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：萬依萍 博士
Advisor: Dr. I-Ping Wan

Spoken Word Recognition in Taiwan Mandarin: Evidence from Isolated Disyllabic Words

臺灣華語的口語詞彙辨識歷程：從雙音節詞來看

研究生：錢昱夫　撰
Student: Yu-Fu Chien
中華民國一百年七月
July, 2011

The members of the Committee approve the thesis of Yu-Fu Chien

defended on Spoken Word Recognition in Taiwan Mandarin:
Evidence from Isolated Disyllabic Words

I-Ping Wan

Advisor

Yow-Yu Lin

Committee member

Naomi Ogasawara

Naomi Ogasawara

Committee Member

Approved:

Huichen Chan (acting)

Kawai Chui, Director, Graduate Institute of Linguistics

誌謝

人生有很多階段，每個階段都分別有重要的使命等著我們去完成。如果把碩士班生涯當成我人生中的一個階段，那碩士論文和取得碩士學位，無疑地，是我這個階段最重要的且必須去完成的使命。

還記得三年前的我，對碩士班充滿憧憬的踏入環境優美的校園，編織著對未來的美夢。但漸漸地，當我學得更多，了解得更多，思考得更多，才知道學術之路不是如我先前所想的那麼平順，這條路是佈滿荊棘、阻礙的。要達成目標，除了需要一顆充滿研究熱忱的心和不斷的努力之外，一路上一雙雙希望之手不斷拉拔我，讓我每每在徬徨無助時，得到繼續走下去的勇氣和力量。

首先，我要感謝政大語言所的萬依萍老師、黃瓊之老師、徐嘉慧老師、蕭宇超老師和何萬順老師，他們孜孜不倦的教導，讓我對語言學充滿興趣，引領我進入語言學的殿堂。尤其感謝我的指導教授萬依萍老師：感謝老師在我理不清論文題目時，像明燈一樣給我指了一條路；感謝老師不厭其煩的幫我改論文；感謝老師好幾次的推薦我，讓我實現了很多不可能的事；感謝老師耐心的指導與無盡的包容，幫我渡過論文的瓶頸；感謝老師啟發了我對心理語言學的興趣，並讓我有往下一個目標前進的勇氣。

另外我要非常感謝我的兩位口試委員，小笠原奈保美老師和林祐瑜老師。感謝他們在百忙之中不辭辛勞的來參加我的口試，並且給我非常多的寶貴意見。尤其是在實驗設計、統計和邏輯思辨上給了我非常多的指導和啟發。這些珍貴的意見不僅僅對我這篇論文有幫助，我相信對我日後的研究也會有非常大的助益。

還要感謝政大語言所的同學們，你們讓我的碩士生涯不只是做研究和上課而已，還充滿了樂趣。你們就像彩色筆，把可能是黑白的我塗成彩色。還要特別感謝超級好友雯琪，如果不是你每天的加油打氣，我可能已經放棄；如果不是你每天跟我聊天解悶，我可能會是個書呆子；如果不是你在學術上給我一些建議，我可能也無法參加一些學術會議。有了大家的陪伴，一路上雖然有風有雨，但卻總是有人幫我撐傘，給我向上的力量。

最後，我要感謝我的家人。謝謝爸爸媽媽給我那麼好的讀書環境，讓我能無後顧之憂的讀書。爸媽不時的叮嚀我要何時畢業以及論文進度，也是我努力奮發不懈怠的動力。謝謝哥哥總是聽我抱怨，好似我的垃圾桶一般。沒有他，我真的不知道這些垃圾要向誰傾訴。謝謝外婆總是默默的替我加油，讓我感到無比的溫馨。家人的支持與關懷永遠都是我最強而有力的後盾。

現在，我完成了這個階段的使命，即將結束碩士班生涯，邁向人生的另一段旅程。我懷抱著感恩的心，夢想著未來。這些關懷的種子，一定會帶領我突破未來的每一個考驗和難關。

Chinese Abstract

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：臺灣華語的口語詞彙辨識歷程：從雙音節詞來看

指導教授：萬依萍 博士

研究生：錢昱夫

論文提要內容：（共一冊，17770 字，分六章）

本研究用雙音節詞來探討不同音段和聲調在臺灣華語的口語詞彙辨識歷程中的重要性。Cohort 模型（1978）非常強調詞首訊息的重要性，然而 Merge 模型（2000）認爲訊息輸入和音韻表徵的整體吻合才是最重要的。因此，本研究企圖探索不同音段和詞首詞尾在臺灣華語的口語詞彙辨識歷程中的重要性。然而，聲調的問題並無在先前的模型裡被討論。因此，聲調在臺灣華語的口語詞彙辨識歷程中所扮演的角色也會在本研究中被討論。另外，詞頻效應也會在本研究中被探索。本研究的三個實驗均由同樣的十五名受試者參加。實驗一是測試不同音段在臺灣華語的口語詞彙辨識歷程中的重要性。實驗一操弄十二個雙音節高頻詞和十二個雙音節低頻詞，每一個雙音節詞的每一個音段都分別被噪音擋住。實驗二是在探索詞首和詞尾在臺灣華語的口語詞彙辨識歷程中的重要性。實驗二操弄十二個雙音節高頻詞和十二個雙音節低頻詞。這些雙音節詞的詞首 CV 或詞尾 VG/N 都分別被雜音擋住。實驗三操弄二十四個雙音節高頻詞和二十四個雙音節低頻詞。這些雙音節詞的聲調都被拉平到 100 赫茲。在這三個實驗中，受試者必須聽這些被操弄過的雙音節詞，並且辨認它們。受試者的反應時間和辨詞的準確率都用 E-Prime 來記錄。實驗結果顯示，傳統的 Cohort 模型不能被完全支持，因爲詞首訊息被噪音擋住的詞仍能被受試者成功的辨識出來。強調聲音訊息和音韻表徵的整體吻合度的 Merge 模型，比較能解釋實驗的結果。然而，Merge 模型必須要加入韻律節點才能處理臺灣華語的聲調辨識的問題。本研究也顯示，雙音節詞的第一個音節的母音在口語詞彙辨識歷程中是最重要的，而雙音節詞的第二個音節的母音是第二重要的。這是因爲母音帶了最多訊息，包括聲調。另外，雙音節詞的詞首和詞尾在臺灣華語的口語詞彙辨識歷程中是扮演差不多重要的角色。母音對於聲調的感知是最重要的。詞頻效應也完全表現在臺灣華語的口語詞彙辨識歷程中。

關鍵詞：口語詞彙辨識歷程、臺灣華語、華語聲調、音段、Cohort 模型、Merge 模型

Abstract

The present study investigated the importance of different segments and the importance of tone in spoken word recognition in Taiwan Mandarin by using isolated disyllabic words. Cohort model (1978) emphasized the absolute importance of the initial information. On the contrary, Merge (2000) proposed that the overall match between the input and the phonological representation is the most crucial. Therefore, this study tried to investigate the importance of different segments and the importance of onsets and offsets in the processing of Mandarin spoken words. However, the issues of tone were not included in the previous models. Thus, the importance of tone was also investigated in this study. The issues about frequency effect were also explored here. Three experiments were designed in this study. Fifteen subjects were invited to participate in all three experiments. Experiment 1 was designed to investigate the importance of different segments in Taiwan Mandarin. In experiment 1, 12 high-frequency disyllabic words and 12 low-frequency disyllabic words were selected. Each segment of each disyllabic word was replaced by the hiccup noise. Experiment 2 was designed to investigate the importance of onsets and offsets. In experiment 2, 12 high-frequency disyllabic words and 12 low-frequency disyllabic words were chosen. The CV of the first syllable and the VG/N of the second syllable were replaced by the hiccup noise. Experiment 3 was designed to investigate the importance of Mandarin tones. In experiment 3, 24 high-frequency disyllabic words and 24 low-frequency disyllabic words were selected. The tones of the disyllabic words were leveled to 100 Hz. In the three experiments, subjects listened to the stimuli and recognized them. The reaction time and accuracy were measured by E-Prime. The results indicated that traditional Cohort model cannot be fully supported because words can still be correctly recognized when word initial information is disruptive. Merge model, which proposed that the overall match between the input and the lexical representation is the most important, was more compatible with the results here. However, Merge model needs to include the prosody nodes, so that it can account for the processing of tones in Taiwan Mandarin. In addition, the current study also showed that the first vowel of the disyllabic word is the most crucial and the second vowel of the disyllabic word is the second influential since the vowel carries the most important information, including tones. The results of experiment 2 demonstrated that the onsets and offsets are almost the same important in Mandarin. Furthermore, vowel is the most influential segment for the perception of Mandarin tones. Finally, frequency effect appeared in the processing of Mandarin words.

Keywords: spoken word recognition, Taiwan Mandarin, Mandarin tones, segments, Cohort, Merge

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 Introduction

1.1 The background of spoken word recognition

Speech perception has been a popular issue for several decades. Researchers in different fields, such as physics, engineering, linguistics, and psychology, keep concerning the issues of speech perception regarding how humans perceive speech sounds effectively and efficiently. In the field of linguistics, the long-term issues that have always raised many scholars' interests regarding speech perception include how acoustic signals map to the phonetic segments, how phonetic segments map to phonemes, and how phonemes are combined to form words.

Concerning the issue of speech perception, the most fundamental problem is how acoustic properties map to phonetic segments. According to many previous studies, acoustic signals vary among individuals, between genders, and even within a particular person. In addition, the acoustic properties of a specific segment also alter from context to context. Therefore, how human perceptual systems decode these various acoustic signals, and how human perceptual systems pick out any invariance among a plethora of variance in speech signals for speech perception, have been the issues tackled by many researchers for many years.

Another unresolved issue as to speech perception is what the elementary unit of perception is. Traditionally, it is assumed that the phonetic segment is the elementary unit of speech perception because it can differentiate one word from another and it is the minimal speech sound unit (Cutler, Norris, & Williams, 1987). However, some proposed that it is the phonetic feature that is the true basic unit of speech perception for the reason that it cannot be broken down further into other smaller linguistic unit (Jakobson, Fant, & Halle, 1952). Different from the above mentioned suggestions, some studies indicated that the syllable is the elementary unit of speech perception since it is impossible to draw a clear-cut boundary between segments in a syllable (Savin and Bever, 1970). The acoustic properties of one segment overlap with those of the preceding or the following segments. Therefore, what the elementary unit of speech perception is still a controversial issue for researchers to explore.

As already noted, in the early field of speech perception, researchers mainly focused on the phonetic segments, including how phonetic segments are discriminated from one another, and how the phonetic segments are categorized. In the 1970s, a new issue concerning the processes and representations for perception of spoken words attracted many researchers' attention. This new issue came from the concerns that a comprehensive theory of speech perception cannot only focus on the consonants, vowels, and syllables. How the hearers perceive and understand the fluent speech is the most crucial issue. Hence, spoken words became the focus of the research.

Prior to the studies on the perception of spoken words, the research regarding the perception of printed words had already been investigated by many scholars. However, the theories about the recognition of visual words could not be applied to the recognition of spoken words, since the theories concerning the recognition of visual words could not explain how the acoustic signals are perceived by listeners and mapped to the hundreds of thousands of representations in the human brain.

One of the most significant models of the recognition of spoken words was the Cohort theory, which was proposed by Marslen-Wilson (1978, 1980). Marslen-Wilson's Cohort theory turned over a new leaf in the history of speech perception and set up the field of spoken word recognition. According to the Cohort model (Marslen-Wilson & Welsh, 1987), word initial information is very crucial for activating an initial set of hypotheses about the acoustic input. However, over-emphasizing the word initial information results in incorrect prediction since humans can still correctly recognize a word even if the word initial phoneme is disrupted. Thus, some models were invented to modify the defects of the cohort theory. In contrast with the cohort model, disruptions of word-initial phonemes in the models such as Race (Cutler & Norris, 1979), TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994b), and Merge (Norris, McQueen, and Cutler, 2000) are not disastrous because the acoustic information of the other phonemes still contributes to the activation of a lexical entry. Although there had already been some studies pointing out the shortcomings of the Cohort theory, it could not deny that Marslen-Wilson’s Cohort theory aroused many concerns for the following decades. Actually, many issues that are still active in the field of spoken word recognition now are either closely related to the Cohort theory or trying to modify the defects of the original model.

1.2 Motivation and research questions

Four groups of questions will be discussed in this study. Questions to be asked involve the following.

(i) Is the word-initial information such important as Cohort theory predicts? If this is the case, then any stimuli in the experiment of this study whose initial segment is replaced by the hiccup noise cannot be perceived correctly. If the word-initial information is not as crucial as what the Cohort theory predicts, then those stimuli whose word-initial segments are disrupted can still be perceived correctly by the listeners. However, if the results show what Cohort theory predicts is wrong, then it raises the question about the status of the acoustic onset and offset in spoken word recognition. That is, is it the onset or the offset that is the most crucial for spoken word recognition in Mandarin?

Among the models of spoken word recognition, there are two different arguments concerning the importance of the initial segment. Cohort theory (Marslen-Wilson & Welsh, 1978) proposed that a set of representations in memory are activated by the acoustic input, known as the word-initial cohort. All of the words which have the same initial acoustic information as the input signals are activated in the listener’s mind. Therefore, the early Cohort theory put much emphasis on the importance of the word-initial input, suggesting that spoken word recognition would break down if the initial input is seriously disturbed. However, other models of spoken word recognition did not emphasize the importance of the word-initial input to such an extent. The Merge model (Norris, McQueen, & Cutler, 2000) focused on the overall similarity between the acoustic inputs and the words being activated. It suggests that word-initial segment is not of critical importance. Even though the word-initial information is severely damaged, the particular word can still be activated and recognized depending on the rest of the acoustic information. Although a great number of studies have been conducted to investigate the importance of the word-initial information in the recognition of spoken words, few studies focused on the role of the final segments in spoken word recognition (Wingfield, Goodglass, & Lindfield, 1997). In addition, most of the studies concerning this issue focused on English or some western languages such as Dutch; very few focused on the role of the initial and final segments in spoken word recognition in Mandarin. Hence, these questions based on the gaps mentioned above will be tackled in this study.

(ii) What is the status of different segments in spoken word recognition? If the initial consonant is the most important segment in spoken word recognition, then the result of the experiment will predict the longest reaction time and lowest accuracy when the initial consonant is replaced by the hiccup noise. However, if it is the prenuclear glide, vowel, postnuclear glide, or the final nasal that occupies the prestigious status in spoken word recognition, then the result of the experiment will presage the longest reaction time and lowest accuracy when the prenuclear glide, vowel, postnuclear glide, or final nasal, is replaced by the hiccup noise.

