Toward a Model of Cantonese Spoken Word Production

WONG, Wing Kuen

A Thesis Submitted in Partial Fulfillment

of The Requirements for the Degree of

Doctor of Philosophy

in

Psychology

The Chinese University of Hong Kong

July 2010

Thesis/ Assessment Committee

Professor Him Cheung (Chair)

Professor Hsuan-Chih Chen (Thesis Supervisor)

Professor Alan Chun-Nang Wong (Committee Member)

Professor Niels Olaf Schiller (External Examiner)

Professor Ellick Kin-Fai Wong (External Examiner)

Acknowledgements

This research was supported by General Research Fund Grant CUHK441008 from the Research Grants Council of the Hong Kong Special Administrative Region, China, to Professor Hsuan-Chih Chen, and by a dissertation grant from the Department of Psychology at the Chinese University of Hong Kong. I would like to express my gratitude to my thesis supervisor, Professor Hsuan-Chih Chen, for his valuable comments on this work. Thanks also go to my thesis committee chair, Professor Him Cheung, my thesis committee member, Professor Alan Wong, and also to the two external examiners, Professor Niels O. Schiller and Professor Ellick Wong for their constructive advice.

TABLE OF CONTENTS

Acknowledgements 3

Table of Contents 4

List of Tables 8

List of Figures 9

English Abstract 10

Chinese Abstract 12

Chapter 1: Introduction 14

1.1 Spoken word production processes 15

1.1.1 *Representational structure* 15

1.1.2 *Processing dynamics* 17

1.2 Phonological encoding 20

1.2.1 *WEAVER++ model* 20

1.2.2 *Dell's model* 22

1.2.3 *A comparison between the two models* 23

1.2.4 *Evidence for the models' assumptions* 24

1.2.4.1 *Segments as important planning units* 24

1.2.4.2 *Frame structure representation* 26

Chapter 2: The Case of Cantonese 29

2.1 Unique features of Cantonese speech 29

2.2 Unresolved issues 32

2.2.1 Role of tone 33

2.2.2 Nature of sub-syllabic representations 34

2.2.3 Relationship between tonal and segmental processes 35

2.2.4 Possible language-specific properties 36

Chapter 3: The Present Study 38

3.1 Motivations 38

3.2 Paradigm 40

3.3 Overview of the five experiments 42

Chapter 4: Individual Segments in Cantonese Spoken Word Planning 45

4.1 Experiment 1 – Vowel and Coda in Cantonese 45

4.1.1 Participants ....

4.1.2 Materials and Design ....

4.1.3 Apparatus ....

4.1.4 Procedure ....

4.1.5 Results and Discussion ....

Chapter 5: Lexical Tone in Cantonese Spoken Word Planning 50

5.1 Experiment 2 – Lexical Tone in Cantonese 50

5.1.1 Participants ....

5.1.2 Materials and Design ....

5.1.3 Procedure ....

5.1.4 Results and Discussion ....

Chapter 6: Relationship between Tonal and Segmental Processes 54

6.1 Experiment 3 – Tones and Segments in Cantonese Speech 54

6.1.1 Participants ....

6.1.2 Materials and Design ....

6.1.3 Procedure ....

6.1.4 Results and Discussion ....

Chapter 7: Onsets and Rhymes in Cantonese Speech Preparation 60

7.1 Experiment 4 – Onsets and Rhymes in Cantonese Speech 60

7.1.1 Participants ....

7.1.2 Materials and Design

7.1.3 Procedure

7.1.4 Results and Discussion

Chapter 8: Potential Influences from Phonetic Features 68

8.1 Experiment 5A – Sonority of the Rhyme Unit 68

8.1.1 Participants

8.1.2 Materials and Design

8.1.3 Procedure

8.1.4 Results

8.2 Experiment 5B – Duration of the Rhyme Unit 73

8.2.1 Participants

8.2.2 Materials and Design

8.2.3 Procedure

8.2.4 Results and Discussion

Chapter 9: General Discussion 79

9.1 Role of tone 84

9.2 Nature of sub-syllabic representations 86

9.3 Relationship between tonal and segmental processes 92

9.4 Possible language-specific properties 94

9.5 A model of Cantonese spoken word production 97

9.6 Other related issues 101

9.7 Future directions 103

9.8 Conclusion 106

References 108

Appendix A: Materials in Experiments 1 and 2 118

Appendix B: Materials in Experiment 3 120

Appendix C: Materials in Experiment 4..........122

Appendix D: Materials in Experiment 5..........124

List of Tables

Table 1. Statistical properties of the distractors in Experiment 1 46

Table 2. Mean naming latencies and error rates per condition in Experiment 1.......48

Table 3. Statistical properties of the distractors in Experiment 2 51

Table 4. Mean naming latencies and error rates per condition in Experiment 2.......52

Table 5. Statistical properties of the distractors in Experiment 3 55

Table 6. Mean naming latencies and error rates per condition in Experiment 3.......56

Table 7. Statistical properties of the distractors in Experiment 4 63

Table 8. Results of analyses of variance in Experiment 4 65

Table 9. Mean naming latencies and error rates per condition in Experiment 4.......66

Table 10. Statistical properties of the distractors in Experiment 5A 70

Table 11. Results of analyses of variance in Experiment 5A....................................72

Table 12. Mean naming latencies and error rates per condition in Experiment 5A..73

Table 13. Statistical properties of the distractors in Experiment 5B.........................75

Table 14. Results of analyses of variance in Experiment 5B....................................76

Table 15. Mean naming latencies and error rates per condition in Experiment 5B..77

Table 16. Summary of the results from all Cantonese PWI studies..........................80

List of Figures

Figure 1. Fundamental frequency profiles of six lexical tones in Cantonese. 30

Figure 2. A model of spoken word production in Cantonese....................................99

Abstract

Five experiments were conducted to investigate how phonological information is processed in Cantonese spoken word production using the picture-word interference (PWI) paradigm. Participants were asked to name aloud individually presented pictures and ignore an accompanying auditory word distractor. In the first three experiments of the present study, the target picture names were Cantonese mono-syllables with a consonant-vowel-consonant (CVC) structure. Participants' picture naming latencies were found faster when the target (e.g., 星 /**sing1**/[1], "Star") and the distractor (e.g., /g**ing2**/, /**si**k6/, or /so**eng**3/) shared two identical segments (irrespective of the segments' syllable-internal position), than when they were unrelated, whereas no reliable effects were obtained when they shared only the vowel (e.g., /d**i**m3/), the coda (e.g., /hu**ng**2/), or together with the tone (e.g., /b**i**t**1**/ or /fu**ng1**/). Furthermore, the facilitation effect observed in the consonant+consonant+tone-related condition (e.g., /**s**oe**ng1**/) was found reliably larger than that in the consonant+consonant-related condition (e.g., /**s**oe**ng**3/). In Experiment 4, the syllable structure of the targets was manipulated such that half of the picture names were mono-syllables with a consonant-vowel (CV) structure (e.g., 梳 /**so1**/, "comb") and the other half a CVC structure (/**sing1**/). A significant syllable (without tone) related facilitation was found irrespective of the target type, whereas no reliable effect was observed when the target and the distractor (e.g., /se6/ and /sam2/, for CV and CVC targets, respectively) shared only the onset consonant. Furthermore, Experiment 5 investigated whether phonetic factors modulate the facilitation effect observed in a PWI task. To this end, the sonority level and duration of the target's rhyme component were manipulated in Experiments 5A and 5B, respectively. A

[1] The transcriptions for Cantonese syllables presented in this paper are based on the Cantonese Romanization Scheme proposed by the Linguistic Society of Hong Kong (Chinese Character Database: With Word-formations, 2003). The number besides each syllable marking denotes the lexical tone.

significant rhyme-related priming effect was found in both Experiments 5A and 5B, and such effect was neither affected by the sonority level nor duration of the target's rhyme component. These results are in line with the notions that 1) a single segment does not have a significant impact on Chinese spoken word planning, 2) sub-syllabic constitutes such as rhyme is an important processing unit, 3) the lexical tone has a unique role to play during phonological encoding, and 4) segmental and tonal retrieval are operated in an interactive manner. Based on all the available results from Cantonese PWI research, an interactive model of Cantonese spoken word production is proposed.

摘要

本論文報告了五個圖詞干擾實驗，旨在研究廣東話口語詞彙產生中語音的加工過程。被試的任務是要爲每次在螢幕中出現的圖片命名，並要忽略與圖片同時出現的干擾詞。在首三個實驗中，目標詞皆爲以子音-母音-子音組成的廣東話單音節詞。結果發現，當聽覺呈現的干擾詞 (如: /**ging**2/[2], /**sik**6/, 或 /**soeng**3/) 與目標詞 (如: 星 /**sing1**/) 之間有兩個音素相同時 (不管該兩音素的音節位置)，被試的圖片命名時間要比當它們兩者無關之時 (即: 無關控制條件) 顯著的快。可是，當干擾詞與目標詞之間只有韻腹相同 (如: /d**i**m3/)、韻尾相同 (如: /hu**ng**2/)，或韻腹及聲調相同 (如: /b**i**t**1**/)，又或韻尾及聲調相同 (如: /fu**ng1**/) 時，被試的圖片命名時間與無關控制條件均無顯著分別。此外，當干擾詞與目標詞爲聲母、韻尾及聲調俱同的情況下 (如: /**s**oe**ng1**/)，被試的圖片命名時間要比當它們兩者之間只有聲母及韻尾相同時 (如: /**s**oe**ng**3/) 顯著的快。實驗四包含了兩種目標詞，分別爲以子音-母音組成的單音節 (如: 梳 /**so1**/)，以及由子音-母音-子音組成的單音節 (如: /**sing1**/)。結果發現，當干擾詞與目標詞爲音節 (不包括聲調) 相同時 (不管目標詞的音節結構)，被試的圖片命名時間要比無關控制條件顯著的快。而當干擾詞與目標詞之間只有聲母相同時 (如: /**s**e6/ 及 /**s**am2/)，不管目標詞的音節結構，被試的圖片命名時間均與無關控制條件相若。另外，實驗五探討了干擾詞本身的語音特質會否對圖詞干擾實驗中的促進效應產生影響。爲此，實驗五-甲及五-乙均包含了與目標詞爲韻母相同的干擾詞，並且實驗五-甲操控了目標詞韻母部份的響亮程度，而實驗五-乙操控了目標詞韻母部份的長度。結果發現，當干擾詞和目標詞爲韻母相同時，被試的圖片命名時間要比無關控制條件顯著的快，而且此促進效應並沒有受到目標詞韻母部份的響亮程度或長度所影響。總括而言，上述結果支持以下論點：(一) 單一音素在中文言語產生過程中並沒有顯著的影響，(二) 次音節單位如韻母是一個重要的加工單位，(三) 聲調信息在語

[2] 本論文採用了香港語言學學會所提出的粵語注音系統 (Chinese Character Database: With Word-formations, 2003)。注音旁的數字爲該音節的聲調。

音編碼過程中扮演著獨特的角色，(四) 音素和聲調信息的提取過程中存在著交互作用。最後，論文的總結部份整合了所有從廣東話圖詞干擾實驗中得來的結果，並提出了一個交互激活的廣東話口語產生模型。

Chapter 1

Introduction

Spoken languages in the world can broadly be categorized into tonal or non-tonal languages depending on whether tone is lexically distinctive (Pike, 1948). However, previous studies on speech production were almost exclusively conducted on non-tonal languages such as Dutch or English, and the related research on tonal languages (e.g., Cantonese Chinese) is relatively scarce. Consequently, two very important theoretical issues are left unresolved (Chen & Zhou, 1999). The first concerns to what extent the existing theories of speech production can be generalized from non-tonal to tonal languages. And the second is how language-specific features in non-tonal languages can be explained. With these concerns, the present study was conducted to investigate phonological processing in Cantonese monosyllabic word production.

In the following, I will first review some of the major findings in the speech production literature concerning the representational structure and processing dynamics of the speech production system, and then describe the phonological encoding processes assumed by two prominent speech production theories, together with their supporting evidence. In the next chapter, I will illustrate the unique properties of Cantonese speech and indicate the important issues that remain to be solved. After that, an overview of the present study will be furnished in Chapter 3. Then five experiments will be reported from Chapter 4 to Chapter 8. And a General Discussion will be provided in Chapter 9 in which the results obtained in this and in the previous related studies will be discussed together. An interactive model describing how phonological information is processed in Cantonese spoken word production is proposed in General Discussion.

1.1 Spoken Word Production Processes

1.1.1 Representational structure

Models of speech production generally agree that there are at least two processing steps in spoken word planning: One concerns the specification of the word's semantic-syntactic properties and the other relates to the generation of the word's phonological form (Butterworth, 1989; Caramazza, 1997; Dell, 1986, 1988; Garrett, 1988; Levelt, 1989, 1992; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992, 1997; Starreveld & La Heij, 1996; Stemberger, 1985). Evidence supporting the distinction between semantic-syntactic and phonological processes comes from various sources. For instance, studies investigating the tip-of-the-tongue phenomenon found that speakers who failed to generate the phonological form of the intended word could nevertheless indicate the meaning or grammatical gender (i.e., syntactic information) of the word (Brown & McNeill, 1966; Caramazza & Miozzo, 1997; Vigliocco, Antonini, & Garrett, 1997). Similarly, patients with impairments in phonological retrieval were relatively intact in semantic retrieval, whereas other cases showed an opposite pattern of results (Caramazza, Papagno, & Ruml, 2000; Cuetos, Aguado, & Caramazza, 2000). This double dissociation pattern has been taken as evidence to support the distinction between semantic-syntactic and phonological processes. Furthermore, it has been reported that speech errors that involve lexical units differ substantially from those involve sub-lexical units (Dell, 1986; Garrett, 1975). For instance, lexical errors typically involve word exchange across different phrases or clauses in which the two words are of the same syntactic category, whereas sub-lexical errors typically involve segment (i.e., phoneme) exchange between adjacent phonologically similar words which belong to different syntactic categories.

The results from chronometric and neuroimaging research also reached a similar conclusion. For instance, previous picture-word interference (PWI) studies have found that participants' picture naming responses were slower if the accompanied word distractor was semantically related to the picture name, relative to when they were unrelated, whereas faster naming responses, relative to the unrelated condition, were observed if the target and the distractor shared similar phonological segments (Damian & Martin, 1999; Jescheniak & Schriefers, 2001; Schriefers, Meyer, & Levelt, 1990). The differential effects between semantic and phonological distractors (i.e., interference vs. facilitation) are in line with the notion that distinct levels are involved for semantic-syntactic and phonological processes. Moreover, there is evidence from event-related brain potential (ERP) studies showing that semantic processes precede syntactic processes (Schmitt, Schiltz, Zaake, Kutas, & Münte, 2001), which in turn precede phonological processes (Schmitt, Münte, & Kutas, 2000; Van Turennout, Hagoort, & Brown, 1997, 1998; but see also Abdel Rahman & Sommer, 2003). In addition, evidence from magnetoencephalograph (MEG) (Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998) and event-related functional magnetic resonance imaging (fMRI) studies (de Zubicaray, McMahon, Eastburn, & Wilson, 2002) indicates that distinct neural networks are responsible for semantic and phonological processes in spoken word production (see also Indefrey & Levelt, 2004; Price, 2010).

As a result, the findings reviewed above all seem to suggest that distinct levels are involved for semantic-syntactic and phonological processing in spoken word production. However, this is where the agreement among models of speech production ends (Caramazza, 1997). One of the controversial issues in the literature is whether syntactic processing mediates between semantic and phonological processes. Proponents of this syntactic-mediation hypothesis include the WEAVER++ (i.e., Word-form Encoding by Activation and VERification) model proposed by Levelt, Roelofs, and colleagues (Levelt, 1999, 2000; Levelt et al., 1999; Roelofs, 1992, 1997, 2000), which is a spreading activation based network model developed from an earlier model proposed by Levelt (1989, 1992), as well as the interactive activation model proposed by Dell and colleagues (Dell, 1986, 1988; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997). However, models such as the Independent Network model proposed by Caramazza (1997) and colleagues do not hold this syntactic-mediation assumption. Instead, Caramazza's model assumes a direct connection between semantic and phonological levels (see also Starreveld & La Heij, 1996a for a similar proposal), and the word-syntactic information is specified along with phonological retrieval. Although it is still controversial concerning the locus of syntactic processing in spoken word production, the broad distinction between semantic-syntactic and phonological processes is nevertheless undisputable.

