選」 was presented via the earphone

(Audio-technicaATH-A700). An array of four characters was presented on the screen before 200msec the auditory target word onset. Participants used mouse to click what they heard on the computer screen with no time pressure. After the clicking, the next trial initiated.

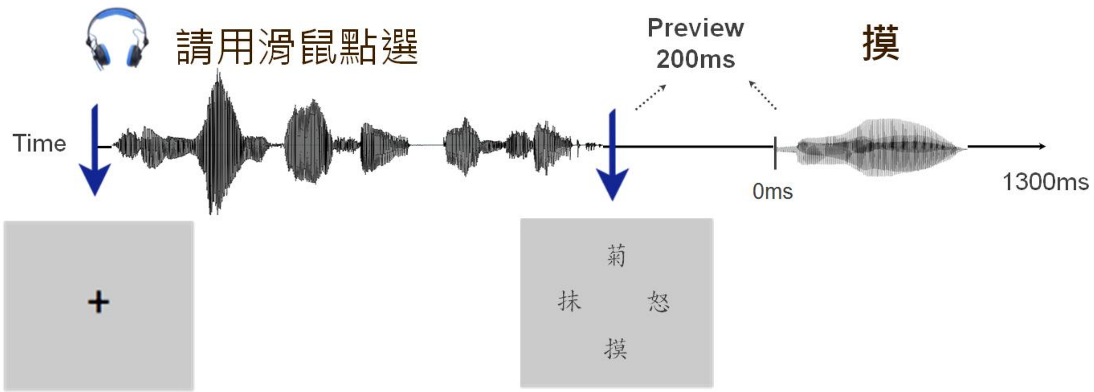


Figure 6. Experimental procedure and examples of a visual stimulus used in Experiment One. The display contained words: 摸 (the target), 抹 (the segmental competitor), 怒 and 菊 (the unrelated distractors).

3.2 Data analysis

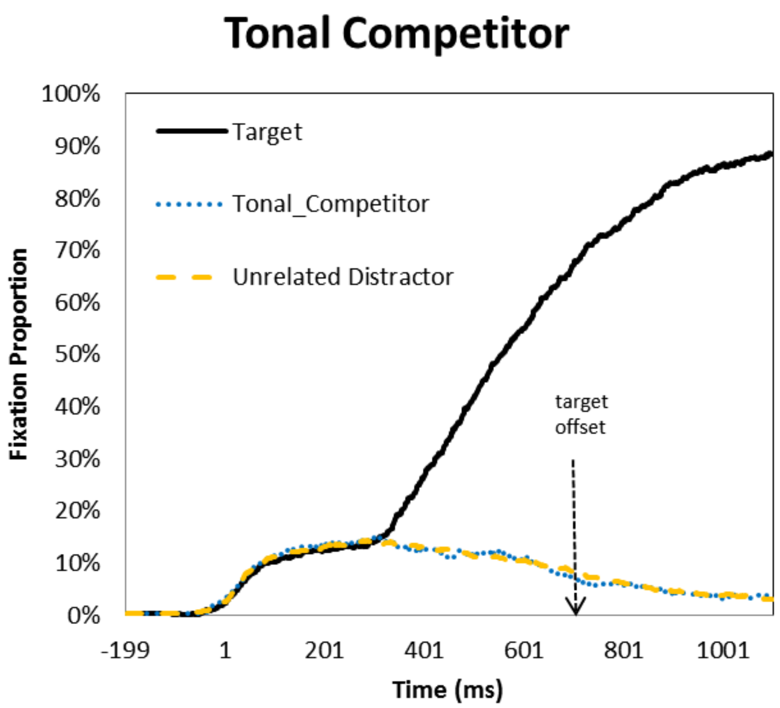
Eye gaze position was coded by defining four square regions in visual display, each occupying 32 pixels height and width around each character. A fixation was counted only as being within these regions. Fixation proportions to target (TAR), tonal competitors (TC), segmental competitors (SC) and unrelated distractors (UR) were then computed for every millisecond in each of the tonal and segmental condition within a recording length of 1,500 msec. The calculation of proportions was also divided into four tones for analyses.

For statistical analysis, the mean fixation proportions of each type of character in the visual display across two experimental conditions were computed during each 100-msec interval time bin. The result of analyses of variance (ANOVA) by participants (*F1*) and items (*F2*) for each time bin in the 1-1000-msec range from the acoustic target onset were reported.

3.3 Results

The mean of reaction time and correct hits on matching the acoustic target character to the character in visual display were computed for each participant. Mean reaction time for response was 1466.4 msec (SD = 526.9) and mean accuracy rate was 98 % (Max = 100%, Min = 95%).

Figure 7 plots the fixation proportions of target, competitor, and unrelated distractor for every millisecond from 200 msec, the time when visual display showing to the target acoustic onset 0 msec until 1000 msec after acoustic target onset in tonal and segmental conditions. The time when the target diverged from the tonal competitor was before the average duration (703msec) of the auditory stimuli. This result suggests that tone affects on-line lexical processing before the ending of the spoken character. Starting from about 301msec, segmental competitor attracted more fixations than the unrelated distractors. However, there was no significant difference between the tonal competitor and unrelated distractors. Furthermore, the time when the curve of target diverged from tonal competitor was earlier than when that of target diverged from segmental competitor about 200msec. In addition, as illustrated in Figure 7, the fixation proportion of segmental competitor was higher than that of tonal competitor.



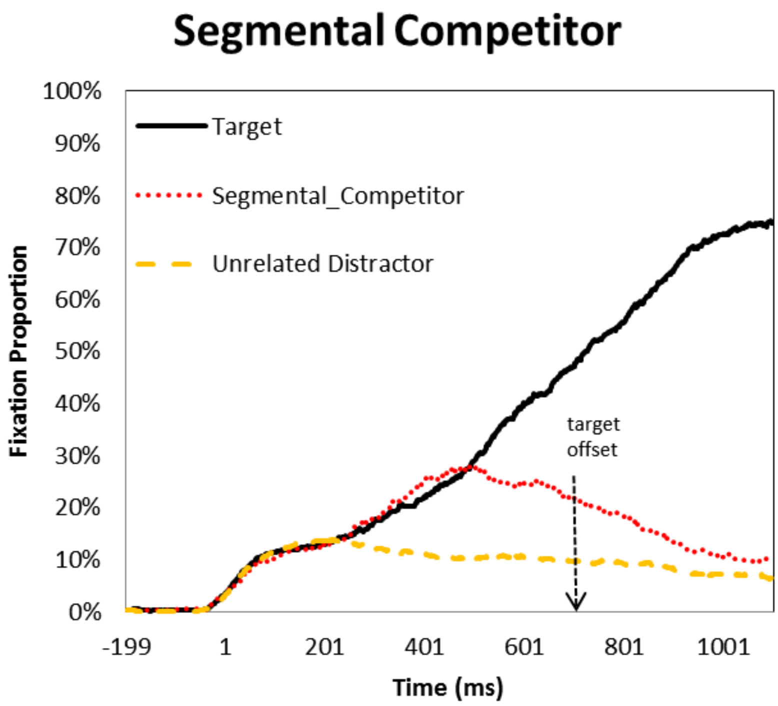


Figure 7. Fixation proportions to targets, competitors, and unrelated distractors for trials with tonal or segmental competitors in Experiment 1. The x-axis shows time in milliseconds from visual display onset, 200 msec before target acoustic onset, for the 1200 msec period.

We performed the analyses of variance (ANOVAs) by participants (F1) and items (F2) for the fixation proportions in time bins of 100 msec, starting from the onset of acoustic target for 1000 msec in tonal and segmental condition (Table 3). As Table 3 illustrated, for the tonal condition, the differences among target, competitor, and distractor were significant during 401 msec to 1000 msec. As for the segmental condition, the differences among target, competitor, and distractor were significant during 301 msec to 1000 msec.

Table 3. Analyses of variance by participant and item comparing mean fixation

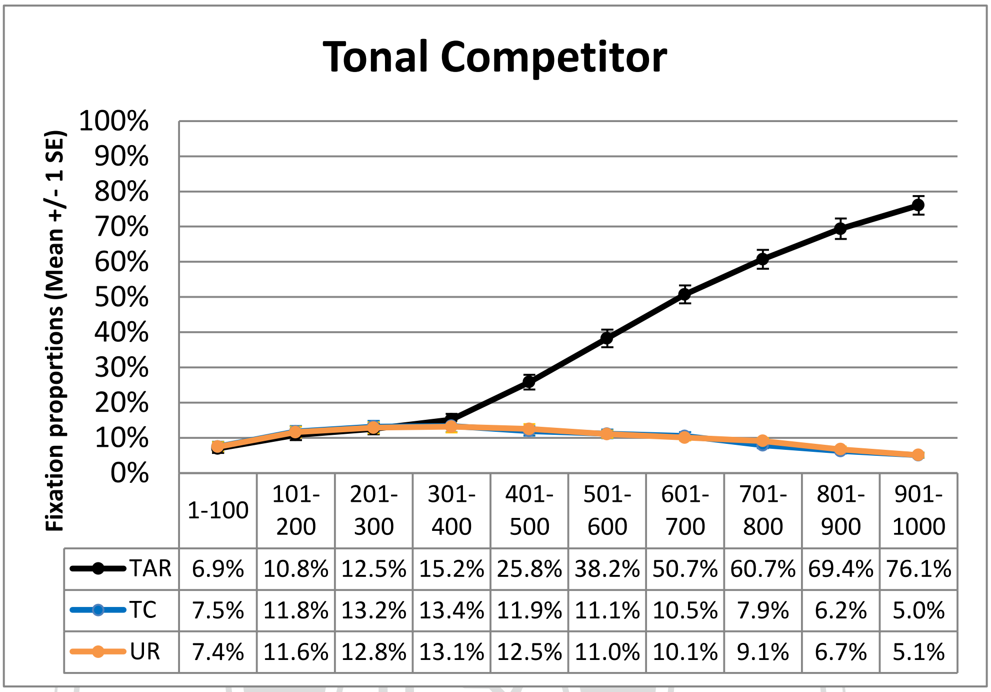
proportions to tonal and segmental competitors with those of the target and unrelated distractors from 1 msec to 1000 msec after acoustic target onset in Experiment 1

		Time bin (ms)									
Condition	Test	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
Tonal	$F_1(2,62)$	0.30	0.54	0.20	1.04	50.21	106.20	217.40	318.00	371.50	550.10
	p	0.743	0.587	0.823	0.361	0.000	0.000	0.000	0.000	0.000	0.000
	$F_2(2,54)$	0.18	0.23	0.09	0.98	29.39	89.72	150.40	263.40	539.70	718.10
	p	0.838	0.797	0.911	0.383	0.000	0.000	0.000	0.000	0.000	0.000
Segmental	$F_1(2,62)$	0.24	0.16	0.13	8.49	35.31	51.58	75.70	131.60	180.30	242.20
	p	0.787	0.852	0.879	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	$F_2(2,54)$	0.20	0.13	0.12	8.07	18.84	18.38	21.42	31.93	48.94	90.44
	p	0.823	0.879	0.887	0.001	0.000	0.000	0.000	0.000	0.000	0.000

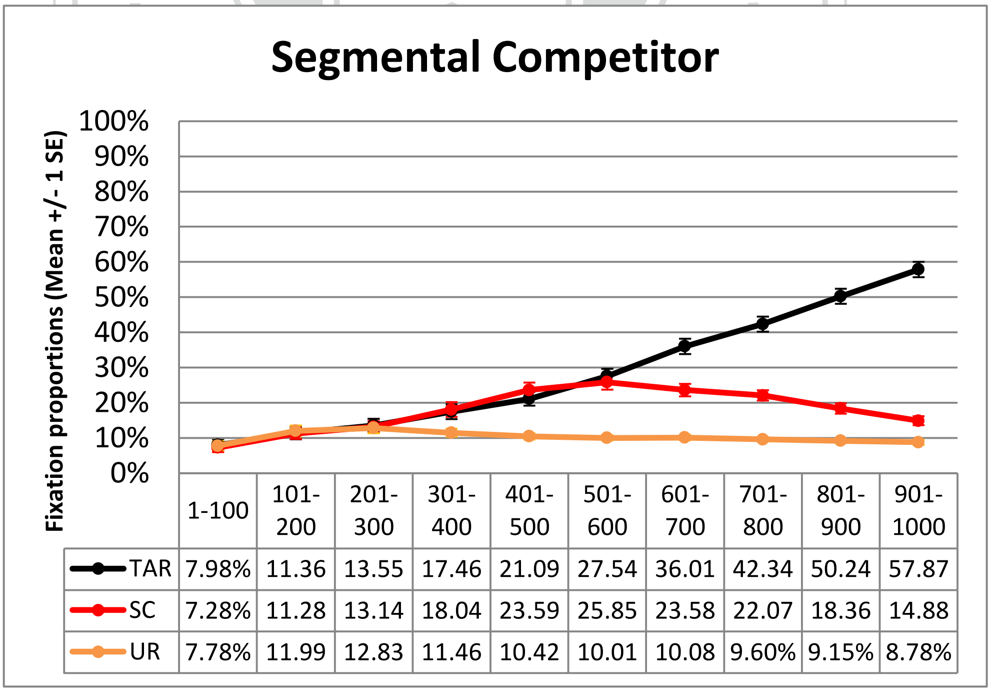
Note: $F1$ = 32 participants, $F2$ = 28 items

The mean fixation proportions to target, competitor, and the unrelated distractors and standard errors were shown in Figure 8. For each time bin, a one-way ANOVA was performed for trials with tonal or segmental competitors and their targets and distractors, followed by the post-hoc comparisons (Table 4). The post-hoc comparisons indicated no significant difference between the tonal competitor and unrelated distractors in the 1000 msec duration. The fixation proportion curves of the tonal competitor and the unrelated distractor were similar. As for segmental condition, the mean fixation proportions to the competitor and the unrelated distractors were significantly different during 301-1000-msec.

The mean fixation proportions to tonal competitor and target was significantly different by participants and by item during 401-1000-msec. The mean fixation proportions to segmental competitor and target were significantly different in both the by participants and by item analyses during 601-1000 msec.