One of the active questions in the field of spoken word recognition is about the nature of lexical and sublexical representations. Research on the lexical competition mainly put emphasis on the competition between the representations of words. Nevertheless, another crucial issue regarding the spoken word recognition is the nature and existence of sublexical representations. Marslen-Wilson and Warren (1994) argued against the existence of sublexical representations. They suggested that phonetic features maps to words directly, without any intermediate sublexical representations. Other researchers, in contrast to Cohort theory, argued for the nature and the existence of sublexical representations though different models proposed different viewpoints about the interactions between segmental and lexical representations. At present, much evidence is in favor of the existence of sublexical representations, in contrast with Cohort theory. However, there is still a gap in that few studies focused on the role of different segments in a word. Thus, the questions concerning this gap will be dealt with in this study.

(iii) Can the spoken words be recognized successfully if the tones of the words are leveled? What is the interaction between Mandarin tones and segments in the recognition of spoken words? Which segment, namely, the initial consonant, prenuclear glide, vowel, postnuclear glide, or final nasal, is the most influential segment for the perception of Mandarin tones? If the segment which is replaced by hiccup noise results in the wrong perception of the particular tone, it can be inferred that the segment carries the most important acoustic information of that tone. If the segment which is replaced by hiccup noise is perceived correctly concerning its tone, it suggests that the acoustic information in that segment is not enough to cause the incorrect perception of Mandarin tone.

In the history of speech perception, the issues regarding how the acoustic signals map to phonetic features, how phonetic features map to phonemes, how phonemes map to syllables, and how syllables map to words, have already been tackled by a number of researchers. These issues are segmental. Other issues concerning suprasegmental have also been explored. For example, studies about segmentation of words in fluent speech suggested the prosodic solution to the segmentation problem, which stated that listeners parse the speech stream by exploiting rhythmic characteristics of their language (Cutler, 1996; Cutler & Norris, 1988). Although a great number of studies have been conducted to investigate both segmental and suprasegmental issues about speech perception, few concern the status of Mandarin tones in speech perception. Therefore, the status of Mandarin tones will be investigated in this study.

(iv) Does frequency effect affect Mandarin spoken word recognition? If frequency effect really exists in Mandarin, then the reaction time of the high frequency words will be shorter than that of the low frequency words. To the contrary, if the frequency effect plays no role in Mandarin spoken word recognition, the reaction time of the high frequency words will not be longer than that of the low frequency words.

A number of previous studies have already proved that low frequency words are more difficult to be picked up by high frequency words in spoken word recognition (Monaco, 2007; Savin, 1963; Broadbent, 1967; Elliott, 1987). Nevertheless, few studies examined this phenomenon in Taiwan Mandarin. As a result, the study will examine this effect in Taiwan Mandarin.

Given the gaps mentioned above, in this study, we intend to investigate several issues concerning Mandarin word recognition more thoroughly and completely.

1.3 Organization

The organization of the following chapters is as follows. Reviews of literature on the models of spoken word recognition, together with a number of issues concerning spoken word recognition, are discussed in chapter 2. Chapter 3 focuses on the methods for conducting this study, including the details of the subjects, segmentation criteria of the initial consonant, prenuclear glide, vowel, postnuclear glide, and final nasal, the equipments used in this study, along with the procedures of the experiment. Chapter 4 introduces the statistical analyses and a brief result. The discussion and theoretical explanations relevant to the study are shown in Chapter 5.

CHAPTER 2 LITERATURE REVIEW

In this chapter, previous studies on word recognition from acoustic signals, and the acoustic-phonetic cues concerning consonants, vowels and tones in Mandarin will be discussed. Section 2.1 introduces the cohort model (Marslen-Wilson & Welsh, 1978), which is a parallel lexical access model, and the Merge model (Norris, McQueen, and Cutler, 2000), which is one kind of connectionist model. Section 2.2 reviews studies concerning auditory word recognition, focusing on the acoustic onsets and acoustic offsets. Section 2.3 puts emphasis on the Mandarin phonological system, including the syllable structure of Mandarin. Section 2.4 displays the acoustic-phonetic cues of Mandarin consonants. Section 2.5 briefly shows the acoustic-phonetic cues of Mandarin vowels. Section 2.6 reviews the perception of Mandarin tones, putting emphasis on the acoustic-phonetic cues and the processing of Mandarin tones.

2.1 Models of spoken word recognition

In this section, we briefly introduce two crucial models of spoken word recognition in recent years, including Cohort (1978), and Merge (2000).

2.1.1 Cohort model (1978)

The Cohort model, although share some basic assumptions with regard to lexical access with the logogen model, was designed to explain the process of auditory word recognition. Marslen-Wilson et al. (Marslen-Wilson & Zwitserlood, 1989; Tyler, 1984; Marslen-Wilson & Tyler, 1980, 1981) proposed that when we hear a word, all of the words which bear the phonological resemblance with the heard word are activated. For example, if we hear the sentence “Paul wants to be a ca-…,” cap, capital, Capricorn, capture, captain, captive, and many others, would be activated, which means that all of these activated words can be the candidate for selection. This set of words is called the “word initial cohort”. Hence, as the assumption of logogen model, possible candidates would be activated until the final candidate is identified. As the other parallel access models, activation of a word in cohort model is based on direct mapping between the speech input and the lexicon.

In Cohort theory, all possible candidates for lexical access would be activated by the auditory input and then eliminated gradually by the following two ways-either the context narrows the word initial cohort or the possible candidates are kicked out as more and more phonological information is perceived. In the latter case, as more of the spoken word is identified, the cohort narrows the window. For instance, if the phoneme /p/ follows the sequence *ca-*, captain, captive, and all the other words that have the same initial letters are the potential candidates from the initial cohort. The pool of candidates continues to narrow as more acoustic signal is received. Only when one single candidate left can the particular word be recognized. The schema of the

Cohort model and how cohort model operates are shown in Figure 1.

Time --→

Input: [trɛspəs] (trespass)	Recognized Phoneme	[t]	[tr]	[trɛ]	[trɛs]	[trɛsp]	Recognized word: [trɛspəs]
	Current Cohort	[tri:], [taɪm], [trɛspəs], [trɛɪn], [trɛnd], [trɛs], ...	[tri:], [trɛspəs], [trɛɪn], [trɛnd], [trɛs], ...	[trɛspəs], [trɛɪn], [trɛnd], [trɛs], ...	[trɛspəs], [trɛs], ...	[trɛspəs]	

Figure 1. Schema of the cohort model.

This figure displays that when the phoneme [t] is recognized, it activates a series of words which begin with [t]. This set of words is called "word-initial cohort." As more and more phonemes are recognized, the activated words become less and less. Finally, when the phoneme [p] is recognized, the word [trɛspəs] is also recognized because the phoneme sequence [trɛsp] can only activate [trɛspəs] but no other candidates.

Originally, the Cohort theory puts heavy emphasis on the absolute match between the perceived auditory signal and the phonological representation in the mental lexicon, which means that the word initial stimulus is of paramount importance and cannot be mispronounced or blocked by a cough or the surrounding noise. If the initial stimulus is disturbed, the word initial cohort cannot be activated. However, subsequent experiments indicate that a word can still be recognized even if the initial input of the word is obstructed (Marslen-Wilson, 1987). Therefore, the

Cohort theory was revised by Marslen-Wilson (1987) so that the system chooses the best match to fit an incoming word. Under this revised Cohort theory, word recognition depend less on the initial auditory input. A word can be recognized as long as the phonological representation of that word shares enough features with the incoming stimulus. Nevertheless, Marlen-Wilson (1989) reemphasized the importance of the word-initial information because lexical activation would be obstructed even if all the other information except word-initial information is consistent with the target words.

2.1.2 Merge (2000)

The Merge model (2000), which is an autonomous model, was proposed by Norris, McQueen, and Cutler. The network of Merge is a simple competition-activation network which is the same as the basic dynamics as Shortlist (Norris, 1994b). In Merge, there are three types of nodes, including input nodes, lexical nodes and phoneme decision nodes. As in Figure 2, the input nodes are associated by facilitatory links to the appropriate lexical nodes and the phoneme decision nodes. The lexical nodes are also connected by facilitatory links to the suitable phoneme decision nodes. But, different from the TRACE model (McClelland & Elman, 1986), an interative model, there is no feedback from the lexical nodes to the prelexical phoneme nodes. Inhibitory activation happens between lexical nodes as well as phoneme decision nodes, but not between input nodes.

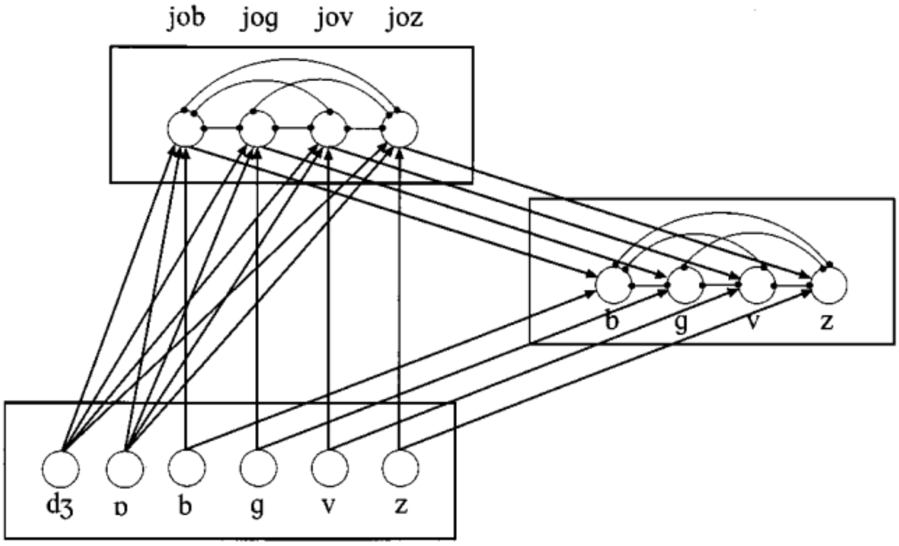

Figure 2. The basic architecture of Merge. The facilitatory connections, which are unidirectional, are displayed by bold lines with arrows; the inhibitory connections, which are bidirectional, are illustrated by fine lines with circles (Norris, McQueen, & Cutler, 2000)

Figure 2 displays the simulation of the subcategorical mismatch in the architecture of Merge. The network was designed with merely 14 nodes, including 6 input nodes (/dʒ/, /ɒ/, /g/, /b/, /v/, and /z/), 4 phoneme decision nodes, and 4 possible word nodes, *job*, *jog*, *jov*, *joz*. The latter two word nodes stand for only the possible combinations of words, rather than the real words.

The Merge model, which is faithful to the basic principles of autonomy, was designed to explain the data which were not compatible with TRACE (McClelland & Elman, 1986). Merge, with the phoneme decision nodes that combine the lexical and phonemic information flows, provides a simple and appropriate account for the data proposed by Marslen-Wilson & Warren (1994), McQueen et al. (1999a), Connine et al. (1997), and Frauenfelder et al. (1990), which cannot be explained either by TRACE appropriately. Therefore, Merge can give a full explanation of the known empirical findings in phonemic decision making.

2.2 The role of acoustic onsets and acoustic offsets

A basic feature of speech signal is its intrinsic directionality in time. When utterances proceed, speech signals are moving along the time line from the beginning to the end of the utterances. This fundamental property of speech signal strongly implies that the initial acoustic signal is of paramount importance, which is in accordance with the claims of Cohort theory (Marslen-Wilson, 1984).

Auditory word recognition is a very complicated language processing issue because of many linguistic and non-linguistic factors that may disrupt the acoustic cues of speech signal. These disruptive factors include speech errors, acoustic phonetic variability under different phonological conditions, and the auditory obstructions due to the surrounding noise. These possible acoustic disruptions can happen at any moment of auditory word recognition. However, human brains can still recognize words with little difficulty most of the time. Moreover, the speech input is a stream of acoustic signal. Hearers do not exactly know whether the particular input is in the initial, medial, or final position of a word.

From the above mentioned difficulties of speech processing, it is clear that the Cohort model, first proposed by Marslen-Wilson & Welsh (1978), putting great emphasis on the importance of initial acoustic cues, cannot account for the fact that speech signal is more or less varied or disrupted under different circumstances. Therefore, the Cohort model was revised by Marslen-Wilson (1987), which rejected the total dependence on the word-initial cues for auditory word recognition. Unlike the old Cohort model (1978), the relatively new Cohort theory claims that the disruptions of the word-initial signal are not the end of the world because the non-word-initial information can still bring about the activation of candidates. Therefore, in the latter Cohort theory, 100 percent match between the word-initial acoustic signal and the phonological representation of a given word is not as crucial as what the original Cohort theory claims given the acoustic information in spoken words. In addition, there are other experiments indicating that auditory word recognition is not blocked even though the word-initial signal is distorted. One such experiment is that an ambiguous phoneme between /d/ and /t/ is presented before the sequence /ajp/. The subjects, after listening the stimulus, have to decide what phoneme they have heard (Connine & Clifton, 1987). The result shows that subjects tend to label the ambiguous phoneme as /t/ when followed the sequence /ajp/ because /t^{h}ajp/ 'type' is a word. This indicates that the word-initial signal is not extremely crucial in auditory word recognition; otherwise the word 'type' cannot be recognized due to the ambiguous word-initial acoustic cues.

Marslen-Wilson and Zwitserlood (1989) conducted experiments to investigate whether a nonword can activate a real word if the nonword is different from the real word only by the initial phoneme. The results of their study indicated that the nonword different from the real word by merely the initial phoneme cannot activate the real word generally. According to the results, Marslen-Wilson and Zwitserlood reemphasized the importance of the word-initial information. They claimed that lexical activation would be barred even if all the other information except the initial phoneme is consistent with the hypothesized words. Therefore, mispronunciation of the initial phoneme of a word cannot facilitate the base but preclude the activation of it. From the result that a nonword derived from a real word by merely changing its initial phoneme cannot facilitate the real word, it is clear that word-initial information is very important in auditory word recognition. In addition to the studies regarding the initial segment of the input, Nooteboom and van der Vlugt (1988) compared the importance of word onsets and offsets. The results indicated that words can be recognized equally well no matter the inputs are heard from the beginnings or from the endings as long as the hearers knows which part of the words they have heard. However, they still claimed that word-beginning priority exists due to the fact that word initial information is more easily to be associated correctly to the lexical representation than the word final information.

Another relevant study concerning the role of the initial segment of a nonword was done by Connie, Blasko, and Titone (1993). The purpose of their research was to demonstrate whether phonetically similar initial phonemes in a derived nonword would be sufficient to produce activation of a base word. They designed the nonwords which was only one or two phonetic features different from the base words. The altered segments of those nonwords were either in the initial position or medial position. The results of the study indicated that a base word can still be activated by a nonword with a similar initial phoneme. The results also showed that the altered position of a nonword is not the factor that influences the priming effect. Connie, Blasko, and Titone (1993) concluded that relative similarity of elements in the input to a lexical representation is critical for auditory word recognition. Furthermore, it is not the exact positional acoustic information of a particular lexical item that is important in spoken word recognition, but the overall acoustic-phonetic similarity between the input and lexical representation that is influential. Therefore, the findings of their study contradict the cohort theory, which claims that the initial segment serves to determine the activated word candidates.