1.1.2 Processing dynamics

Another highly controversial issue within the speech production literature concerns how activation flows across different levels. Regarding this issue, models of speech production can broadly be categorized into three different types, namely discrete serial, cascaded, and interactive. The difference between discrete serial and the other two types of model rests on the number of lexical nodes that can proceed to the phonological encoding process. For instance, according to the WEAVER++ model, spoken word production begins with activating the conceptual representation of the to-be-spoken message, followed by selecting the corresponding lemma node (i.e., a representation for lexical-syntactic information such as word class, gender, and number, see also Kempen & Huijbers, 1983), and then only the selected lemma node can pass activation onto the word-form (i.e., lexeme, according to WEAVER++) level for morpho-phonological encoding. In other words, only the target's phonological information would be activated in spoken word planning based on WEAVER++ (Levelt, et al., 1999; Levelt, Schriefers, et al., 1991). Since the phonological encoding process would not start until the final product of the semantic-syntactic processes is generated, models such as WEAVER++ is considered as discrete serial. On the contrary, cascaded or interactive models posit that all activated lemma nodes (i.e., including the target and the semantically related lemma nodes) would activate their corresponding phonological forms to a certain extent (Cutting & Ferreira, 1999; Dell, 1986, 1988; Peterson & Savoy, 1998). Furthermore, the major difference between interactive and the other two types of model is that feedback activations from the phonological to the prior processing level are allowed in the interactive model (e.g., Dell, 1986, 1988), but not in discrete serial or cascaded models (Peterson & Savoy, 1998; see also Rapp & Goldrick, 2000, for a review).

In the PWI study conducted by Schriefers et al. (1990), the stimuli onset asynchrony (SOA) between the picture and the word distractor was manipulated in three conditions in which the distractor word was either presented 150 ms before (SOA = -150 ms), simultaneously with (SOA = 0 ms), or 150 ms after (SOA = 150 ms), the onset of picture presentation. The crucial finding from that study is that a semantic interference effect was observed exclusively in the earliest SOA condition (i.e., SOA = -150 ms), whereas phonological facilitations were found only at later SOAs (i.e., SOA = 0 or 150 ms). The time course and the non-overlapping nature of the two observed effects have been argued in favor of the assumption that activation flows uni-directionally from a semantic-syntactic level to a phonological level during spoken word production. Furthermore, by combining a picture naming and an auditory lexical decision task, Levelt et al. (1991) observed significant inhibitory effects on lexical decision times when the picture name (e.g., sheep) and the auditory probe were either semantically (e.g., goat) or phonologically (e.g., sheet) related, relative to when they were unrelated. However, no reliable effect was found when the auditory probe (e.g., goal) was phonologically related to a semantic associate of the target (e.g., goat). With these results, Levelt et al. (1991) argued that only the phonological information of the selected target lemma would be activated in the course of spoken word production (i.e., the discrete serial assumption).

Nevertheless, later picture naming studies reported evidence showing that synonyms (or near-synonyms) of the target would also activate their corresponding phonological codes to a certain degree (e.g., Peterson & Savoy, 1998). In addition, by using a similar PWI paradigm, other researchers reported phonological facilitation effects in early SOA conditions where semantic interference effects were typically observed (Damian & Martin, 1999; Jescheniak & Schriefers, 2001; Starreveld, 2000). As demonstrated by Starreveld (2000), whether an early phonological effect would be observed is subject to factors such as the experimental design (e.g., between vs. within participants design) and the content of the item set (e.g., presence or absence of semantically related distractors). Consequently, the non-overlapping nature of the two effects observed by Schriefers et al (1990) did not give unequivocal support to the discrete serial assumption. Furthermore, there is evidence showing that interactivity does exist between semantic-syntactic and phonological levels (e.g., Damian & Martin, 1999, Starreveld & La Heij, 1996a). For instance, by factorially manipulating the semantic and phonological[3] relationship between targets and distractors in a PWI study, Starreveld & La Heij (1996a) found a significant Semantic x Phonological Relatedness interaction effect, which was, according to the authors, not expected by the discrete serial model (see Damian & Martin, 1999, for similar results with auditory distractors, see also Roelofs, Meyer, & Levelt, 1996, Starreveld & La Heij, 1996b). The interaction effects obtained by Starreveld and La Heij (1996a), as well as those by Damian and Martin (1999), have been taken as evidence showing that

[3] Note that as confessed by the authors, the phonological distractors used in Starreveld and La Heij's (1996a) study also possessed orthographic similarity with the targets.

interactively exists between semantic-syntactic and phonological levels.

The debate between discrete and interactive models is still ongoing. However, the general findings from the literature appear that there is at least certain degree of interactivity exists in the word production system (Rapp & Goldrick, 2000, 2004; Goldrick, 2006). For the most extreme case, even discrete models such as WEAVER++ would allow feedbacks from the lemma to the conceptual level. Therefore, the presence of interactivity in the production system is a widely accepted notion, although the degree of interactivity postulated differed substantially across models.

Having reviewed the past findings on the general characterization of the speech production system, I will focus on a more specific part of the word production processes. In the following section, I will discuss the phonological encoding process assumed by two prominent models in the field, namely the WEAVER++ model proposed by Levelt and colleagues and the interactive activation model proposed by Dell and colleagues. These two models are chosen because they are perhaps the two most widely examined models in the production literature. More importantly, these two models have detailed the mechanisms involved in phonological encoding which are of particular relevance to the present study.

1.2 Phonological Encoding

1.2.1 WEAVER++ model (Levelt, 1999, 2000; Levelt et al., 1999; Roelofs, 1997)

According to WEAVER++, the processes of spoken word production start with the retrieval of the target's conceptual representation, followed by the selection of the corresponding lemma node (i.e., syntactic properties of the target). After the target lemma is selected, the corresponding lexeme (i.e., morpheme) node would be activated, and the phonological encoding process starts afterward. Note that, according to WEAVER++, after the target lemma has been selected, activation spreads through the network uni-directionally. Furthermore, attaching with each node in the network is a procedure which verifies the property of the link between a node and its target node one level up, such that an active but inappropriate node would not be selected.

At the beginning of phonological encoding, the target segments of the selected morpheme would be activated and selected in parallel, and the order of the segments is indicated by the links developed between morpheme and the segmental units. In addition, a metrical frame which specifies the word's prosodic information such as its number of syllables and stress pattern would be retrieved along with the segmental retrieval process, and the two processes operate independently from each other. Note, however, that the metrical information of a lexical unit is stored only for multi-syllabic words which do not have primary stress on word-initial syllable. However, this kind of "exceptional" word only constitutes around 10% of the words in Dutch. Therefore, WEAVER++ assumes that the metrical structure of a word is computed online following the default setting (i.e., stress on the word-initial syllable with a full vowel for multi-syllabic words), and is retrieved only for exceptional words. For the production of an exceptional word, the retrieved segments are then associated with the metrical frame incrementally in a rightward direction which is called syllabification. And that the syllabification process follows the language universal as well as the language specific phonological rules such as assigning a consonant to a syllable's onset position whenever possible, and one vowel for each syllable. The products of the syllabification process are phonological syllables which are then used to address the motor programs stored in the phonetic syllabary (i.e., mental syllabary) for articulation. Therefore, according to WEAVER++, phonological syllables are computed on-the-fly (rather than retrieved) but phonetic syllables are stored in, and retrieved from, the mental syllabary (see also Cholin, Levelt, & Schiller, 2006; Levelt & Wheeldon, 1994).

1.2.2 Dell's model (Dell, 1986, 1988, 1990; Dell et al., 1997)

According to the interactive activation model proposed by Dell and colleagues, the phonological encoding process starts right after the target morpheme is activated. Activation then spreads through the network from the morpheme node to the corresponding syllable nodes and then to the segment nodes. Meanwhile, a word-shape header node which specifies the syllable structure of the word (e.g., CV or CVC, as standing for consonant-vowel, and consonant-vowel-consonant, respectively) would also be activated. Note that although Dell's (1986) model does not refute the possibility that phonological units other than syllables and segments (e.g., phonological features, and syllable constituents including onset clusters and rhymes) might have their own representations in the lexical network, the basic units in the model are segmental units (Dell, 1986, p.296). Furthermore, the syllable internal position (e.g., onset, nucleus, or coda) of the segments are specified at the segmental level. Hence, the segment /k/ which appears in the syllable onset position is represented differently from the segment /k/ which appears in the syllable coda position. In addition, the nodes across levels are connected by equally weighted facilitatory links which are bi-directional in nature, hence feedback activations (or bottom-up influences) are allowed in Dell's model.

Phonological encoding, according to this model, is achieved by selecting the node with the highest level of activation at each level after a set period of time which is determined by the speaker's speech rate. In other words, the faster the speech rate is, the lesser the time is allowed for the system to settle. A special feature of this model is that syllables are selected serially with the word-initial syllable first. However, segments of the currently active syllable are activated and selected in parallel. Segmental units with the highest level of activation are then associated to their corresponding phonemic category nodes in the word-shape header frame. Note that although representations for phonological features are allowed in Dell's model, how these units are used for guiding the articulation process has not yet been explicitly specified.

1.2.3 A comparison between the two models

The WEAVER++ model and Dell's model differed largely from each other with regard to their assumptions on phonological encoding. The differences between the two models are reflected in both representational and processing aspects.

For the representational aspect, the two models made very different assumptions on the nature of phonological units. For WEAVER++, segments are the only legitimate phonological units in the network, with an exception that the metrical frames for "exceptional" words are also being stored. On the contrary, Dell's model holds that phonological units other than segments (i.e., syllables, onset clusters, rhymes, or phonological features) also have their own representations in the network, although segments are the fundamental processing units. Moreover, the segmental units portrayed in WEAVER++ are positional-free whereas they are positional-specific according to Dell's model. Furthermore, prosodic information such as the stress pattern of the word is specified by the metrical frame (or by the default rule) according to WEAVER++ but the frame structure assumed in Dell's model specifies the syllable internal structure of the word.

For the processing aspect, the major differences between the two models concern how activation flows and how target nodes are selected. Consistent with their postulations on the speech production system in general (e.g., discrete serial vs.

interactive), the two models make very different claims on how activation flows across different phonological encoding stages, in which WEAVER++ assumes that activation flows uni-directionally only whereas Dell's model posits that activation flows in a bi-directional manner. Furthermore, the mechanisms used for node selection are different between the two models. For Dell's model, the node with the highest level of activation at each level is selected after a set period of time which is dependent on the speech rate. On the contrary, for WEAVER++, a checking procedure is attached with each node which verifies the property of the link between the current node and the node at one level up, and a node is selected only after the local verification process is completed. In other words, dissimilar to Dell's model in which the degree of activation is the main determinant for node selection, locally embedded verification mechanisms are involved for node selection according to WEAVER++.

Although the two models differed largely from each other, however, they do share a few crucial similarities concerning their assumptions on phonological encoding. Among other things, the two models posit a prominent role for segmental units and a slot-filling-in process (i.e., segment-to-frame association). Evidence related to these two assumptions is reviewed below.

1.2.4 Evidence for the models' assumptions

1.2.4.1 Segments as important planning units

Evidence supporting an important role for segments in phonological encoding is ample. For instance, about 70 to 90% of the phonological slips errors involved a single segment (Dell, 1986; Nooteboom, 1969; Stemberger, 1982). Also, findings from other speech error studies suggest that a majority of the phonological errors was induced by the movement of segment-sized units (Dell, 1988; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1982; see also Meyer, 1992, and Roelofs, 1999, for a comprehensive review).

Apart from the speech error data, findings from chronometric studies also point to a similar conclusion. For instance, in Sevald and Dell's (1994) study, participants were asked to utter a sequence of four CVC monosyllables as many times as possible within a period of 8 seconds. It was found that the syllable production rate was slower when the syllables shared an identical onset (i.e., syllable-initial consonant), relative to the unrelated control condition. And a reliably stronger inhibitory effect was found when the syllables also shared the same vowel. In addition, a significant facilitation effect was observed when the syllables shared an identical coda (i.e., syllable-final consonant) only. These results indicate that individual segments, irrespective of their syllable internal positions, all have a significant role to play in spoken word production.

Furthermore, by using the form-preparation paradigm firstly introduced by Meyer (1990), Roelofs (1999) had Dutch participants uttered various sets of words across different blocks in the experiment. A significant facilitation effect on naming latencies was observed when the response words within a set shared the word-initial segment relative to when they were unrelated, but no significant effect was found when they shared only the word-initial phonological feature, indicating that segments but not features are the planning units in speech preparation.

In a masked priming study conducted by Schiller (2008), Dutch-speaking participants were asked to name aloud individually presented pictures which were preceded by a briefly displayed (50ms) masked prime. In two conditions the prime was either the first or the final segment of the picture name. Despite the prime was presented in a way that participants could barely identify it, significant priming effects on picture naming latencies, relative to the control condition, were still obtained with either the first or the final-segment primes, indicating that individual segments have a prominent role to play in phonological encoding, at least in Dutch.

In addition, Damian and Dumay (2009) investigated the influence of repeated segments on short phrase production. In their study, English-speaking participants were asked to utter an adjective-noun phrase (i.e., a color adjective + an object name) upon seeing a colored object picture. A reliable facilitation effect on response latencies was obtained when the color adjective and the object name shared a similar segment than when they were unrelated. Notably, comparable facilitation effects were obtained whether the shared segment was located at the word-initial (e.g., black boat), word-medial (e.g., black pan), or word-final (e.g., black monk) positions. More importantly, a reliable priming effect was also found when the shared segments did not occupy the same syllable internal position (e.g., green flag). These results suggest that segments are important planning units which are represented in a position-nonspecific manner (see also Levelt et al., 1999).

1.2.4.2 Frame structure representation

Both WEAVER++ and Dell's model posit a kind of frame structure representation for supra-segmental information although very different assumptions have been made concerning its content. Evidence from speech error studies shows that speech errors that involve segment movements generally follow the syllable position constraint (e.g., onsets exchange with onsets, or codas exchange with codas), indicating that an abstract representation for syllable structure might involve in phonological encoding (Fromkin, 1971, but see also Shattuck-Hufnagel, 1987; Meyer, 1992, for an alternative account). Furthermore, there were speech errors which involve only an incorrect stress assignment but the sequencing of the word internal segments was preserved (Cutler, 1980). In addition, Nickels and Howard (1999) found a significant effect of word stress pattern on word production among a group of

English-speaking aphasic patients, in which the patients showed a higher error rate when repeating bisyllabic words with a stressed second syllable, relative to those with a stressed first syllable. Taken the above reviewed studies together, it appears that a unique representation for metrical information (i.e., information apart from segmental contents) does exist although the content of information being represented (e.g., syllable structure or word stress pattern) is still debatable.

In addition, although there is evidence from speech monitoring experiments showing that metrical stress, similar to segments, is encoded in a serial manner for multi-syllabic words (Schiller, 2006; Schiller, Jansma, Peters, & Levelt, 2006), the results from priming studies suggest that metrical information is processed distinctively from its segmental counterparts (Schiller, Fikkert, & Levelt, 2004). For instance, despite past PWI studies have repeatedly demonstrated a significant facilitation effect for distractors which shared similar segmental contents with the targets (Meyer & Schriefers, 1991; Schriefers et al., 1990), no trace of priming was induced by the distractors which shared an identical stress pattern with the targets only (Schiller et al., 2004). Furthermore, in a form-preparation study conducted by Roelofs and Meyer (1998), no priming effect was found when the response words shared only the same word initial segments but had a different metrical structure (i.e., number of syllables and stress pattern). However, a reliable facilitation effect on word production was found when the words shared both the same initial-segments AND metrical structure, indicating that the metrical information (for exceptional words) is retrieved independently from segmental information (Roelofs & Meyer, 1998).