	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR	6.9%	10.8%	12.5%	15.2%	25.8%	38.2%	50.7%	60.7%	69.4%	76.1%
TC	7.5%	11.8%	13.2%	13.4%	11.9%	11.1%	10.5%	7.9%	6.2%	5.0%
UR	7.4%	11.6%	12.8%	13.1%	12.5%	11.0%	10.1%	9.1%	6.7%	5.1%



	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR	7.98%	11.36	13.55	17.46	21.09	27.54	36.01	42.34	50.24	57.87
SC	7.28%	11.28	13.14	18.04	23.59	25.85	23.58	22.07	18.36	14.88
UR	7.78%	11.99	12.83	11.46	10.42	10.01	10.08	9.60%	9.15%	8.78%

Figure 8. Mean fixation proportions to targets, competitors, and unrelated distractors for tonal and segmental conditions in the 1,000 msec period following acoustic target onset in Experiment 1. Each data point represents the average of fixation proportions across participants in the time bin of 100 msec and the error bars show the standard

error of the data.[1]

Table 4. Analyses of variance by participant and item comparing mean fixation proportions to competitors with those of the target and unrelated distractors from 1 msec to 1000 msec after acoustic target onset in Experiment 1

		Time bin (ms)									
Condition	*Test*	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR - TC	z_1	-0.72	-0.98	-0.64	1.18	9.00	12.80	18.24	22.44	24.08	29.20
		p =1	*p* =.976	*p* =1	*p* =.713	*p* < .001	*p* < .001	*p* < .001	*p* < .001	*p* < .001	*p* < .001
	z_2	-0.56	-0.64	-0.44	1.15	6.90	11.79	15.21	20.47	29.09	33.44
		p =1	*p* =1	*p* =1	*p* =.750	*p* < .001	*p* < .001	*p* < .001	*p* < .001	*p* < .001	*p* < .001
TAR - SC	z_1	0.68	0.07	0.29	-0.33	-1.53	0.90	5.99	10.11	14.26	17.96
		p =1	*p* =1	*p* =1	*p* =1	*p* =.381	*p* =1	*p* < .001	*p* < .001	*p* < .001	*p* < .001
	z_2	0.62	0.06	0.28	-0.33	-1.12	0.54	3.20	4.99	7.45	11.00
		p =1	*p* =1	*p* =1	*p* =1	*p* =.791	*p* =1	*p* = .004	*p* < .001	*p* < .001	*p* < .001
TC - UR	z_1	0.09	0.17	0.32	0.16	-0.37	0.04	0.21	-0.51	-0.19	-0.02
		p =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1
	z_2	0.07	0.11	0.22	0.15	-0.29	0.04	0.17	-0.47	-0.23	-0.03
		p =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1	*p* =1
SC - UR	z_1	-0.49	-0.53	0.22	3.78	8.04	8.45	6.50	6.22	4.12	2.55
		1.000	1.000	1.000	*p* < .001	*p* < .001	*p* < .001	*p* < .001	*p* < .001	*p* < .001	*p* =.032
	z_2	-0.45	-0.48	0.22	3.70	5.89	5.06	3.47	3.07	2.15	1.56
		p =1	*p* =1	*p* =1	*p* =.001	*p* < .001	*p* < .001	*p* =.002	*p* =.006	*p* =.095	*p* =.355

Note: $z1$ = 32 participants, $z2$ = 28 items

For segmental condition, the mean fixation proportions to segmental competitor and target were significantly different by participants analyses on 601-700-msec in Tone 1, Tone 2, and Tone 4 [Tone 1:z1= 4.49, p < .001; Tone 2:z1 = 4.12, < .001; Tone 4:z1= 2.76, p = .017]. However, in the time period 601-700-msec, there was no significant difference by participants analyses between the segmental competitor and target in Tone 3. The late time of divergence between the target and the segmental competitor in Tone 3 suggests that tonal information might influence the on-line lexical processing. Additionally, this result was consistent to the result of Lai and Zhang (2008) suggesting that Tone 3 had the latest isolation point for the size of the

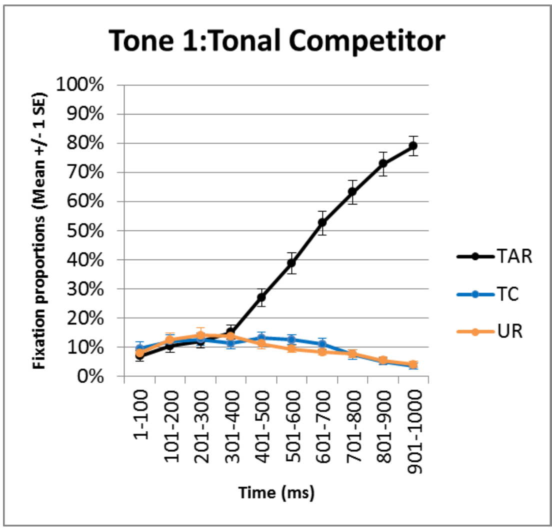
[1] The calculation of mean fixation proportions in each time bin included the area of target, competitor, unrelated distractors, cross , and the remaining areas outside the above-mentioned areas in the visual display

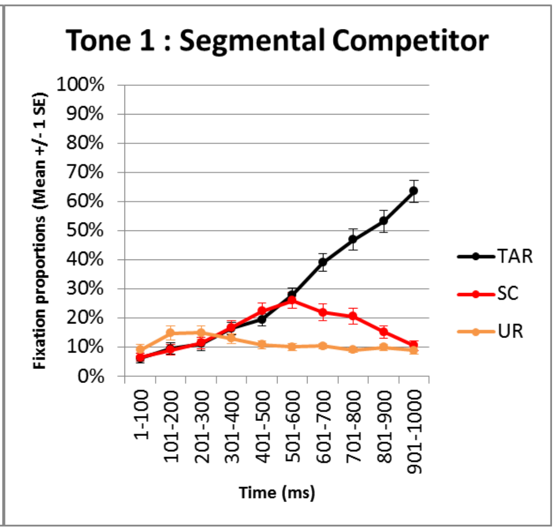
segments needed to correctly identify the stimulus without further changes in the gating paradigm.

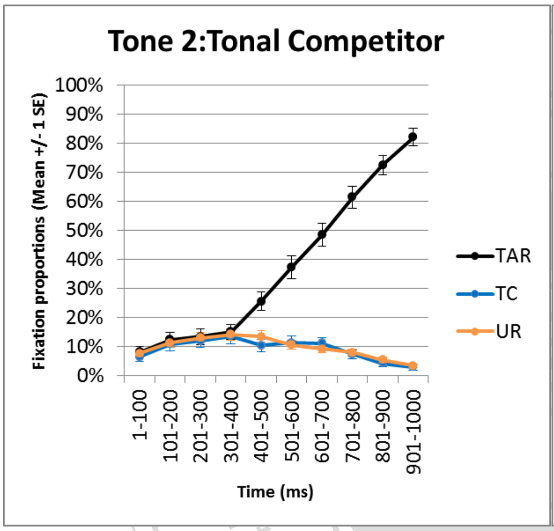
For tonal condition, the mean fixation proportions to tonal competitor and target was significantly different by participants analyses on 401~500msec in Tone 1, Tone 2, Tone 3, and Tone 4 [Tone 1:z_1 =4.24, p < .001; Tone 2:z_1 = 4.54, p < .001; Tone 3: z_1 =3.88, p < .001; Tone 4:z_1 = 3.97, p < .001].

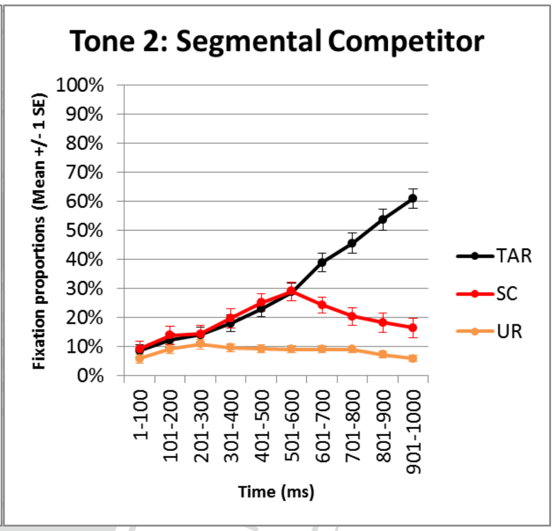
There was no significant difference of the time of divergence between the tonal competitor and the unrelated distractors among four tones. However, the mean fixation proportions to the segmental competitor and unrelated distractors was significantly different by participants analyses on 301-400-msec in Tone 2 and Tone 4 [Tone 2: z_1 = 2.84, p = .014; Tone 4: z_1 = 2.76, p = .018]. There was no significant difference by participants analyses on 301-400-msec in Tone 1 and Tone 3.

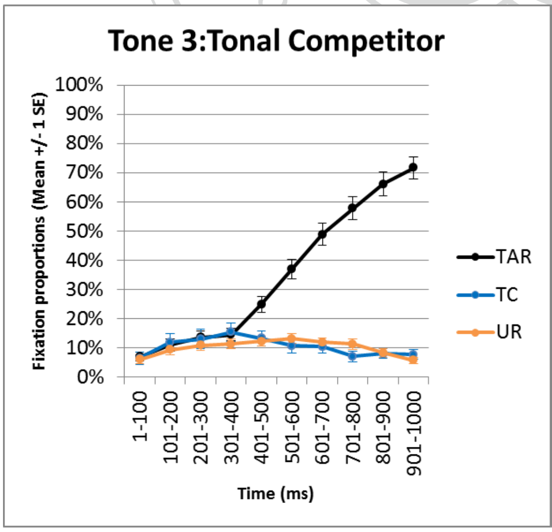
Figure 9 plots the fixation proportion over time in two conditions among four tones. Table 5 shows the time period when mean fixation proportion had significant difference by participants analysis between competitor and its target and unrelated distractors in two conditions of Experiment 1.

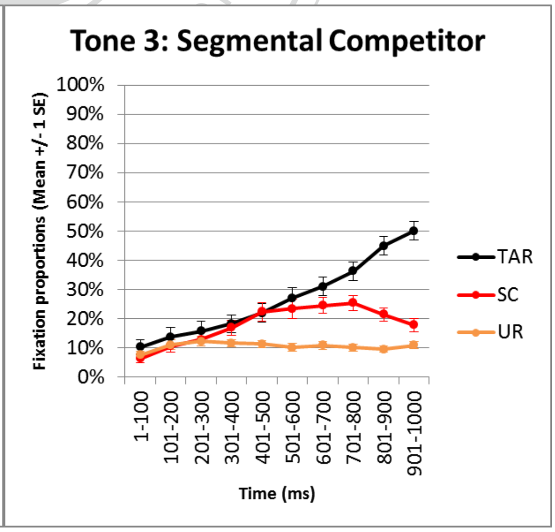

Tone 1:Tonal Competitor
Fixation proportions (Mean +/- 1 SE)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1-100
101-200
201-300
301-400
401-500
501-600
601-700
701-800
801-900
901-1000
Time (ms)
TAR
TC
UR


Tone 1 : Segmental Competitor
Fixation proportions (Mean +/- 1 SE)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1-100
101-200
201-300
301-400
401-500
501-600
601-700
701-800
801-900
901-1000
Time (ms)
TAR
SC
UR


Tone 2:Tonal Competitor
Fixation proportions (Mean +/- 1 SE)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1-100
101-200
201-300
301-400
401-500
501-600
601-700
701-800
801-900
901-1000
Time (ms)
TAR
TC
UR


Tone 2: Segmental Competitor
Fixation proportions (Mean +/- 1 SE)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1-100
101-200
201-300
301-400
401-500
501-600
601-700
701-800
801-900
901-1000
Time (ms)
TAR
SC
UR


Tone 3:Tonal Competitor
Fixation proportions (Mean +/- 1 SE)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1-100
101-200
201-300
301-400
401-500
501-600
601-700
701-800
801-900
901-1000
Time (ms)
TAR
TC
UR


Tone 3: Segmental Competitor
Fixation proportions (Mean +/- 1 SE)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1-100
101-200
201-300
301-400
401-500
501-600
601-700
701-800
801-900
901-1000
Time (ms)
TAR
SC
UR

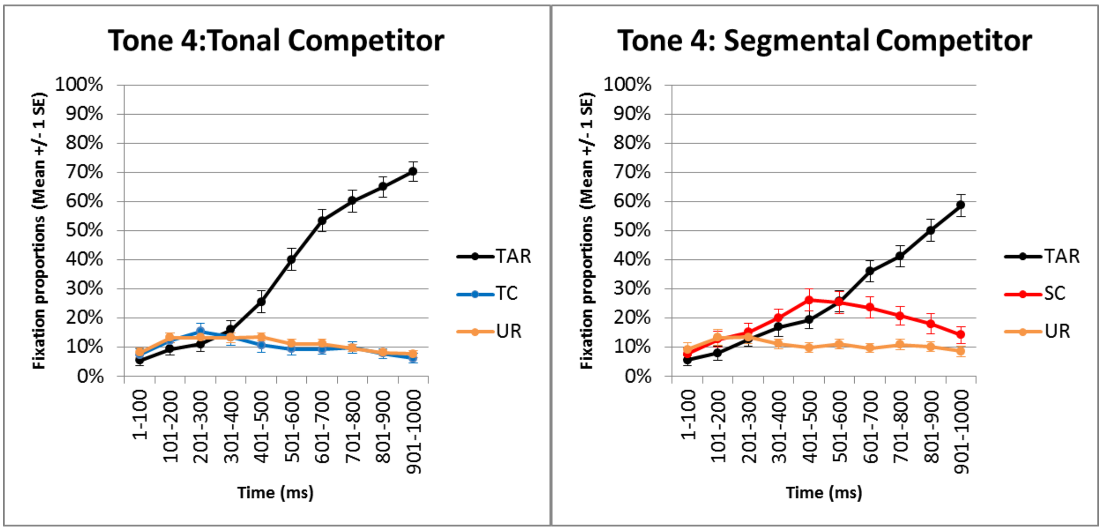


Figure 9. Mean fixation proportions to targets, competitors, and unrelated distractors for tonal and segmental conditions in the 1,000 msec period following acoustic target onset in Experiment 1. Each data point represents the average of fixation proportions across participants in the time bin of 100 msec and the error bars show the standard error of the data. Fixation proportions to target, competitors and unrelated distractors over time for 4 tones are shown respectively.

Table 5. The time period when mean fixation proportion had significant difference between TAR-TC, TAR-SC, TC-UR and SC-UR by participants in Experiment 1

	Experiment 1			
	TAR-TC	TAR-SC	TC-UR	SC-UR
All	401-1000	601-1000	--	301-1000
Tone 1	401-1000	601-1000	--	401-800
Tone 2	401-1000	601-1000	--	301-800
Tone 3	401-1000	701-1000	--	401-900
Tone 4	401-1000	601-1000	--	301-700

Note. – no significant differences

3.4 Discussion

There was no significant difference on the proportion of looks between the tonal competitor and the unrelated distractor. The mismatch of the first segments may override the possible effect of the tone in the activation stage. Therefore, subjects would nearly consider the tonal competitor as the unrelated distractors.

When could tone be accessed during the spoken character recognition? The divergent time of the curves between target and the segmental competitor could be extrapolated as the time point when listeners could distinguish the tonal difference of segmental competitor from target. The result showed that the divergent time of the segmental competitor and the target was at about 601-700-msec. This divergent time was earlier than the offset of the target (at about 703msec), which suggests that tone may begin to affect the spoken character processing before subjects catch the whole acoustic information.

While the results suggest that tonal information could affect the spoken character processing before the end of the auditory stream, however, the divergent time of the curves between the tonal competitor and target was the same among four tones. If tonal information had impact on the processing, fixation proportions of tonal competitor should have significant differences in different time bin among four tones. Different tonal characteristics could not affect listeners' decision on choosing the target. This result could imply that the tonal information has weak influence on the spoken character processing.

Additionally, weak tonal effect could be the reason why the curves of segmental competitor diverged from the target in later time, comparing to the time when the curves of target diverging from unrelated distractor. Owing to the totally different syllable structure between target and unrelated distractor, the time when listeners are able to distinguish two characters could consider to be the earliest time listeners could identify the tonal and segmental information. It was worth suspecting that tonal information could affect spoken character processing in an earlier time.