Wingfield et al. (1997) used gating technique to investigate the interaction among the acoustic onsets and offsets, the cohort size, and syllabic stress in English. Their analysis on the cohort sizes from both forward and backward gating showed that the cohort size is significantly larger at the recognition point from forward gating than from backward gating for two and three syllable words and for all stress patterns. This finding depicted a great advantage of forward gating over backward gating for two and three syllable words and for all stress patterns, indicating that acoustic onset information is much more important than acoustic offset information for all stress patterns though words can be identified from both beginning and ending directions. However, Wingfield et al. (1997) degraded the absolute word-onset priority principle when taking the stress patterns into consideration. They assumed that stress patterns can restrict the cohort size. They showed that the cohort sizes at recognition point were not only significantly reduced, but the cohort sizes at recognition point were also equal in both forward and backward gating directions. This analysis supported the claim that cohort reduction is a very crucial mechanism in auditory word recognition, regardless of direction of gating, which supported the overall goodness-of-fit hypothesis, rather than the absolute word-onset priority principle. Nevertheless, Wingfield et al. did not deny the fact that more acoustic information is needed for word recognition if a word is gated from its ending. That is possibly due to the fact that, for any given cohort size, a longer gate duration is needed in the backward-gating condition than in forward-gating in English.

2.3 Mandarin phonological system

There are 12 combinations of Mandarin syllable structure, including V, CV, GV, VG, VN, CVG, CVN, CGV, GVG, GVC, CGVG, and CGVN. In Mandarin, a syllable is traditionally divided into three parts, including an optional initial, a final and a tone (C. Cheng, 1973). The initial part can be a nasal or a consonant. The final part contains an optional prenuclear glide, a vowel, and an optional postnuclear glide or a nasal. However, during the past two decades, the status of the prenuclear glides in Mandarin syllable has raised many debates (Bao, 2002; Yip, 2002; Duanmu, 2002; Wan, 2002a). Under the study, since the status of the prenuclear glide is not the focus, the prenuclear glide was not grouped with the onset or the rhyme and was replaced by the hiccup noise alone just as the initial consonant and the vowel. Last but not least, in order not to let the duration of the rime be much longer than that of the prenuclear glide and that of the initial consonant, the rime was further divided into a vowel plus a postnuclear glide, or a vowel plus a final nasal. Each part of the rime could be replaced by the hiccup noise individually.

2.4 The acoustic-phonetic cues of the consonants in Taiwan Mandarin

In Taiwan Mandarin, there are overall 21 onset consonants, namely, six oral stops /p/, /pʰ/, /t/, /tʰ/, /k/, /kʰ/, two nasal stops, /m/, /n/, six fricatives /f/, /ɕ/, /x/, /ʂ/, /s/, /ʐ/, six affricates /tɕ/, /tʰɕ/, /tʂ/, /tʰʂ/, /ts/, /tʰs/, and one liquid /l/. In the following sections, the acoustic-phonetic characteristics of those onset consonants are introduced. These characteristics serve as the criteria of segmentation in experiment 1 and 2.

2.4.1 The acoustic-phonetic cues of stops

There are three acoustic-phonetic cues for distinguishing stops. They are formant transitions, burst amplitude, and duration.

First, formant transitions are crucial for detecting the place of articulation of stops. The F2 and F3 transitions from the bilabial stops into the following vowels are rising. The F2 and F3 transitions from the alveolar stops into the following vowels are almost flat. The F2 and F3 transitions from velar stops into the following vowel come together. Second, previous research (Repp, 1984) indicated that the burst amplitude of labial stops is weaker than that of the alveolar and velar stops. Perceptual experiments have shown that burst amplitude can influence the identification of labial and alveolar stops. This effect can be better realized on voiceless stops than voiced stops. Third, VOT is of paramount importance for the detection of voicing. Stops, which have relatively long VOT, tend to be perceived as voiceless stops; in contrast, stops, which have relatively short VOT, are prone to be recognized as voiced stops. In addition, voiceless aspirated stops have the longest VOT compared with voiced stops, and voiceless unaspirated stops. In Mandarin, the mean VOTs for /p/, /p^h/, /t/, /t^h/, /k/, and /k^h/ are 14 ms, 82 ms, 16 ms, 81 ms, 27 ms, and 92 ms, respectively (Chao et al., 2006).

2.4.2 The acoustic-phonetic cues of nasals

According to Ladefoged (2000), there are four acoustic-phonetic cues for recognizing nasals. First, there is a sharp change in the spectrogram at the time of the formation of the articulatory closure. Second, the bands of the nasal are lighter than those of the vowel, which indicates that the intensity of the nasal is weaker than that of the vowel. Third, the F1 of the nasal is often very low, centered at around 250 Hz. Fourth, there is a large space above the F1 with no energy. Based on these acoustic-phonetic cues, nasals can be identified.

2.4.3 The acoustic-phonetic cues of fricatives

The most crucial acoustic-phonetic cue for separating voiceless fricatives from voiced fricatives is by examining the extended period of noise (Borden *et al*., 1994). The extended period of noise can be easily detected on the spectrogram. Voiceless fricatives have longer duration and stronger intensity. To the contrary, voiced fricatives (i.e., /ʐ/ in Mandarin) are shorter in duration and weaker in intensity, but their formant frequencies are clearer than those of voiceless fricatives.

Fricatives are known for their high-frequency noise in the spectrum, which is an acoustic-phonetic cue for distinguishing the place of articulation of fricatives. Another acoustic-phonetic cue for distinguishing the place of articulation of fricatives is the intensity of frication. Sibilants (i.e., /s/, /ʂ/, /ɕ/, /ʐ/, /tɕ/, /t^{h}ɕ/, /tʂ/, /t^{h}ʂ/, /ts/, and /t^{h}s/in Mandarin) are noted for relatively steep, high-frequency spectral peaks, whereas nonsibilants (i.e., /f/ and /x/ in Mandarin) are famous for relatively flat and wider band spectra. Moreover, alveolar sibilants (i.e., /s/ in Mandarin) can be distinguished from palatal sibilants (i.e., /ʂ/...) by the location of the lowest spectral peak. The lowest spectral peak of the alveolar sibilants is around 4000 Hz, while the lowest spectral peak of the palatal sibilants is around 2500 Hz. Furthermore, the intensity shown on the spectrogram can also differentiate the place of articulation of fricatives. Stronger intensity is the feature of sibilants; weaker intensity, the feature of nonsibilants. This is because the resonating cavity in front of the alveolar or the palatal constrictions results in high intensity. However, there is no resonating cavity in front of the labio-dental constriction, which brings about the relatively weak intensity. The acoustic-phonetic characterization of fricatives is illustrated in Figure 3.

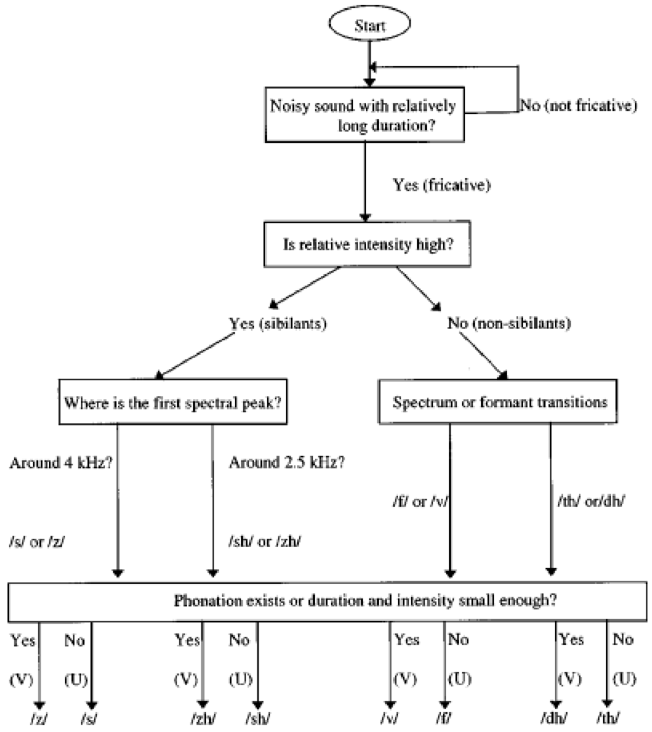

Figure 3. Acoustic-phonetic characteristics of fricatives (Borden et al., 1994)

Figure 3 shows how listeners perceive fricatives. When the listener hears an input, it enters the first filter and is judged by whether it has noisy sound with relatively long duration. If the answer is yes, the input is regarded as a fricative and sent to the next filter. In the second filter, the input is examined by whether its intensity is relatively high. If the answer is yes, the input is considered a sibilant and sent to the next filter. In the third filter, the input is investigated by its first spectral peak. If the first spectral peak of the input is around 4kHz, it is viewed as /s/ or /z/ and sent to the next filter. In the fourth filter, the input is judged by “phonation exists or duration and intensity small enough?” If the answer is yes, the input is perceived as /z/; if the answer is no, it is perceived as /s/. By those filters, the input is examined step by step and finally recognized by the listener.

2.4.4 The acoustic-phonetic cues of affricates

There are three pairs of affricates in Mandarin, /tɕ/, /tʰɕ/, /tʂ/, /tʰʂ/, /ts/, and /tʰs/. According to Ladefoged (2000), an affricate is simply a sequence of a stop followed by a homorganic fricative. Therefore, it can be inferred that affricates have the acoustic-phonetic characteristics of both stops and fricatives.

2.5 The acoustic-phonetic cues of the vowels in Taiwan Mandarin

Phonetically speaking, there are overall 12 vowels in Taiwan Mandarin, including 4 high vowels ([i], [u], [y], and [ɨ]), 2 low vowels ([a] and [ɒ]), as well as 6 mid vowels ([e], [ɛ], [ə], [ɤ], [o], and [ɔ]). Vowels have very different phonetic cues from consonants. First of all, vowels have much longer duration than consonants. Second, the formants of vowels are much clearer than those of consonants. Third, the energy of vowels is stronger than that of consonants, causing darker spectrogram. Fourth, the F0 in vowels displays the tones in Mandarin. From the acoustic-phonetic cues, vowels can be distinguished from consonants.

2.6 Mandarin tone

2.6.1The perception of Mandarin Chinese tones

Lexical tones are pitch patterns that can distinguish lexical meanings in a given language. In Mandarin Chinese, tones, like the aspirated and unaspirated stops, are phonemic features that can differentiate word meanings. Mandarin Chinese phonemically distinguishes four tones, which are Tone 1, with high-level pitch, Tone 2, with high-rising pitch, Tone 3, with low falling-rising pitch, and Tone 4, with high-falling pitch (Chao, 1948). The same syllable structure can have different meanings if it carries different tones. For instance, *ma* with Tone 1 has the meaning of ‘mother’; *ma* with Tone 2 has the meaning of ‘numbness’; *ma* with Tone 3 has the meaning of ‘horse’; *ma with* Tone 4 has the meaning of ‘scold’.

There are several factors that can affect the perception of Mandarin Chinese tones. First, fundamental frequency plays a role in the Mandarin Chinese tone perception. Previous acoustic studies have found that the F0 height and F0 contour are the acoustic cues for Mandarin Chinese tone perception. Howie (1976) performed the tone perception experiments to test whether the participants could identify the correct tones of the stimuli. Howie designed three contrasted conditions, which were synthetic speech with natural F0 patterns, synthetic speech with the monotonic F0 contour, and synthetic speech sounding like a whisper. The results showed that subjects easily recognized the synthetic speech which F0 patterns were maintained. Gandour (1984) and Tseng & Cohen (1985) indicated that both F0 height and F0 contour are very crucial acoustic cues for Mandarin tone perception. Neither one can be missed. Moore and Jongman (1997) differentiated Tone 2 from Tone 3 in terms of two characteristics. One is turning point, which is the point in time at which the tone changes from falling to rising, and the other is ΔF0, which is the F0 change from the onset to the turning point. Moore and Jongman found that the turning point of Tone 2 is earlier than that of Tone 3, and the ΔF0 of Tone 2 is smaller than that of Tone 3. Comparing the acoustic cues of Tone 3 and Tone 4, Garding et al. (1986) found that the stimuli which have the early peak of pitch and fall dramatically after the turning point tend to be perceived as Tone 4. The stimuli which stay at low F0 range and have long duration tend to be recognized as Tone 3. This study demonstrates that F0 contour is of paramount importance for Mandarin tone perception.

The second factor that can influence the perception of Mandarin Chinese tones is the temporal properties of tones. According to the production data, Nordenhake and Svantesson (1983) found that the duration of Tone 3 is the longest, which is only slightly longer than that of Tone 2, while the duration of Tone 4 is the shortest. Given that the F0 contours are similar between Tone 2 and Tone 3, Nordenhake and Svantesson (1983) further indicated that Tone 2 could be perceived as Tone 3 if it is lengthened.

In addition to F0 and duration, amplitude can also affect the perception of Mandarin Chinese tones though only to a small extent. Whalen and Xu (1992) designed stimuli whose formant structures and F0 contours were removed, but the amplitude cues of the stimuli were reserved, and then they asked the participants to identify the stimuli. The results demonstrated that participants could successfully identify Tone 2, Tone 3, as well as Tone 4, but fail to recognize Tone 1.

From the above mentioned studies on acoustic phonetic characteristics of Mandarin Chinese tones, it is clear that fundamental frequency, turning point, ∆F0, duration, and amplitude are the acoustic cues which play a critical role in the perception of Mandarin Chinese tones. Nevertheless, the acoustic quality of tones can be influenced by the surrounding context, which may also affect the perception of tones.

Shen (1990) studied the tonal coarticulation of Mandarin Chinese and found that tonal coarticulation not only affects the F0 height of the onset or offset, but it affects the F0 height of the entire word. The tones that are most easily to be affected are those which follow Tone 1 and Tone 2. Both Tone 1 and Tone 2 have high offset F0 value, which can raise the entire F0 value of the following tones. In addition, the high onset F0 value of Tone 4 also has the effect of raising the whole F0 value of the preceding tones. Unlike Tone 1, the offset of Tone 2, and onset of Tone 4, the onset of Tone 2 as well as Tone 3, whose onset F0 value sits on the middle of the frequency range, do not have the ability of raising the entire F0 value of the preceding tones. In addition to all these findings above, shen also found that the tonal contour does not change even if the tone's entire F0 value has risen. Shen finally pointed out that tonal coarticulation cannot extend beyond one syllable.