Nevertheless, evidence available in support of the two models' assumptions comes mostly, if not all, from studies on stress languages such as Dutch or English. Therefore, it is still unclear whether the same assumptions also apply to the processing of other languages with distinct properties (Chen, 1992, 1999, 2006; Chen

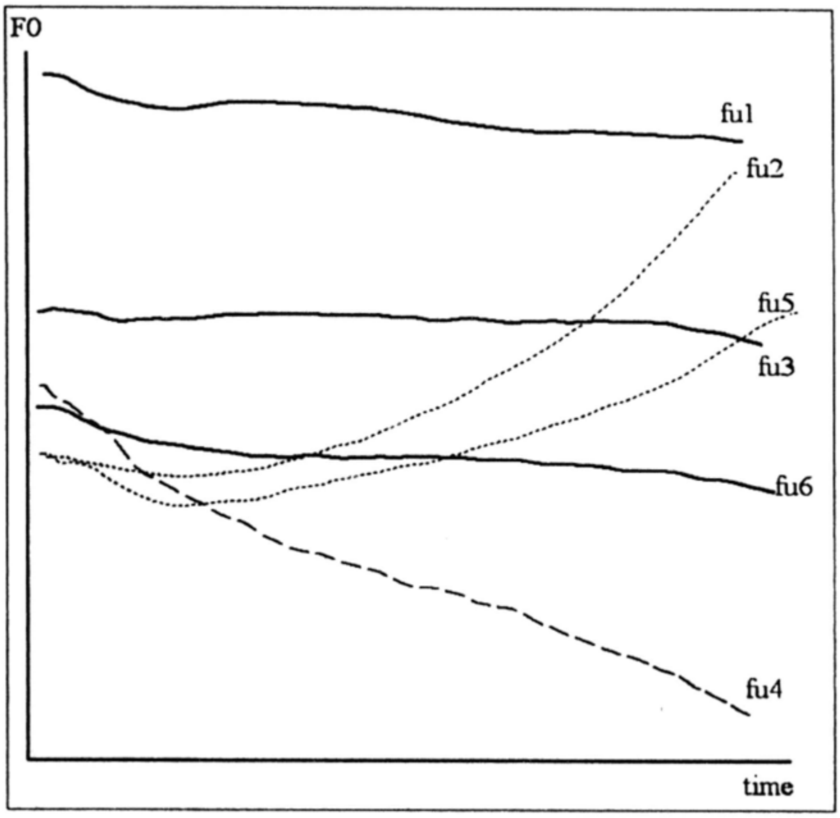
& Zhou, 1999). With this concern, the present study was conducted to investigate the processes of phonological encoding in Cantonese Chinese, a tonal language which is distinctively different from Dutch and English. In the following Chapter, I will discuss some of the unique features of Cantonese, followed by the motivations of the present study.

Chapter 2

2.1 Unique features of Cantonese speech

Cantonese is a Chinese dialect spoken by the people in Southern China. Although major Chinese dialects differ substantially from each other to an extent that could be mutually unintelligible (DeFrancis, 1984; Li & Thompson, 1981), the use of prosodic information such as tone to differentiate lexical meanings is widely adopted. In the case of Cantonese, there are six lexical tones (but four in Mandarin). When combining with different lexical tones, a syllable in Cantonese can be used to denote very different meanings. For instance, the syllable /fu/ has the meaning of "husband" when it combines with Tone 1, "tiger" with Tone 2, "trousers" with Tone 3, and so on. In addition, dissimilar to the stress in stress languages in which cues such as amplitude, duration, and segmental contents can be used to distinguish a stressed syllable from an unstressed one, the lexical tone in Cantonese is determined mostly by the fundamental frequency level and contour realized on a syllable (Cutler and Chen, 1997). Since tones in Cantonese are lexically contrastive, Cantonese is considered as a tonal language. Figure 1, which is adopted from Tsang (2009), plots the fundamental frequency profiles of the six lexical tones in Cantonese.

Figure 1. Fundamental frequency (F0) profiles of the six lexical tones in Cantonese.



Note: Level tones are represented by solid lines (Tone 1, Tone 3, and Tone 6); Rising tones are represented by dotted lines (Tone 2 and Tone 5); Falling tone is represented by dashed line (Tone 4).

Another interesting feature in Cantonese concerns the relatively small number of syllables in the language. In Dutch, for instance, there are over 12000 existing syllables (Schiller, Meyer, Baayen, & Levelt, 1996) but only about 600 (1900 when tone is counted) in Cantonese. In addition, dissimilar to many stress languages using alphabetic scripts, in which the basic orthographic units (i.e., letters) were made to correspond to the basic units in speech such as phonemes, Chinese adopts a logographic writing system in which the basic orthographic units (i.e., characters) were made to represent the basic meaning units such as morphemes, and each Chinese character corresponds to a syllable in speech (see Chen, 1992, 2006, for details). Given the relatively small number of syllables in the language, together with the explicit relationship among characters, morphemes, and syllables, it has been proposed that syllables might have their own representations and an important role to play in processing Cantonese speech (Yip, 2000; see also Chen, 2000, for a similar notion in Mandarin).

However, controversies exist with regard to the issue of sub-syllabic representations in Cantonese. The traditional view on Chinese phonology posits that a syllable (without tone), is generally composed of an initial and a final component (Bruche-Schulz, 1997; Li & Thompson, 1981). An initial is the consonantal beginning of a syllable, and because there are no consonant clusters in Cantonese (so as in Mandarin), it is usually a single consonant[4]. And the final is the rest of the syllable. Accordingly, there are 19 initials and 53 finals in Cantonese (but 21 initials and 37 finals in Mandarin). Note that the initials and finals in Cantonese have been considered as the minimal contrastive units similar to the phonemes in other languages according to this view (Bruche-Schulz, 1997). Although the initial-final dichotomy is relatively undisputable in describing Chinese phonology, some linguists, however, proposed that the finals can further be decomposed into nuclei (i.e., vowels) and codas (or guides) (DeFrancis, 1984; Kao & Hoosain, 1986). According to this latter conjecture, the syllables in Cantonese can generally be described with a (C)V(C/G)[5] structure (see also Chen, Chen, & Dell, 2002, for a similar proposal in Mandarin). Nevertheless, the issues of whether these proposed phonological units (e.g., initials and finals, or nuclei and codas) have their psychological realities in processing Cantonese speech still remain open.

---

4 Except the case when the syllable begins with a vowel (e.g., "zero" initial).

5 C stands for syllable-initial consonant, V stands for vowel, and G stands for glide. Optional components are shown in parentheses. Note, however, that exceptions to this general rule are syllabic nasals such as /m/ (i.e., 唔) and /ng/ (i.e., 吳) in which the nasal itself can stand alone as a syllable unit. Nevertheless, cases like these are extremely rare and a majority (more than 99%) of the syllables in Cantonese contains a vowel component.

2.2 Unresolved issues

Given the unique features of Cantonese speech as discussed above, it is not clear to what extent that the existing theories of speech production can explain the processes of phonological encoding in Cantonese. If tones in Cantonese are processed like the stress in stress languages, one would predict that a kind of "tonal frame" would be independently retrieved along with the segmental retrieval process under the context of the WEAVER++ model. However, dissimilar to languages such as Dutch or English in which prosodic information such as stress is only infrequently used to distinguish lexical items[6], prosodic information such as tone is crucial in lexical distinction in Cantonese. Therefore, it is still unclear whether the role of tone in Cantonese is similar (or equivalent) to that of stress in stress languages, although both of them can be considered as prosodic. In addition, even if one assumes that the lexical tone is handled by a kind of representation similar to the metrical frame assumed in WEAVER++, one needs to face a set of related questions. For instance, dissimilar to the case of Dutch in which a majority (~90%) of the lexical items follow the default stress pattern (Schiller, Meyer, & Levelt, 1997), there is no such an equivalent default "tone value" in Cantonese. As a result, does a metrical frame retrieval process required for all words in Cantonese speech production? Similarly, as WEAVER++ assumes that segmental and metrical retrieval processes run independently from each other, does this conjecture also apply to the case of segmental and tonal retrieval? Answers for these, as well as other related, issues would nevertheless be crucial to the development of a universal theory of speech production. Consequently, the present study was conducted to investigate the

---

[6] Indeed, there is evidence to suggest that the role of stress is minimal in lexical access, see for instance, Cutler, 1986; Cutler & Clifton, 1984.

processes of phonological encoding in Cantonese. In particular, the present study will focus on the following four major issues:

1) *The role of tone in phonological encoding.*

As tones and segments are both lexically contrastive, are they processed or represented similarly? Can tone be handled by a kind of representation similar to the metrical frame assumed in WEAVER++?

2) *The nature of sub-syllabic representation.*

What is the nature of sub-syllabic representations (if any) in Cantonese? Are they in the forms of initials and finals? Or units such as nuclei and codas also have their own representations?

3) *The relationship between tonal and segmental processes.*

What is the relationship between tonal and segmental retrieval (if they are indeed different)? Is it independent or interactive?

4) *Possible language-specific properties.*

Regarding the salience of various phonological units, would different languages show differential patterns? For instance, would segments show a more prominent effect than syllables in alphabetic languages such as Dutch, whereas the opposite is true in Chinese? In addition, would certain phonological units show stronger effects than the others in Cantonese?

In the following section, I will review the major findings concerning the above four issues. Since the corresponding evidence from Cantonese studies is scarce and a very large degree of phonological similarity is shared between Cantonese and

Mandarin, the results from previous Cantonese and Mandarin speech production studies will be covered together.

2.2.1 Role of tone

The available evidence concerning the role of tone in Chinese speech production comes mostly from speech error studies. Research in this line relies heavily on comparing the patterns of segmental and tonal errors observed among patients or normal participants, but the results obtained were mixed. Earlier studies in the literature reported similar patterns of segmental and tonal speech errors among Chinese speakers which suggest that tones and segments are processed or represented similarly (Moser, 1991; Packard, 1986). This proposal has been coined as the *segmental view* in which tones were treated similarly as segments (Fromkin, 1980). However, recent studies using a similar approach gave rise to an opposite pattern of results (Chen, 1999; Liang & van Heuven, 2004). For instance, by analyzing naturalistic slips of the tongue in Mandarin, Chen (1999) found that tonal errors occurred only very rarely which was far less frequently than that of segmental errors, and that a majority of the former were perseverations while the latter were anticipations. The author thus argued that the tone in Mandarin is represented or processed differently from segments, and is similar to the stress in English (Chen, 1999). Similarly, by using an acoustic analysis method, Liang and van Heuven (2004) found that the tonal errors exhibited by a Mandarin speaking aphasic patient behaved independently from the segmental errors that the same patient made. Liang and Heuven therefore concluded that the tone in Mandarin, and in tonal languages in general, is processed or represented differently from segments. The latter proposal was considered as the *autosegmental view* according to Goldsmith's (1976) definition, which posits that tone has its unique representation separate from segmental units.

2.2.2 Nature of sub-syllabic representations

There is evidence to suggest that sub-syllabic units such as onsets and rhymes are represented in the Cantonese lexicon and the effect of rhyme is more salient than the onset. For instance, in a PWI study conducted by Wong and Chen (2008), a significant phonological facilitation effect was found when the target (e.g., 星 /sing1/[7], meaning "Star") and the distractor shared the same rhyme component (e.g., /ging2/) but no significant effects were observed if they shared only the same onset component (e.g., /sam3/). More importantly, the observed rhyme-related facilitation effect was not affected by the complexity of the target's rhyme component (e.g., V or VC rhymes), indicating that the rhyme unit might have its own representation in Cantonese spoken word planning.

Furthermore, in a later Cantonese PWI study, Wong and Chen (2009) investigated whether a complex rhyme (e.g., a rhyme with a VC structure) in Cantonese can further be decomposed into its subparts such as vowels and codas. In that study, a CV-related condition (i.e., targets and distractors shared the same initial-consonant, vowel, but have a different tone and coda, e.g., /sing1/ and /sim4/) was included, and the rationale was that if the rhyme in Cantonese is a non-decomposable minimal unit, then a CV-related distractor would be equivalent to an onset related distractor (because they both share the same onset but have a different rhyme with the target), and a null effect would be expected in such condition. However, a significant facilitation effect was found in the CV-related condition by Wong and Chen (2009), indicating that phonological units such as vowels and codas also have their own representations in the Cantonese lexicon.

2.2.3 Relationship between tonal and segmental processes

Mixed results have been reported regarding this issue. The results from speech error and patient studies, as the ones reviewed above (Chen, 1999; Liang & van Heuven, 2004), seem to suggest that tonal and segmental processes are operated independently from each other but there is also chronometric evidence suggesting that interactivity might exist between the two. For instance, in Wong and Chen's (2008) Cantonese PWI study, null effects were obtained when the target and the distractor shared only the same tone, whereas a significant facilitation effect was found when they shared the same syllable without tone. More importantly, a reliably stronger priming effect, relative to the syllable (without tone) priming, was found when the target and the distractor shared both the same syllable and tone. The null effect of the Tone-related condition, and the reliable difference between Syllable+Tone related and Syllable (without tone) related conditions, seem to suggest that tonal and segmental processes are interactively related to each other and hence resulted for the non-additive pattern across the three conditions. However, it is plausible that the Syllable+Tone-related facilitation was arising from both lexical (assumed that homophones shared the same word-form representation at the lexical level, see for example, Jescheniak, Meyer, & Levelt, 2003) and sub-lexical levels, whereas the Syllable-related distractors could only induce a sub-lexical priming. Consequently, a reliable difference was observed between the two conditions. As such, there is no clear chronometric evidence yet available which shows the relationship between tonal and segmental processes.

2.2.4 Possible language-specific properties

Previous studies on non-tonal languages (e.g., Dutch and English) have frequently demonstrated a significant role for individual segments in speech production (see Chapter 1 above). However, similar research conducted on tonal languages such as Mandarin or Cantonese showed a different picture. For instance, Chen et al. (2002) investigated the process of phonological encoding in Mandarin word production using a similar word-form preparation task as the one used by Roelofs (1999) but failed to obtain any reliable effect even when the words within a set shared the same word-initial consonant. Likewise, in the Cantonese PWI study conducted by Wong and Chen (2008), no reliable effect was found when the target and the distractor shared only the same word-initial consonant (i.e., onset) or together with the same tone. In another PWI study, Wong and Chen (2009) examined the nature of effective phonological units in Cantonese spoken word planning using mono-syllabic words with a consonant-vowel-consonant (CVC) structure as targets and distractors. Significant facilitation effects, relative to the unrelated control, were found when the target (e.g., /sing1/) and the visually presented distractor (e.g., 食 /sik6/ or 境 /ging2/) shared (at least) the same vowel and an identical consonant, but no significant effect was observed when they shared only a single segment (i.e., the vowel[8]). Taken together, it appears that the salience of individual segments in spoken word planning is subject to cross-language differences.

[8] Note, however, that a significant vowel-related facilitation was observed in Wong and Chen (2008) when the target was a CV syllable. Although this result seems contradictory to the notion that a single segment (e.g., vowel) does not have a significant impact on Cantonese spoken word planning, the vowel-related distractors included in Wong and Chen (2008) also shared the same rhyme-component with the targets. Therefore, it is still unclear whether the vowel-related priming observed in Wong and Chen (2008) was due to the effect of a single segment or due to the effect of the rhyme-component. On the contrary, the targets used in Wong and Chen (2009) were all CVC syllables. Hence, the null effect of the vowel-related distractors showed more clearly that the effect of a single segment is minimal in Cantonese spoken word preparation.

Chapter 3

3.1 Motivations of the present study

Given the unique features of Cantonese speech and the open issues reviewed above, five experiments were therefore conducted in the present study using the PWI paradigm. The motivations for the present study are illustrated in the following.

First, since visual distractors were used in Wong and Chen's (2009) PWI study, and there is evidence showing that orthography and phonology are not as closely linked in Chinese as that in alphabetic languages (e.g., Chen & Juola, 1982; Chen & Shu, 2001), it is plausible that the null effect of vowel-related distractors observed was due to the mode of distractor presentation, but was not to the characteristics of vowels in Cantonese. Furthermore, to my knowledge, no prior studies exist which explored the role of coda in Cantonese spoken word production. Therefore, the issues of whether vowels and codas have an important role to play in Cantonese spoken word planning, as they do in non-tonal language production (e.g., Damian & Dumay, 2009), are yet to be addressed.

Second, as significant facilitation effects were observed in Wong and Chen (2009) when the target and the distractor shared the same vowel and an identical consonant, the authors therefore proposed that an effective phonological unit in Cantonese speech planning is a Vowel+Consonant combination. However, an alternative to this claim is that whenever the target and the distractor shared two similar segments, which does not necessarily include a similar vowel, a significant priming effect would then be observed. Hence, it is an empirical question of whether reliable priming would still be observed in a Cantonese PWI task if the target and the distractor shared only the same onset-consonant and the same final-coda.