The weak and late tonal effect might be because of the great segmental disparity between the tonal competitor and target. As the auditory stimuli unfold, the initial segments of the character of tonal competitor make participants consider it different from the target and thus fixate less on it in an early time. The great segmental differences of tonal competitor might be the reason why tone could not take effect during spoken character processing.

The last issue was the characteristics of tone in processing spoken character. There was a latest time of divergence between the segmental competitor and the target in Tone 3, in which of Lai and Zhang (2008) that the last isolation point in a gating paradigm is in Tone 3. However, in the present study, the order when the curves of competitor and target diverged in Tone 1, Tone 2, and Tone 4 were inconsistent with the order of Isolation Point in Lai and Zhang (2008).

To sum up, the result of experiment one showed that tone would constrain before the unfolding of the whole syllable information. However, the totally different segmental structure between the target and the tonal competitor caused listeners barely consider the tonal competitor as unrelated distractor. Experiment two examined the earlier time point when tone would affect lexical processing when the initial segments were the same.

Chapter 4

Experiment Two

In Experiment 2, the design was similar as experiment one. However, the targets shared the same initial cohort (two segments) with two types of competitors. One type of the cohort competitors shared the initial cohort and has the same tone with the target (cohort-tone competitor, CTC), the other shared the cohort but the tone is different from targets (cohort-only competitor, COC). For the unrelated distractor, similar as experiment one, the tonal and segmental structure was totally different from the target and the competitors.

The time point of divergence between the curves of target and the competitors, and between the competitors and the unrelated was examined. Thus, if tone affected the processing in early phase of spoken character recognition, the time point of divergence between the curves target and the competitor would be earlier in cohort-only competitor than in cohort-tone competitor. In addition, the proportion of looks to cohort-only competitor would be lower than that of the cohort-tone competitor. On the contrary, if the tone information did not affect the spoken character processing in early time, the curve of the cohort-only competitor would be similar as the cohort-tone competitor. Table 6 illustrates the predictions for the accounts whether tone is accessed in initial character in Experiment 2.

Table 6. Predictions of two accounts for initial tonal processing in Experiment 2

Accounts	Predictions	
	Fixation Proportion	Divergent time between TAR and competitors
1. Tone is accessed early	CTC > COC	COC is earlier than CTC
2. Tone is accessed late	CTC = COC	Comparable time point

4.1 Method

4.1.1Participants

Thirty-two participants, including 21 females and 11 males were recruited through on-line sign-up sheets and paid to participate in the experiment. Their mean age was 21.7 years old, ranging from 19 to 25 years old. All participants had normal or correct-to-normal vision and were native speakers of Mandarin Chinese.

4.1.2 Material

4.1.2.1 Stimuli

There were 140 monosyllabic Chinese characters in the experimental stimuli, including 28 target characters, 28 cohort-tone competitor character, 28 cohort-only competitor characters, and 56 unrelated characters. The segmental structure of these stimuli comprised 70 CVC, and 70 CGVC. The stimuli was consisted of 36 characters with Tone 1, 35 characters with Tone 2, 35 characters with Tone 3, and 34 characters with Tone 4. The features of target and competitors were controlled as follows. First of all, the frequency of these characters were controlled in a range between 7~200, as computed from the CKIP Electronic Dictionary (The CKIP Electronic Dictionary is an electronic lexicon for Mandarin Chinese containing 88,000 entries). There was no significant difference in character frequency across the target and the cohort-tone and cohort-only competitors ($F(2, 81) = 2.105$, $p=.128$). Secondly, the average stroke of these characters were controlled in a range between 5~20. There was no significant difference of the stroke across the target and two types of competitors were balanced. ($F(2, 81) = 0.122$, $p=.885$). Lastly, the average number of homophone was under 10. There was no significant difference in the average number of homophone across the target and the two types of competitors ($F(2, 81) = 1.172$, $p=.315$).

Table 7. Means and SDs of character frequency, strokes, and homophone number for target, cohort-tone competitor and cohort-only competitor

	Frequency		Stroke		Homophone	
	Mean	SD	Mean	SD	Mean	SD
TAR	42.93	25.30	12.46	2.86	2.64	1.47
CTC	64.25	49.01	12.21	3.00	3.14	1.21
COC	49.68	41.16	12.04	3.84	3.11	1.40
UR	47.03	32.15	11.66	3.12	3.13	1.74

Note. TAR: target; CTC: cohort-tone competitor; COC: cohort-only competitor; UR: unrelated distractor

4.1.2.2 Recording

The target and the competitors were recorded by a 25-year-old female Chinese native speaker through the Audio-technica MB 4k/c cardioid condenser microphone. The recording data was digitalized at a sampling rate of 44100 Hz, 16 bits through the software Praat. The mean durations of target, cohort-tone competitor, and cohort-only competitor characters were 745.3 msec, 741.5 msec, and 755.2 msec, respectively.

4.1.2.3 Auditory Stimuli Pretest

To ensure that the tone of the auditory stimuli was clear for the subject to recognize, a survey was conducted. Four females and one male whose average age was 26.4 years old took part in the survey. The participants listened to all of the auditory stimuli and then ticked off the tone (Tone 1 to Tone 4) and typed the Chinese character they considered to be. The result showed that average of the accuracy of each experimental stimulus was 99%.

4.1.3 Design

Two conditions were manipulated in this experiment. There were two syllable structures CVC and CGVC in the present study. The initial cohort in CVC syllable structure was CV, and the initial cohort in CGVC syllable structure was CG. The cohort-tone competitors shared cohort structure and tone with the target. The cohort-only competitors shared only the cohort structure with the target. For each target, the two types of competitors shared the same rhyme. For example, a stimuli set included a target / *tang1* / ‘soup’, a cohort-tone competitor / *tai1* / ‘fetus’, a cohort-only competitor / *tai4* / ‘peaceful’, and two unrelated distractors (the segmental and tone were different from target: / $p^{h}ow3$ / ‘to cut open’, and / *xən2* / ‘scar’). An experimental trial comprised one target, one competitor, which was either cohort-tone competitor or cohort-only competitor, and two unrelated distractors. The entire experiment consisted of 62 trials, including 56 experimental trials, 4 filler trials and 2 practices. The filler trials and practice trials were not included for analysis. The experimental trials were mixed and randomly distributed into four lists. In each of the lists, the number of each condition was equal and the conditions were counterbalanced across subjects. There were two blocks of 32 trials, of which the first two trials were fillers. The relationship between the target and the competitor in the first block was exchanged in the second block.

4.1.4 Layout of visual stimuli

The layout of visual stimuli was same as Experiment 1.

4.1.5 Apparatus & procedure

The apparatus and procedure follows that of Experiment 1.

4.2 Data analysis

All the analyses and measures were the same as in Experiment 1.

4.3 Result

The mean of reaction time and correct hits on matching the acoustic target character to the character in visual display were computed for each participant. Mean reaction time for response was 1315.3 msec (SD = 275.3) and mean accuracy rate was 99 % (Max = 100%, Min = 98%).

Figure 10 plots the fixation proportions of target, competitor, and unrelated distractor for every millisecond from 200 msec, the time when visual display showing to the target acoustic onset 0 msec until 1000 msec after acoustic target onset in cohort-tone and cohort-only conditions. Starting from about 301msec, both of the two competitors attracted more fixations than the unrelated distractors. As expected, the time when the curve of target diverged from cohort-only competitor was earlier than when the target curve diverged from cohort-tone competitor curve. In addition, as illustrated in Figure 11, the fixation proportions of cohort-tone competitor was higher than that of cohort-only competitor.

Figure 12 and Figure 13 plots the fixation proportions of target, competitor, and unrelated distractor in CVC and CGVC syllable structure respectively for every millisecond from 200 msec, the time when visual display showing to the target acoustic onset 0 msec until 1000 msec after acoustic target onset in cohort-tone and cohort-only conditions. For the syllable structure of CVC, the time when the curve of target diverged from cohort-only competitor was earlier than when the target curve diverged from cohort-tone competitor curve. However, as for CGVC, the time when the curve of target diverged from cohort-only competitor was similar when the target curve diverged from cohort-tone competitor curve.

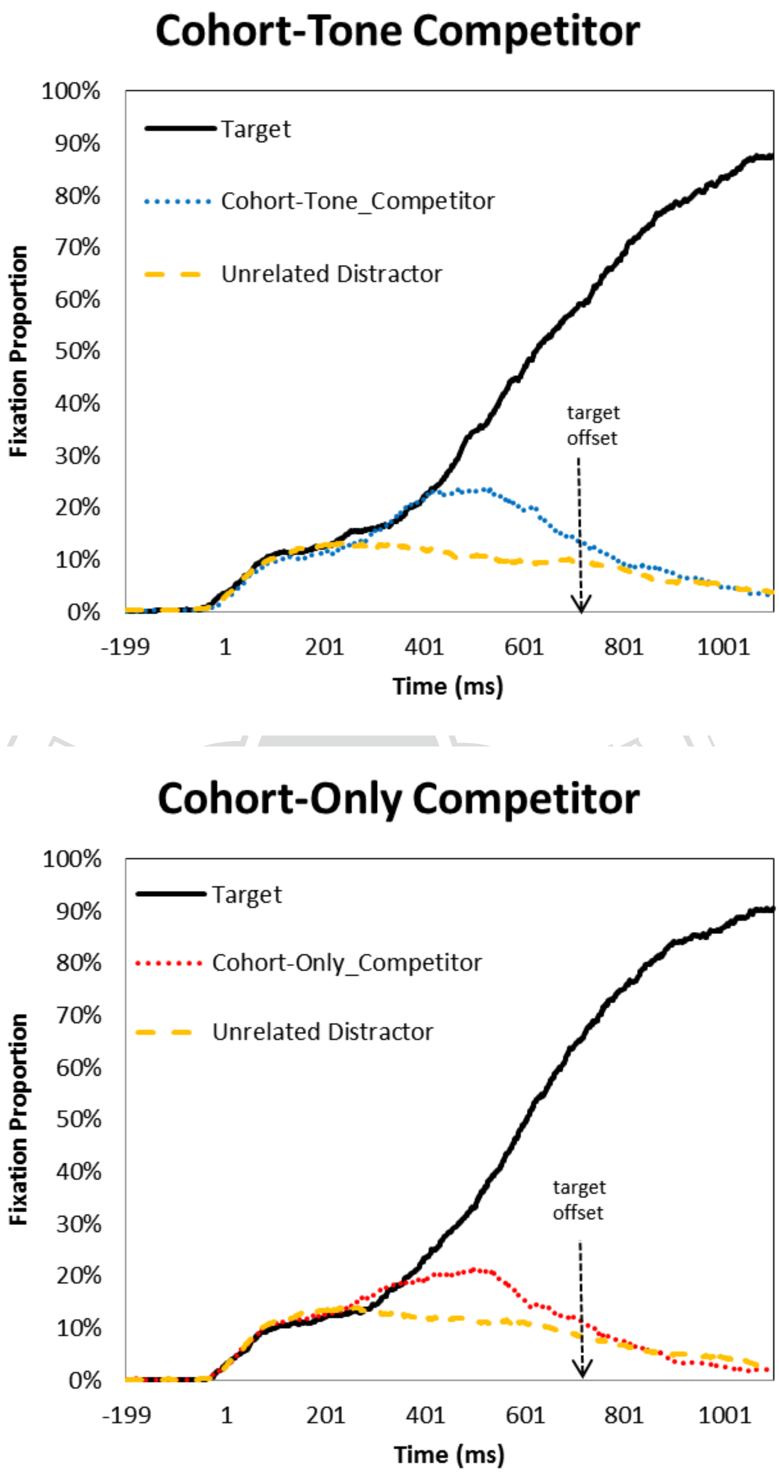


Figure 10. Fixation proportions to targets, competitors, and unrelated distractors for trials with cohort-tone or cohort-only competitors in Experiment 2. The x-axis shows time in milliseconds from visual display onset, 200 msec before target acoustic onset, for the 1200 msec period.

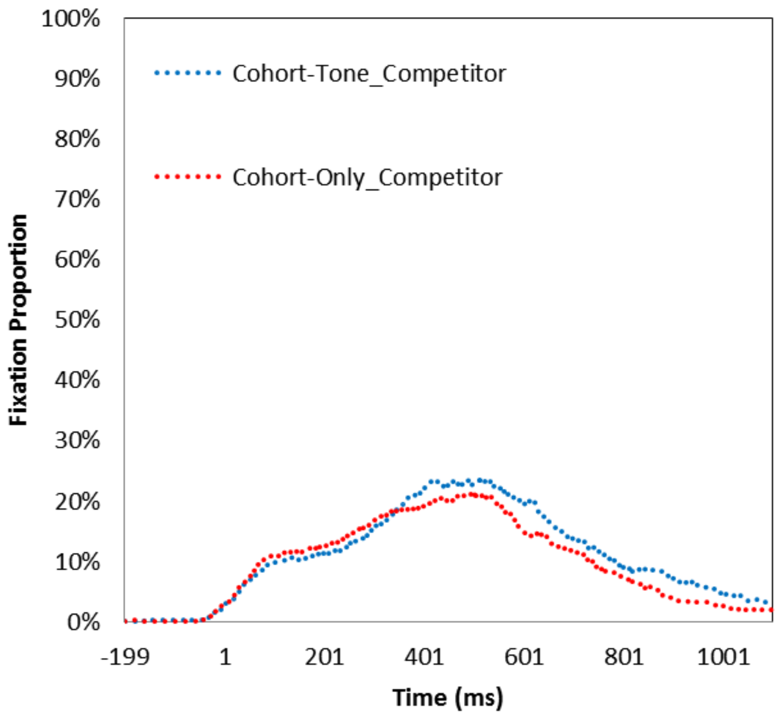


Figure 11. Fixation proportions to cohort-tone/cohort-only competitors across two experimental conditions in Experiment 2. The x-axis shows time in milliseconds from the display onset, for 1200 msec.

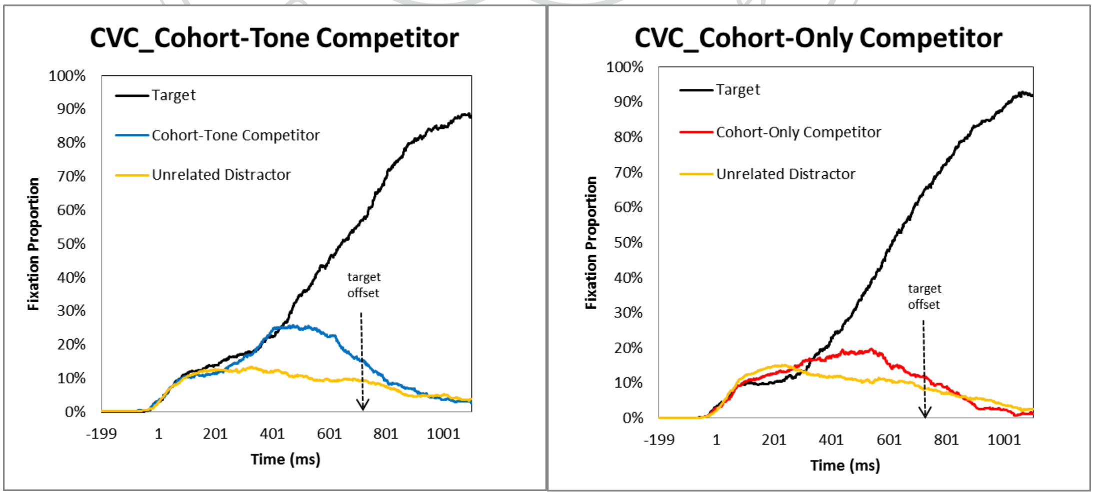


Figure 12. Fixation proportions to targets, competitors, and unrelated distractors in CVC syllable structure for trials with cohort-tone or cohort-only competitors in Experiment 2. The x-axis shows time in milliseconds from visual display onset, 200 msec before target acoustic onset, for the 1200 msec period.