2.6.2 The processing of Mandarin tone

Lexical tone is of paramount importance in processing spoken words in tone languages. Fox and Unkefer (1985) asked subjects to identify the tone of each stimulus in a continuum. The results displayed that the responses were much easier to make an ambiguous word a real word rather than a nonword. By the results, Fox and Unkefer indicated that lexical tone is an integral part of lexical representation in Mandarin. Cutler and Chen (1997) asked the subjects to judge whether the monosyllabic words and nonwords in pairs in Cantonese, differing only by onset consonant, vowel, or tone, were the same or different. The results showed that responses were slower and more inaccurate when the words and nonwords differed by tone than by onset consonant or vowel. Cutler and Chen proposed that tonal information arrives later than segmental information, which results in the slower process of tone. Moreover, Ye and Connie (1999) performed a tone monitoring task. In the experiment, the penultimate syllable of the idiom (*tɕin55 y51 ljɑŋ35 jɛn35*) was changed to the close tone (ie. Tone 3 and tone 2 are acoustically close) and far tone (ie. Tone4 and tone 2 are acoustically far) to the final tone, respectively. That is, *ljɑŋ35* was changed to *ljɑŋ21* and *ljɑŋ51*. The results demonstrated that responses to far tones were significantly slower than to close tones. The results indicated that tonal information maps to lexical representation in a graded style.

Lee (2000) investigated the processing of lexical tone and segment in Mandarin by using the direct form priming task. In the experiment, eighty monosyllabic Mandarin words were elected as targets. For each target, there were four primes having different kinds of relationship to it, including ID prime, Same Seg prime, Same Tone prime, and Unrelated prime. ID prime means that the prime and the target share not only the same segments but also the same tone, such as the prime *pʰaw21* ("to run" in English) and the target *pʰaw21*. Same Seg prime refers to the prime and target sharing only the segments, but not tone, such as the prime *pʰaw55* ("to fling" in English) and the target *pʰaw21*. Same Tone prime concerns the prime and target sharing merely the tone, but not segments, such as the prime *fej21* ("a bandit" in English) and the target *pʰaw21*. Unrelated prime means that the prime and target are not similar at all, such as the prime *tɕyn51* ("handsome" in English) and the target *pʰaw21*. During the experiment, subjects heard the monosyllabic prime first and then judged whether the second monosyllabic word, the target, is a real Mandarin word or not. By this experiment, Lee (2000) found that significant facilitatory priming effect appeared for targets following ID primes. Non-significant facilitatory priming effect occurred for targets following Same Seg primes, while non-significant inhibitory priming effect happened for targets following Same Tone primes. From the results, Lee (2000) indicated that

Same Seg primes cannot fully activate the targets although their segments overlap totally. Therefore, it can be inferred that lexical tone in Mandarin is used on-line to resolve lexical identity. Lee also demonstrated that the decrease of activation level from ID prime, Same Seg prime, to Same Tone prime may be due to the degree of phonological match between the input and lexical representation. That is, Segmental information has higher degree of phonological match to lexical representation than tonal information, so that Same Seg prime has stronger power of activation than Same Tone prime. It can be further implied that tone is more like a phonetic segment than an independent tier in the processing of Mandarin words.

2.7 Summary

In this chapter, two models regarding spoken word recognition and the past studies concerning the acoustic-phonetic cues for the word recognition in Mandarin were discussed. In general, a gap can be observed; that is, previous studies concerning spoken word recognition mainly focused on western languages, whereas only a few studies focused on Mandarin. It is one of our main goals to fill this gap by investigating the status of different segments and tones in spoken word recognition in Taiwan Mandarin.

Two spoken word recognition models, Cohort model and Merge model, were presented as well. As a result, in this study, we are going to examine the two models to see which model can best explain the spoken word recognition in Taiwan Mandarin.

CHAPTER 3 METHODS

This chapter shows the research methods used in this study. Section 1 introduces the subjects' backgrounds. Section 2 describes the recording and broadcasting equipments. Section 3 introduces the details of the stimuli. Section 4 illustrates the design and procedures of the study.

3.1 Subjects

Thirty subjects were recruited in this study. They all lived in Taipei City or Taipei County. The 15 subjects, 7 males and 8 females, were all native Mandarin speakers and were not good at Taiwan Southern Min. They were at the age of 22 to 30 at the time of participating in the experiment.

3.2 Equipments

During the experiment, participants listened to the stimuli played by ACER Aspire One Series computer. E-Prime was used to record the participants' responses. The reaction time of subjects' responses was also measured by E-Prime.

3.3 Stimuli

All of the 48 stimuli in the experiment were disyllabic words embedded in the carrier sentence "tʂɤ51-kɤ tsɨ51 ʂɨ51 ____", "This word is ____". The stimuli were all recorded by Praat with mono channel and 11kHz sampling rate. The 48 disyllabic words were selected from Academia Sinica Balanced Corpus of Modern Chinese. The details of the stimuli are listed in appendix 1, and 2.

3.3.1 Word frequencies

There were totally 48 disyllabic words, which included 24 high frequency words and 24 low frequency words. The 48 disyllabic words were chosen from Word List with Accumulated Word Frequency in Sinica Corpus 3.0. The average frequency of the 24 high frequency words was 2503 occurrences out of 5 million tokens. The average frequency of the 24 low frequency words was 16 occurrences out of 5 million tokens.

3.3.2 Segmentation

There are overall 12 combinations of Mandarin syllable structure, namely V, CV, GV, VG, VN, CVG, CVN, CGV, GVG, GVN, CGVG, and CGVN. In this paper, the 48 disyllabic words contain all of the possible syllable structures in Mandarin, except for one structure, "V". The reason why the syllable structure "V" is excluded in this study is that if the single vowel word is replaced by the hiccup noise, there is nothing left to be heard by the subjects. Therefore, those words having only one vowel and nothing else are excluded in this study.

Since speech is continuous, it is really challenging to make a clear-cut distinction between segments. Nonetheless, for the purpose of finding out which part of the syllable to be of paramount importance for auditory word recognition in Mandarin, segmentation needs to be conducted. The following acoustical principles were the criteria for segmentation in this study.

3.3.2.1 Segmentation of the initial consonant

The first way to segment the initial consonant from the following prenuclear glide or vowel was to see the waveform. The initial consonant was measured from the starting point of the vibration on the waveform to the beginning of the intense vibration on the waveform. In order to further eliminate the quality of the initial consonant, the segmentation boundary between the initial consonant and the prenuclear glide or vowel moved forward about 13 milliseconds, as in Figure 4.

Another way to distinguish the first cut-off part from the following prenuclear glide or vowel was to examine the spectrogram. The boundary between the initial consonant and the following prenuclear glide or vowel lied on the first relatively dark vertical striation. For the purpose of further excluding the quality of the initial consonant, the cut-off part moved backward about 13 milliseconds. After the cut-off part had been decided, we eliminated that part and pasted the hiccup noise, whose duration was the same as the cut-off part, to the position that was originally occupied by the cut-off part, as exemplified in Figure 4.

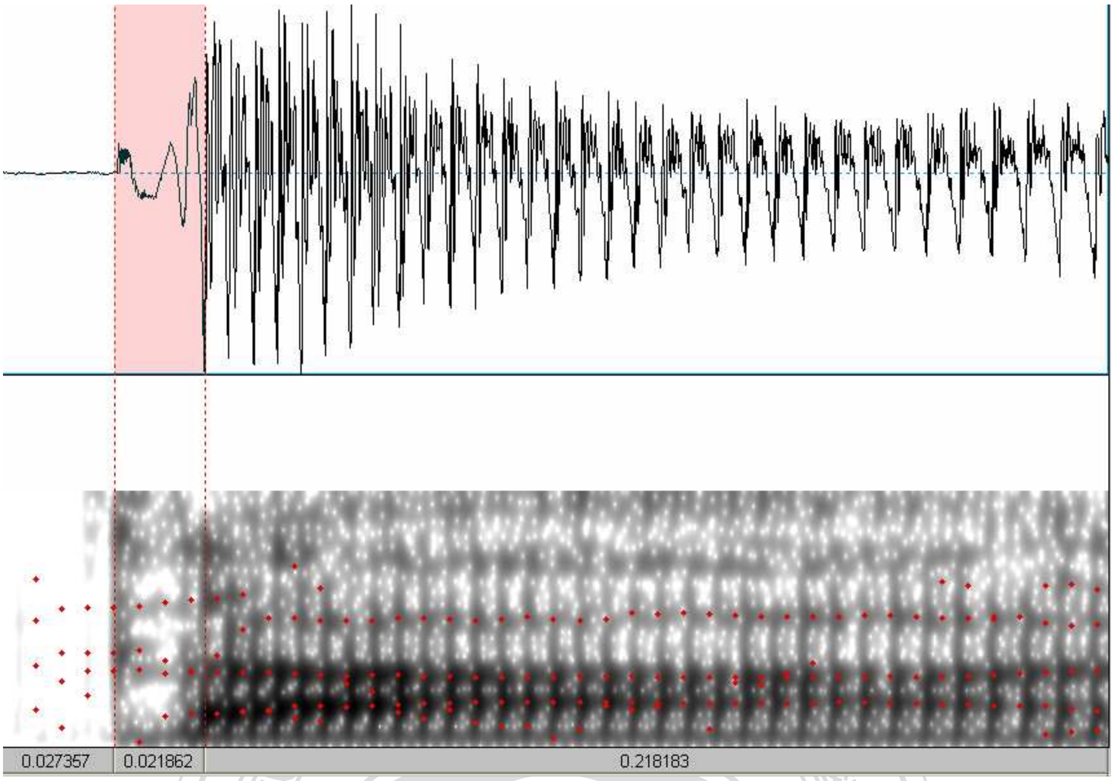

Figure 4. The marked part designates the initial consonant /t/ in /ta51 ɕɥɛ35/. In order to further eliminate the quality of /t/, the hiccup noise replaces the part starting from the first dotted line to 13 milliseconds after the second dotted line.

Figure 4 marked the initial consonant /t/ in /ta51 ɕɥɛ35/ ‘university’ (大學). The first red-dotted line, which was situated at the start of the vibration on the waveform, designated the start of the initial consonant /t/. The second red dotted line, which was located at the first relatively dark vertical striation on the spectrogram, marked the end of the initial consonant /t/. The space between the two red dotted lines was the initial consonant /t/.

3.3.2.2 Segmentation of prenuclear glides

There were two points in segmentation of prenuclear glide. One was where the boundary between the initial consonant and the prenuclear glide is; another was where the boundary between the prenuclear glide and the vowel is. The boundary between the initial consonant and the prenuclear glide could be generally defined by the waveform and the spectrogram. Nevertheless, in order to fully eliminate the quality of prenulear glide, the starting point of the cut-off part was situated slightly before the beginning of the intense vibration on the waveform or the first relatively dark vertical striation on the spectrogram. In terms of the boundary between the prenuclear glide and the vowel, it is relatively difficult to define. Chang (2009) investigated the vowels in Taiwan Mandarin acoustically. The principle for him to analyze the vowel quality of a diphthong was to examine the energy and the comparatively steady formants. In the spectrogram, the darker area represents the stronger energy; the lighter, the weaker. Vowels usually have stronger energy than prenuclear glides. In addition, the formants of the vowel are steady compared with those of the prenuclear glide. Therefore, Chang (2009) proposed that the main vowel of a diphthong should be the section having both the strongest energy and the comparatively steady formants. In this study, following Chang's (2009) methods, the boundary between the prenuclear glide and the vowel could be decided, as illustrated in Figure 5.

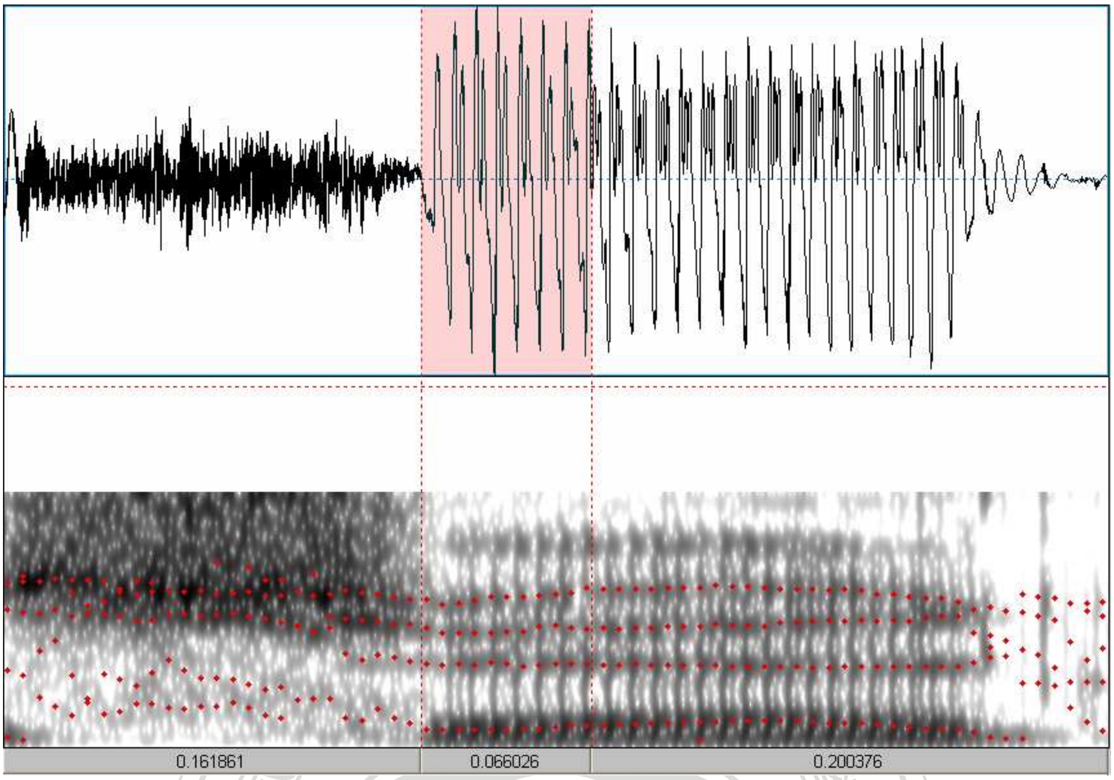

Figure 5. The marked part designates the prenuclear glide /ɥ/ in /ta51 ɕɥɛ35/. The part between the two dotted lines is then replaced by the hiccup noise.

Figure 5 marked the prenuclear glide /ɥ/ in /ta51 ɕɥɛ35/(大學, university in English). The first red dotted line, which was situated at the first relatively dark vertical striation on the spectrogram, designated the start of the prenuclear glide /ɥ/. The second red dotted line, which was located at the start of the comparatively steady formants on the spectrogram, marked the end of the prenuclear glide /ɥ/. The space between the two red dotted lines was the prenuclear glide /ɥ/.

After determining the starting point and the ending point of the prenuclear glide, the prenuclear glide was replaced by the hiccup noise, which had the same duration as the prenuclear glide.

3.3.2.3 Segmentation of vowels

The vowel in Mandarin is either preceded by the initial consonant or by the prenuclear glide. The ways to distinguish the initial consonant and the prenucleaer glide from the vowel have already been discussed above. What has yet to be discussed is about distinguishing the vowel from the postnuclear glide or final nasal. The ways to distinguish the vowel from the postnuclear glide were similar to the ways to distinguish the vowel from the prenuclear glide. The relatively dark area represents the strong energy section, which designates the position of the main vowel. The comparatively light area represents the weak energy section, which shows the position of the postnuclear glide. Furthermore, the comparatively steady formants can also designate the position of the main vowel.