Third, mixed results were obtained in the past concerning the role of tone in tonal language production. Therefore, it is still an open issue whether tone is processed or represented similarly as segments (e.g., segmental view vs. autosegmental view). Furthermore, how tone is processed in conjunction with segmental information is still largely unclear. Regarding this issue, two proposals have been put forward.

One proposal, which is derived from the assumption that tone is processed similarly as stress in stress languages (Chen, 1999), holds that tonal and segmental units are independently retrieved during phonological planning. For instance, according to the WEAVER++ model, the segmental contents and the metrical structure of the target (i.e., exceptional word) are spelled-out in parallel, and independently at the beginning of phonological encoding (see also Levelt, 1993; Levelt & Wheeldon, 1994), followed by a segment-to-frame association stage during which the retrieved ordered segments associated with the frame incrementally from left to right. Accordingly, if one assumes that lexical tone is specified by a kind of structure similar to the metrical frame proposed in WEAVER++ (Chen et al., 2002), one would predict that segmental and tonal processes are operated independently, at least at the beginning stage of phonological encoding.

An alternating proposal is that segmental and tonal processes are operated in an interactive manner in Chinese speech processing (Wong & Chen, 2008; Ye & Connine, 1999). For instance, according to the Mandarin speech recognition model proposed by Ye & Connine (1999), there are lexical, phoneme, and toneme representations (for syllabic, segmental, and tonal information, respectively) in the phonological lexicon, in which different representations are connected by bi-directional excitatory links. Hence, activation of a toneme would influence the activation of its corresponding segmental nodes and vice versa. The idea that interactivity exists between segmental and tonal processes was also incorporated in the Cantonese speech production model proposed by Wong and Chen (2008). This model posits a three-layer architecture in which activation flows from a semantic to a lexical level, and then to a sub-lexical level (which includes separate representations for onsets, rhymes, and tones) during spoken word production. Note that feedbacks are allowed from the sub-lexical to the lexical level (see also Dell, 1986, for a similar proposal). Hence, the phonological priming effects observed in a PWI experiment were accounted for by both the pre-activation of the sub-lexical nodes and the enhanced feedback sending from the sub-lexical to the lexical level (Wong & Chen, 2008). Since the links between lexical and sub-lexical nodes are bi-directional, activation of a tonal node can thus, in principle, influence the activation of the related segmental nodes, and vice versa. Accordingly, Wong and Chen's (2008) model predicts that segmental and tonal processes are interactively related to each other in Cantonese spoken word planning.

Fourth, a null effect of onset-relatedness and a reliable facilitation effect of rhyme-relatedness were observed in Wong and Chen's (2008) PWI study, the authors therefore argued that the rhyme unit is more important than the onset in Cantonese spoken word planning. However, since the targets used in Wong and Chen were mostly composed of CVC (or CVG) syllables, it is still unclear whether the null effect of onset-related distractors was due to the (relatively) low degree of segmental overlap with the target, or due to the role of onset. Furthermore, the differential effects between onsets and rhymes might not be due to their phonological properties but due to their phonetic differences. For instance, it is plausible that the rhyme component is more noticeable than the onset and hence leading to a stronger priming effect in a PWI task. However, this possibility has not been addressed before.

3.2 Paradigm

The present study employs the PWI paradigm, which has its roots in the classical Stroop color-naming task (Stroop, 1936; see also van Maanen, van Rijn, & Borst, 2009, for a theoretical discussion on the similarities between the two tasks). After the seminal work reported by Rosinski, Golinkoff, and Kukish (1975), the PWI task has since been widely used as a tool to investigate the processes of spoken word production. In a typical PWI task, participants are asked to name aloud individually presented pictures (e.g., line-drawings of single objects) and ignore an accompanying visual or auditory word distractor. The general findings from the PWI research are that participants' picture naming responses would be slower if the distractor word and the picture name were members of the same semantic category, relative to when they were unrelated (i.e., semantic interference), but faster naming responses would be observed, relative to the unrelated control, if the distractor shared similar segmental contents with the picture name (i.e., phonological facilitation). Although the locus of semantic interference is still controversial (Damian & Bowers, 2003; Janssen, Schirm, Mahon, & Caramazza, 2008), the received view is that the phonological facilitation effect is arising from the phonological encoding process in spoken word production.

For instance, WEAVER++ assumes that "active phonological segments in the perceptual network can also directly affect the corresponding segment nodes in the production lexicon" (Levelt et al., 1999, p.7). Accordingly, the phonological facilitation effect observed in a PWI task is attributed to the pre-activation of the target segments by the phonologically related distractors. In particular, WEAVER++ assumes that the locus of such effect is situated at the segmental retrieval stage during which the target segments are spelled-out in parallel. And the reason for situating the effect at the segmental retrieval stage is based on the results from two lines of research. Studies employing the PWI task have reported significant phonological effects with both word-initial related and word-final related distractors (Meyer & Schriefers, 1991), whereas studies employing the form-preparation task indicated that a reliable phonological priming effect could be observed only when the response words shared the same word-initial but not the same word-final segments (Meyer, 1990, 1991). With these results, Levelt et al. (1999) argued that the priming effect observed in a PWI task is originated at the segmental retrieval stage during which segments are retrieved in parallel (i.e., position does not matter), while the priming effect in a form-preparation task is arising from the syllabification process during which segments are associated with the frame incrementally from left to right (i.e., only word-initial segments can be prepared in advance). In addition, there were experiments conducted to examine the locus of the two above mentioned phonological effects (e.g., Levelt et al., 1999, p.25), and the two effects were found to be additive, indicating that they have different loci in the course of phonological encoding.

3.3 Overview of the five experiments

Experiment 1 was designed to test whether reliable effects would be observed when the target and the distractor shared only the same vowel (i.e., V-related) or the same coda (i.e., c-related). Based on the results from Experiment 1, Experiment 2 was set to investigate whether tones in Cantonese are processed or represented similarly as segments by including distractors which shared with the target a similar vowel and tone (i.e., VT-related), or a similar coda and tone (i.e., cT-related). Given it has been demonstrated in the past Cantonese PWI studies that a reliable priming effect would be observed when the target and the distractor shared the same vowel and coda (i.e., Vc-related, see Wong & Chen, 2008, 2009), so if tone is processed or represented similarly as a segmental unit (e.g., a vowel or a coda), a comparable priming effect should also be observed in the VT-related or the cT-related conditions in Experiment 2.

Experiment 3 was conducted to investigate whether a similar vowel component is necessary for a reliable priming effect to arise. To this end, a Cc-related condition was included in which the target and the distractor shared only the same onset-consonant and the same coda. In addition, a CcT-related condition was also included in Experiment 3 in which the target and the distractor shared not only two similar consonants but also an identical tone. If tone is processed similarly as stress in stress languages (Chen, 1999; Chen et al., 2002), and if it is retrieved independently from segments at the beginning of phonological encoding (Levelt, 1993; Levelt & Wheeldon, 1994), comparable results would be expected across the Cc-related and CcT-related conditions in Experiment 3, because it has been shown that tone by itself was not strong enough to induce any reliable priming (e.g., Chen et al., 2002, Wong & Chen, 2008). On the contrary, if tonal and segmental processes are operated in an interactive manner (Wong & Chen, 2008; Ye & Connine, 1999), it is plausible that a significant difference would be observed across the Cc-related and CcT-related conditions in Experiment 3, similar to the difference observed between the Syllable-related and Syllable+Tone-related distractors in Wong and Chen (2008)[9].

Furthermore, Experiment 4 was conducted to address an unsettled issue concerning the possible roles of onsets and rhymes in Cantonese. As mentioned above, a significant rhyme-related facilitation and a null effect of onset-related distractors were observed by Wong and Chen (2008), however, the rhyme-related distractors included generally possessed more similar segments with the targets, relative to the onset-related distractors. Consequently, the results of Wong and Chen (2008) did not give unequivocal support to the claim that the rhyme, as a processing unit, is more

---

[9] Note that one may argue that tone may interact with segments in a later stage of phonological planning although they are retrieved independently at the initial stage. Therefore, even if a significant difference is observed between CcT and Cc related conditions, it is not in contrast to the proposal that tone and segments are initially processed independently. However, given there is evidence showing that the PWI paradigm is sensitive mostly to the early stage of phonological planning, the proposal that tonal and segmental information is independently retrieved at the beginning of phonological planning would have difficulties in accounting for the difference between CcT and Cc related conditions if such a difference is to be found significant.

important than the onset. To investigate the possible roles of onsets and rhymes in Cantonese speech planning, the same set of picture stimuli with CV mono-syllabic names used in Wong and Chen (2008, Experiment 5) was adopted in Experiment 4, and an onset-related distractor condition was included. Note that a reliable rhyme-related facilitation was found in Wong and Chen (2008, Experiment 5) even when the target was a CV syllable. Hence if the onset unit is as salient as the rhyme unit in Cantonese spoken word planning, a reliable onset-related priming effect should also be observed in Experiment 4 because in this case (i.e., when target is a CV syllable), both onset-related and rhyme-related distractors shared the same number of segment with the target.

Finally, Experiment 5 was designed to examine whether phonetic factors such the segment's sonority level (i.e., loudness) and duration would affect the priming effects observed in a PWI task. To this end, the sonority level and the duration of the target's rhyme component were manipulated in Experiments 5A and 5B respectively, and a rhyme-related distractor condition was included in both sub-experiments. If phonetic factors such as the segment's sonority and duration would affect the size of rhyme-related priming, the expected rhyme-related facilitation effects should be modulated by the type of targets (e.g., targets with a more sonorant rhyme vs. targets with a less sonorant rhyme) in both Experiments 5A and 5B.

Chapter 4

Individual Segments in Cantonese Spoken Word Planning

4.1 Experiment 1 – Vowel and Coda in Cantonese

The targets and the distractors used in the first three experiments of the present study were Cantonese mono-syllables with a Consonant-Vowel-Consonant (CVC) structure[10]. Four types of distractor with varying degrees of phonological overlap with the target (e.g., /sing1/, meaning *Star*) were included in Experiment 1, namely Vc-related (e.g., /ging2/), V-related (e.g., /dim3/), c-related (e.g., /hung2/), and unrelated control (e.g., /gok3/). If individual segments such as vowels or codas are important in Cantonese speech preparation, significant facilitation effects should be observed in both V-related and c-related conditions in this experiment.

4.1.1 Participants

Twenty-eight undergraduates (12 males) from the Chinese University of Hong Kong who were native speakers of Cantonese and had normal or corrected-to-normal vision participated in this experiment. None of them reported having any kind of hearing deficits and they were paid for their participation (HK$30[11] each).

4.1.2 Materials and Design

Twenty pictures were used as stimuli in this experiment (see Appendix A). They were all black-on-the-white line drawings adopted from Wu (2001) with high naming agreements (above 80%). The distractors were all auditorily presented and were

---

[10] Due to the constraints in material construction, the names of four pictures used in Experiments 1 and 2, and also the names of three pictures used in Experiment 3, were Cantonese mono-syllables with a consonant-vowel-glide (CVG) structure. For instances, target syllables such as /coi3/ (*vegetable*) and /paau3/ (*leopard*), are ended with a glide (or semi-vowel) /i/ and /u/ respectively. However, since a glide can also be considered as a unit which occupies a slot within a syllable, CVG syllables were thus adopted in the present study. More importantly, as the accompanying distractors for those CVG targets were also CVG syllables (i.e., they were matched in their syllabic structures), hence the possible influence of syllabic structure on the present study is minimal at best. For the sake of simplicity, the term CVC syllable is used throughout the whole paper in referring to the types of syllable having a 3-slot structure including both CVC and CVG syllables in Cantonese.

[11] Approximately equal to US$4.

orthographically unrelated to the name of the accompanying picture in their written form. In addition, the distractors were matched in frequency and number of homophones across the four conditions (Chinese Character Database: With Word-formations, 2003), *F*s $\leqq$ 1.12 and *ps* $\geqq$ .35. The mean log frequencies and average number of homophones across the four distractor conditions are shown in Table 1.

Table 1. Mean Log Frequencies (*Log Freq*) and Number of Homophones (*NOH*) across the four Distractor Conditions in Experiment 1 (standard deviations in parentheses).

Condition	*Log Freq*	*NOH*
Vc-related	3.41 (0.7)	3.4 (2.3)
V-related	3.49 (0.7)	3.6 (2.5)
c-related	3.53 (0.3)	2.9 (1.5)
Control	3.72 (0.5)	3.5 (1.7)

During the experiment, each participant received two blocks with 40 trials each. There were equal numbers of trials from each distractor condition and each picture appeared twice in each block. The order in which participants received the blocks was counter-balanced across participants. The items within each block were presented in a randomized order but the same picture would not repeat in consecutive trials.

4.1.3 Apparatus

The pictures were presented on a Flex Scan 17-in. monitor, with a Dell Pentium-IV computer running the DMDX program[12]. The pictures were digitized and

[12] The experiments reported in this paper were run using the DMDX experimental software programmed by Jonathan Forster at the University of Arizona.

displayed at a size of 9cm x 12cm. The auditory distractors used were made by a female Cantonese native speaker who was naïve to the design of the experiment, digitized with a sampling rate of 44.1 kHz, and presented to the participants through a pair of Sony MDR-NC 20 Noise Canceling headphones. Vocal responses were recorded via a microphone connected to the computer. The computer recorded the onset of the responses to the nearest millisecond by a voice onset relay.

4.1.4 Procedure

Participants were tested individually. Each participant was first familiarized with the pictures by viewing each picture for 2 seconds with the picture name appeared below it. A practice block was then administered in which each picture was presented once and the participant was asked to name the picture. Any response other than the expected one (extremely rare though) was corrected by the experimenter.

Before each trial block, participants were instructed to ignore the accompanying distractor, and to name the picture through the microphone as quickly and accurately as possible. In each trial, a centered-fixation-cross was first presented for 500 ms followed by a blank of 500 ms. After that, the target picture and the distractor were presented simultaneously. The picture disappeared once a vocal response was detected or after a lapse of 2000 ms. Participant's naming latency was measured from the time when the picture was shown. The next trial began after a 1000-ms blank.

4.1.5 Results and Discussion

The same criteria were used for data trimming and data analysis across the first three experiments reported here. Three following types of response were excluded from data analyses: Incorrect responses (0.3%), trials in which the voice key malfunctioned or mistriggered (1.7%), and response times which were 2.5 standard deviations beyond either participant or item mean (4.6%).

A one-way analysis of variance (ANOVA) with Distractor Type (4 levels) as the within-participants and within-items factor was performed on the rest of the data. A significant effect of Distractor Type was found in both by participants [F_1 (3, 81) = 3.39, MSE = 793.1, p = .039] and by items analyses [F_2 (3, 57) = 2.88, MSE = 347.3, p = .044]. Then paired t-tests were performed to compare each phonologically related condition with the unrelated control condition. A significant facilitation effect, relative to the control condition, was found in the Vc-related condition in both by participants [$t_1(27) = 2.4$, $p = .023$] and by items analyses [$t_2(19) = 2.25$, $p = .037$]. However, no significant effects were found in the other two conditions (all ts $\leqq$ 1.62, and ps $\geqq$ .12).

In addition, a reliable difference was also observed between Vc-related and c-related conditions [$t_1(27) = 2.12$, $p = .043$; $t_2(19) = 3.18$, $p = .005$]. No other significant comparisons were found. The mean naming latencies (by participants) in various conditions are shown in Table 2. A similar analysis was conducted on the error data but no significant results were observed (Fs < 1).

Table 2. Mean Naming Latencies (M, in ms) and Error Rates (*Errors*, in %) per condition in Experiment 1 (standard deviations in parentheses).

	M	*Errors*
Vc-related	551(73)	0.54(1.58)
V-related	561(85)	0.18(0.95)
c-related	569(90)	0.36(1.32)
Unrelated Control	568(78)	0.18(0.95)

The significant facilitation obtained in the Vc-related condition in Experiment 1 replicated the previous findings (Wong & Chen, 2008, 2009). Furthermore, the null effects obtained in the V-related and the c-related conditions add to the evidence that a single segment alone is not strong enough for a reliable priming effect to arise in a Chinese spoken word production task (Chen et al., 2002; Wong & Chen, 2008, 2009). These results are consistent with the notion that an individual segment does not have a significant impact on Chinese spoken word preparation.