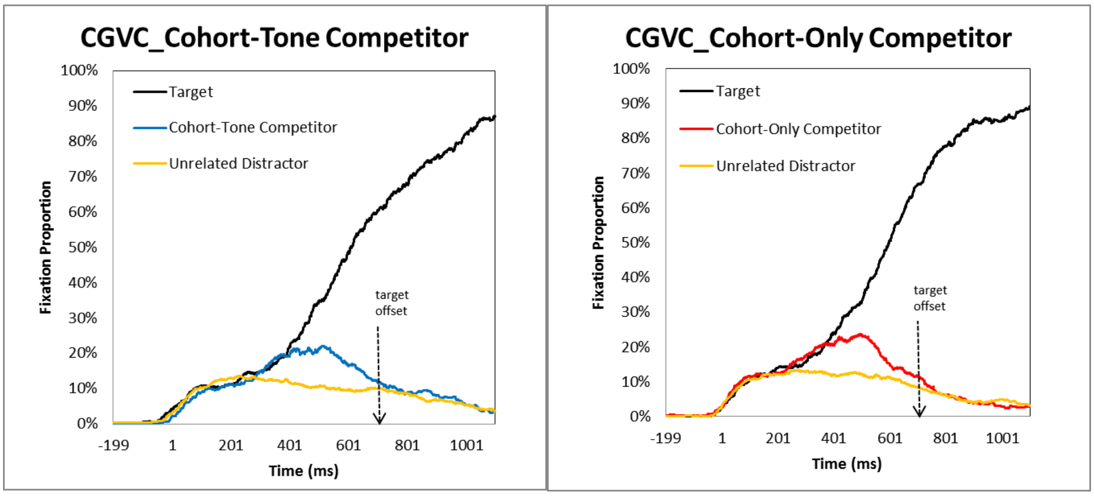


Figure 13. Fixation proportions to targets, competitors, and unrelated distractors in CGVC syllable structure for trials with cohort-tone or cohort-only competitors in Experiment 2. The x-axis shows time in milliseconds from visual display onset, 200 msec before target acoustic onset, for the 1200 msec period.

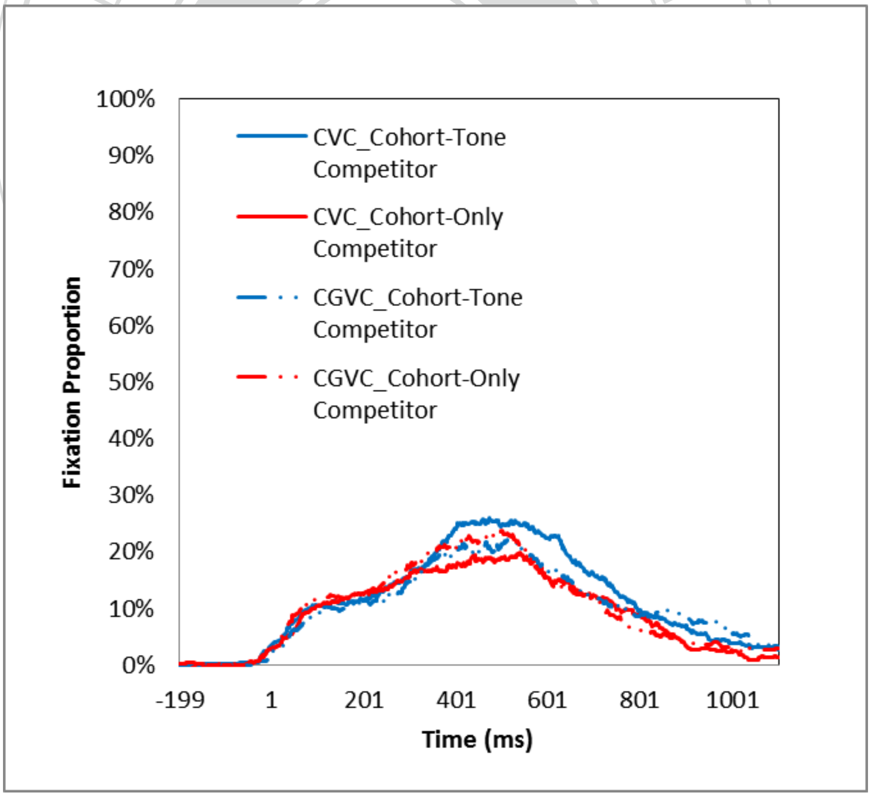


Figure 14. Fixation proportions to cohort-tone/cohort-only competitors in CVC and CGVC syllable structure across two experimental conditions in Experiment 2. The x-axis shows time in milliseconds from the display onset, for 1200 msec.

We performed the analyses of variance (ANOVAs) by participants (F1) and items (F2) for the fixation proportions in time bins of 100 msec, starting from the onset of acoustic target for 1000 msec in cohort-tone and cohort-only condition (Table 8). Additionally, we performed the same analyses as above-mentioned in CVC and CGVC respectively (Table 9 and Table 10). As Table 8 illustrated, in both of the two conditions, the differences among target, competitor, and distractor were significant during 301 msec to 1000 msec. Table 9 showed that the differences among target, competitor, and distractor were significant during 401 msec to 1000 msec in both conditions. Table 10 showed that the differences among target, competitor, and distractor were significant during 401 msec to 1000 msec in cohort-tone condition but during 301 msec to 1000ms in cohort-only condition.

Table 8. Analyses of variance by participant and item comparing mean fixation proportions to tonal and segmental competitors with those of the target and unrelated distractors from 1 msec to 1000 msec after acoustic target onset in Experiment 2

All											
		Time bin (ms)									
Condition	Test	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
Cohort-Tone	$F1(2,62)$	0.66	0.90	0.91	5.70	30.72	81.58	116.20	249.00	364.60	609.10
	p	0.521	0.414	0.410	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	$F2(2,54)$	0.45	0.54	0.57	3.33	19.41	77.86	162.20	183.70	321.10	485.70
	p	0.642	0.584	0.569	0.043	0.000	0.000	0.000	0.000	0.000	0.000
Cohort-Only	$F1(2,62)$	0.36	0.62	0.36	4.32	28.56	66.35	193.50	390.90	599.20	1055.00
	p	0.701	0.541	0.702	0.018	0.000	0.000	0.000	0.000	0.000	0.000
	$F2(2,54)$	0.24	0.96	0.49	6.05	34.81	63.29	180.50	327.20	741.50	1758.00
	p	0.790	0.390	0.614	0.004	0.000	0.000	0.000	0.000	0.000	0.000

Note: $F1$ = 32 participants, $F2$ = 28 items

Table 9. Analyses of variance by participant and item comparing mean fixation proportions to tonal and segmental competitors with those of the target and unrelated distractors in CVC syllable structure from 1 msec to 1000 msec after acoustic target onset in Experiment 2

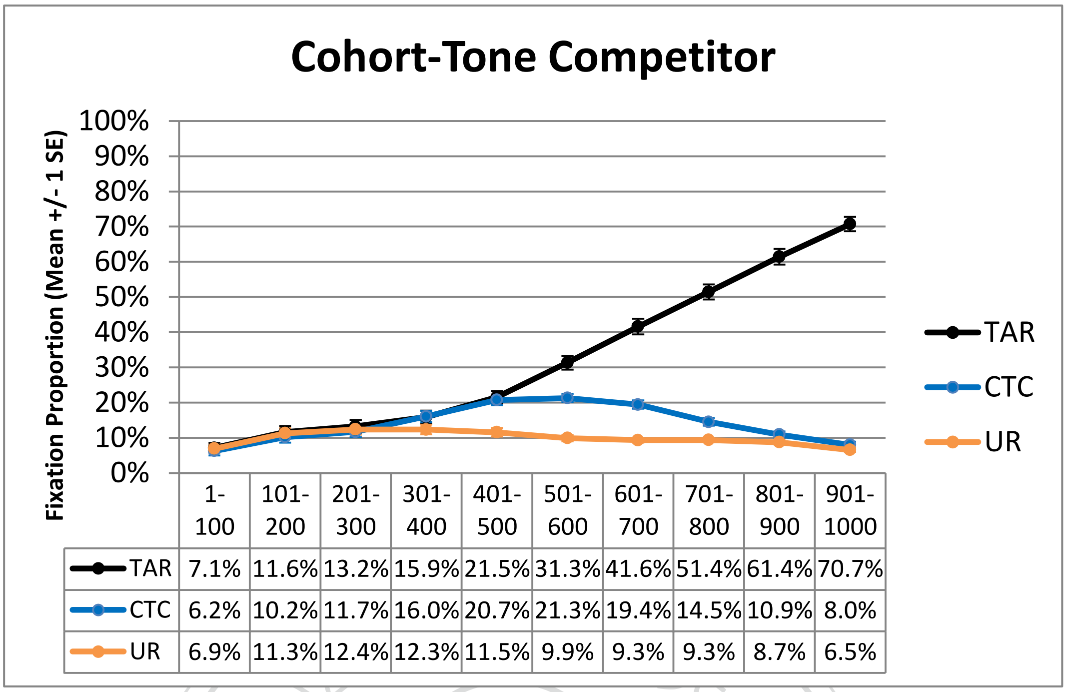
CVC											
		Time bin (ms)									
Condition	Test	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
Cohort-Tone	*F*1(2,62)	0.06	0.68	0.92	2.67	11.32	31.64	-0.39	119.40	236.70	454.90
	p	0.938	0.512	0.403	0.077	0.000	0.000	1.000	0.000	0.000	0.000
	*F*2(2,54)	0.04	0.54	0.69	2.08	7.72	37.54	71.36	81.06	227.80	659.10
	p	0.958	0.590	0.509	0.145	0.002	0.000	0.000	0.000	0.000	0.000
Cohort-Only	*F*1(2,62)	0.08	1.30	1.40	0.85	10.59	45.30	89.00	140.90	277.00	699.60
	p	0.925	0.281	0.253	0.431	0.000	0.000	0.000	0.000	0.000	0.000
	*F*2(2,54)	0.12	2.17	2.91	2.07	19.71	30.83	73.82	117.30	323.00	702.90
	p	0.892	0.134	0.072	0.147	0.000	0.000	0.000	0.000	0.000	0.000
Note: *F*1 = 32 participants, *F*2 = 28 items											

Table 10. Analyses of variance by participant and item comparing mean fixation proportions to tonal and segmental competitors with those of the target and unrelated distractors in CGVC syllable structure from 1 msec to 1000 msec after acoustic target onset in Experiment 2

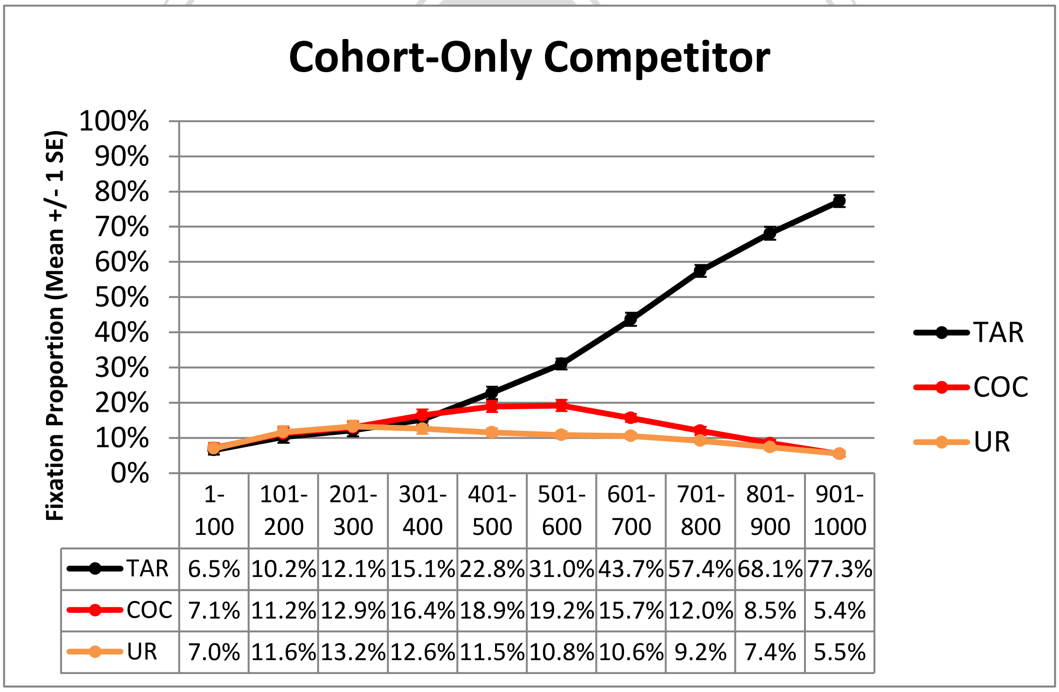
CGVC											
		Time bin (ms)									
Condition	Test	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
Cohort-Tone	*F*1(2,62)	1.66	0.47	0.45	2.18	11.81	27.89	50.15	120.60	184.30	327.80
	p	0.198	0.629	0.638	0.121	0.000	0.000	0.000	0.000	0.000	0.000
	*F*2(2,54)	1.35	0.23	0.35	1.28	14.02	39.77	102.20	109.20	118.50	138.20
	p	0.276	0.796	0.711	0.294	0.000	0.000	0.000	0.000	0.000	0.000
Cohort-Only	*F*1(2,62)	0.67	0.08	0.30	3.16	13.32	24.41	85.99	281.90	504.10	673.70
	p	0.517	0.925	0.742	0.049	0.000	0.000	0.000	0.000	0.000	0.000
	*F*2(2,54)	0.41	0.12	0.52	3.88	16.23	31.47	105.30	238.00	462.10	1172.00
	p	0.665	0.888	0.603	0.034	0.000	0.000	0.000	0.000	0.000	0.000
Note: *F*1 = 32 participants, *F*2 = 28 items											

The mean fixation proportions to target, competitor, and the unrelated distractors and standard errors were shown in Figure 15. For each time bin, a one-way ANOVA was performed for trials with cohort-tone or cohort-only competitors and their targets and distractors, followed by the post-hoc comparisons (Table 11). The mean fixations proportions to cohort-tone competitor and target was significantly different by participants and by item analyses from 501-1000-msec. The mean fixations proportions to cohort-only competitor and target was significantly different by participants and by item analyses from 401-1000-msec. When the cohort-only competitor diverged from the target was earlier than when the cohort-tone competitor diverged from the target.