The vowel in Mandarin can also be followed by the final nasal. The ways to distinguish the vowel from the final nasal was based on the acoustic cues proposed by Ladefoged (2006). There are four acoustic cues used in this study to differentiate the vowel from the final nasal. First, a clear mark of a nasal consonant is a sharp change in the spectrogram at the time of the formation of the articulatory closure. Second, the bands of the nasal are fainter than those of the vowel. Third, the first formant of the nasal consonant is usually very low, which is centered at about 250 Hz. Fourth, there is a large region above the first formant with no energy. According to the acoustic cues mentioned above, the vowel and the final nasal can be distinguished. After determining the starting point and the ending point of the vowel, the vowel was replaced by the hiccup noise, as illustrated in Figure 6.

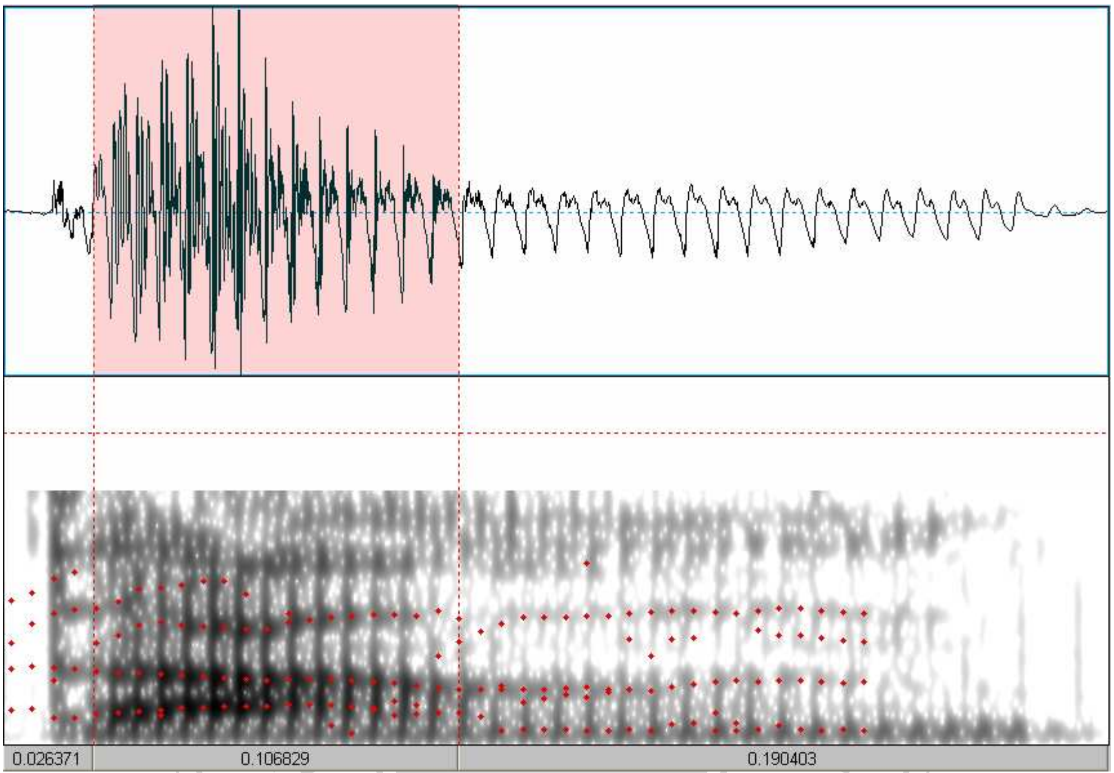

Figure 6. The marked part designates the vowel /ɑ/ in /taŋ21 t^{h}wan35/. The part between the two dotted lines is then replaced by the hiccup noise.

Figure 6 marked the vowel /a/ in /taŋ21 t^{h}wan35/ ‘political party’ (黨團). The first red dotted line, which was situated at the first relatively dark vertical striation on the spectrogram, designated the start of the vowel /a/. The second red dotted line, which was located at the sharp change on the spectrogram, marked the start of the final nasal /ŋ/. The space between the two red dotted lines was the vowel

/ɑ/.

3.3.2.4 Segmentation of postnuclear glides, and final nasals

The ways of designate the boundary between the postnuclear glide and the preceding vowel, as well as the final nasal and the previous vowel have already been addressed above. The ways of designating the ending points of the postnuclear glide and the final nasal are the same, namely, the point where the waveform shows no vibration and the spectrogram displays no energy, as illustrated in Figure 7. After determining the starting point and the ending point of the prenuclear glide or the final nasal, the section occupied by the postnuclear glide or the final nasal was cut off and was replaced by the hiccup noise with the same duration.

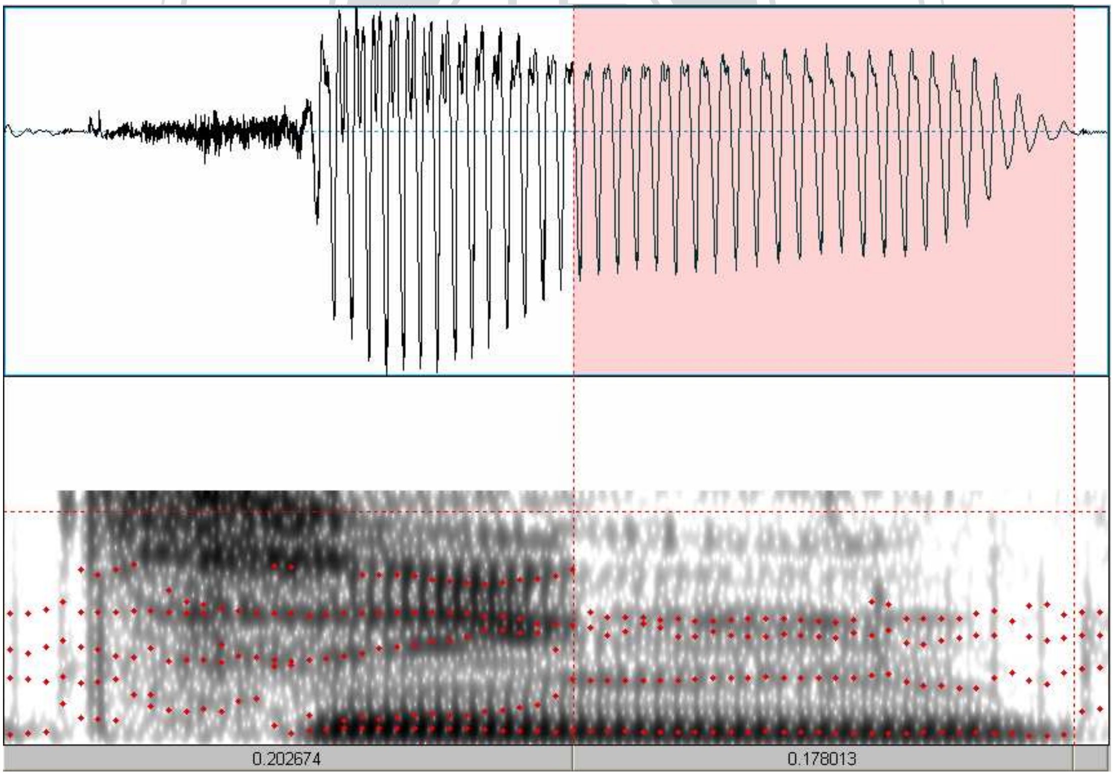

Figure 7. The marked part designates the final nasal /n/ in /xwan35 tɕin51/. The first dotted line displays the beginning of /n/; the second dotted line shows the end of /n/. This marked part is then replaced by the hiccup noise.

Figure 7 marked the final nasal /n/ in /xwan35 tɕin51/ ‘the environment’ (環境).

The first red dotted line, which was situated at the sharp change on the spectrogram, designated the start of the final nasal /n/. The second red dotted line, which was located at the place showed no vibration and energy, marked the end of the final nasal /n/. The space between the two red dotted lines was the final nasal /n/.

3.3.3 Leveling of tones

The tones of the 48 stimuli, including both high and low frequency words, were leveled, which means that the pitch contours of the disyllabic words disappear, resulting in the robot-like sounds. The F0 of the stimuli centers around 100Hz, as illustrated in figure 8.

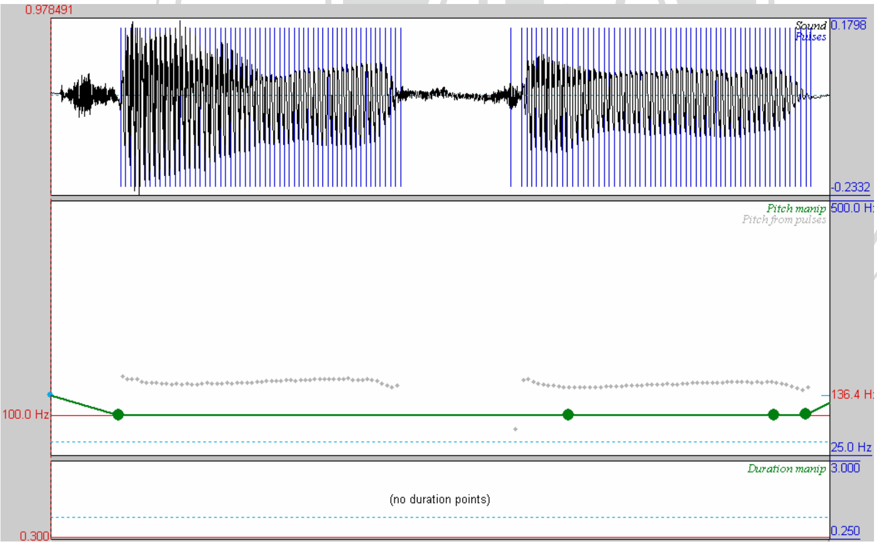

Figure 8. The test item 中心(/tʂoŋ55 ɕin55/, a center) whose tones are leveled at around 100Hz.

In this figure, the original tones of the test item 中心(/tʂoŋ55 ɕin55/, a center) were designated by the two gray dotted lines. The green straight line marked the manipulated tones, centered at 100Hz.

3.4 Design

Experiment 1

Thirty linguistically naïve subjects, 7 males and 8 females, were recruited. There were two groups of disyllabic words, including 12 high-frequency words and 12 low-frequency words, in experiment 1 (Appendix 1). In this experiment, the phonetic segments in each disyllabic word were replaced by the hiccup noise. For example, the disyllabic word /*ta51 ɕɥɛ35*/, which means university, could produce 5 test items, namely, **N***a51 ɕɥɛ35* (**N** stands for the substitution of the hiccup noise), *t***N***51 ɕɥɛ35*, *ta51* **N***ɥɛ35*, *ta51 ɕ***N***ɛ35*, and *ta51 ɕɥ***N***35*. The total number of the test items in experiment 1 was 168. The order of broadcasting the test items was at random during the whole experiment.

This experiment was designed to explore the research questions 1, 2, and 3. By measuring the reaction time and the accuracy of the responses, the importance of the initial consonant, and the status of different segments can be investigated. If the stimuli whose initial consonants are replaced by the hiccup noise can still be recognized by the subjects, it indicates what Cohort model claimed is not compatible with the findings here; if the stimuli whose particular segment is replaced by the hiccup noise produce longest reaction time and lowest accuracy, it displays that it is the particular segment that occupies the most prestigious status in the processing of

Mandarin. Moreover, this experiment can also deal with the question regarding which segment is the most influential segment for the perception of Mandarin tones. If the stimulus whose particular segment is replaced by the hiccup noise results in the highest rate of the misperception of tones, it demonstrates that the particular segment bears the most crucial tonal information in the processing of Mandarin words. If the stimulus whose particular segment is replaced by the hiccup noise causes lowest rate of the misperception of tones, it indicates that the particular segment does not carry the most critical tonal information in Mandarin.

Experiment 2

Thirty linguistically naïve subjects, 7 males and 8 females, were recruited. There were two groups of disyllabic words, including 12 high-frequency words and 12 low-frequency words, in experiment 2 (Appendix 2). In this experiment, the CV of the first syllable and the VG/N of the second syllable of each disyllabic word were replaced by the hiccup noise respectively. All 24 disyllabic words were CVX-CVX structure (X can be either a nasal or a glide). Therefore, each word yielded two test items. For example, if the disyllabic word is /tʂoŋ55 ɕin55/(meaning “a center” in Mandarin), it yields two test items, including /**NN**ŋ55 ɕin55/, and /tʂoŋ55 ɕ**NN**55/. The total number of the test items in experiment 2 was 48. The order of broadcasting the test items was at random during the whole experiment.

This experiment was aim to resolve the question regarding the role of onsets and offsets in research question 1. If subjects' responses to the stimuli with the initial CV replaced by the hiccup noise are faster and more accurate than those with the final rime replaced by the hiccup noise, it illustrates that the onsets are more important than the offsets in the processing of Mandarin, which implies that initial effect happens in spoken word recognition in Mandarin. In contrast, if subjects' responses to the stimuli with the final rime replaced by the hiccup noise are faster and more accurate than those with the initial CV replaced by the hiccup noise, it depicts that the offsets play more crucial role in the processing of Mandarin.

Experiment 3

Thirty linguistically naïve subjects, 7 males and 8 females, were recruited. There were two groups of disyllabic words, including 24 high-frequency words and 24 low-frequency words. The 48 disyllabic words were the same as those in experiment 1 and 2, but they were manipulated differently. In this experiment, the tones of the 48 disyllabic words were all leveled to 100 Hz, like low robotic sounds. Each disyllabic word yielded one test item. For example, the disyllabic word /tʂoŋ55 ɕin55/(meaning "a center" in Mandarin) yielded the test item /tʂoŋ**100Hz** ɕin**100Hz**/. The total number of the test items in experiment 3 was 48. The order of broadcasting the test items was at random.

Experiment 3 was designed to tackle the questions concerning the role of tone in research question 3. By measuring subjects' responses to the tone-leveled stimuli, the status of tone can be investigated. This experiment also handles the issue concerning whether the segmental information alone is enough for the spoken word recognition in Mandarin.

The issue regarding frequency effect was explored in all three experiments. In the three experiments, responses of high-frequency words and low-frequency words were compared. If the responses of the high-frequency words are faster and more accurate than those of the low-frequency words, it means that frequency effect plays a role in spoken words recognition in Mandarin.

3.5 Procedures

Fifteen subjects, participating in experiment 1, 2, and 3, were invited to a quiet and comfortable room one by one and sat in front of the computer. The experimenter told them about how the experiment was going and gave them two examples to practice. After practicing, the subjects were asked to listen to the test items one after another. The subjects had to identify what the disturbed word was. The subjects could interrupt the utterance anytime they recognized the target word. Once the subjects recognized the words, they needed to repeat the words and then write them down in order to make sure that the subjects' responses were lexical words rather than nonwords. If they could not recognize the words, they could also say “I don’t know”. The reaction time was measured from the end of the target word to the beginning of the subjects’ utterances. All of the test items were broadcasted once. If the subjects could not recognize the target words or recognized the target words incorrectly, the reaction time of the target words was designated as “fail.”