As mentioned in Introduction, tone is lexically distinctive in Cantonese, similar to its segmental counterparts. An interesting but yet unsettled issue is whether tones are processed or represented similarly as segments in Cantonese (i.e., segmental view), or if they are represented separately (i.e., autosegmental view). If tone plays a similar role as segments (e.g., vowel and coda) in Cantonese spoken word planning, a reliable priming effect, similar to that observed in the Vc-related condition in Experiment 1, should also be observed in a VT-related or cT-related conditions. Experiment 2 was thus conducted to address this issue.

Chapter 5

Lexical Tone in Cantonese Spoken Word Planning

5.1 Experiment 2 – Lexical Tone in Cantonese

The design of Experiment 2 was largely similar to that of Experiment 1 except that the phonologically related distractors also shared the same lexical tone with the target (e.g., /sing1/). Consequently, the distractors employed in Experiment 2 included VcT-related (i.e., distractors and targets shared the same vowel, coda, and tone, e.g., /ging1/), VT-related (e.g., /bit1/), cT-related (e.g., /fung1/), and unrelated control.

5.1.1 Participants

Twenty-six undergraduates (14 males) from the Chinese University of Hong Kong who were native speakers of Cantonese and had normal or corrected-to-normal vision participated in this experiment. None of them had participated in the previous experiment or reported any kind of hearing deficits. They were paid for their participation (HK$30 each).

5.1.2 Materials and Design

The twenty pictures used in Experiment 1 were adopted as stimuli in this experiment. Similar to Experiment 1, the distractors were all auditorily presented and were orthographically unrelated to the name of the accompanying picture in their written form (see Appendix A). In addition, the distractors were matched in frequency and number of homophones across the four conditions in this experiment (Chinese Character Database: With Word-formations, 2003), $Fs \leqq 1.51$ and $ps \geqq .22$. The mean log frequencies and average number of homophones across the four distractor conditions are shown in Table 3. Furthermore, the overall design, apparatus, and procedure adopted in Experiment 2 were identical to those in Experiment 1.

Table 3. Mean Log Frequencies (*Log Freq*) and Number of Homophones (*NOH*)

across the four Distractor Conditions in Experiment 2 (standard deviations in parentheses).

Condition	*Log Freq*	*NOH*
VcT-related	3.65 (0.6)	4.4 (3.1)
VT-related	3.47 (0.5)	3.9 (3.7)
CT-related	3.22 (0.8)	4.2 (2.4)
Control	3.72 (0.6)	3.5 (1.7)

5.1.3 Results and Discussion

Overall, 0.5% (incorrect responses), 1.8% (voice key mistriggered), and 3.7% (2.5 SD beyond either participant or item mean) of the original data were eliminated.

The results of ANOVA showed a significant effect of Distractor Type though such effect was found significant in the by participants analysis only [F_1 (3, 75) = 3.15, MSE = 398.5, p = .03; F_2 (3, 57) = 1.72, MSE = 539.4, p = .17]. Subsequent paired t-tests showed that a significant facilitation effect, relative to the control, was found in the VcT-related condition in both analyses [$t_1(25) = 3.01$, $p = .006$; $t_2(19) = 2.46$, $p = .024$], but no reliable effects were obtained in the remaining phonological conditions (all ts $\leqq$ 1.24, and ps $\geqq$ .22).

In addition, a reliable difference was also observed between VcT-related and cT-related conditions in the by participants analysis [$t_1(25) = 2.56$, $p = .017$], though was not in the by items analysis [$t_2(19) = 1.14$, $p = .27$]. No other significant comparisons were found. The mean naming latencies (by participants) in various conditions are shown in Table 4. A similar analysis was conducted on the error data but no significant results were obtained (Fs < 1.98, ps > .1).

Table 4. Mean Naming Latencies (M, in ms) and Error Rates (*Errors*, in %) per

condition in Experiment 2 (standard deviations in parentheses).

	M	*Errors*
VcT-related	573(85)	0.20(0.99)
VT-related	584(75)	0.39(1.36)
CT-related	583(89)	0.39(1.36)
Unrelated Control	590(83)	1.16(2.15)

Consistent with the previous findings, a reliable facilitation effect was observed in the VcT-related condition in Experiment 2 (Wong & Chen, 2008, 2009). Furthermore, the null effects observed in the VT-related and the cT-related conditions provide convergent evidence to the claim that a single segment alone, even together with the tone, does not have a significant role to play in Cantonese spoken word planning. More importantly, the discrepant results observed across the Vc-related condition in Experiment 1 and the VT-related or cT-related condition in Experiment 2 indicate that the lexical tone is processed or represented differently from segments in Cantonese, because otherwise one would expect to see reliable facilitation effects in these two condition.

Taken together, the null effects obtained in the VT-related and cT-related conditions in Experiment 2, together with the reliable facilitation effect obtained in the Vc-related condition in Experiment 1, indicate that the effect of tone behaves differently from that of the segments. These results are inconsistent with the segmental view which holds that tones are the same as segments (Fromkin, 1980; Moser, 1991; Packard, 1986), but are consistent with the autosegmental view which posits that tone is represented separately from segmental units (e.g., Liang & van Heuvan, 2004). Nevertheless, a related but yet unsettled issue is how tonal and segmental information is retrieved in the process of phonological encoding. Experiment 3 was conducted in an attempt to address this issue using a similar PWI task.

Chapter 6

Relationship between Tonal and Segmental Processes

6.1 Experiment 3 – Tone and Segments in Cantonese Speech

One unanimous finding from the first two experiments and from the previous studies is that a significant facilitation effect could be observed only when the target and distractor shared the same vowel together with an identical consonant, consistent with the claim that an effective planning unit in Cantonese speech is (minimally) a vowel + consonant combination (Wong & Chen, 2009). Yet, an alternative to this hypothesis is that a similar vowel is not necessary, but that two similar segments shared between the target (e.g., /sing1/) and the distractor is essential for a reliable effect to arise. Experiment 3 was thus designed to examine this possibility by including Cc-related (e.g., soeng3/), CV-related (e.g., /sik6/), and unrelated control conditions.

Moreover, the issue of how tone is processed in conjunction with segments was also addressed in Experiment 3 by including a CcT-related condition (e.g., /soeng1/). Note that it has been reported that tone alone was not sufficient enough to produce any reliable priming in a PWI task (Wong & Chen, 2008). Therefore, if segmental and tonal processes are operated independently of each other in the early stage of Cantonese spoken word planning, one would expect to see a similar result across Cc-related and CcT-related conditions in the present experiment. Conversely, if the two processes are interactively related to each other, it is plausible that a significant difference would be found across the two conditions, similar to the difference between Syllable-related and Syllable+Tone-related conditions observed in Wong and Chen (2008).

6.1.1 Participants

Twenty-six undergraduates (12 males) from the Chinese University of Hong

Kong who were native speakers of Cantonese and had normal or corrected-to-normal vision participated in this experiment. None of them had participated in the previous experiments or reported any kind of hearing deficits. They were paid for their participation (HK$30 each).

6.1.2 Materials and Design

Since the original set of target pictures used in Experiments 1 and 2 could not yield all of the three types of phonologically related distractors required in this experiment, a slightly different set of target pictures was adopted in Experiment 3 (see Appendix B). Note, however, that the twenty pictures used in Experiment 3 were also adopted from Wu (2001) and had comparably high naming agreements (above 80%). Similar to the previous two experiments, the distractors were all auditorily presented and were orthographically unrelated to the name of the accompanying picture in their written form in Experiment 3. In addition, distractors across the four conditions were matched in their frequency and number of homophones (Chinese Character Database: With Word-formations, 2003), $Fs \leqq 1.9$ and $ps \geqq .13$. The mean log frequencies and average number of homophones across the four distractor conditions are shown in Table 5. Furthermore, a similar design, apparatus set, and procedure were adopted in Experiment 3 as those in the first two experiments.

Table 5. Mean Log Frequencies (*Log Freq*) and Number of Homophones (*NOH*) across the four Distractor Conditions in Experiment 3 (standard deviations in parentheses).

Condition	*Log Freq*	*NOH*
CcT-related	3.32 (0.7)	3.8 (2.6)
Cc-related	3.31 (0.7)	4.1 (2.7)

CV-related	3.27 (0.8)	2.7 (2.1)
Control	3.73 (0.6)	3.0 (1.6)

6.1.3 Results and Discussion

Overall, 0.6% (incorrect responses), 3.1% (voice key mistriggered), and 3.4% (2.5 SD beyond either participant or item mean) of the original data were eliminated.

The results of ANOVA showed a significant effect of Distractor Type in both analyses [F_1 (3, 75) = 14.0, *MSE* = 629.1, $p < .001$; F_2 (3, 57) = 9.37, *MSE* = 668.5, $p < .001$]. Subsequent paired *t*-tests showed that significant facilitation effects, relative to the control condition, were obtained in the CcT-related [$t_1(25) = 5.01, p < .001$; $t_2(19) = 5.31, p < .001$], Cc-related [$t_1(25) = 3.19, p = .004$; $t_2(19) = 2.13, p = .047$], and CV-related [$t_1(25) = 4.8, p < .001$; $t_2(19) = 4.15, p = .001$] conditions.

In addition, significant differences were also observed between CcT-related and Cc-related conditions [$t_1(25) = 2.72, p = .012$; $t_2(19) = 2.13, p = .047$], and between CV-related and Cc-related conditions [$t_1(25) = 2.97, p = .006$; $t_2(19) = 2.1, p = .05$], but no reliable difference was found between CcT-related and CV-related conditions (both *t*s < 1). The mean naming latencies (by participants) in various conditions are shown in Table 6. A similar analysis was conducted on the error data but no significant results were found (*F*s < 1.19, *p*s > .1).

Table 6. Mean Naming Latencies (*M*, in ms) and Error Rates (*Errors*, in %) per condition in Experiment 3 (standard deviations in parentheses).

	M	*Errors*
CcT-related	552(76)	0.39(1.36)
Cc-related	568(77)	0.77(1.84)

CV-related	549(85)	0.20(0.99)
Unrelated Control	589(76)	0.97(2.01)

A reliable facilitation was found in the Cc-related condition in Experiment 3, indicating that a similar vowel shared between target and distractor is not necessary for a significant priming effect to arise in a Cantonese PWI task. More importantly, a more robust facilitation was observed in the CcT-related condition than that in the Cc-related condition. Since previous research has demonstrated that no reliable effects were obtained when the target and the distractor shared only the tone (Wong & Chen, 2008), the difference observed between Cc-related and CcT-related conditions in the present experiment appears to suggest that segmental and tonal processes are operated in an interactive manner. I will further discuss the implications of these results together with the findings from previous Cantonese PWI research in General Discussion.

Apart from the above, the priming effect observed in the CV-related condition was found reliably larger than that in the Cc-related condition. This result is rather unexpected since the degree of segmental overlap between target and distractor was the same across these two conditions. However, two plausible accounts, which possess different assumptions on the locus of the phonological effects, are available for this result. The first account is based on the general assumption that the phonological effect observed in a PWI task is arising from the stage during which segments are retrieved in parallel (Levelt et al., 1999). Accordingly, the difference observed between CV-related and Cc-related conditions suggests that the vowel in Cantonese carries a heavier weight than that of a coda, and thus leading to a stronger priming effect in the CV-related than that in the Cc-related condition. However, such a conclusion is in contrast to the assumption that segments are the basic processing units which are of equally important in the course of phonological encoding (Dell, 1986, 1988; Levelt et al., 1999). Regarding the issues of whether segments are the only legitimate phonological units in Cantonese (apart from prosodic information such as the tone), and if different processing units posses different weights in phonological encoding will be further investigated in Experiment 4.

Alternatively, one might argue that the effects observed in the present experiment might not only arise from the segmental retrieval stage, but also from the subsequent stage in which segmental units are encoded serially during the segment-to-frame association process. Specifically, previous studies using the form-preparation paradigm have shown that the segment-to-frame association process can be facilitated when the response words within a set shared the word-initial segment, and that the size of priming is dependent on the number of word-initial segments shared among them, whereas no reliable effect was found when they shared only the word-final segments (Meyer, 1991; Roelofs, 1999). Therefore, if the PWI paradigm presently adopted is also sensitive to the segment-to-frame association process, one would expect to see a stronger facilitation produced by the CV-related distactors than by the Cc-related distractors, because the former can (possibly) facilitate the segment-to-frame association process up to the target's vowel position, but for the latter, only the association between the target's initial-segment and the frame can be facilitated.

However, the second account discussed above faces a number of difficulties. Among other things, it is not clear whether a frame retrieval process is needed for mono-syllabic word production (Levelt et al., 1999). Also, even if a segment-to-frame association process is involved in Cantonese mono-syllabic word production, it is not clear what kind of information is represented by the frame structure. If one assumes that the frame specifies the tonal information of the syllable (Chen et al., 2002), then neither CV-related nor Cc-related distractors would benefit from the segment-to-frame association process because they both have a different tone with the target, and there is evidence to suggest that the segment-to-frame association process could be facilitated only when both metrical structure and word-initial segments were primed (Roelofs & Meyer, 1998). On the contrary, one might argue that the frame specifies the CVC structure of the syllable which is similar to the frame representation proposed by Dell (1986, 1988). Based on this assumption, however, one would predict no difference between CV-related and Cc-related conditions because the segments within a syllable are encoded in parallel according to Dell (1986, 1988).

Although the present data do not differentiate the above two proposed accounts, it appears that the first proposed account is relatively more plausible and parsimonious. If this is the case that certain phonological units in Chinese carry a heavier weight than the others (see Tong, Francis, & Gandour, 2008; Wong & Chen, 2008, for a similar notion), then a crucial but yet unresolved issue is what units are more important and what units are less so. Experiment 4 was hence conducted to address this issue.

Chapter 7

Onsets and Rhymes in Cantonese Speech Preparation

7.1 Experiment 4 – Onsets and Rhymes in Cantonese Speech

The results from the first three experiments, as well as those from Wong and Chen (2009), seem to suggest that an individual segment is not strong enough to produce any reliable priming in a Cantonese PWI task. However, in Wong and Chen's (2008) study, a significant facilitation effect was observed in a vowel-related condition where the target was a CV syllable in Cantonese (i.e., the target and the distractor shared only one segment). In Experiment 5 of Wong and Chen's study (2008), half of the target picture names were Cantonese mono-syllables with a CV structure and the other half a CVC structure. A syllable (without tone) related and a rhyme-related condition were included in that experiment, and the distractors were all auditorily presented. Significant priming effects were observed in both syllable-related and rhyme-related conditions, and more importantly, the facilitation effects observed in the two conditions were not affected by the syllable structure (i.e., CV or CVC) of the targets, indicating that a reliable phonological effect was found even when the target (i.e., a CV syllable) and the distractor (i.e., a rhyme-related distractor) shared only a single segment (i.e., the vowel).

There are two possible accounts for this result. One is that the degree of segmental overlap between target and distractor is the crucial factor which determines whether a significant facilitation effect would arise in a Cantonese PWI task. In particular, if one assumes that a reliable priming effect would be observed when the target and the distractor shared at least 50% of similar segments, one can easily explain why a significant vowel-related priming effect was observed when the target was a CV syllable (Wong & Chen, 2008), as well as the findings that a reliable facilitation would be observed only when the distractor and a CVC target shared at least two similar segments (e.g., CV-related, Cc-related, and Vc-related). This is because the distractors in all these conditions shared at least 50% of their segmental contents with their corresponding CV or CVC targets. According to this segmental overlap hypothesis, one would also predict that a reliable onset-related priming effect would be observed if the target is a CV syllable. However, since the targets used in Wong and Chen (2008, Experiments 1 and 4) for investigating the effect of onset-related distractors were mostly composed of CVC syllables in Cantonese, it is still unclear whether a reliable onset-related facilitation would be observed if the target is a CV syllable.

Apart from the segmental overlap hypothesis, an alternating account for the observed results is that the rhyme component of a syllable in Cantonese has its own representation and is an important unit in phonological planning (Wong & Chen, 2008). Accordingly, although a vowel-related distractor shared with its CV target only a single segment (i.e., the vowel), they also shared the same rhyme component. Since the rhyme is an important phonological unit, the priming of such unit could be the reason for the significant vowel-related facilitation effect observed among the targets with a CV syllable structure. Note, however, that this proposal does not exclude the possibility that a Vc- rhyme can possibly be decomposed into its vowel and coda sub-parts. The issue of whether vowels and codas are uniquely represented in the Cantonese phonological lexical will further be discussed in General Discussion. Based on the alternating account proposed here, the rhyme unit has a more privileged role to play than other sub-lexical phonological units such as onset or tone. Accordingly, one would predict that a null effect of onset-relatedness would be observed even among the targets with a CV structure.