The mean fixations proportions to cohort-tone competitor and unrelated distracter was significantly different by participants analyses from 301-1000-msec and by item analyses from 401-1000-msec. The mean fixations proportions to cohort-only competitor and unrelated distracter was significantly different by participants and by item analyses from 301-1000-msec. When the unrelated distractors diverged from two types of competitors was on a consistent time bin by participants analysis.



	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR	7.1%	11.6%	13.2%	15.9%	21.5%	31.3%	41.6%	51.4%	61.4%	70.7%
CTC	6.2%	10.2%	11.7%	16.0%	20.7%	21.3%	19.4%	14.5%	10.9%	8.0%
UR	6.9%	11.3%	12.4%	12.3%	11.5%	9.9%	9.3%	9.3%	8.7%	6.5%



	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR	6.5%	10.2%	12.1%	15.1%	22.8%	31.0%	43.7%	57.4%	68.1%	77.3%
COC	7.1%	11.2%	12.9%	16.4%	18.9%	19.2%	15.7%	12.0%	8.5%	5.4%
UR	7.0%	11.6%	13.2%	12.6%	11.5%	10.8%	10.6%	9.2%	7.4%	5.5%

Figure 15. Mean fixation proportions to targets, competitors, and unrelated distractors for cohort-tone and cohort-only conditions in the 1,000 msec period following acoustic target onset in Experiment 2. Each data point represents the average of fixation proportions across participants in the time bin of 100 msec and the error bars show the standard error of the data.[2]

---

[2] The calculation of mean fixation proportions in each time bin included the area of target, competitor, unrelated distractors, cross , and the remaining areas outside the above-mentioned areas in the visual display

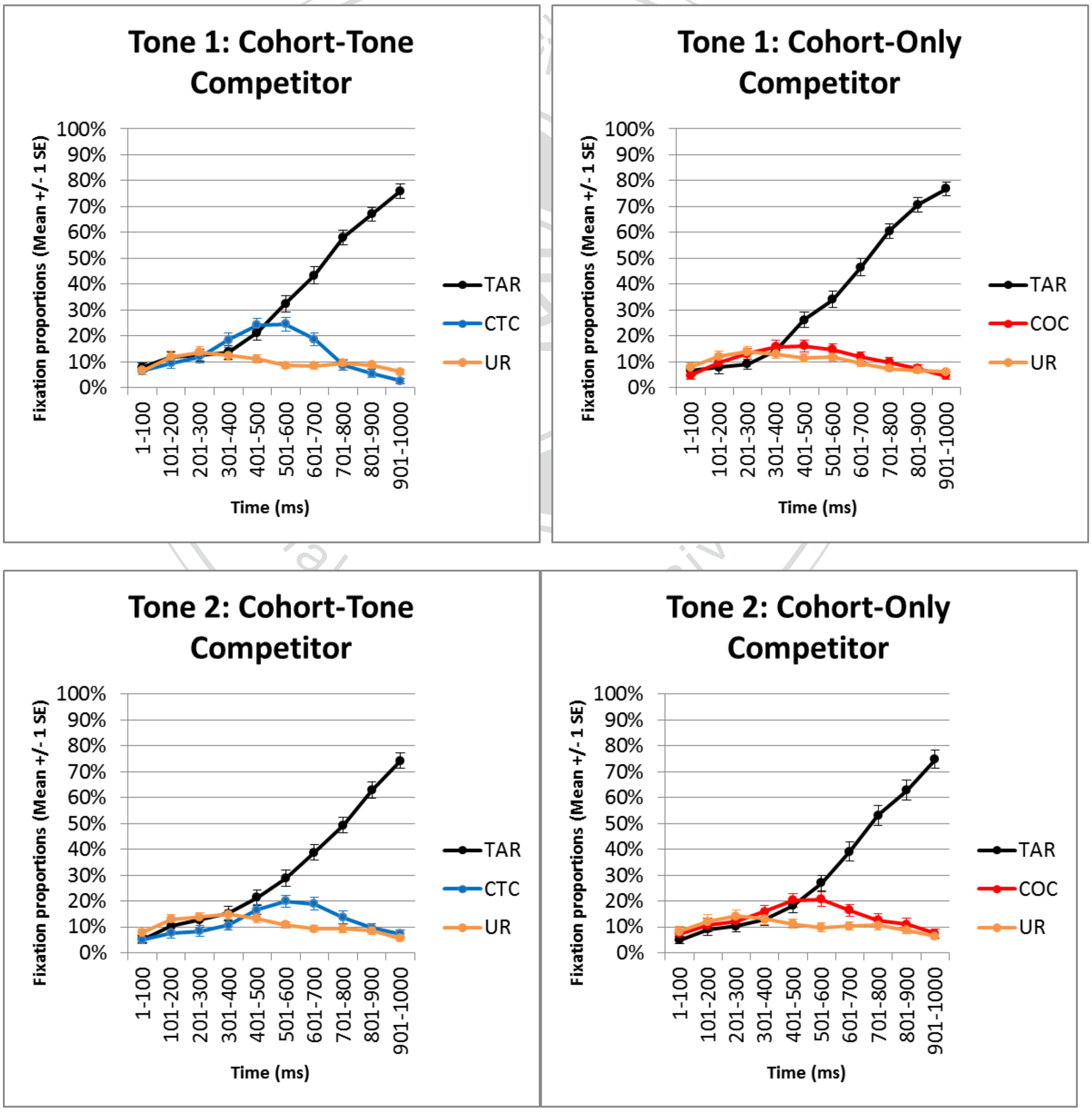
Table 11. Analyses of variance by participant and item comparing mean fixation proportions to competitors with those of the target and unrelated distractors from 1 msec to 1000 msec after acoustic target onset in Experiment 2

		Time bin (ms)									
Condition	*Test*	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR - CTC	z_1	1.13	1.30	1.36	< 1	< 1	6.07	10.40	18.23	23.24	30.35
		p =.778	p =.583	p =.517	p =1	p =1	p < .001	p < .001	p < .001	p < .001	p < .001
	z_2	0.93	1.01	1.08	< 1	< 1	5.94	12.31	15.70	21.86	27.16
		p =1	p =.932	p =.835	p =1	p =1	p < .001	p < .001	p < .001	p < .001	p < .001
TAR -COC	z_1	-0.81	-0.78	-0.62	< 1	2.62	6.82	15.70	23.84	30.17	40.45
		p =1	p =1	p =1	p =.987	p =.026	p < .001	p < .001	p < .001	p < .001	p < .001
	z_2	-0.66	-0.97	-0.73	< 1	2.90	6.67	15.20	21.86	33.64	52.31
		p =1	p =.99	p =1	p =.741	p =.011	p < .001	p < .001	p < .001	p < .001	p < .001
CTC - UR	z_1	-0.82	-1.00	-0.60	3.04	6.60	6.90	4.75	2.55	1.00	< 1
		p =1	p =.952	p =1	p =.007	p < .001	p < .001	p < .001	p =.032	p =.952	p =1
	z_2	-0.679	-0.781	-0.479	2.33	5.26	6.76	5.62	2.2	< 1	< 1
		p =1	p =1	p =1	p =.059	p < .001	p < .001	p < .001	p =.084	p =1	p =1
COC - UR	z_1	0.14	-0.32	-0.21	2.93	4.94	4.83	2.86	1.46	< 1	< 1
		p =1	p =1	p =1	p =.01	p < .001	p < .001	p =.013	p =.433	p =1	p =1
	z_2	0.12	-0.40	-0.24	3.48	5.47	4.73	2.77	1.34	< 1	< 1
		p =1	p =1	p =1	p =.002	p < .001	p < .001	p =.017	p =.542	p =1	p =1

Note : $z1$ = 32 participants, $z2$ = 28 items

The mean fixations proportions to cohort-only competitor and target was significantly different by participants analyses from 401-500-msec in Tone 1[401-500-msec: $z_1 = 3.32$, p = .003]. The mean fixations proportions to cohort-only competitor and target was significantly different by participants analyses from 501-600-msec in Tone 4[501-600-msec: $z_1 = 4.21$, $p < .001$]. Lastly, the mean fixations proportions to cohort-only competitor and target was significantly different by participants analyses from 601-700-msec in Tone 2 and Tone 3 [601-700-msec :Tone 2: $z_1 = 5.47$, $p < .001$; Tone 3: $z_1 = 6.60$, $p < .001$].

The result indicated that in Tone 1, the divergent time between the curve of cohort-only competitor and target was significantly earlier than Tone 4, which was also earlier than Tone 2 and Tone 3. Figure 16 plots the fixation proportion over time in two conditions among four tones.



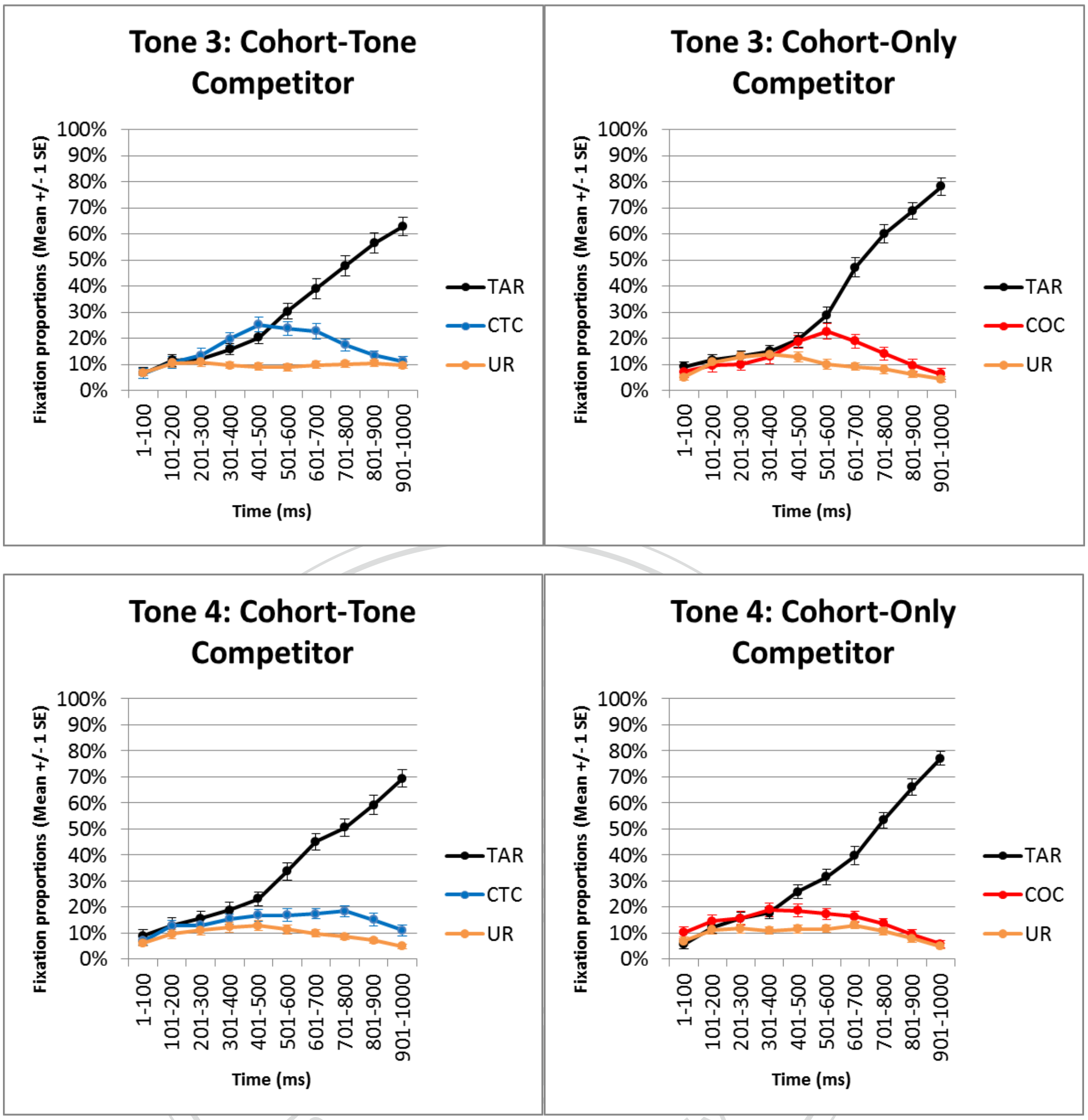


Figure 16. Mean fixation proportions to targets, competitors, and unrelated distractors for cohort-tone and cohort-only conditions in the 1,000 msec period following acoustic target onset in Experiment 2. Each data point represents the average of fixation proportions across participants in the time bin of 100 msec and the error bars show the standard error of the data. Fixation proportions to target, competitors and unrelated distractors over time for 4 tones are shown respectively.

The mean fixation proportions to target, competitor, the unrelated distractors and standard errors were shown in Figure 17 for CVC syllable structure and Figure 18 for

CGVC syllable structure. A one-way ANOVA was performed for trials with cohort-tone or cohort-only competitors and their targets and distractors for each time bin, followed by the post-hoc comparisons (CVC for Table 12 and CGVC for Table 13). In CVC, the mean fixations proportions to cohort-tone competitor and target was significantly different by participants and by item analyses from 501-1000-ms. The mean fixations proportions to cohort-only competitor and target was marginally significant different by participants analyses and significantly different by item analyses from 401-1000-ms. In CGVC, the mean fixations proportions to cohort-tone competitor and target was significantly different by participants and by item analyses from 501-1000-ms. The mean fixations proportions to cohort-only competitor and target was also significantly different by participants and by item analyses from 501-1000-ms. Table 14 showed the time when mean fixation proportion had significant difference between the competitor and its target and unrelated distractors by participants in two conditions of Experiment 2.

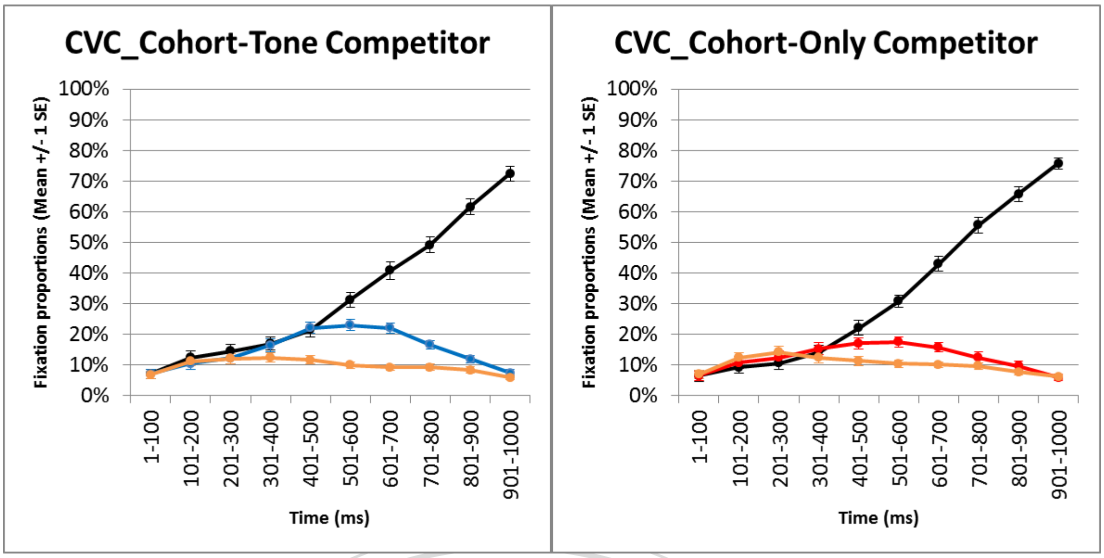


Figure 17. Mean fixation proportions to targets, competitors, and unrelated distractors in CVC syllable structure for cohort-tone and cohort-only conditions in the 1,000 msec period following acoustic target onset in Experiment 2. Each data point represents the average of fixation proportions across participants in the time bin of 100 msec and the error bars show the standard error of the data.