CHAPTER 4 Results

In this chapter, the results of the three experiments will be presented. Section 4.1 displays the results of experiment 1 with some possible explanations. Section 4.2 shows the results of experiment 2 with some possible solutions. In Section 4.3, the results of experiment 3 are demonstrated with brief discussions. The reasons why the results occurred are also explained.

4.1 Experiment 1: one-segment disruption

The results of experiment 1 are shown in Table 1 and 2.

Table1. High frequency words: one-segment disruption

Position	1C	1Pre	1V	1Po	1N	2C	2Pre	2V	2Po	2N
RT(msec.)	596	616	659	661	591	581	615	659	554	551
Pass	137	90	152	15	78	178	101	169	15	41
Fail	7	0	28	0	6	2	1	11	0	1
Total	144	90	180	15	84	180	102	180	15	42
Acc(%)	95.14	100	84.44	100	92.86	98.89	99.02	93.89	100	97.62
Invalid	6	0	0	0	6	0	3	0	0	3

(Position=the position replaced by the hiccup noise
RT=the average reaction time
Acc=accuracy
1C=the initial consonant in the first syllable replaced by the hiccup noise
1Pre=the prenuclear glide in the first syllable replaced by the hiccup noise
1V=the vowel in the first syllable replaced by the hiccup noise
1Po=the postnuclear glide in the first syllable replaced by the hiccup noise
1N=the final nasal in the first syllable replaced by the hiccup noise
2C=the initial consonant I in the second syllable replaced by the hiccup noise
2Pre=the prenuclear glide in the second syllable replaced by the hiccup noise
2V=the vowel in the second syllable replaced by the hiccup noise
2Po=the postnuclear glide in the second syllable replaced by the hiccup noise
2N=the final nasal in the second syllable replaced by the hiccup noise)

Table 1 shows the results of experiment 1 (high-frequency words). According to the table, the first row designates the position replaced by the hiccup noise. The first column illustrates the reaction time (written in millisecond), the number of the test items which are successfully recognized by the subjects (Pass), the number of the test items which weren't recognized by the subjects (Fail), the total number of the test items (Total), the rate of the test items which can be correctly recognized (Acc%), and the number of the invalid responses (Invalid). For instance, the number situated in the second row and the second column is 596. This means that in average subjects need 596 milliseconds after the end of the targets to recognize the targets whose initial segments of the first syllable are replaced by the hiccup noise. The data located in the fourth row and the second column is 7. This means that there are 7 test items whose initial segment of the first syllable is replaced by the hiccup noise not able to be recognized or correctly recognized. In addition, the table displays that 1Po has the longest reaction time. 1V and 2V have the second longest reaction time. 2N has the shortest reaction time. Last but not least, the lowest accuracy of the test items is 84.44%, nestled in the 1V column.

In this table, it is obvious that the vowel in the first syllable is the most important segment in the processing of spoken words. The words whose first vowel is replaced by the hiccup noise need 659 milliseconds after the end of the stimuli to be recognized. The reaction time of the stimuli whose first vowel is replaced by the hiccup noise is just slightly faster than that of the stimuli whose first postnuclear glide is replaced by the hiccup noise (661 ms) and the same as that of the stimuli whose second vowel is replaced by the hiccup noise (659 ms). Although the reaction time is the second longest when the first vowel is replaced by the hiccup noise, there are 28 test items that cannot be correctly recognized by the subjects. The accuracy for 1V is 84.44%, which is much lower than the accuracy for 1Po (100%) and the accuracy for 2V (93.89%). Consequently, the first vowel in the disyllabic word is the most important in spoken word recognition in Taiwan Mandarin.

Table2. Low frequency words: one-segment disruption

Position	1C	1Pre	1V	1Po	1N	2C	2Pre	2V	2Po	2N
RT(msec.)	632	640	715	690	694	625	640	706	622	601
Pass	145	100	141	26	101	147	41	141	28	88
Fail	18	3	37	0	10	17	1	36	0	1
Total	163	103	178	26	111	164	42	177	28	89
Acc(%)	88.96	97.09	79.21	100	90.99	89.63	97.62	79.66	100	98.88
Invalid	2	2	2	4	9	1	3	3	2	1

Table 2 shows the results of experiment 1, including low-frequency words. According to the table, the first row designates the position replaced by the hiccup noise. The first column illustrates the reaction time (written in millisecond), the number of the test items which are successfully recognized by the subjects (pass), the number of the test items which cannot be recognized by the subjects (fail), the total number of the test items, the rate of the test items which can be correctly recognized, and the number of the invalid responses. For instance, the data situated in the second row and the second column is 632. This means that in average subjects need 632 milliseconds after the end of the targets to recognize the test items whose initial segments of the first syllable are replaced by the hiccup noise. The data located in the fourth row and the second column is 18. This means that there are 18 test items whose initial segment of the first syllable is replaced by the hiccup noise not able to be recognized or correctly recognized. In addition, the table displays that 1V has the longest reaction time; 2N has the shortest reaction time. Last but not least, the lowest rate of the unrecognizable test items is about 79%, nestled in the 1V and 2V columns.

The results of the low frequency words show the similar results as the high frequency words, which indicates that the vowel in the first syllable is the most important for spoken word recognition and the vowel in the second syllable is the second crucial segment in the processing of Mandarin words. The results display that the 1V stimuli need 715 milliseconds to be correctly recognized by the subjects, which takes the longest reaction time. The 2V stimuli need 706 milliseconds to be successfully recognized, which takes the second longest reaction time. The results also illustrates that there are 37 test items (Accuracy: 79.21%) which cannot be identified correctly because of the disruption of the first vowel and 36 test items

(Accuracy: 79.66%) which cannot be recognized successfully when the vowel in the second syllable is corrupted. The longest reaction time and lowest accuracy for both 1V and 2V indicate that the first vowel and the second vowel are very important, so subjects need more time to identify the word whose first and second vowel are disruptive. Furthermore, similar to the results of high-frequency words, the results of low-frequency words show that the segments in the first syllable are more important (longer reaction time and lower accuracy) than their corresponding segments in the second syllable. This finding suggests that the perceived order in time has some effect on the spoken word recognition in Mandarin.

As for the frequency effect, the results depict that subjects have more difficulties in recognizing the low frequency words. The results show that the disruptive segments of the low-frequency words cause longer reaction time and lower accuracy compared with their corresponding disruptive segments of the low-frequency words. This result shows that frequency effect appears here. Therefore, it can be inferred that frequency effect exists in spoken word recognition in Mandarin.

Table 3. Incorrect responses of tones (high-frequency words): one-segment disruption

Position	1C	1Pre	1V	1Po	1N	2C	2Pre	2V	2Po	2N
Fail	0	0	7	0	0	0	0	3	0	0
Total	3	0	13	0	4	1	1	5	0	0
Percentage	0	0	53.85	0	0	0	0	60	0	0

Table 4. Incorrect responses of tones (low-frequency words): one-segment disruption

Position	1C	1Pre	1V	1Po	1N	2C	2Pre	2V	2Po	2N
Incorrect	0	0	4	0	1	0	0	2	0	0
Total	7	2	14	0	6	9	0	10	0	0
Percentage	0	0	28.57	0	16.67	0	0	20	0	0

Table 3 and Table 4 display the incorrect perception of tone among the incorrect responses in experiment 1. The incorrect responses here mean that subjects did say a word when they heard the particular stimulus, but the tone of the response to the particular stimulus was wrong. From these two tables, we know that vowels carry most tonal information in Mandarin, so when the vowels are replaced by the hiccup noise, the percentages of the incorrect responses of tones are higher. It is also noticeable that there is one misperception of tone of 1N. Although coda nasal dose not occupy a long period of time in words, it still carries tonal information because it belongs to rime. Therefore, tone can still be misperceived when the coda nasal is replaced by the hiccup noise.

4.2 Experiment 2: two-segment disruption

The results of experiment 2 are displayed in table 2.

Table 5. Two-segment disruption

Position	H_CV	H_VG/N	L_CV	L_VG/N
RT(msec.)	902	924	1139	1117
Pass	85	98	36	23
Fail	95	80	144	157
Total	180	178	180	180
Acc(%)	47.22	55.06	20	12.78
Invalid	0	2	0	0

(Position=the position replaced by the hiccup noise

RT=reaction time
Acc=accuracy
H_CV=high-frequency words with the initial CV replaced by the hiccup noise
H_VN/G=high-frequency words with final VG/N replaced by the hiccup noise
L_CV=low-frequency words with initial CV replaced by the hiccup noise
L_VG/N=low-frequency words with final VG/N replaced by the hiccup noise)

Table 5 shows the results of experiment 2. According to the table, the first row designates the parts of the test items replaced by the hiccup noise. The first column illustrates the reaction time (written in millisecond), the number of the test items which are successfully recognized by the subjects (pass), the number of the test items which cannot be recognized by the subjects (fail), the total number of the test items, and the rate of the test items which are recognized correctly. For instance, the data situated in the second row and the second column is 902 milliseconds. This means that in average subjects need 902 milliseconds after the end of the targets to recognize the targets whose initial CVs are replaced by the hiccup noise. The data located in the fourth row and the second column is 95 milliseconds. This means that there are 95 test items whose initial CVs are replaced by the hiccup noise not able to be recognized or correctly recognized. In addition, the table displays that initial CV has the shorter reaction time for high-frequency words compared with the final rime, and initial CV has slightly longer reaction time for low-frequency words compared with the final rime. Last but not least, the lower percent of the accurate responses is 47.22% for high-frequency words, nestled in the H_CV column, and 12.78% for low-frequency words, seated in the L_VG/N column.

From the results above, it is clear that the reaction time of the stimuli whose CV and VG/G are replaced by the hiccup noise is not greatly different. It implies that the CV of the first syllable and the VG/N of the second syllable are almost the same important, which means that the onsets and offsets of the disyllabic words play the same role in the processing of Mandarin words. However, it seems to be a paradox that the accuracy of CV is lower than that of VG/N for high-frequency words, but higher than that of VG/N for low-frequency words. It may be due to the fact that some high-frequency stimuli whose CV of the first syllable is replaced by the hiccup noise activate very prominent candidates. Those prominent candidates are very easy to be selected by the subjects, resulting in the lower accuracy of CV for high-frequency words. To the contrary, concerning the low-frequency words, the stimuli whose VG/N in the second syllable is disruptive activate some prominent candidates which are very easy to be selected by the subjects. Therefore, it causes lower accuracy of VG/N for low-frequency words. The more prominent activated words can be proved by the incorrect responses of the subjects. The incorrect responses mean that the subjects did say a word when they heard a stimulus, but the word is not the correct one. Those incorrect responses are the candidates activated by the stimuli, which can disturb the correct selection of the target. The numbers of the incorrect responses of H_CV, H_VG/N, L_CV, and L_VG/N, are 56 (56/180, 31.11%), 46 (46/178, 25.84%), 81 (81/180, 45%), and 84 (84/180, 46.67%), respectively. The more incorrect responses of H_CV and L_VG/N may give the answer to the paradox regarding why the accuracy of CV for high-frequency words is lower than that of VG/N, but higher for low-frequency words than that of VG/N.

Furthermore, frequency effect reveals here. The reaction time of the high frequency words is shorter than that of the low frequency words and the accuracy of CV together with VG/N for high-frequency words is greatly higher than that for low-frequency words.

Table 6. Incorrect responses of tones: two-segment disruption

Position	H_CV	H_VG/N	L_CV	L_VG/N
misperception	15	34	12	57
Total	56	46	81	84
Percentage	26.79	73.91	14.81	67.86

Table 6 shows the incorrect responses of tones by the subjects. “Total” means the overall number of the incorrect responses. “Misperception” means that the incorrect responses are not only wrong regarding the segmental level, but also tonal level. Concerning the misperception of tones, there are 118 incorrect perceptions of tones, including 15 for H_CV, 34 for H_VG/N, 12 for L_CV, and 57 for L_VG/N. The incorrect tone perception most frequently happens when the vowel is interrupted. If both vowel and postnuclear glide/coda nasal are disrupted, the rate of the misperception of tones will be even higher since postnuclear glide and coda nasal also carry tonal information, while the onset consonant does not.

The results indicated that tone 4 is most likely to be misperceived if some part of tone 4 is disruptive and the other tones can also be misperceived as tone 4. Among the 118 misperceptions of tones, 48 involve tone 4 (40.68%), 30 involve tone 1 (25.42%), 20 involve tone 2 (25.42%), and 20 involve tone 3 (25.42%). This may be due to the fact that tone 4 ranges from low pitch to high pitch. If the final part of the tone 4 is replaced by the hiccup noise, the high pitch at the beginning of the tone 4 can be misperceived as tone 1 or 2. For instance, the test item /min35 tʂoŋ51/ 'the common people' (民眾), as illustrated in Figure 9, is perceived as /lin35 tɕy55/ 'the neighbor' (鄰居) when the second rime of the test item is replaced by the hiccup noise.

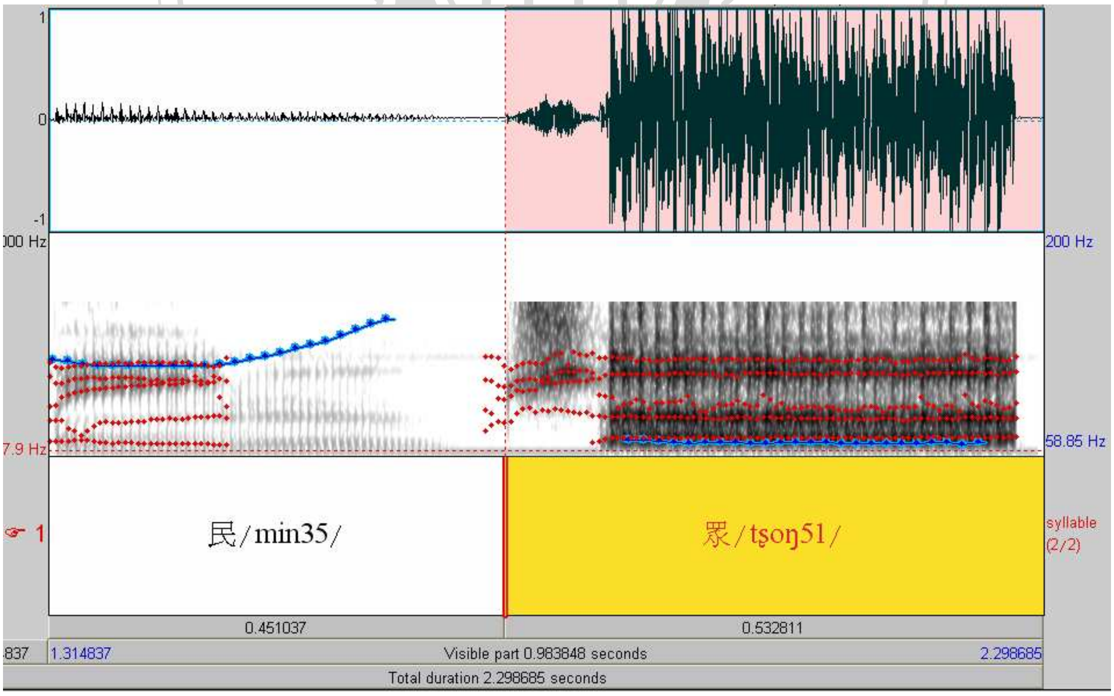

Figure 9. The test item /min35 tʂoŋ51/ 'the common people' (民眾) whose second rime is replaced by the hiccup noise

The original tone 4 of the second syllable is misperceived as tone 1. It is because the pitch of the beginning of tone 1 is similar to that of tone 4. Thus, when the final part of tone 1 is replaced by the hiccup noise, the original tone 1 may be misperceived as tone 4. However, more complicated situation occurs. For instance, the test item /fan51 t^{h}saj51/ ‘a meal’ (飯菜) is misperceived as /fan51 t^{h}ʂɑŋ21/ (a nonword) when the final rime of the test item is replaced by the hiccup noise. It is because the final part of the first tone 4 is a low-falling pitch. The beginning of the second tone 4 is a high pitch. The combination of the final part of the first tone 4 and the initial part of the second tone 4 results in a full tone 3, which forms the impression that the tone of the second syllable is 3 rather than 4.