To verify the two alternating hypotheses mentioned above, Experiment 4 was thus conducted to investigate whether a significant onset-related priming effect would be observed among the targets with a CV structure. To better compare the effects of onset-relatedness and rhyme-relatedness, the same set of pictures used in Experiment 5 of Wong and Chen (2008) was adopted in the present experiment. Accordingly, half of the targets used in the present experiment were Cantonese mono-syllables with a CV structure and the other half a CVC structure. Similar to Wong and Chen (2008), a syllable (without tone) related condition was included in this experiment as a check to see if a similar syllable-related facilitation effect would also be observed here. More importantly, an onset-related condition was included in the present experiment. If the segmental overlap hypothesis holds, one would expect to see a significant onset-related facilitation effect among the targets with a CV structure but not among the targets with a CVC structure. Therefore, an Onset-relatedness x Target Type interaction would be expected in the present experiment according to the segmental overlap hypothesis. Alternatively, if the vowel-related priming observed among the CV targets previously reported by Wong and Chen (2008) was not due to the overall segmental overlap between target and distractor, but was due to the reason that the rhyme in Cantonese has a unique role to play in phonological encoding, then a null effect of onset-relatedness would be expected in the present experiment irrespective of the target's syllable structure. This is because past studies using different word production paradigms have demonstrated that the onset-consonant by itself does not have a significant impact on Chinese spoken word planning (Chen et al., 2002; Wong & Chen, 2008).

7.1.1 Participants

Similar to Wong and Chen (2008, Experiment 5), 20 undergraduates (8 males) from the Chinese University of Hong Kong who were native speakers of Cantonese and had normal or corrected-to-normal vision participated in this experiment. None of them had participated in the previous experiments or reported any kind of hearing deficits. They were paid for their participation (HK$30 each).

7.1.2 Materials and Design

The twenty target pictures used in Wong and Chen (2008, Experiment 5) were adopted as stimuli in the present experiment (see Appendix C). Ten of the twenty target picture names were Cantonese mono-syllables with a CV structure and the other ten with a CVC structure. Two types of phonologically related distractors were included in the present experiment, including syllable (without tone) related (i.e., the target and the distractor shared the same syllable but not the same lexical tone) and onset-related (i.e., the target and the distractor shared the same onset-consonant only) distractors. Two unrelated control conditions were also included in this experiment by re-arranging the target-distractor pairings in the syllable-related and the onset-related conditions. Consequently, the present experiment adopted a 2 (Target Type: CV vs. CVC syllables) by 2 (Distractor Type: Syllable based vs. Onset based) by 2 (Relatedness: Phonologically related vs. Unrelated) experimental design.

Similar to the previous experiments, the distractors were all auditorily presented and were orthographically unrelated to the name of the accompanying picture in their written form in Experiment 4. Note that although the same set of distractors was used across the phonologically related and the unrelated control conditions in this experiment, the two types of distractor (i.e., syllable based and onset based) used in the two target conditions were also matched in their frequencies and number of homophones (Chinese Character Database: With Word-formations, 2003), $F\text{s} < 1$. The mean log frequencies and average number of homophones across conditions are shown in Table 7. Apart from the above, the apparatus and procedure adopted in this experiment were the same as those employed in the previous experiments.

Table 7. Mean Log Frequencies (*Log Freq*) and Number of Homophones (*NOH*)

across Distractor Conditions in Experiment 4 (standard deviations in parentheses).

Target Type	Distractor Type	*Log Freq*	*NOH*
CV	Syllable-based	3.30 (0.7)	3.3 (1.3)
	Onset-based	3.32 (0.6)	3.3 (2.6)
CVC	Syllable-based	3.48 (0.9)	4.3 (3.1)
	Onset-based	3.22 (0.7)	2.4 (2.0)

7.1.3 Results and Discussion

Three following types of response were excluded from data analyses: Incorrect responses (0.6%), trials in which the voice key malfunctioned or mistriggered (2.4%), and response times which were 2.5 standard deviations beyond either participant or item mean (1.9%).

An ANOVA with Target Type (2 levels: CV vs. CVC targets), Distractor Type (2 levels: Syllable-based vs. Onset-based distractors), and Relatedness (2 levels: Phonologically Related vs. Unrelated Control) as the three within-participants factors was performed on the rest of the data. A similar ANOVA was conducted for items analysis except that Target Type was treated as a between-items factor in the by items analysis. The main effect of Relatedness was found significant in both by participants and by items analyses [F_1 (1, 19) = 15.1, *MSE* = 1004.3, *p* = .001; F_2 (1, 18) = 8.23, *MSE* = 715.2, *p* = .01], showing that participants' naming latencies were significantly shorter if the target and the distractor were phonologically related than if they were unrelated. Also, the Target Type x Distractor Type interaction was found significant in both analyses [F_1 (1, 19) = 8.23, *MSE* = 1457.7, *p* = .01; F_2 (1, 18) = 5.81, *MSE* = 1056.4, *p* = .027]. In addition, a reliable Distractor Type x Relatedness interaction was found in the by participants analysis though was not in the by items analysis [F_1 (1, 19) = 7.19, *MSE* = 976.1, *p* = .015; F_2 (1, 18) = 2.03, *MSE* = 1243.3, *p* = .17]. No other significant effects were observed. A summary of the ANOVA results is shown in Table 8.

Table 8. Analyses of Variance with Target Type, Distractor Type, and Relatedness as variables in Experiment 4.

	By participants				By items			
Source	F_1	*MSE*	*p*		F_2	*MSE*	*p*	
Target Type (TT)	1.96	1542.2	0.18		0.26	9284.9	0.62	
Distractor Type (DT)	0.4	2693.2	0.54		0.3	1056.4	0.59	
Relatedness	15.14	1004.3	0.001	***	8.23	715.2	0.01	***
TT x DT	8.23	1457.7	0.01	***	5.81	1056.4	0.03	**
TT x Relatedness	0.24	1922.6	0.63		1.92	715.2	0.18	
DT x Relatedness	7.19	976.1	0.015	**	2.03	1243.3	0.17	
TT x DT x Relatedness	0.62	1376.1	0.44		0.06	1243.3	0.81	

Since a significant Target Type x Distractor Type interaction was observed, the data from the related and the unrelated conditions were collapsed, and the results pattern indicates that participants' mean naming latencies towards the CV targets were slower if the distractor was a Syllable-based distractor relative to an Onset-based distrator, and the opposite was true if the target was a CVC syllable. However, subsequent *t*-tests yielded no significant comparisons (*ts* $\leq$ 2.08, and *ps* $\geq$ .065), indicating that the Target Type x Distractor Type interaction observed was induced by the overall pattern of the results.

In addition, as the Distractor Type x Relatedness interaction was also found significant in the by participants analysis in this experiment, subsequent t-tests were conducted to examine the effects of Relatedness across the two types of distractor. The results indicate that participants' naming latencies in the Syllable-related condition were found significantly shorter than that in the Syllable-unrelated control condition [$t_1(19) = 3.9, p = .001; t_2(19) = 2.42, p = .026$]. However, no significant difference was observed between the Onset-related condition and its unrelated control condition [$t_1(19) = 1.17, p = .26$; $t_2(19) = .80, p = .44$].

A similar analysis was also conducted on the error data. The main effect of Distractor Type was found significant in the by items analysis [F_2 (1, 18) = 7.71, *MSE* = 1.46, $p = .012$], but was not in the by participants analysis [F_1 (1, 19) = 2.11, *MSE* = 2.66, $p = .16$], indicating that the error rates were generally higher when the distractor was an Onset-based distractor relative to a Syllable-based distractor. However, no other significant results were obtained (Fs ≤ 2.92, and $ps \geq .104$). The mean naming latencies and error rates (by participants) in various conditions are shown in Table 9.

Table 9. Mean Naming Latencies (*M*, in ms) and Error Rates (*Errors*, in %) per condition in Experiment 4 (standard deviations in parentheses).

Target Type	Condition	*M*	*Errors*
CV	Syllable related	648 (154)	0.25 (1.12)
	Control	689 (157)	0 (0)
	Onset related	654 (146)	0.75 (1.84)
	Control	659 (158)	0.25 (1.12)
CVC	Syllable related	630 (145)	0 (0)
	Control	655 (116)	0.25 (1.12)
	Onset related	661 (163)	0.5 (2.24)
	Control	668 (143)	0.5 (1.54)

The results of Experiment 4 are clear: A reliable syllable-related facilitation effect was found in both CV and CVC target conditions, whereas no onset-related priming was observed irrespective of the target's syllable structure. The null effect of onset-related distractors on CV targets stands in an interesting contrast to the reliable rhyme-related priming observed in Wong and Chen (2008) with the use of an identical set of picture materials. The differential effects of onset-related and rhyme-related distractors on CV targets are inconsistent with the segmental overlap hypothesis, which would otherwise predict a similar result across the two types of distractor. On the contrary, this pattern of results is in line with the notion that the rhyme in Cantonese is an important processing unit which carries a heavier weight than that of the onset or the tone components (e.g., Wong & Chen, 2008).

Nevertheless, it is important to note that onsets and rhymes are different from each other in their phonetic properties. For instance, vowels (i.e., rhymes, for CV syllables) are in general more sonorant (i.e., louder) and have a longer duration than the consonants (i.e., onsets, for CV syllables). As a result, it is plausible that the stronger effect observed in the rhyme-related condition was simply due to the reason that the rhyme component is relatively more noticeable such that participants could be more easily to detect the similarity between target and the rhyme-related distractor. Experiment 5 was therefore designed to verify this hypothesis.

Chapter 8

Potential Influences from Phonetic Features

The present experiment investigated whether the discrepant effects produced by onset-related and rhyme-related distractors on CV targets were due to the *phonological* differences between onsets and rhymes (i.e., an output effect) or due to their *phonetic* differences (i.e., an input effect). In particular, two sub-experiments were carried out in an attempt to investigate whether phonetic factors such as the segment's sonority level and duration would affect the phonological effects observed in a PWI task.

8.1 Experiment 5A – Sonority of the Rhyme Unit

According to Ladefoged (2001), the sonority level of a phone reflects the resonance of the phone relative to other phones in the speech. In other words, a more sonorant phone is generally perceived louder than a less sonorant phone. Generally speaking, the sonority levels of consonants are lower than that of high vowels (e.g., /i/, /u/, and /yu/), which in turn lower than low vowels (e.g., /a/, /o/, and /e/). To investigate whether the sonority level of the phonological units shared between target and distractor would affect the facilitation effects observed in a PWI task, Experiment 5A was conducted using the PWI paradigm in which a vowel related condition was included and the sonority level of the targets' vowel component was manipulated. Specifically, the names of the target pictures used in the present experiment were all mono-syllables in Cantonese with a CV structure, and half of them were composed of a low vowel (i.e., a more sonorant vowel) whereas the other half a high vowel (i.e., a less sonorant vowel). If this is the case that the sonority level of the to-be-primed segment plays an important role in determining whether a significant phonological effect would arise in a Cantonese PWI task, then a significant Target Type (i.e., targets with a low vowel vs. targets with a high vowel) x Vowel Relatedness interaction effect would be expected in the present experiment. On the contrary, if the sonority level of the to-be-primed segment only has a minimal role to play, comparable vowel-related priming effects would be expected across the two types of targets in this experiment.

In addition, the purpose of using CV syllables as targets in this experiment is two-fold. First is to ensure that the comparison between low vowels and high vowels is valid by using targets with a CV but not a CVC structure. This is because the phonetic features of a vowel would be changed by a subsequent segment (e.g., a coda). Second is to replicate the vowel-related facilitation effect previously reported in Wong and Chen (2008) in which the targets were CV syllables in Cantonese. Since the major conclusion made in Experiment 4 is very much dependent on such a vowel-related priming effect, a replication of this finding would be of necessary in the present study.

8.1.1 Participants

Twenty-six undergraduates (18 males) from the Chinese University of Hong Kong who were native speakers of Cantonese and had normal or corrected-to-normal vision were recruited to participate in both Experiments 5A and 5B. Since the time taken for each sub-experiment was rather short (about 15 minutes for each participant) and different sets of materials were adopted across the two sub-experiments, a same group of participants was thus recruited to participate in both Experiments 5A and 5B. Nevertheless, none of the participants in Experiment 5 had participated in the previous experiments or reported any kind of hearing deficits. And they were paid for their participation (HK$30 each). In addition, the presentation order of the two sub-experiments was randomly assigned to the participants such that about half of them (11 males and 4 females) did Experiment 5A prior to Experiment 5B, and a reversed order was applied to the remaining participants.

8.1.2 Materials and Design

Eighteen pictures were used as stimuli in Experiment 5A (see Appendix D). They were all black-on-the-white line drawings adopted from Wu (2001) with high naming agreements (above 80%). The names of the pictures were all Cantonese monosyllables with a CV structure. Among the eighteen target picture names, nine of them consisted of a low vowel (i.e., a vowel with relatively high sonority level) such as /a/, /o/, and /e/, whereas the remaining nine picture names contained a high vowel (i.e., a vowel with relatively low sonority level) such as /i/, /u/, and /yu/. The distractors were all auditorily presented and were orthographically unrelated to the name of the accompanying picture in their written form. Two distractor conditions were included in this experiment, namely vowel-related and unrelated control. Similar to Experiment 4, an identical set of distractors was used across the vowel-related and the unrelated conditions only that the target-distractor pairings were different across the two conditions. Furthermore, the distractors used across the two target type conditions were matched in their frequency and number of homophones (Chinese Character Database: With Word-formations, 2003), $Fs \leq 1.18$ and $ps \geq .29$. The mean log frequencies and average number of homophones of the distractors across conditions are shown in Table 10.

Table 10. Mean Log Frequencies (*Log Freq*) and Number of Homophones (*NOH*) across Distractor Conditions in Experiment 5A (standard deviations in parentheses).

Rhyme Sonority	*Log Freq*	*NOH*
High	3.29 (0.6)	2.89 (1.6)
Low	3.46 (0.8)	2.89 (3.2)

During the experiment, each participant received all the 36 trials within a single block. The items were presented in a randomized order but the same picture would not repeat in consecutive trials. Apart from the above, the procedure of each trial and the apparatus used for stimuli presentation in this experiment were largely identical to those used in the previous experiments.

8.1.3 Results

Three following types of response were excluded from data analyses: Incorrect responses (0.3%), trials in which the voice key malfunctioned or mistriggered (4.38%), and response times which were 2.5 standard deviations beyond either participant's or item's mean (2.67%). The rest of the data were then used for further analyses.

Since the same group of participants was recruited to participate in both Experiments 5A and 5B, and the presentation order of the two sub-experiments was randomly assigned to the participants, the variable Experiment Sequence was included in the following analyses as a check to see if the effect of interest in the two sub-experiments (e.g., whether the size of rhyme-related priming would be modulated by the sonority level or the duration of the target's rhyme component) would be affected by the participants' experience in a prior experiment. For Experiment 5A, an ANOVA with Experiment Sequence (2 levels: Experiment 5A first vs. Experiment 5B first) as a between-participants factor, together with Target Type (2 levels: Targets with a more sonorant rhyme unit vs. Targets with a less sonorant rhyme unit) and Target-Distractor Relatedness (2 levels: Related vs. Unrelated Control) as the two within-participants factors was performed. A similar ANOVA was conducted for the items analysis except that the Experiment Sequence and the Target Type were treated as a within-items and a between-items factor respectively. The main effect of Target-Distractor Relatedness was found significant in both by participants and by items analyses [F_1 (1, 24) = 10.37, MSE = 821.7, p = .004; F_2 (1, 16) = 8.78, MSE = 770.9, p = .009], indicating that participants' naming latencies were significantly shorter if the target and the distractor were phonologically related than if they were unrelated. In addition, the main effect of Experiment Sequence was found significant in the by items analysis but was not in the by participants analysis [F_1 (1, 24) = 1.23, *MSE* = 10296.8, *p* = .28; F_2 (1, 16) = 15.89, *MSE* = 469.3, *p* = .001]. A closer inspection into the data from Experiment 5A revealed that the overall mean response latency of the participants who did Experiment 5A first (i.e., 614 ms) was numerically larger than that of the participants who did Experiment 5A after completing Experiment 5B (i.e., 591 ms), and hence resulted for the reliable effect of Experiment Sequence in the by items analysis. Furthermore, the main effect of Target Type was found marginal in the by participants analysis [F_1 (1, 24) = 4.16, *MSE* = 1353, *p* = .053] though was not in the by items analysis [$F_2 < 1$], and the trend of the effect indicates that participants' mean naming latencies towards the targets with a more sonorant rhyme unit (i.e., 597 ms) were generally shorter than to their less sonorant counterparts (i.e., 612 ms). Apart from the above, no significant results were observed in Experiment 5A. A summary of the ANOVA results is shown in Table 11.