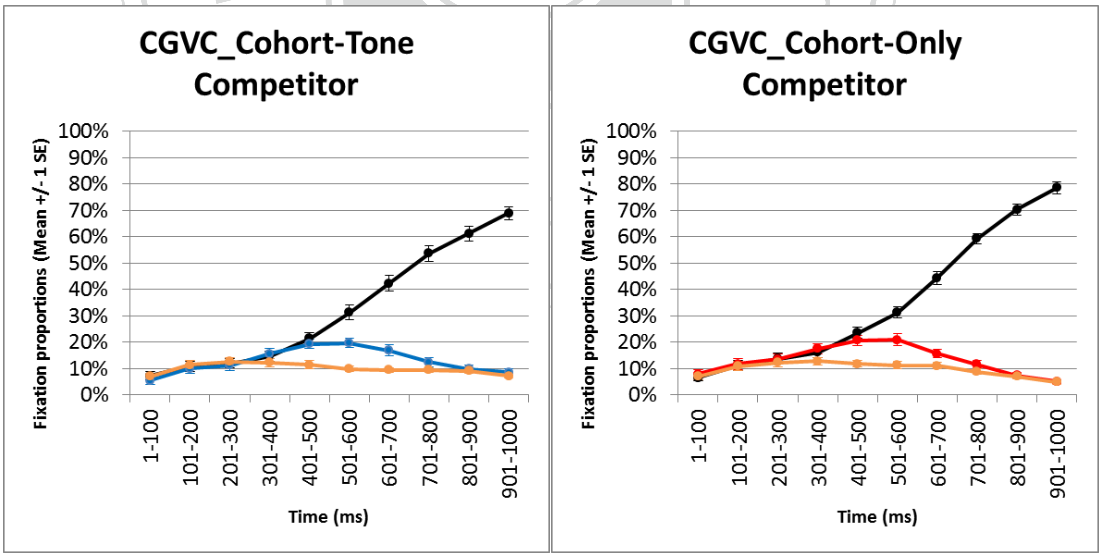


Figure 18. Mean fixation proportions to targets, competitors, and unrelated distractors in CGVC syllable structure for cohort-tone and cohort-only conditions in the 1,000 msec period following acoustic target onset in Experiment 2. Each data point represents the average of fixation proportions across participants in the time bin of 100 msec and the error bars show the standard error of the data.

Table 12. Analyses of variance by participant and item comparing mean fixation proportions to competitors with those of the target and unrelated distractors in CVC syllable structure from 1 msec to 1000 msec after acoustic target onset in Experiment 2

		CVC									
		Time bin (ms)									
Condition	*Test*	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR - CTC	*z1*	-0.03	1.17	1.15	0.26	-0.25	3.15	6.25	12.01	18.45	26.22
	p	1.000	0.725	0.746	1.000	1.000	0.005	0.000	0.000	0.000	0.000
	z2	-0.03	1.07	1.02	0.24	-0.21	3.50	7.30	10.11	18.49	32.24
	p	1.000	0.859	0.922	1.000	1.000	0.001	0.000	0.000	0.000	0.000
TAR - COC	*z1*	0.06	-0.82	-0.86	-0.52	2.18	6.21	10.50	14.30	20.34	32.98
	p	1.000	1.000	1.000	1.000	0.088	0.000	0.000	0.000	0.000	0.000
	z2	0.07	-1.08	-1.26	-0.82	3.04	5.23	9.77	13.32	22.43	33.76
	p	1.000	0.835	0.621	1.000	0.007	0.000	0.000	0.000	0.000	0.000
CTC - UR	*z1*	0.33	-0.44	0.08	1.89	4.31	4.87	4.30	2.75	1.33	0.63
	p	1.000	1.000	1.000	0.177	0.000	0.000	0.000	0.018	0.554	1.000
	z2	0.28	-0.40	0.07	1.70	3.63	5.42	5.03	2.31	1.33	0.78
	p	1.000	1.000	1.000	0.266	0.001	0.000	0.000	0.062	0.552	1.000
COC - UR	*z1*	-0.37	-0.82	-0.85	1.32	2.49	3.32	2.17	0.91	0.73	-0.13
	p	1.000	1.000	1.000	0.565	0.038	0.003	0.089	1.000	1.000	1.000
	z2	-0.46	-1.08	-1.24	2.09	3.47	2.80	2.02	0.84	0.80	-0.13
	p	1.000	0.843	0.642	0.109	0.002	0.016	0.130	1.000	1.000	1.000

Note : *z* 1 = 32 participants, *z* 2 = 28 items

Table 13. Analyses of variance by participant and item comparing mean fixation proportions to competitors with those of the target and unrelated distractors in CGVC syllable structure from 1 msec to 1000 msec after acoustic target onset in Experiment 2

		CGVC									
		Time bin (ms)									
Condition	*Test*	1-100	101-200	201-300	301-400	401-500	501-600	601-700	701-800	801-900	901-1000
TAR - CTC	*z1*	1.66	0.54	0.54	-0.55	1.00	4.12	7.53	13.18	16.77	22.27
	p	0.293	1.000	1.000	1.000	0.952	0.000	0.000	0.000	0.000	0.000
	z2	1.53	0.39	0.48	-0.43	1.11	5.02	10.98	12.81	13.73	14.77
	p	0.382	1.000	1.000	1.000	0.797	0.000	0.000	0.000	0.000	0.000
TAR - COC	*z1*	-1.17	-0.26	0.11	-0.74	1.18	3.67	10.61	20.27	27.88	32.27
	p	0.730	1.000	1.000	1.000	0.710	0.001	0.000	0.000	0.000	0.000
	z2	-0.94	-0.33	0.15	-0.84	1.33	4.25	11.99	19.02	27.26	43.47
	p	1.000	1.000	1.000	1.000	0.547	0.000	0.000	0.000	0.000	0.000
CTC - UR	*z1*	-1.55	-0.98	-0.97	2.05	3.69	3.46	2.16	0.92	0.25	0.51
	p	0.364	0.982	1.000	0.120	0.001	0.002	0.093	1.000	1.000	1.000
	z2	-1.43	-0.70	-0.86	1.61	4.10	4.22	3.15	0.90	0.21	0.34
	p	0.460	1.000	1.000	0.324	0.000	0.000	0.005	1.000	1.000	1.000
COC - UR	*z1*	0.69	0.40	0.62	2.49	3.83	3.43	1.68	1.20	0.12	0.06
	p	1.000	1.000	1.000	0.038	0.000	0.002	0.279	0.687	1.000	1.000
	z2	0.56	0.50	0.83	2.81	4.32	3.98	1.90	1.13	0.12	0.08
	p	1.000	1.000	1.000	0.015	0.000	0.000	0.173	0.777	1.000	1.000

Note : *z* 1 = 32 participants, *z* 2 = 28 items

Table 14. The time when mean fixation proportion had significant difference between TAR-CTC, TAR-CC, CTC-UR and COC-UR by participants in Experiment 2

	Experiment 2			
	TAR-CTC	TAR-COC	CTC-UR	COC-UR
All	501-1000	401-1000	301-800	301-700
CVC	501-1000	401-1000	401-800	401-600
CGVC	501-1000	501-1000	401-600	301-600
Tone 1	601-1000	401-1000	401-700	--
Tone 2	501-1000	601-1000	501-700	401-600
Tone 3	601-1000	601-1000	301-700	501-600
Tone 4	501-1000	501-1000	701-800	301-400

Note. -- no significant differences

4.4 Discussion

The result of Experiment 2 indicates that the tonal information had influence on processing the initial phonemes of Chinese characters. The fixation proportions of cohort-tone competitor were higher than that of cohort-only competitor. In addition, the time when the curve of cohort-tone diverged from target was later than the curve divergence between cohort-only competitor and target. In CVC syllable structure, when the cohort-tone competitor diverged from target was also later than when cohort-only competitor diverged from target. Lastly, the curve of target diverged from cohort-only competitor was at about 400ms, indicating the time when the tonal information can start to affect the lexical processing.

Comparing to the cohort-only competitors, the cohort-tone competitors had higher fixation probability and needed more time to be discriminated from target. This result is also consistent with the result of CVC. In CVC syllable structure, only when competitor and target shared the same vowel could tonal information be influential. It could be the evidence that tone starts to influence the processing since the initial phonemes presents. The processing of tone might be inconsistent with the studies of showing late tonal processing (Cutler & Chen, 1997; Ye & Connine, 1999). The relatively early impact of tone on the spoken character processing could because of the auditory features such as onset F0 height and onset F0 contour of lexical tone.

When the unrelated distractors diverged from two types of competitors were in a similar time interval. It might because that the tonal and segmental difference between two competitors and its unrelated distractors cause the difficulty in distinguishing whether the two competitors share the same tonal information with auditory target. Whether cohort-tone competitor and cohort-only competitor shared the same tonal information with the auditory target or not could not result in different time when two types of competitors diverged from its unrelated distractors. Although the cohort-tone competitor shared the same tone with auditory target, when its curve diverged from unrelated distractor did not later than the time when cohort-only competitor diverged from unrelated distractor. Thus, tonal information makes no impact on the discrimination from competitor and unrelated distractors under two conditions. As for target and two competitors, which shared the same initial segmental structure, the tonal disparity could influence the spoken character processing and result in the divergence at different time point. The comparable time when two competitors diverged from its unrelated distractors suggests the relatively weak effect of tone since the great disparity between the auditory target and the unrelated distractors. This result could also indicate that tone might not affect spoken character processing independently.

According to the result, the curves of two types of competitors were similar in CGVC syllable structure but dissimilar with each other in CVC syllable structure. There are several possible reasons to explain the above-mentioned result. Firstly, it could result from the integral role in lexical processing with vowel. Only when the competitors shared the initial CV instead of initial CG with target could tonal information have impact on the processing. Tonal information seems affect processing when it integrates with vowel. However, this claim needs to be further examined. The salience of vowel makes participants rely more segmental information than tonal information; consequently, they discriminated the competitor and the target as soon as the vowel presents. Secondly, it is possible that the tonal information needs more time to be processed. Therefore, in CVC syllable structure, tone seems to affect the processing when the initial two segments presents. Thirdly, in CGVC syllable structure, the duration of glide is so short that tonal information became effective until the vowel showed. If the duration of glide became longer, perhaps it is possible that the effect of tonal information could be shown.

When the curves of cohort-only competitor diverged from that of target was inconsistent among four tones. The first divergent time between the curves of cohort-only competitor and target among four tones was Tone 1, which then followed by Tone 4. The last was Tone 2 and Tone 3. This order was consistent with the findings of Lai and Zhang (2008) with the gating paradigm, showing the four tones' order of Isolation Point, which means the size of the segment needed to be correctly identify the stimulus without further changes. However, the present study did not manipulate the onset acoustic detail to show its effect on lexical processing as Lai and Zhang (2008) did. In the present study, there was no tonal effect when the competitors shared CG with target in CGVC syllable structure. This result could imply that the tonal information could not affect lexical processing as early as the onset presents. Therefore, the acoustic features of onset may occur too early to have impact on lexical processing.

Chapter 5

General Discussion

This thesis examined the time course of accessing tonal information of Chinese spoken word recognition. In Experiment 1, two types of competitors differed in sharing the same tone or segmental structure with targets. The results showed that starting at about 600 msec, the fixation proportions of the targets were statistically distinguishable from segmental competitors, which had different tone from target. However, there was no significant difference between tonal competitor and unrelated distractors. Experiment 2 manipulated two types of cohort competitors differed in sharing the same tone or not with targets. Comparing to Experiment 1, the results provided the evidence of an earlier time around 400 msec when the tonal information could be effective during processing spoken characters.

5.1 The relatively early effect of tonal information

It seems that tonal information could affect spoken character processing when the initial two segments present. Not until the last consonants shows could tone has impact on lexical processing. The result is inconsistent with the studies that tone is accessed in the later time of lexical processing (Cutler & Chen, 1997) . Cutler and Chen (1997) suggested that the prosodic information became usable to process relatively slowly. They claimed that tone may become usable only when more of the vowel that carries it is available than is needed for identification of the vowel itself. Ye and Connine (1999) also suggested that the perceptual disadvantage prevents an early role during lexical processing. Lee (2007) maintained that tonal information could not prevent the priming between the tone minimal pairs in the early phase of lexical activation by the priming experiment. He suggested that there was no prosodic influence at early time during spoken processing. In contrast to the findings of previous studies, the present study has shown relatively early tonal influence on spoken character processing when the initial two segments are processed.

The relatively early effect of tonal information is consistent with previous studies which suggested that tone and segment would be accessed at the comparable time point (Malins & Joanisse, 2010; Schirmer et al., 2005; Zhao et al., 2011). Why tonal information contributes to the lexical processing relatively early before the whole syllable unfolds? The reasons are as follows. Firstly, lexical tone spans over the segments from the onset with pitch movement. The f0 height is different among the four tones since the onset of the character. Tone 2 and Tone 3 have a lower f0 height while Tone 1 and Tone 4 have a higher f0 height (Chang, 2010; Jongman et al., 2006; Lee, 2009). It is likely that the different onset might start to have an effect on the lexical processing. In addition, some studies have shown that Chinese listeners use f0 height to discriminate low-onset tones, such as Tone 2 and 3from high-onset tones, such as Tone 1 and Tone 4 (Chang, 2010; Lee, 2009). Chinese tonal contrast was significant when disambiguating lexical meanings; thus, the different onset f0 height among four tones might be the evidence that tonal information affected lexical processing relatively early.

Another reason is relate to tone's phonemic function in lexical processing. Tone plays a significant role in Chinese in helping listeners to identify the meaning of words. Because there was a high degree of segmental homophony in Chinese, tone is important in disambiguate lexical meaning. Chinese listeners might heavily rely on tonal information when facing a small number of syllables, which share the same segmental structure. To take the word /ma/ for example, it has distinctive meanings with different tones. In this way, tone actively contrasts the meanings during lexical processing. Due to the influential characteristic of lexical tone, tone would have impact on lexical processing relatively early.

5.2 The dependence of tonal effect

Our results are inconsistent with those of previous studies, which have shown that tones are processed separately as segments (Cutler & Chen, 1997; Liu & Samuel, 2007; Ye & Connine, 1999). In Experiment 1, there was no significant difference between the fixations proportions of tonal competitors and unrelated distractors. The null effect could be due to the tonal competitors used in Experiment 1 had different initial segments with the targets. In addition, the divergent time of the curves between the tonal competitor and target was similar among four tones. In Experiment 2, the time when cohort-tone and cohort competitors diverged from unrelated distractors was the same. Taken the results together, we suspect that tone might not affect lexical processing independently. With the segmental information coming one after another during the processing, the effect of tonal information seems to be weakened. If tonal information influences lexical process independently, lexical tone should not be affected by segmental information.