In addition, if the initial part of the tone 4 is replaced by the hiccup noise, the low pitch at the end of the tone 4 can be misperceived as tone 3. For example, the test item /nej51 ʐoŋ35/ ‘contents’ (內容), as exemplified in figure 10, is perceived as /mej21 ʐoŋ35/ ‘to improve one’s looks’ (美容) when its initial CV is replaced by the hiccup noise. The tone of the first syllable is perceived as tone 3 rather than the original tone 4. This is because the low-falling pitch of the final part of tone 4 gives the impression that it is tone 3 instead of tone 4. Therefore, tone 4 is the easiest tone to be misperceived if it is partly disruptive and the other tones are also easy to be misperceived as tone 4.

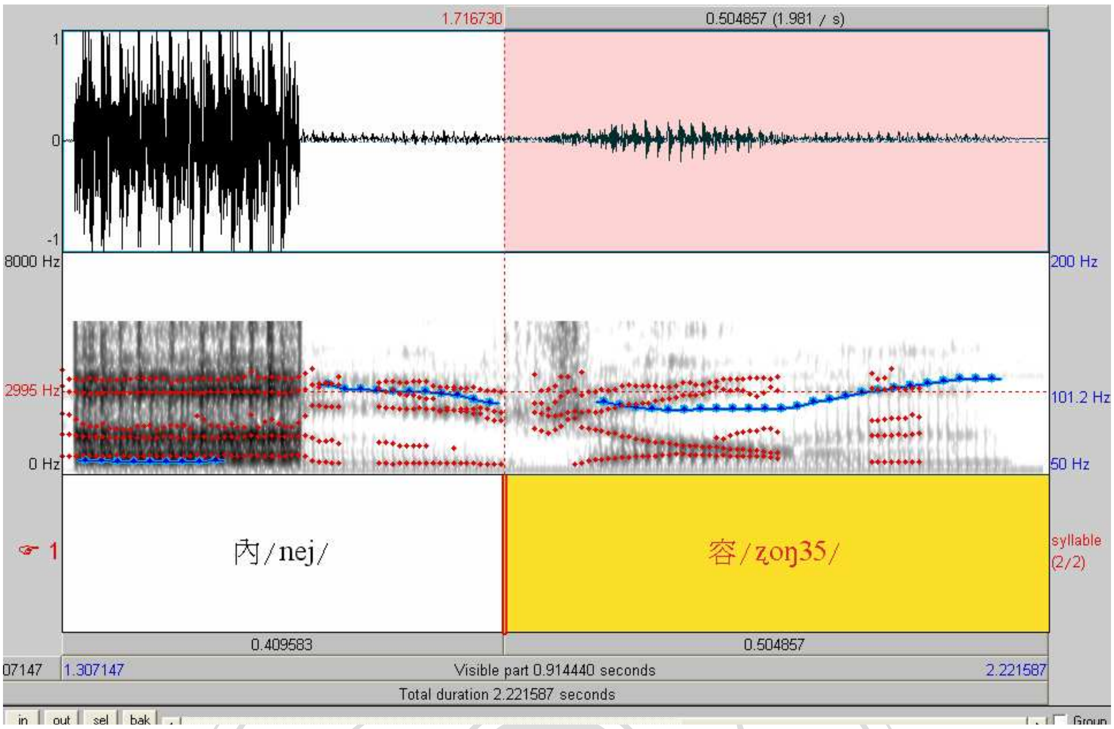

Figure 10. The test item /nej51 ʐoŋ35/ 'contents' (內容) whose initial CV is replaced by the hiccup noise

Apart from the misperception of tone 4, tone 1 is the second easiest tone to be misperceived. There are 30 test items whose original tone is tone 1 and is perceived as tone 3. The tone 2 test items whose initial CV is replaced by the hiccup noise are easy to be perceived as tone 1. It is because tone 1 is a high-level tone. The pitch of tone 2 rises from mid to high. Once the initial part of tone 2 is replaced by the hiccup noise, the high pitch at the end of tone 2 is easy to be perceived as tone 1, which is a high tone. However, there are also some test items perceived as tone 1 whose second VG/N is replaced by the hiccup noise. For example, the test item /piŋ35 tʂəŋ51/ 'a certificate' (憑證) is perceived as /tiŋ35 t^hʂɤ55/ 'to park a car' (停車). It may result from the high pitch at the beginning of the second syllable. The high pitch of tone 4 at the beginning of the second syllable is similar to the high pitch tone 1. Tone 4 is a high-falling tone.

When the final part of tone 4 is replaced by the hiccup noise, the beginning high pitch of tone 4 is likely to be perceived as tone 1. Therefore, tone 4 can also be perceived as tone 1under the above mentioned circumstances.

In short, tone 4 is most likely to be misperceived as the other tones if some part of it is replaced by the hiccup noise, and the other tones are also easy to be misperceived as tone 4 if some part of them is disruptive. The reason is that tone 4 has the widest pitch range.

4.3 Experiment 3: tone leveling

The results of experiment 3 are illustrated in Table 7.

Table 7. Results of tone-leveled stimuli

	High Frequency words	Low Frequency words
RT(msec.)	693	970
Pass	258	202
Fail	102	154
Total	360	356
Acc(%)	71.67	56.74
Invalid	0	4

Table 7 displayed the results of experiment 3. In the experiment, the test items were deprived of their tonal contour and were given a new level tone which centered at around 100 Hz. The results indicated that the rates of correctly recognizing the test items are very low. Only 71.67 percent of the test items, namely 258 out of 360, can be correctly recognized for the high frequency words given a new level tone. The rate of correctly recognizing the test items of the low frequency words is even lower than that of the high frequency words. Only 56.74 percent of the test items, namely 202 out of 356, can be recognized appropriately. The results depicted that the rates of correctly recognizing the test items whose tones are leveled are lower compared with the rates of correctly recognizing the test items whose single segment is disruptive. The reason why the tone-leveled stimuli result in the lower accuracy and longer reaction time than the stimuli whose single segment is replaced by the hiccup noise is because the manipulated tone of the tone-leveled stimuli spans the entire word. The stimuli whose single segment is replaced by the hiccup noise only have a short disrupted part though the disrupted part may carry tone. Comparing the reaction time and accuracy of the stimuli whose CV or rime is replaced by the hiccup noise with the reaction time and accuracy of the tone-leveled stimuli, it reveals that the reaction time and accuracy of the stimuli whose CV or rime is disruptive are longer and lower. It is because the stimuli whose CV or rime is disrupted not only lack the segmental information but also tonal information. It is very hard for subjects to recognize a disyllabic word lacking the acoustic information of almost a whole syllable. Although the tonal information of the tone-leveled stimuli is disruptive, the segmental information can still play a role in recognizing the words. Thus, the tone-leveled stimuli cause longer reaction time and lower accuracy than the stimuli whose single segment is replaced by the hiccup noise, but result in shorter reaction time and higher accuracy than the stimuli whose CV or rime is corrupted.

Considering the frequency effect in this experiment, it is very obvious that the accuracy is much lower for low-frequency words (56.74%) than for high-frequency words (71.67%). Moreover, the reaction time for low-frequency words (970 ms) is much longer than that for high frequency words (693 ms). Thus, frequency effect still exists in this experiment.

In short, the results of experiment 1 show that the vowel in the first syllable is the most important in spoken word recognition in Mandarin because the 1V disrupted stimuli cause lowest accuracy and long reaction time. The results of experiment 2 display that the CV in the first syllable and the VG/N in the second syllable are almost the same influential since the initial CV disrupted stimuli and final VG/N disrupted stimuli bring about nearly the same reaction time and accuracy. The results of experiment 3 demonstrate that tone is more important than a single segment but less crucial than two segments in the processing of spoken words because the tone leveled stimuli result in the lower accuracy and longer reaction time than the single-segment-disrupted stimuli; the tone leveled stimuli cause higher accuracy and shorter reaction time than the two-segment-disrupted stimuli. Frequency effect appears in the three experiments; namely, the stimuli of high-frequency words are easier to be recognized by the subjects than those of the low-frequency words. The results will be further explained based on the Cohort model and Merge model in the next chapter.

CHAPTER 5 Discussion

In this chapter, the results are investigated based on the two models, Cohort model and Merge model. Section 5.1 will be the connection between the results and the two models, Cohort and Merge. Section 5.2 concerns the validity of the Cohort model and Merge model in the spoken word recognition in Taiwan Mandarin. Section 5.3 provides a simple demonstration for spoken word recognition in Taiwan Mandarin under the framework of Merge model.

5.1 The results and the two models (Cohort and Merge)

According to the results of experiment 1, it is obvious that there are some problems regarding Cohort model. In terms of the initial consonant replaced by the hiccup noise in the first syllable for high-frequency words, the reaction time is not very long (596 ms) and there are only seven test items (4.86%) that cannot be recognized by the subjects, which means that most of the test items can still be successfully recognized. As for the results of the low-frequency words, the average reaction time of 1C is 632 milliseconds and the accuracy of 1C is 88.96%. It shows that most of the targets can still be recognized though the initial consonant of the first syllable is disruptive. This finding infers what Cohort theory predicts is not right because the words can still be correctly recognized even if the initial consonant in the first syllable is replaced by the hiccup noise. However, although the reaction time is not the longest when the initial consonant of the first syllable is replaced by the hiccup noise, there are still some test items which cannot be recognized correctly. This may infer that the initial consonant still plays a role in the processing of spoken words though it is not the most important.

Although the results above are not compatible with Cohort theory, they are not completely compatible with the Merge model, which indicates that the overall match between the input and the lexical representation is the most important. It is the vowel in the first syllable that is the most important. It may be due to the fact that the duration of the vowel is the longest. The replacement of the vowel gives rise to the longest interruption of the word. In addition, vowels carry the most important information in the syllable, including the tone. The lost of vowels brings about the lost of tones, which is very critical in Taiwan Mandarin. Moreover, from the results that the reaction time and the accuracy of 1C, 1Pre, 1V, 1Po, together with 1N are longer and lower than the reaction time and the accuracy of 2C, 2Pre, 2V, 2Po along with 2N, we can infer that the perceived order of the spoken words is very important. The vowel in the first syllable not only carries much acoustic information, but also occupies the front position in disyllabic words. Therefore, we can propose that the vowel in the first syllable is the most important segment in processing of the disyllabic words in Taiwan Mandarin. Although the vowel in the second syllable also carries the tone and occupies a relatively long period of time in the disyllabic words, it lacks the advantage of being processed first. Hence, the vowel in the second syllable is less influential than the vowel in the first syllable.

5.2 Cohort and Merge models in Taiwan Mandarin

According to the Cohort model (Marslen-Wilson & Zwitserlood, 1989; Tyler, 1984; Marslen-Wilson & Tyler, 1980, 1981), word initial input is of paramount importance. Therefore, once the initial input is disrupted by the noise, the word can hardly be recognized. This claim is not true on the basis of the results in this study. The results of experiment 1 display that the accuracies of 1C for the high-frequency and low-frequency words are 95.14% and 88.96%, respectively. This demonstrates that even if the word initial information (the initial consonant) is replaced by the hiccup noise, the words can still be successfully recognized in most of the cases.

Unlike the Cohort theory, the Merge model (Norris, McQueen, and Cutler, 2000) proposed that it is the overall match between the acoustic input and the representations that is the most crucial for spoken word recognition. This model greatly reduces the importance of the initial input. However, the results of the study are not fully compatible with the model. The results depict that it is the first vowel in the disyllabic words that is the most important and the second vowel in the disyllabic words that is the second important, but it is not stated in the model which segment is the most important. Therefore, the test items whose first vowels are replaced by the hiccup noise cause the lowest rate of successful recognition.

According to the results, the word initial information is not the most crucial in the spoken word recognition of Taiwan Mandarin. The first vowel in the disyllabic word is the most important. This is because the vowel occupies the longest period of time in words and carries much important information which is very crucial for spoken word recognition in Taiwan Mandarin. One of the important acoustic-phonetic cues in vowel is tone. Tones are very important in Taiwan Mandarin since it can distinguish the meaning of words. The results of the study are compatible with this claim. The results of experiment 3 show that if the tones of the disyllabic words are leveled to around 100Hz, subjects merely have 71.67 percent chance to recognize the high-frequency words and 56.74 percent chance to recognize the low-frequency words, which is much lower than the chance to recognize the words whose single segment is replaced by the hiccup noise. The fact indicates that the whole tone of the disyllabic words is more important than one single segment though some segments also carry tone, such as the vowel and coda nasal.

According to the results in experiment 3, tone should be added to the processing of spoken words in Merge. Nevertheless, it cannot be inferred only by the experiments in this study whether Mandarin tones should be processed before the segments or after the segments. Cutler and Chen (1997) asked the subjects to judge whether the word and nonword in a pair differing only by the initial consonant, vowel, or tone were the same or different. The results displayed that subjects' responses to the pair differing by tone were slower and more inaccurate than those differing by the onset consonant and vowel. Therefore, they proposed that tone is processed slower than segment. According to Cutler and Chen (1997), it can be proposed that tonal level can be added to the Merge model after the phoneme level. The acoustic-phonetic cues of tones can be processed in tonal level and sent to the lexical nodes by the excitatory connections. The candidates activated in the lexical level not only need to match the segmental information but the tonal information of the input as well. The candidate having the best match to the acoustic input wins the lexical competition.

In addition, the results of experiment 1 can serve as the support for Merge. In merge, the phoneme decision level is designed to resolve the issues regarding phoneme decision making. The integration approach of Merge allows the prelexical information to proceed independently of lexical processing. Both prelexical and lexical processing information proceed to the phoneme decision level and then merge together. In the Merge model, the prelexical processing activates some compatible lexical candidates. At the same time, the prelexical processing also sends the excitatory information to the phoneme decision level. The phoneme decision nodes also keep accepting the facilitatory information from the lexical nodes and merge the two inputs from different levels together. The merged information competes with each other by inhibitory connections and decides which phonemes are actually present in the input.