Table 11. Analyses of Variance with Block Sequence, Target Type, and Relatedness as variables in Experiment 5a.

	By participants				By items			
Source	F_1	*MSE*	*p*		F_2	*MSE*	*p*	
Block Sequence (BS)	1.23	10296.8	0.28		15.9	469.3	0.001	***
Target Type (TT)	4.16	1353	0.053		0.16	19112.3	0.69	
Relatedness	10.37	821.7	0.004	***	8.78	770.9	0.009	***
BS x TT	0.003	1353	0.96		0.01	469.3	0.92	
BS x Relatedness	0.84	821.7	0.37		0.34	507.4	0.57	
TT x Relatedness	0.02	930.6	0.89		0.46	770.9	0.51	
BS x TT x Relatedness	0.1	930.6	0.76		0.6	507.4	0.45	

A similar analysis was performed on the error data but no significant results were obtained (all *ps* > .05). The mean naming latencies and error rates in various conditions (by participants) are shown in Table 12.

Table 12. Mean Naming Latencies (*M*, in ms) and Error Rates (*Errors*, in %) per condition in Experiment 5A (standard deviations in parentheses).

Rhyme Sonority	Distractor Type	*M*	*Errors*
	Experiment 5A first		
High	Rhyme related	601 (57)	0 (0)
	Control	612 (64)	0 (0)
Low	Rhyme related	613 (72)	1.5 (3.9)
	Control	629 (55)	0 (0)
	Experiment 5B first		
High	Rhyme related	572 (41)	0 (0)
	Control	596 (42)	0 (0)
Low	Rhyme related	587 (51)	0 (0)
	Control	610 (66)	1.0 (3.4)
	Overall		
High	Rhyme related	588 (52)	0 (0)
	Control	605 (55)	0 (0)
Low	Rhyme related	602 (64)	0.9 (3.1)
	Control	621 (60)	0.5 (2.2)

8.2 Experiment 5B – Duration of the Rhyme Unit

Experiment 5B was conducted to investigate whether the duration of the to-be-primed phonological unit would affect the phonological facilitation effect observed in a PWI task. To this end, a rhyme-related distrtactor condition was included in Experiment 5B and the duration of the targets' rhyme component was manipulated. Two types of target picture names which were different in the duration of their rhyme component were included in this experiment. For the target group which has a shorter rhyme component, the target syllable was always ended with a stop-coda such as the segment /p/, /t/, or /k/. But for the remaining targets which have a longer rhyme component, they were either ended with a glide such as /i/ or /u/, or a coda such as /n/ or /ng/. Note that a rhyme with a stop coda is generally shorter than a rhyme without a stop coda in Cantonese. Therefore, if phonetic factors such as duration matters in accounting for the discrepant results previously observed between onset-related and rhyme-related distractors, one would expect to see a reliable Target Type x Rhyme Relatedness interaction in this experiment. On the contrary, if the influence of phonetic factors is minimal, one would expect to see comparable rhyme-related priming across the two types of targets in Experiment 5B.

8.2.1 Materials and Design

Twenty pictures were used as stimuli in Experiment 5B (see Appendix D). They were all black-on-the-white line drawings adopted from Wu (2001) with high naming agreements (above 80%). The names of the pictures were all Cantonese monosyllables with a CVC (or CVG) structure. Among the twenty target picture names, ten of them consisted of a stop coda (i.e., having a shorter rhyme component), whereas the remaining ten picture names did not (i.e., having a relatively longer rhyme component). The distractors were all auditorily presented and were orthographically unrelated to the name of the accompanying picture in their written form. Two distractor conditions were included in this experiment, namely rhyme-related and unrelated control. Similar to Experiment 4, an identical set of distractors was used across the rhyme-related and the unrelated conditions only that the target-distractor pairings were different across the two conditions. Furthermore, the distractors used across the two target type conditions were matched in visual character frequency and in their number of homophones (Chinese Character Database: With Word-formations, 2003), *F*s < 1. The mean log frequencies and average number of homophones of the distractors across conditions are shown in Table 13.

Table 13. Mean Log Frequencies (*Log Freq*) and Number of Homophones (*NOH*)

across Distractor Conditions in Experiment 5B (standard deviations in parentheses).

Rhyme Duration	*Log Freq*	*NOH*
Short	3.07 (0.8)	2.0 (0.9)
Long	3.16 (0.7)	2.2 (1.6)

Furthermore, the mean duration of the distractors' rhyme component in the Long Rhyme Condition (Mean = 544 ms; SD = 56 ms) was found reliably longer than that in the Short Rhyme Condition (Mean = 183 ms; SD = 74 ms), $t(18) = 12.2$, $p < .001$; whereas the mean duration of the onset component was matched across the two types of distractor (Mean = 139 ms, SD = 65 ms, and Mean = 109 ms, SD = 82 ms, for the Long Rhyme and the Short Rhyme conditions, respectively), $t(18) < 1$, $p > .36$.

During the experiment, each participant received a total of 40 trials within a single block. The items were presented in a randomized order but the same picture would not repeat in consecutive trials. Apart from the above, the procedure of each trial and the apparatus used for stimuli presentation in this experiment were largely identical to those used in the previous experiments.

8.2.2 Results and Discussion

Three following types of response were excluded from data analyses: Incorrect responses (0.87%), trials in which the voice key malfunctioned or mistriggered (1.73%), and response times which were 2.5 standard deviations beyond either participant's or item's mean (3.08%). The rest of the data were then used for further analyses.

Similar to Experiment 5A, an ANOVA with Experiment Sequence (2 levels: Experiment 5B first vs. Experiment 5A first) as a between-participants factor, together with Target Type (2 levels: Targets with a shorter rhyme unit vs. Targets with a longer rhyme unit) and Target-Distractor Relatedness (2 levels: Related vs. Unrelated

Control) as the two within-participants factors was performed. Also, a similar ANOVA was conducted for the items analysis except that the Experiment Sequence and the Target Type were treated as a within-items and a between-items factor respectively. The main effect of Target-Distractor Relatedness was found significant in both by participants and by items analyses [F_1 (1, 24) = 6.07, *MSE* = 1147, p = .002; F_2 (1, 18) = 4.88, *MSE* = 1282.3, p = .004], indicating that participants' naming latencies were significantly shorter if the target and the distractor were phonologically related than if they were unrelated. In addition, the main effect of Experiment Sequence was found significant in the by items analysis but was not in the by participants analysis [$F_1 < 1$; F_2 (1, 18) = 6.27, *MSE* = 676.3, p = .002]. However, dissimilar to Experiment 5A, the data from Experiment 5B indicate that the overall mean naming latency of the participants who did Experiment 5B first (i.e., 577 ms) is shorter than those who did Experiment 5B after completing Experiment 5A (i.e., 596 ms). Hence, although both Experiments 5A and 5B showed a significant effect of Experiment Sequence in the by items analysis, an opposite pattern of results was observed, indicating that such a difference was due to the differences across participant groups (e.g., mean speech rate) but not to the effect of experience in a prior experiment. Furthermore, the main effect of Target Type was also found significant in the by participants analysis [F_1 (1, 24) = 30.8, *MSE* = 978.1, $p < .001$] though was not in the by items analysis [F_2 (1, 18) = 3.88, *MSE* = 5949.5, p = .064], and the trend indicates that participants' naming responses towards the targets with a shorter rhyme unit (i.e., 601 ms) were generally slower than to those with a longer rhyme unit (i.e., 568 ms). Apart from the above, no significant results were observed in Experiment 5B and a summary of the ANOVA results is shown in Table 14.

Table 14. Analyses of Variance with Block Sequence, Target Type, and Relatedness as

variables in Experiment 5b.

	By participants				By items			
Source	F_1	*MSE*	*p*		F_2	*MSE*	*p*	
Block Sequence (BS)	0.78	5897.2	0.39		6.27	676.3	0.022	**
Target Type (TT)	30.8	978.1	0.001	***	3.88	5949.5	0.064	
Relatedness	6.07	1147	0.021	**	4.88	1282.3	0.04	**
BS x TT	3.04	978.1	0.094		4.07	676.3	0.059	
BS x Relatedness	0.05	1147	0.83		0.01	793.2	0.93	
TT x Relatedness	0.6	664.9	0.44		0.15	1282.3	0.71	
BS x TT x Relatedness	0.001	664.9	0.99		0.02	793.2	0.89	

A similar analysis was performed on the error data but no significant results were obtained (all *ps* > .05). The mean naming latencies and error rates in various conditions (by participants) are shown in Table 15.

Table 15. Mean Naming Latencies (*M*, in ms) and Error Rates (*Errors*, in %) per condition in Experiment 5B (standard deviations in parentheses).

Rhyme Duration	Distractor Condition	*M*	*Errors*
	Experiment 5B first		
Short	Rhyme related	590 (42)	1.8 (4.1)
	Control	609 (32)	0 (0)
Long	Rhyme related	548 (46)	1.8 (4.1)
	Control	560 (66)	0.9 (3.1)
	Experiment 5A first		
Short	Rhyme related	591 (42)	1.3 (3.5)
	Control	613 (48)	0.7 (2.6)
Long	Rhyme related	571 (43)	0.7 (2.6)
	Control	585 (49)	0 (0)
	Overall		
Short	Rhyme related	590 (41)	1.6 (3.7)
	Control	611 (41)	0.4 (1.9)
Long	Rhyme related	562 (45)	1.2 (3.3)

Control	575 (57)	0.4 (1.9)

The results of Experiment 5 are clear. A significant rhyme-related priming effect was observed in both Experiments 5A and 5B. More importantly, the observed rhyme-related facilitation effect was not affected by phonetic factors such as the sonority level and the duration of the target's rhyme component. In fact, the size of priming was found numerically larger in the Low Sonority (19 ms) than that in the High Sonority (17 ms) condition, and also larger in the Short Duration (21 ms) than the Long Duration (13 ms) condition. These results suggest that the influence of phonetic factors on the phonological effects obtained in a PWI task is minimal at best. In addition, the results of Experiment 5 support the notion that the discrepant results observed across onset-related and rhyme-related distractors were not due to their phonetic differences, but were due to the reason that the rhyme unit carries a heavier weight than the onset in Cantonese. This latter conjecture is augmented by the significant rhyme-related priming observed in Experiment 5A in which the target and the distractor shared only a single segment (i.e., the vowel), replicating a similar vowel-related effect reported by Wong and Chen (2008). The contrasting results between onset-related and vowel-related distractors on CV targets will further be discussed in General Discussion.

Chapter 9

General Discussion

In this study, five PWI experiments were conducted to investigate how phonological information is retrieved in Cantonese spoken word production. In the first three experiments, the targets were Cantonese mono-syllable with a CVC (or CVG) structure, and significant facilitation effects, relative to the control, were observed when the target and the distractor shared two similar segments regardless of their syllable-internal positions (e.g., Vc-related, VcT-related, CV-related, Cc-related, or CcT-related), whereas no reliable effects were found when they shared only one single segment (e.g., V-related or c-related), or together with an identical tone (e.g., VT-related or cT-related). Furthermore, the priming effects observed in the CcT-related and CV-related conditions were found reliably larger than that in the Cc-related condition. In Experiment 4, the syllable structure (i.e., CV vs. CVC) of the target picture name was manipulated, and a significant syllable (without tone) related facilitation effect was observed, whereas no reliable effect was found if the target and the distractor shared only the same initial consonant (i.e., onset-related), irrespective of the target's syllable structure. Moreover, Experiment 5 investigated whether the contrasting results between onset-related and rhyme-related distractors in a Cantonese PWI experiment were due to the phonetic differences between onsets and rhymes. To this end, the relative sonority level and the duration of the target's rhyme component were manipulated in Experiments 5A and 5B respectively. Significant rhyme-related priming effects were observed in both sub-experiments but the facilitation effects obtained were not affected by the rhyme's sonority level or its duration, indicating that the influence of phonetic factors is minimal in the present PWI study.

Table 16. A Summary of the Results from all Cantonese PWI Studies.

		Mode of Distractor Presentation		
Target Type	Distractor Type	Visual	Auditory	Source
CVc	CVc + Tone	38ms		Wong & Chen, 2008, Exp 3
	CVc	27ms		Wu, 2001, Exp 4
		20ms		Wong & Chen, 2008, Exp 3
		23ms		Wong & Chen, 2009, Exp 1
			62ms	Wong & Chen, 2008, Exp 5
			25ms	Present study, Exp 4
	CV + Tone	15ms		Wong & Chen, 2009, Exp 2
	Vc + Tone	13ms		Wong & Chen, 2008, Exp 2
			15ms	Wong & Chen, 2008, Exp 5
		17ms		Wong & Chen, 2009, Exp 2
			17ms	Present study, Exp 2
	Cc + Tone		37ms	Present study, Exp 3
	[redacted]	1ms		Wong & Chen, 2008, Exp 1
			5ms	Wong & Chen, 2008, Exp 4
		-2ms		Wong & Chen, 2009, Exp 2
			6ms	Present study, Exp 2
			7ms	Present study, Exp 2
	Cc		21ms	Present study, Exp 3
	CV	19ms		Wong & Chen, 2009, Exp 1
			40ms	Present study, Exp 3
	Vc	3ms		Wong & Chen, 2008, Exp 2
			6ms	Wong & Chen, 2008, Exp 5
		16ms		Wong & Chen, 2009, Exp 1
			17ms	Present study, Exp 1
			17ms	Present study, Exp 5B
	[redacted]	6ms		Wong & Chen, 2008, Exp 1
			1ms	Wong & Chen, 2008, Exp 4
			7ms	Present study, Exp 4
			7ms	Present study, Exp 1
			-1ms	Present study, Exp 1
		6ms		Wong & Chen, 2008, Exp 1
		1ms		Wong & Chen, 2008, Exp 2

	[redacted] 3ms		Wong & Chen, 2008, Exp 3
		-2ms	Wong & Chen, 2008, Exp 4
CV	CV	64ms	Wong & Chen, 2008, Exp 5
		41ms	Present study, Exp 4
	V + Tone	29ms	Wong & Chen, 2008, Exp 5
	[redacted]	5ms	Present study, Exp 4
	V	22ms	Wong & Chen, 2008, Exp 5
		18ms	Present study, Exp 5A

Note: The labels given under the column Distractor Type refer to the similar phonological component shared between target and the corresponding distractor type. Capital letter C = onset consonant; V = vowel; Small letter c = coda. The numeral values shown in the table indicate the difference in mean response time between the distractor condition and the unrelated baseline condition (i.e., +ve means facilitation while –ve means inhibition). Distractor types which are shaded refer to those conditions in which no reliable effects have been reported so far.

Table 16 shows a summary of all the documented results of phonological priming from previous Cantonese PWI studies together with the results obtained in the present study. The two general types of target syllables (i.e., CVC and CV syllables) that have been studied so far are listed separately in the table. And the relative degrees of phonological similarity between target and distractor are shown in a descending order beginning from the top of the table. An overview of the results listed in Table 16 indicates several consistent and interesting patterns. First, the size of phonological priming (i.e., in terms of the mean reaction time difference between phonologically related and unrelated conditions) drops gradually along the way that the number of similar phonological components shared between target and distractor decreases. Such a pattern of results is consistent with the general findings from the past speech production research suggesting that the size of priming obtained was dependent on the degree of phonological overlap between the prime and the target (e.g., Meyer, 1990, 1991; Schiller, 1998, 2000).