One possible explanation for no independent role of early tonal processing needs to take the tone's auditory features into account. First, belongs to the suprasegmental dimension, tone extends and spans beyond segments. Therefore, tone could not be separated individually like segments when processing the tonal information. Mandarin tones differ in terms of overall duration(Jongman et al., 2006). It appears that only be attached on segments could tone take effects on lexical process. Second, the nature and features of Mandarin tone could cause this inseparable attribute in terms of lexical processing. Generally, Mandarin tones could be discriminated by the relative F0 height (Tone 1 and 4: high; Tone 2 and 3: low), F0 movement (e.g. the rising or falling contour), and duration (Tone 2 and 3 tend to be the longest, Tone 4 the shortest) (Jongman et al., 2006; Lin & Repp, 1989). The three features category the different patterns of the four tones. The continuum characteristics of F0 contour, F0 height, and duration also prove the continuous feature of the tone. The acoustic features of tone imply that tone could not affect the spoken character processing independently.

Previous studies have tried to provide some insights to the existing models such as the TRACE, the Cohort, the NAM, and the Shortlist/Merge model (Malins & Joanisse, 2010; Zhao et al., 2011). Malins and Joanisse (2010) claimed that their results fit well with the view of the TRACE model. Additionally, both Zhao et al. (2011) and Malins and Joanisse (2010) suggested that a new “toneme” node could be incorporated into the modified TRACE model, which was firstly proposed by Ye and Connine (1999). The reason to add a new node was because they consider tone was a separate representation like phoneme during speech processing. However, this separate characteristic of tone was contradictory with the present study. During spoken processing, tonal information could not take effect independently. In this way, the present study is not consistent with the modified TRACE model.

5.3 Tonal processing in visual world paradigm

In the present study, the different time points when the four tones could be distinguished suggest that the phonetic detail could be shown in the visual world paradigm. This paradigm has been broadly used in spoken language comprehension such as the sentence processing, lexical meaning processing, phonetic processing, or prosodic processing (Dahan, Magnuson, & Tanenhaus, 2001; Malins & Joanisse, 2010; Bob McMurray et al., 2002; M. Tanenhaus, Magnuson, Dahan, & Chambers, 2000). In the current study, the results show not only the prosodic processing, but also the effect of the tonal features. Both Experiment 1 and 2 suggested different time points when the curve of target diverged from that of competitor among four tones.

5.4 Suggestions for future research

There is a continuing need for considering the contextual effect on the spoken character processing. The context includes another character in a word, idioms, or sentences. According to Tyler (1984), context should have a major effect on the rate at which the pool of word candidates is reduced over time. First, lexical processing might be affected by the tone-sandhi mechanism if the stimulus was in a word context. Second, the spoken character processing might also be influenced by idiom or sentence context. In addition, previous studies have shown that when the stimulus is presented in context, tonal information would become more important than the stimulus is in isolation (Liu & Samuel, 2007; Ye & Connine, 1999). Another reason to consider the contextual effect might because the result may be closer to a typical speech context. Lastly, perhaps future research could examine the contrast of Mandarin four tones more detailed and the number of four tones should be balanced.

Reference

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language, 38*(4), 419-439. doi: 10.1006/jmla.1997.2558

Carroll, D. W. (2008). *Psychology of language*: Thomson/Wadsworth.

Chang, C.-Y. (2010). *Dialect Differences in the Production and perception of Mandarin Chinese Tones.* Ohio State University.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review, 81*(4), 348-374. doi: 10.1037/h0036656

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84-107. doi: 10.1016/0010-0285(74)90005-x

Cutler, A. (1986). Forbear is a homophone: lexical prosody does not constrain lexical access. [Article]. *Language & Speech, 29*(3), 201-220.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception&Psychophysics, 59*(2), 165-179.

Cutler, A., & Clifton, C. (1984). The use of prosodic information in word recognition. In H. B. D. G. Bouwhuis (Ed.), *Control of language processes : (proceedings of the tenth international symposium on attention and performance* (pp. 183~196). Venlo, the Netherlands,: London : Lawrence Erlbaum associates.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the Comprehension of Spoken Language: A Literature Review. [Article]. *Language & Speech, 40*(2), 141-201.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time Course of Frequency Effects in Spoken-Word Recognition: Evidence from Eye Movements. *Cognitive Psychology, 42*(4), 317-367. doi: 10.1006/cogp.2001.0750

Epstein, W. (1961). The Influence of Syntactical Structure on Learning. *The American Journal of Psychology, 74*(1), 80-85.

Foss, D. J., & Hakes, D. T. (1978). *Psycholinguistics: an introduction to the psychology of language*: Prentice-Hall.

Frauenfelder, U. H., & Tyler, L. K. (1987). *Spoken word recognition*: MIT Press.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. [Article]. *Language & Speech, 5*(4), 171-189.

Gandour, J. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics, 12*(2), 235-261.

Garman, M. (1990). *Psycholinguistics*: Cambridge University Press.

Gleason, J. B., & Ratner, N. B. (1998). *Psycholinguistics*: Harcourt Brace College Publishers.

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition, 25*(1–2), 135-155. doi: 10.1016/0010-0277(87)90007-2

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460-482. doi: 10.1016/j.jml.2007.02.001

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and Production of Mandarin Chinese Tones. In L. H. T. E. Bates, and O.J.L Tzeng (Ed.), *Handbook of Chinese Psycholinguistics*: Cambridge University Press.

Jusczyk, P. W., & Luce, P. A. (2002). Speech perception and spoken word recognition: past and present. *Ear and Hearing, 23*(1), 2-40. doi: citeulike-article-id:6136467

Ladefoged, P. (2005). *Vowels and consonants: an introduction to the sounds of languages*: Blackwell Pub.

Lai, Y., & Zhang, J. (2008). Mandarin Lexical Tone Recognition: The Gating Paradigm. *Kansas Working Papers in Linguistics, 30*.

Lass, N. J. (1976). *Contemporary issues in experimental phonetics*: Academic Press.

Lee, C. Y. (2007). Does Horse Activate Mother?Processing Lexical Tone in Form Priming. *Language and Speech, 50*(1), 101-123.

Lee, C. Y. (2009). Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study. [Article]. *Journal of the Acoustical Society of America, 125*(2), 1125-1137. doi: 10.1121/1.3050322

Leena, M. (2012). Prosodic Features for Speaker Recognition. In A. Neustein & H. A. A. Patil (Eds.), *Forensic Speaker Recognition* (pp. 365-388): Springer New York.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358-368.

Lin, H.-B., & Repp, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech, 32*, 25--45.

Liu, S., & Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes,* *22*(4), 566-594. doi: 10.1080/01690960600989600

Lyn, F. (1987). Structure in auditory word recognition. *Cognition, 25*(1–2), 157-187. doi: 10.1016/0010-0277(87)90008-4

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 62*(4), 407-420. doi: 10.1016/j.jml.2010.02.004

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1-71. doi: 10.1016/0010-0277(80)90015-3

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63. doi: 10.1016/0010-0285(78)90018-x

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86. doi: 10.1016/0010-0285(86)90015-0

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic bulletin & review, 15*(6), 1064-1071.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33-B42. doi: 10.1016/s0010-0277(02)00157-9

Myers, T., Laver, J., & Anderson, J. M. (1981). *The Cognitive representation of speech*: North-Holland Pub. Co.

Roca, I., & Johnson, W. (1999). *A course in phonology*: Blackwell Publishers.

Schirmer, A., Tang, S., Penney, T. B., Gunter, T. C., & Chen, H. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *J Cogn Neurosci, 17*(1), 1-12.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-303.

Studdert-Kennedy, M. (1976). speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243-293). New York: Academic Press.

Taft, M., & Chen, H.-C. (1992). Judging Homophony in Chinese: The Influence of Tones. In C. Hsuan-Chih & J. L. T. Ovid (Eds.), *Advances in Psychology* (Vol. Volume 90, pp. 151-172): North-Holland.

Tanenhaus, M., Magnuson, J., Dahan, D., & Chambers, C. (2000). Eye Movements and Lexical Access in Spoken-Language Comprehension: Evaluating a Linking Hypothesis between Fixations and Linguistic Processing. *Journal of Psycholinguistic Research, 29*(6), 557-580. doi: citeulike-article-id:789099

Tanenhaus, M. K. (2007). Eye movements and spoken language processing. In M. H.

F. Roger P.G. Van Gompel, Wayne S. Murray and Robin L. Hill (Ed.), *Eye movements: a window on mind and brain*: Elsevier.

Tsang, Y.-K., Jia, S., Huang, J., & Chen, H.-C. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height. *Neuroscience Letters, 487*(3), 268-272. doi: 10.1016/j.neulet.2010.10.035

Tyler, L. (1984). The structure of the initial cohort: Evidence from gating. *Attention, Perception, & Psychophysics, 36*(5), 417-427. doi: 10.3758/bf03207496

Uli H, F., & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition, 25*(1-2), 1-20. doi: 10.1016/0010-0277(87)90002-3

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A, 58*(2), 251-273. doi: 10.1080/02724980343000927

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception And Psychophysics, 50*(4), 351-360.

William D, & Marslen-Wilson. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1–2), 71-102. doi: 10.1016/0010-0277(87)90005-9

Ye, Y., & Connine, C. M. (1999). Processing Spoken Chinese: The Role of Tone Information. *Language and Cognitive Processes, 14*, 609-630. doi: 10.1080/016909699386202

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia, 49*(7), 1761-1770. doi: 10.1016/j.neuropsychologia.2011.02.054

吳聲弘. (2012). *音韻組成與中文口語感知*. 碩士論文, 國立交通大學, 新竹.

Appendixes

A. Experiment materials of Experiment 1

Set	Type	Character	IPA	Tone	Frequency	Stroke	Homophone number	Syllable Structure
	TAR	摸	/mɔ 1/	1	65	14	2	CV
	TC	挖	/wa1/	1	62	9	6	CV
1	SC	抹	/mɔ 3/	3	27	8	1	CV
	UR	怒	/nu 4/	4	61	9	1	CV
	UR	菊	/tɕy 2/	2	14	12	7	CV
	TAR	奢	/ʂɤ 1/	1	17	11	1	CV
	TC	禿	/tʰu 1/	1	14	7	2	CV
2	SC	捨	/ʂɤ 3/	3	45	11	2	CV
	UR	握	/wɔ 4/	4	143	12	7	CV
	UR	逆	/ni 4/	4	28	10	7	CV
	TAR	紮	/tsa 1/	1	18	11	2	CV
	TC	喝	/xɤ 1/	1	154	12	3	CV
3	SC	砸	/tsa 2/	2	10	10	3	CV
	UR	庫	/kʰu 4/	4	200	10	3	CV
	UR	努	/nu 3/	3	162	7	2	CV
	TAR	粗	/tsʰu 1/	1	61	11	1	CV
	TC	踢	/tʰi 1/	1	21	15	3	CV
4	SC	促	/tsʰu 4/	4	153	9	8	CV
	UR	娶	/tɕʰy 3/	3	20	11	4	CV
	UR	麻	/ma 2/	2	119	11	4	CV
	TAR	遮	/tʂɤ 1/	1	26	15	1	CV
	TC	潑	/pʰɔ 1/	1	45	15	3	CV
5	SC	浙	/tʂɤ 4/	4	9	10	4	CV
	UR	魯	/lu 3/	3	84	15	4	CV
	UR	擠	/tɕi 3/	3	65	17	8	CV

6	TAR	拋	/pʰɑw 1/	1	29	8	1	CVC
	TC	暈	/ɥən 1/	1	14	13	1	CVC
	SC	跑	/pʰɑw 3/	3	190	12	1	CVC
	UR	綜	/tsoŋ 4/	4	69	14	4	CVC
	UR	慶	/tɕʰiŋ 4/	4	151	15	4	CVC
7	TAR	猜	/tsʰaj 1/	1	40	11	1	CVC
	TC	兵	/piŋ 1/	1	159	7	2	CVC
	SC	采	/tsʰaj 3/	3	47	8	6	CVC
	UR	鬥	/tow 4/	4	104	10	7	CVC
	UR	柔	/ʐow 2/	2	58	9	3	CVC
8	TAR	圈	/tɕʰɥɛn 1/	1	99	11	1	CGVC
	TC	吹	/tʂʰwej 1/	1	80	7	2	CGVC
	SC	勸	/tɕʰɥɛn 4/	4	48	20	2	CGVC
	UR	錶	/pjɑw 3/	3	25	16	3	CGVC
	UR	涼	/ljɑŋ 2/	2	65	11	7	CGVC
9	TAR	鼻	/pi 2/	2	54	14	1	CV
	TC	佛	/fɔ 2/	2	152	7	1	CV
	SC	逼	/pi 1/	1	50	13	2	CV
	UR	埔	/pʰu 3/	3	51	10	7	CV
	UR	瓦	/wa 3/	3	99	5	1	CV
10	TAR	殼	/kʰɤ 2/	2	51	12	2	CV
	TC	俗	/su 2/	2	131	9	1	CV
	SC	渴	/kʰɤ 3/	3	25	12	3	CV
	UR	搭	/ta 1/	1	113	13	4	CV
	UR	訝	/ja 4/	4	24	11	3	CV
11	TAR	婆	/pʰɔ 2/	2	80	11	1	CV
	TC	拔	/pa 2/	2	75	8	2	CV
	SC	頗	/pʰɔ 3/	3	124	14	1	CV
	UR	呼	/xu 1/	1	161	8	5	CV
	UR	督	/tu 1/	1	74	13	3	CV
12	TAR	徐	/ɕy 2/	2	79	10	1	CV
	TC	奴	/nu 2/	2	23	5	2	CV
	SC	虛	/ɕy 1/	1	98	12	8	CV
	UR	葛	/kɤ 3/	3	44	13	2	CV
	UR	臥	/wɔ 4/	4	28	8	7	CV

13	TAR	盆	/pʰən 2/	2	33	9	1	CVC
	TC	肥	/fej 2/	2	62	8	1	CVC
	SC	噴	/pʰən 1/	1	49	15	1	CVC
	UR	頌	/soŋ 4/	4	11	13	5	CVC
	UR	抖	/tow 3/	3	19	7	4	CVC
14	TAR	柴	/tʂʰaj 2/	2	28	10	2	CVC
	TC	熟	/ʂow 2/	2	156	15	2	CVC
	SC	拆	/tʂʰaj 1/	1	60	8	2	CVC
	UR	忍	/ʐən 3/	3	118	7	3	CVC
	UR	挺	/tʰiŋ 3/	3	38	10	4	CVC
15	TAR	描	/mjɑw 2/	2	81	12	3	CGVC
	TC	熊	/ɕɥoŋ 2/	2	106	14	2	CGVC
	SC	廟	/mjɑw 4/	4	88	15	2	CGVC
	UR	嘴	/tswej 3/	3	83	15	1	CGVC
	UR	吞	/tʰwən 1/	1	40	7	1	CGVC
16	TAR	扯	/tʂʰɤ 3/	3	23	7	1	CV
	TC	堵	/tu 3/	3	19	11	5	CV
	SC	徹	/tʂʰɤ 4/	4	61	14	5	CV
	UR	鴉	/ja 1/	1	16	15	8	CV
	UR	窩	/wɔ 1/	1	25	14	3	CV
17	TAR	塔	/tʰa 3/	3	68	13	1	CV
	TC	虎	/xu 3/	3	112	8	4	CV
	SC	踏	/tʰa 4/	4	81	15	9	CV
	UR	滴	/ti 1/	1	46	14	3	CV
	UR	膜	/mɔ 2/	2	34	15	7	CV
18	TAR	掃	/sɑw 3/	3	69	11	2	CVC
	TC	董	/toŋ 3/	3	103	13	2	CVC
	SC	艘	/sɑw1/	1	21	16	3	CVC
	UR	佩	/pʰej 4/	4	51	8	4	CVC
	UR	賊	/tsej 2/	2	21	13	1	CVC
19	TAR	窄	/tʂaj 3/	3	29	10	1	CVC
	TC	鼎	/tiŋ 3/	3	25	13	3	CVC
	SC	摘	/tʂaj 1/	1	31	14	2	CVC
	UR	謀	/mow 2/	2	57	16	4	CVC
	UR	憤	/fən 4/	4	46	15	6	CVC