Both Merge and TRACE (McClelland & Elman, 1986) can account for why the disruptive targets can still be recognized in experiment 1. However, TRACE would overlook the disruptive segment because the interactive models run the stake of hallucinating. Especially when the input is degraded or disruptive, the input information tends to be abandoned. In TRACE, phoneme decision can mainly depend on the lexical information from the lexical level. This is because top-down activation can function as the distortion to the prelexical processing of the acoustic input. The strong top-down feedback would override the disruptive segment and the disruptive segment could be ignored. For example, the hiccup noise in /**N**a51 ɕ$^{\text{ɥ}}$ɛ35/ would be overlooked because of the strong top-down feedback. In reality, subjects can still notice the hiccup noise. In contrast, the hiccup noise would not be overlooked in Merge. The prelexical phoneme nodes are independent of the lexical nodes; that is, there is no top-down feedback from the lexical nodes to the prelexical level. The prelexical nodes accept the hiccup noise and keep sending the prelexical processing information of the following segments to the lexical nodes. The lexical nodes then activate the possible lexical candidates and send excitatory information to the phoneme decision nodes. Therefore, although phoneme decision nodes cannot receive the excitatory information of the segment replaced by the hiccup noise from the prelexical nodes to decide what the disrupted segment is, they can still accept the information from the lexical nodes and do the phoneme decision. Contrary to TRACE, Merge can do the phoneme decision without overlooking the hiccup noise. Hence, Merge is a better model than TRACE in this facet.

5.3 Merge model: Spoken word recognition in Taiwan Mandarin

Given that the results of this study display that tones are more important than a single segment in the recognition of spoken words in Taiwan Mandarin, the information of tones should be added in the spoken word recognition models. In this section, Merge model is used in the spoken words recognition in Taiwan Mandarin, but not Cohort model because Cohort model implies that if the word initial information is disruptive, the word cannot be recognized, which is wrong. Therefore, only Merge model is used in this section.

In the Merge model, there are three kinds of nodes, including phoneme nodes, lexical nodes, and phoneme decision nodes. In the beginning, a sequence of phonemes is activated based on the acoustic-phonetic input. Then the phonemes send the excitatory information into the lexical nodes. Some possible words are activated and they compete against each other through the inhibitory networks. The phoneme decision nodes are designed to resolve the problems concerning phoneme monitory and phoneme restoration. The model functions as follows in the spoken word recognition in Taiwan Mandarin.

For example, the input /ta51 ɕɥɛ35/ activates some phonemes first and then these phonemes send excitatory information to the lexical nodes and activate some possible candidates having the same acoustic information as some part of the input. Therefore, the following lexical candidates are activated, including /ta51/ ‘big’ (大), /ɕɥɛ35/ ‘learn’ (學), /a51/ ‘exclamation marker’ (啊), /ta51 ɕɥɛ35/ ‘university’ (大學), and so on. For clarity, the candidates which merely partially match the input, such as /la51/ ‘spicy’ (辣) and /tʰɕɥɛ35/ ‘lame’ (瘸), are not shown here. After the candidates are activated, they compete against each other. The candidate having the most overlapping phonemes with the input has the highest activation level. In this case, the word /ta51 ɕɥɛ35/ ‘university’ (大學) has the highest activation level.

The illustration mentioned above is the normal situation, in which the input is not disturbed by the noise. What if the situation in which a part of the input is disturbed by the noise? The input /ta51 ɕɥɛ35/ ‘university’ (大學) whose initial consonant is replaced by the hiccup noise can still activate a sequence of phonemes and the phonemes send excitatory information to the lexical nodes, activating a shortlist of candidates, such as /ɕɥɛ35/ 'learn' (學), /a51/ 'exclamation marker' (啊), /ta51/ 'big' (大), /ta51 ɕɥɛ35/ 'university' (大學), and so forth. Although the initial consonant of the input is disruptive, the other parts of the input can still support the possible candidates after they are sent into the lexical nodes. The candidate /ta51 ɕɥɛ35/ 'university' (大學) matching most of the phonemes with the input has the strongest inhibitory power. However, in experiment 2, as more input information is replaced by the hiccup noise, there are less acoustic clues to activate candidates. Sometimes the input information is not enough for spoken word recognition, so the stimuli cannot be recognized. According to the results of experiment 2, two consecutive disruptive segments of disyllabic words would be enough to result in the serious problem of spoken word recognition in Taiwan Mandarin.

Experiment 3 proves that segmental information alone is not sufficient for spoken word recognition in Taiwan Mandarin sometimes. Take /tɕjɑŋ21 ɕi35/ 'lecture' (講習) as an illustration. The words in experiment 3 are deprived of its original tone and given a new tone centering around 100Hz. In Merge model, when the modified input /tɕjɑŋ100Hz ɕi100Hz/ enters, it activates a set of possible candidates in the lexical level, including /tɕjɑŋ21 ɕi35/ 'lecture' (講習), /tɕjɑŋ55 ɕi55/ 'a province in Mainland China' (江西), /tɕjɑŋ51 ɕin55/ 'ingenuity' (匠心), and so forth. These candidates have a variety of tones. According to Merge model, the words /tɕjɑŋ21 ɕi35/ ‘lecture’ (講習) and /tɕjɑŋ55 ɕi55/ ‘a province in Mainland China’ (江西) have the same inhibition power because the segments of the acoustic input map perfectly to the input /tɕjɑŋ100Hz ɕi100Hz/. However, frequency effect may play a role here. The word having higher frequency has higher chance to be recognized. In this case, the frequency of /tɕjɑŋ55 ɕi55/ ‘a province in Mainland China’ (江西), 22 occurrences out of 5 million tokens, is just a little higher than that of /tɕjɑŋ21 ɕi35/ ‘lecture’ (講習), 17 occurrences out of 5 million tokens, so it is hard to say which candidate finally wins. Therefore, Merge model has to take tones into account, or it cannot correctly select the particular word. Both segmental information and tonal information need to be considered in Merge. Segmental information alone is not enough for spoken word recognition in Taiwan Mandarin.

CHAPTER 6　Conclusion

The current study has demonstrated that traditional Cohort model cannot be fully supported because words can still be correctly recognized when word initial information is disruptive. In general, the overall match between the input and the lexical representation plays an important role. However, the Merge model, which proposes that the overall match between the input and the lexical representation is the most important, also cannot provide a thorough explanation on spoken word recognition in Taiwan Mandarin since tonal information is not included in the model. If tonal information is taken into consideration, Merge can account for the spoken word recognition in Taiwan Mandarin. The results of experiment 1 also display that Merge is a better model than TRACE because the strong top-down feedback of TRACE can override the perception of the hiccup noise.

In addition, the current study also showed that the first vowel of the disyllabic word is the most crucial and the second vowel of the disyllabic word is the second influential in spoken word recognition in Taiwan Mandarin since the vowel carries the most important information needed for spoken word recognition in Taiwan Mandarin, including tones. The vowels also occupy the longest period of time in the disyllabic words. Thus, if the vowel is disruptive, the rate of correctly recognizing the spoken words will be lower than the rate of successfully recognizing the spoken words whose consonants are disturbed. The results of experiment 2 also demonstrate that the onsets and offsets are almost the same important in Mandarin. Although the vowel in the first syllable is more influential than that in the second syllable, the coda nasal or postnuclear glide usually occupy longer period of time than the initial consonant and carry more information than the initial consonant, including tone. Therefore, the onsets and offsets are almost the same crucial in mandarin.

Furthermore, the results of this study show that the vowel is the most influential segment for the perception of Mandarin tones. Although the whole rime carries tonal information in Mandarin, the vowel occupies the longest period of time compared with the postnuclear glide and coda nasal. Therefore, if the vowel is replaced by the hiccup noise, it will be the most devastating to the perception of Mandarin tones.

Last but not least, frequency effect appears in experiment 1, 2, and 3, which means that it takes shorter reaction time for high frequency words to be recognized. In the three experiments, it also displays that low frequency words have higher chances to be incorrectly recognized than high frequency words when the words are partially disruptive. Hence, frequency effect is an important factor in spoken word recognition in Taiwan Mandarin.

So for the study has shown some general findings from the experiments, I hope I can recruit more participants for all three experiments to support or revise the results in the future study. In addition, the issues concerning the processing of Mandarin tones will also be more subtly dealt with. Moreover, the question about which tone is the most frequent one to interact with the other tones will be further examine in the future study. Last but not least, the issue regarding Mandarin tone in the Merge model will be investigated more thoroughly in the further study.

References

Connine, C. M., Blasko, D., & Wang, J. (1994). Vertical similarity and spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception and Psychophysics*, *56*, 624-636.

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

Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, *13*, 69-90.

Frauenfelder, U. H.; Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition, 25*, 1-20.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*, 267-283.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics, 38*, 299-310.

Lee, C-Y. (2000). *Lexical tone in spoken word recognition: A view from Mandarin Chinese*. Doctoral dissertation, Brown University.

Jongman, A.; Wang, Y.; Moore, C.; Sereno, J. A. Perception and production of Mandarin Chinese tones. *Handbook of Chinese Psycholinguistics*. E. Bates, L. H.; Tan, & Tzeng, O. J. L. (eds.). Cambridge University Press.

Jusczyk, P. W. & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear & Hearing, 23*, 2-40.

Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions during word recognition in continuous speech. *Cognition, 10*, 29-63.

Marslen-Wilson, W. D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1-71.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 576-585.

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Milberg, M.; Blumstein, S.; & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society, 26*, 305-308.

Nooteboom, S. G.; van der Vlugt, M. J. (1988). A search for a word-beginning superiority effect. *The Journal of the Acoustical Society of America, 84*, 2018-2032.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

Norris, D.; McQueen, J. M.; Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23,* 299-370.

Salasoo, A., & Pisono, D. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language*, *24*, 210-231.

Slowiaczek, L., M., Nusbaum, H., C., Pisoni, D., B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 64-75.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics, 34,* 409-420.

Tyler, L. K., (1984). The structure of the initial cohort: Evidence from gating. *Perception and Psychophysics, 36*, 417-427.

Tyler, L. K., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception and Psychophysics, 38,* 217-222.

Wayland, S. C.; Wingfield, A.; Goodglass, H. (1989). Recognition of isolated words: The dynamics of cohort reduction. *Applied Psycholinguistics*, *10*, 475-487.

Wingfield, A.; Goodglass, H.; Lindfield, K. C. (1997). Word recognition from acoustic onsets and acoustic offsets: Effects of cohort size and syllabic stress. *Applied Psycholingustics, 18,* 85-100.

Ye, Y. & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14(5/6),* 609-630.

Appendix 1.

High frequency words for experiment 1and 3

(from Academia Sinica Balanced Corpus of Modern Chinese)

	Word	IPA	Frequency	Percent	Cumulation
1	美國	mej21 kwɔ35	3806	0.078	35.980
2	研究	jɛn35 tɕjow51	3693	0.076	36.284
3	系統	ɕi51 t^{h}oŋ21	3600	0.074	36.358
4	國家	kwɔ35 tɕja55	3541	0.073	36.650
5	生活	ʂəŋ55 xwɔ35	3533	0.072	36.867
6	大學	ta51 ɕɥɛ35	3492	0.072	37.010
7	活動	xwɔ35 toŋ51	3406	0.070	37.221
8	世界	ʂɨ51 tɕjɛ51	3356	0.069	37.359
9	方式	faŋ55 ʂɨ51	3328	0.068	37.564
10	環境	xwan35 tɕiŋ51	3261	0.067	38.037
11	文化	wən35 xwa51	3097	0.063	38.556
12	關係	kwan55 ɕi55	2931	0.060	39.301

Low frequency words for experiment 1and 3

(from Academia Sinica Balanced Corpus of Modern Chinese)

	Word	IPA	Frequency	Percent	Cumulation
1	竊賊	t^{h}ɕjɛ51 tsej35	17	0.000	90.746
2	黨團	taŋ21 t^{h}wan35	17	0.000	90.748
3	鐘表	tʂoŋ55 pjɑw21	17	0.000	90.749
4	雜音	tsa35 in55	17	0.000	90.755
5	講習	tɕjɑŋ21 ɕi35	17	0.000	90.761
6	臀部	t^{h}wən35 pu51	17	0.000	90.763
7	聲色	ʂəŋ55 sə51	17	0.000	90.763
8	檔名	tɑŋ21 miŋ35	17	0.000	90.766
9	優選	jow55 ɕɥɛn21	17	0.000	90.768
10	隨從	swej35 t^{h}soŋ35	17	0.000	90.769
11	錢幣	t^{h}ɕjɛn35 pi51	17	0.000	90.770

12	選情	ɕɥɛn21 tʰɕiŋ35	17	0.000	90.772

Appendix 2.

High frequency words for experiment 2 and 3

(from Academia Sinica Balanced Corpus of Modern Chinese)

	Word	IPA	Frequency	Percent	Cumulation
1	中心	tʂoŋ55 ɕin55	3011	0.062	39.057
2	產品	tʰʂan35 pʰin21	2455	0.050	42.269
3	生命	ʂəŋ55 miŋ51	1627	0.033	46.943
4	民眾	min35 tʂoŋ51	1573	0.032	47.337
5	功能	koŋ55 nəŋ35	1442	0.030	47.986
6	內容	nej51 ʑoŋ35	1415	0.029	48.219
7	人類	ʑən35 lej51	1368	0.028	48.588
8	情形	tʰɕiŋ35 ɕiŋ35	1316	0.027	48.998
9	精神	tɕiŋ55 ʂən35	1289	0.026	49.345
10	廠商	tʰʂɑŋ21 ʂɑŋ55	1204	0.025	50.187
11	工程	koŋ55 tʰʂəŋ35	1162	0.024	50.695
12	人民	ʐən35 min35	1155	0.024	50.790

Low frequency words for experiment 2 and 3

(from Academia Sinica Balanced Corpus of Modern Chinese)

	Word	IPA	Frequency	Percent	Cumulation
1	台胞	tʰaj35 pɑw55	20	0.000	91.402
2	韓戰	han35 tʂan51	17	0.000	90.760
3	憑證	pʰiŋ35 tʂəŋ51	17	0.000	90.778
4	銅牌	tʰoŋ35 pʰaj35	17	0.000	90.795
5	新手	ɕin55 ʂow21	17	0.000	90.823
6	飯菜	fan51 tʰsaj51	17	0.000	90.832
7	總分	tsoŋ21 fən55	15	0.000	91.402
8	範本	fan51 pən21	15	0.000	91.425
9	領帶	liŋ21 taj51	15	0.000	91.433

10	膽囊	tan21 nɑŋ35	14	0.000	91.738
11	禪心	tʰʂan35 ɕin55	14	0.000	91.740
12	性感	ɕiŋ51 kan21	2	0.000	98.161