20	TAR	粉	/fən 3/	3	69	10	1	CVC
	TC	榜	/pɑŋ 3/	3	33	14	3	CVC
	SC	焚	/fən 2/	2	15	12	2	CVC
	UR	湊	/tsʰow 4/	4	21	12	1	CVC
	UR	淚	/lej 4/	4	56	11	4	CVC
21	TAR	腿	/tʰwej 3/	3	59	14	1	CGVC
	TC	倆	/ljɑŋ 3/	3	29	10	2	CGVC
	SC	頹	/tʰwej 2/	2	11	16	1	CGVC
	UR	捐	/tɕɥɛn 1/	1	64	10	5	CGVC
	UR	迅	/ɕɥən 4/	4	56	7	7	CGVC
22	TAR	柳	/ljow 3/	3	40	9	1	CGVC
	TC	拐	/kwaj 3/	3	11	8	2	CGVC
	SC	溜	/ljow 1/	1	29	13	2	CGVC
	UR	旋	/ɕɥɛn 2/	2	80	11	7	CGVC
	UR	窮	/tɕʰɥəŋ 2/	2	80	15	4	CGVC
23	TAR	帕	/pʰa 4/	4	21	8	2	CV
	TC	兔	/tʰu 4/	4	23	8	3	CV
	SC	爬	/pʰa 2/	2	61	8	5	CV
	UR	米	/mi 3/	3	138	6	4	CV
	UR	剝	/pɔ 1/	1	28	10	7	CV
24	TAR	聘	/pʰin 4/	4	62	13	1	CVC
	TC	菜	/tsʰaj 4/	4	152	12	3	CVC
	SC	拼	/pʰin1/	1	31	9	2	CVC
	UR	桶	/tʰoŋ 3/	3	26	11	3	CVC
	UR	盟	/məŋ 2/	2	90	13	8	CVC
25	TAR	鬧	/nɑw 4/	4	95	15	2	CVC
	TC	孟	/məŋ 4/	4	67	8	2	CVC
	SC	惱	/nɑw 3/	3	40	12	3	CVC
	UR	陪	/pʰej 2/	2	65	11	4	CVC
	UR	淵	/ɥɛn 1/	1	23	11	4	CVC
26	TAR	趟	/tʰɑŋ 4/	4	22	15	2	CVC
	TC	奏	/tsow 4/	4	121	9	3	CVC
	SC	湯	/tʰɑŋ 1/	1	50	12	1	CVC
	UR	墾	/kʰən 3/	3	22	16	4	CVC
	UR	痕	/xən 2/	2	33	11	1	CVC

27	TAR	弄	/noŋ 4/	4	81	7	1	CVC
	TC	脈	/maj 4/	4	70	10	4	CVC
	SC	濃	/noŋ 2/	2	89	16	5	CVC
	UR	寢	/tɕʰin 3/	3	15	14	1	CVC
	UR	沈	/ʂən 3/	3	107	7	4	CVC
28	TAR	碰	/pʰəŋ 4/	4	81	13	1	CVC
	TC	債	/tʂaj 4/	4	31	13	4	CVC
	SC	捧	/pʰəŋ 3/	3	19	11	1	CVC
	UR	欽	/tɕʰin 1/	1	22	12	3	CVC
	UR	猴	/xow 2/	2	43	12	3	CVC

Note. TAR=Target; TC=Tonal Competitor; SC=Segmental Competitor; UR=Unrelated Distractor

B. Experiment materials of Experiment 2

Set	Type	Character	IPA	Tone	Frequency	Stroke	Homophone Number	Syllable Structure
	TAR	潘	/pʰan 1/	1	28	15	2	CVC
	CTC	拋	/pʰɑw 1/	1	29	8	1	CVC
1	COC	泡	/pʰɑw4/	4	52	8	4	CVC
	UR	雷	/lej 2/	2	86	13	4	CVC
	UR	董	/toŋ3/	3	103	13	2	CVC
	TAR	湯	/tʰɑŋ 1/	1	50	12	1	CVC
	CTC	胎	/tʰaj 1/	1	37	9	2	CVC
2	COC	泰	/tʰaj 4/	4	128	10	4	CVC
	UR	剖	/pʰow 3/	3	32	10	1	CVC
	UR	痕	/xən 2/	2	33	11	1	CVC
	TAR	猜	/tsʰaj 1/	1	40	11	1	CVC
	CTC	操	/tsʰɑw 1/	1	135	16	2	CVC
3	COC	曹	/tsʰɑw 2/	2	30	11	3	CVC
	UR	炳	/piŋ 3/	3	10	9	7	CVC
	UR	孟	/məŋ 4/	4	67	8	2	CVC
	TAR	堆	/twej 1/	1	81	11	1	CGVC
	CTC	端	/twan 1/	1	158	14	1	CGVC
4	COC	鍛	/twan 4/	4	6	17	4	CGVC
	UR	窘	/tɕɥoŋ 3/	3	11	12	3	CGVC
	UR	瓊	/tɕʰɥəŋ 2/	2	16	19	4	CGVC
	TAR	荒	/xwɑŋ 1/	1	69	10	2	CGVC
	CTC	昏	/xwən 1/	1	46	8	3	CGVC
5	COC	魂	/xwən 2/	2	41	14	3	CGVC
	UR	謬	/mjow 4/	4	12	18	2	CGVC
	UR	紐	/njow 3/	3	76	10	4	CGVC
	TAR	虧	/kʰwej 1/	1	41	17	3	CGVC
	CTC	框	/kʰwɑŋ 1/	1	17	10	3	CGVC
6	COC	礦	/kʰwɑŋ 4/	4	47	20	6	CGVC
	UR	捲	/tɕɥɛn 3/	3	33	11	3	CGVC
	UR	黏	/njɛn 2/	2	21	17	4	CGVC

	TAR	吹	/tʂʰwej 1/	1	80	7	2	CGVC
	CTC	窗	/tʂʰwɑŋ 1/	1	102	12	3	CGVC
7	COC	床	/tʂʰwɑŋ 2/	2	110	7	2	CGVC
	UR	唸	/njɛn 4/	4	34	11	3	CGVC
	UR	淺	/tɕʰjɛn 3/	3	51	11	4	CGVC
	TAR	萊	/laj 2/	2	40	12	3	CVC
	CTC	牢	/lɑw 2	2	37	7	4	CVC
8	COC	撈	/lɑw 1/	1	13	15	2	CVC
	UR	鄧	/təŋ 4/	4	37	15	4	CVC
	UR	墾	/kʰən 3/	3	22	16	4	CVC
	TAR	逢	/fəŋ 2/	2	53	11	4	CVC
	CTC	焚	/fən 2/	2	15	12	2	CVC
9	COC	芬	/fən 1/	1	45	8	5	CVC
	UR	噪	/tsaw 4/	4	20	16	6	CVC
	UR	咬	/jaw 3/	3	38	9	4	CVC
	TAR	埋	/maj 2/	2	60	10	2	CVC
	CTC	盲	/mɑŋ 2/	2	45	8	5	CVC
10	COC	莽	/mɑŋ 3/	3	4	11	2	CVC
	UR	拼	/pʰin 1/	1	31	9	2	CVC
	UR	偷	/tʰow 1/	1	88	11	1	CVC
	TAR	袍	/pʰɑw 2/	2	13	11	5	CVC
	CTC	牌	/pʰaj 2/	2	166	12	3	CVC
11	COC	拍	/pʰaj 1/	1	167	8	1	CVC
	UR	鼎	/tiŋ 3/	3	25	13	3	CVC
	UR	弄	/noŋ 4/	4	81	7	1	CVC
	TAR	柴	/tʂʰaj 2/	2	28	10	2	CVC
	CTC	巢	/tʂʰɑw 2/	2	25	11	4	CVC
12	COC	炒	/tʂʰɑw 3/	3	35	8	2	CVC
	UR	碰	/pʰəŋ 4/	4	81	13	1	CVC
	UR	噴	/pʰən 1/	1	49	15	1	CVC
	TAR	裘	/tɕʰjow 2/	2	6	13	6	CGVC
	CTC	牆	/tɕʰjɑŋ 2/	2	141	17	3	CGVC
13	COC	搶	/tɕʰjɑŋ 3/	3	89	13	2	CGVC
	UR	桂	/kwej 4/	4	29	10	6	CGVC
	UR	催	/tsʰwej 1/	1	28	13	4	CGVC

	TAR	頹	/t^{h}wej 2/	2	11	16	1	CGVC
	CTC	豚	/t^{h}wən 2/	2	11	11	4	CGVC
14	COC	吞	/t^{h}wən 1/	1	40	7	1	CGVC
	UR	糗	/tɕ^{h}jow 3/	3	3	16	1	CGVC
	UR	朽	/ɕjow 3/	3	10	6	2	CGVC
	TAR	擺	/paj 3/	3	101	18	3	CVC
	CTC	榜	/pɑŋ 3/	3	33	14	3	CVC
15	COC	邦	/pɑŋ 1/	1	75	7	4	CVC
	UR	蚊	/wən 2/	2	34	10	6	CVC
	UR	漏	/low 4/	4	41	14	4	CVC
	TAR	郝	/xɑw 3/	3	47	10	2	CVC
	CTC	喊	/xan 3/	3	56	12	2	CVC
16	COC	憨	/xan 1/	1	5	16	3	CVC
	UR	佩	/p^{h}ej 4/	4	51	8	4	CVC
	UR	鄧	/təŋ 4/	4	37	15	4	CVC
	TAR	逮	/taj 3/	3	26	12	2	CVC
	CTC	擋	/tɑŋ 3/	3	26	16	3	CVC
17	COC	蕩	/tɑŋ 4/	4	29	16	4	CVC
	UR	盆	/p^{h}ən 2/	2	33	9	1	CVC
	UR	鉤	/kow 1/	1	8	13	5	CVC
	TAR	秒	/mjɑw 3/	3	39	9	5	CGVC
	CTC	勉	/mjɛn 3/	3	35	9	6	CGVC
18	COC	眠	/mjɛn 2/	2	34	10	3	CGVC
	UR	潰	/k^{h}wej 4/	4	34	15	5	CGVC
	UR	洶	/ɕɥoŋ 1/	1	10	9	6	CGVC
	TAR	柳	/ljow 3/	3	40	9	1	CGVC
	CTC	倆	/ljɑŋ 3/	3	29	10	2	CGVC
19	COC	諒	/ljɑŋ 4/	4	29	15	6	CGVC
	UR	孫	/swən 1/	1	117	10	2	CGVC
	UR	圈	/tɕ^{h}ɥɛn 1/	1	99	11	1	CGVC
	TAR	錶	/pjɑw 3/	3	25	16	3	CGVC
	CTC	貶	/pjɛn 3/	3	11	12	3	CGVC
20	COC	鞭	/pjɛn 1/	1	19	18	4	CGVC
	UR	葵	/k^{h}wej 2/	2	8	13	6	CGVC
	UR	跪	/kwej 4/	4	10	13	6	CGVC

	TAR	毀	/xwej 3/	3	61	13	6	CGVC
	CTC	謊	/xwɑŋ 3/	3	14	17	4	CGVC
21	COC	晃	/xwɑŋ 4/	4	24	10	1	CGVC
	UR	裙	/tɕʰɥən 2/	2	16	13	2	CGVC
	UR	丟	/tjow 1/	1	50	6	1	CGVC
	TAR	綴	/tʂwej 4/	4	10	14	4	CGVC
	CTC	賺	/tʂwan 4/	4	70	17	5	CGVC
22	COC	磚	/tʂwan 1/	1	39	16	2	CGVC
	UR	熊	/ɕɥoŋ 2/	2	106	14	2	CGVC
	UR	窮	/tɕʰɥəŋ 2/	2	80	15	4	CGVC
	TAR	棍	/kwən 4/	4	11	12	1	CGVC
	CTC	貢	/kwoŋ 4/	4	66	10	3	CGVC
23	COC	拱	/kwoŋ 3/	3	12	9	3	CGVC
	UR	瓢	/pʰjɑw 2/	2	4	16	2	CGVC
	UR	挑	/tʰjɑw 1/	1	126	9	1	CGVC
	TAR	潤	/ʐwən 4/	4	56	15	2	CGVC
	CTC	瑞	/ʐwej 4/	4	127	13	4	CGVC
24	COC	蕊	/ʐwej 3/	3	4	16	1	CGVC
	UR	糾	/tɕjow 1/	1	50	8	4	CGVC
	UR	娘	/njɑŋ 2/	2	81	10	1	CGVC
	TAR	漂	/pʰjɑw 4/	4	59	14	3	CGVC
	CTC	遍	/pʰjɛn 4/	4	138	13	3	CGVC
25	COC	篇	/pʰjɛn 1/	1	104	15	4	CGVC
	UR	迴	/xwej 2/	2	59	10	3	CGVC
	UR	嘴	/tswej 3/	3	83	15	1	CGVC
	TAR	耐	/naj 4/	4	84	9	2	CVC
	CTC	鬧	/nɑw 4/	4	95	15	2	CVC
26	COC	惱	/nɑw 3/	3	40	12	3	CVC
	UR	恆	/xəŋ 2/	2	55	9	5	CVC
	UR	疼	/tʰəŋ 2/	2	40	10	5	CVC
	TAR	燦	/tsʰan 4/	4	19	17	3	CVC
	CTC	蔡	/tsʰaj 4/	4	84	15	3	CVC
27	COC	裁	/tsʰaj 2/	2	92	12	4	CVC
	UR	挺	/tʰiŋ 3/	3	38	10	4	CVC
	UR	冰	/piŋ 1/	1	86	6	2	CVC

	TAR	趟	/tʰɑŋ 4/	4	22	15	2	CVC
	CTC	嘆	/tʰan 4/	4	47	14	5	CVC
28	COC	坦	/tʰan 3/	3	75	8	4	CVC
	UR	勒	/lej 1/	1	93	11	1	CVC
	UR	柔	/ʐow 2/	2	58	9	3	CVC

Note. TAR=Target; CTC=Cohort-Tone Competitor; COC=Cohort-Only Competitor; UR=Unrelated Distractor