Second, the size of phonological priming obtained is generally larger when the distractor was auditorily presented than when it was visually presented. For instances, a syllable (without tone) related distractor could possibly produce a facilitation effect on naming latencies up to 62 ms (given the target was a CVC syllable) but a corresponding visual distractor could only yield around 20 to 27 ms of facilitation across different studies. Even when a visual distractor was homophonous to the target (i.e., they shared the same syllable and tone), a facilitation of only 38 ms has been reported in the literature. The contrasting results obtained between auditory and visual distractors are actually in accord with the findings from Chinese visual word recognition studies which demonstrated that the orthography and phonology in Chinese are not as closely linked as that in alphabetic languages such as English (Chen & Juola, 1982), and that phonology only has a subsidiary role to play in Chinese visual word processing (Chen, Flores d'Arcais, & Cheung, 1995; Chen & Shu, 2001; Zhou & Marslen-Wilson, 2000). Therefore, it is conceivable that the size of phonological priming produced by an auditory distractor is in general larger than that by a printed distractor in Chinese.

Third, consistent with the previous PWI studies on stress languages such as Dutch which found significant facilitation effects with both word-initial related distractors and word-final related distractors (Meyer & Schriefers, 1990), significant phonological priming effects were obtained in Cantonese PWI experiments irrespective of the syllable-internal position of the shared segments (i.e., CV-related, Cc-related, or Vc-related distractors). These findings are consistent with the notions that the segments within a syllable are activated and retrieved in parallel (Dell, 1986, 1988), and that the locus of the phonological effects observed in a PWI task is situated at the segmental retrieval stage (Levelt et al., 1999).

Fourth, dissimilar to the previous studies on Dutch or English in which a significant priming effect has repeatedly been demonstrated even when the prime and target shared only a single segment (see Chapter 1), no significant phonological effect has been observed in a Cantonese PWI task when the target and the distractor shared only a single segment (e.g., onset-consonant, vowel, or coda). One exception is the reliable vowel-related priming effect on targets with a CV syllable structure. However, since such a distractor condition does not separate vowel-relatedness from rhyme-relatedness, it is not clear whether the priming observed in such condition was due to the effect of a single segment or due to the effect of rhyme as a processing unit. Further discussion will be furnished for this result.

Fifth, the priming effects induced by different types of sub-syllabic components are not the same. For instance, given the target is a CV syllable in Cantonese, reliable priming effects were found when the target and the distractor shared the same rhyme (or vowel) but not when they shared only the same onset component. Since the distractors in these two conditions have the same degree of segmental overlap with their CV targets, the discrepant results obtained across the two conditions indicate that the rhyme (or vowel) component has its own representation and has a relatively more important role to play than the onset in Cantonese spoken word planning.

Sixth, although Cantonese is a tonal language where prosodic information such as tone is crucial in lexical distinction (Cutler & Chen, 1997; Gandour, 1981), no significant priming effect has been observed in a Cantonese PWI task if the target and the distractor shared only the tone. This result is consistent with a previous Mandarin spoken word production study using the implicit priming task in which a null effect was also found when the response words within a set shared only the same tone (Chen et al., 2002). In addition, previous research on Chinese speech perception have also obtained a null effect of tone in isolated spoken word recognition (Liu & Samuel, 2007; Tsang, 2009; Ye & Connine, 1999), however, a significant effect of tone has been observed when a prior context exists (Liu & Samuel, 2007; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Ye & Connine, 1999).

Seventh, the priming effects generated by segmental and tonal information are not equivalent. For instance, reliable priming effects have been reported whenever the target and the distractor shared two similar segments whereas no significant results have been found even when the target and the distractor shared a single segment together with an identical tone (e.g., CT-related, VT-related, and cT-related conditions, given the target was a CVC syllable).

Nevertheless, the points listed above are some general patterns observed among the available results from Cantonese PWI studies. The implications of these findings, as well as their relations with other relevant studies, will be discussed in details in the following sections. In the following, I will first focus on the four unsettled issues in Cantonese spoken word production highlighted in Chapter 2, namely 1) the role of tone, 2) the nature of sub-syllabic representation, 3) the relation between tonal and segmental processes, and 4) possible language-specific properties. Each of these four issues will be discussed in the light of the evidence coming from PWI experiments. After that, a sketch of a model on Cantonese spoken word production will be furnished together with an explanation of how this model can possibly account for the available results. Lastly, some possible future research directions based on the model proposed will be discussed.

9.1 Role of Tone in Cantonese Spoken Word Planning

As reviewed in Chapter 2, one of the unresolved issues concerning the role of tone in Chinese speech production is whether tones should be considered as similar to segments (i.e., segmental view; e.g., Fromkin, 1980; Moser, 1991; Packard, 1986) or if they are different phonological units represented separately (i.e., autosegmental view; e.g., Liang & van Heuven, 2004). The discrepant results observed between Vc-related (Experiment 1) and VT- or cT-related conditions (Experiment 2), and those between Cc-related (Experiment 3) and cT-related conditions in the present study, indicate that tone is processed or represented differently from segmental units in Cantonese spoken word planning, because otherwise reliable effects should also be found in VT-related and cT-related conditions. Furthermore, the reliable facilitation effect obtained in the CV-related condition in Experiment 3 (and in Wong & Chen, 2009, but with visual distractors), together with the null effects of the CT-related condition in Wong and Chen (2008, Experiments 1 & 4), further supports the claim that the role of lexical tone is different from that of the segments. Interestingly, findings from the neuropsychological research on Cantonese and Mandarin also reach a similar conclusion. For instance, Law and Or (2001) reported a case of acquired dyslexia who showed errors in Cantonese written character naming in which the segmental content of the monosyllable was preserved but an incorrect tone was assigned to the syllable (see also Luo & Weekes, 2004, for a similar case in Mandarin). With these findings, the authors therefore argued that the tone in Cantonese is represented separately from segmental information, and the error pattern observed was either due to the impairment in tonal representation, or to the connection between tonal and segmental representations (Law & Or, 2001; Luo & Weekes, 2004). Note, however, that past research on the role of tone in Chinese speech production relied heavily on the speech error data from patient or normal participants (Chen, 1999; Fromkin, 1980; Moser, 1991; Liang & van Heuven, 2004; Packard, 1986; see also Meyer, 1992, for a critical review of this approach), the present Cantonese PWI experiments provide clear chronometric evidence in support of the autosegmental view, but not the segmental view, regarding the role of tone in spoken word planning.

In addition, the autosegmental view is in line with the speech perception model forwarded by Ye and Connine (1999), as well as the speech production model proposed by Wong and Chen (2008), in which both assume that lexical tones are represented separately from segmental representations. More importantly, a stronger facilitation effect was observed in the CcT-related condition than that in the Cc-related condition (Experiment 3), indicating that tone has a unique role to play in phonological encoding since the only difference between these two types of distractor is on their tone.

9.2 Nature of sub-syllabic representations in Cantonese

The null effects of V-related and c-related distractors observed in Experiment 1, together with the null effects of C-related distractors reported by Wong and Chen (2008), indicate that a single segment alone does not have a significant impact on the target syllables with a CVC structure. However, as shown in Table 16, significant facilitation effects on naming latencies were found in Cantonese PWI experiments when the target (i.e., CVC syllable) and the distractor shared at least two segments irrespective of the segments' syllable internal positions (i.e., CV-related, Cc-related, or Vc-related), supporting the assumption that the phonological effects were arising from a stage during which segmental contents were retrieved in parallel (i.e., position of the primed segments does not matter).

These results are inconsistent with the view that phonological units such as onsets (a.k.a. initials) and rhymes (a.k.a. finals) are the minimal contrastive units in Cantonese (e.g., Bruche-Schulz, 1997). If that is the case that the rhyme of a CVC syllable is a minimal and inseparable phonological unit, the result obtained in the

CV-related and Cc-related conditions should be similar to that observed in the C-related condition, in which a null effect would be expected. This is because both CV-related and Cc-related distractors shared only a similar onset-consonant but have a different rhyme with the target. As a result, the reliable effects observed in the CV-related and Cc-related conditions in this study pose challenges to the idea that onsets and rhymes are the only legitimate phonological units in Cantonese speech processing

On the contrary, these results are apparently consistent with the assumption that segments are the basic phonological units (Dell, 1986, 1988; Levelt et al., 1999), and that the degree of segmental overlap between target and distractor is the main determinant for a reliable priming effect to arise. Accordingly, although a single segment alone was not strong enough to facilitate the segmental retrieval process in Cantonese spoken word production to an observable extent, a facilitation effect would be observed when more than one segments were primed by the distractor. Indeed, past studies with the use of a variety of speech production paradigms have demonstrated that the degree of segmental overlap between target and prime would significantly affect the size of phonological priming (Meyer, 1990, 1991; Schiller, 1998, 2000; Sevald & Dell, 1994). As a result, most of the existing models of word production agree upon that segments are the important and basic processing units in spoken word planning (i.e., segment based; e.g., Caramazza, 1997; Dell, 1986; Levelt et al., 1999).

Nevertheless, there are two results from the present study which might not easily be accounted for by the segmental overlap hypothesis. First, the facilitation effect observed in the CV-related condition was found reliably larger than that in the Cc-related condition (Experiment 3). Since these two types of distractor posses the same degree of segmental overlap with their targets, a similar result would be expected based on the segmental overlap hypothesis. However, a significant difference was found across the two conditions. This result can be explained if one assumes that the vowel unit carries a heavier weight than the coda. However, this proposed account violates the assumption in most existing speech production models in that the weight assigned to each segmental unit is the same (Caramazza, 1997; Dell, 1986, Levelt et al., 1999).

Second, a significant vowel-related facilitation (Experiment 5A of the present study; Wong & Chen, 2008) together with a null effect of onset-related distractors (Experiment 4) were found when the target was a CV syllable. These results, again, would not be expected by the segmental overlap hypothesis account because the degree of segmental overlap between target and distractor was the same across these two conditions. In addition, since a vowel-related distractor shared only a single segment with the CV target, the reliable effect of vowel-related distractors observed in Experiment 5A seems inconsistent with the conclusion that a reliable priming effect can be observed in a Cantonese PWI experiment only when the target and the distractor shared two similar segments. In sum, it appears that a "pure" segmental overlap hypothesis might not be sufficient enough to explain the present Cantonese data. It is plausible that phonological units other than segments are also involved in the course of phonological encoding in Cantonese. Possible candidates are sub-syllabic units such as onsets and rhymes. Since the vowel-related distractors are also equivalent to rhyme-related distractors in the case when the target is a CV target, the vowel-related priming might thus be an effect of the rhyme unit but not merely the effect of a single segment.

Indeed, there is evidence showing that sub-syllabic units such as onsets and rhymes are involved in speech processing. For instances, developmental studies conducted by Treiman and her colleagues indicated that children can recognize word-initial segments better if they constitute the whole onset component of a word than if they are only part of the onset even though the number of segments to be recognized was the same (Treiman, 1985). Similarly, Treiman and Danis (1988) examined the error patterns produced by child and adult participants when they were asked to recall lists of CVC syllables. The authors found that the errors were mostly composed of a recombination of the onsets and rhymes of the syllables (Treiman & Danis, 1988). Together with some other similar findings (Treiman, 1995; Treiman & Zukowski, 1996), Treiman therefore proposed that the phonological representation for a syllable is composed of a three-layer hierarchical structure, at the top is the representation for syllable, followed by a level of representation for onset and rhyme, and then the bottom is a level for phoneme representations. A similar proposal can also be found in Dell's (1986) model. Although Dell's model, similar to other prominent models of speech production, holds that segments are the important planning units in spoken word production, yet, this model also assumes representations for sub-syllabic units such as onsets and rhymes. In addition, Wong and Chen (2008) found that the size of rhyme-related priming was not dependent on the number of segments included in the rhyme component, Wong and Chen therefore argued that the rhyme unit has its own representation in Cantonese spoken word planning.

Nevertheless, even if one accepts that there is a sub-syllabic representational level for units such as onsets and rhymes, one needs to explain why rhyme-related distractors produced a stronger effect than that of onset-related distractors in Cantonese PWI experiments, given that both onsets and rhymes are situated at the same sub-syllabic level within the hierarchy (Treiman, 1985). However, convergent evidence comes from a Mandarin speech perception study which also suggests that the rhyme unit has a relatively more prominent role to play than the onset in processing Chinese speech. In that study, Tong, Francis, and Gandour (2008) investigated the relative importance of segmental and tonal information in Mandarin speech perception by using the Garner speeded classification task (Garner, 1976). In this classification task, participants were asked to classify auditory sound stimuli on the basis of a designated dimension (e.g., pitch or segmental properties), and there were two block conditions, namely the baseline condition and the filtering condition. During the baseline block condition, the stimuli presented across trials varied only along the target (i.e., attended) dimension, whereas in the filtering block condition, the stimuli also varied along a non-target (i.e., unattended) dimension. Note that participants' responses were generally slower in the filtering block relative to the baseline block condition and the difference between the two was called Garner interference. Using a similar paradigm, Tong et al. (2008) asked their participants to classify Mandarin CV syllables on basis of the syllable's onset consonant, rhyme or tone, and found that the interference effect exerted on a tonal classification task (i.e., with varying segmental information across trials) was reliably larger than that on a segmental classification task (i.e., with varying tonal information across trials). More importantly, the authors observed that the interference effect on onset classification generated by varying the rhyme unit across trials was reliably larger than the interference effect on rhyme classification by varying the onsets. With these results, Tong et al. argued that the vowel (i.e., the rhyme) has a more prominent role to play than consonants (i.e., the onset) in Mandarin speech processing. In addition, the authors suggested that such a difference between onset and rhyme can be accounted for by at least two factors, namely informational values and acoustic features.

As argued by Tong et al (2008), the number of rhymes is much larger than that of the onsets in Mandarin. Consequently, the informational value of rhymes is generally higher than the onsets. The long term experience of Mandarin usage might thus shape the participants to attend more to the rhyme, than to the onset, of the syllables automatically, and hence resulted for the larger influence from rhyme to onset than the other way around (see also Wong & Chen, 2008, for a similar proposal in Cantonese). Alternatively, Tong et al. (2008) suggested that the discrepant results might be due to the acoustic differences between onsets and rhymes. Since rhymes are in general louder and longer than onsets, they might retain in the working memory longer and thus have a higher chance to exert its influence across trials. Consequently, the size of the Garner interference obtained was larger when the rhyme was varied across trials than when the onset was varied across trials.

However, as Tong et al (2008) did not manipulate the acoustic factors in their study, it is still unclear whether the discrepant results between onset and rhyme were due to their differences in informational values or acoustic properties. Nevertheless, the null effects of Rhyme Relatedness x Target Type (i.e., targets with a more sonorant / longer rhyme component vs. targets with a less sonorant / shorter rhyme component) interactions in Experiment 5 of the present study showed clearly that the acoustic feature hypothesis by itself is not sufficient enough to account for the discrepant results between onset-related distractors and rhyme-related distractors in Cantonese PWI studies. In other words, the present results offer evidence in support of the view that rhymes have a more prominent role to play than onsets in processing Chinese speech, and that the difference between rhymes and onsets lies in the higher level phonological values.

However, to propose that rhyme is an important processing unit in Cantonese spoken word planning does not necessarily mean that the vowel and coda components of a VC rhyme do not have their own representations. As argued above, the reliable effects observed in the CV-related and Cc-related conditions show clearly that a VC rhyme can further be decomposed into its sub-parts in Cantonese.

Taken the above discussed results together, it appears that neither the segmental overlap hypothesis (Levelt et al., 1999), nor the initial-final dichotomy hypothesis (e.g., Bruche-Schulz, 1997) alone can fully account for the accumulated findings from Cantonese PWI experiments. A plausible solution is to combine the two accounts in that both sub-syllabic components (i.e., onsets and rhymes) and segmental units are uniquely represented in the Cantonese lexicon, similar to the hierarchical model proposed by Treiman (1985, 1995). Furthermore, in order to account for the differential effects between onsets and rhymes, a heavier weight would be assigned to the rhyme unit relative to the onset unit (see also Tong et al., 2008).

9.3 Relationship between segmental and tonal processes in spoken word planning

Note that the reliable difference observed between CcT-related and Cc-related conditions in this study, together with the null effects of the Onset+Tone-related distractors previously observed by Wong and Chen (2008), poses a challenge to the models which assume that prosodic information is processed independently from segmental information at the early stage of phonological encoding (Levelt, 1993; Levelt & Wheeldon, 1994; Levelt et al., 1999). Although models such as the WEAVER++ does not make an explicit claim about how tone is processed in tonal language production, it has been proposed that the tonal information is handled by a frame-like structure similar to the metrical frame that WEAVER++ assumes (Chen et al., 2002). If this is the case that lexical tone is handled by the metrical frame which is constructed independently from segmental retrieval at the early stage of phonological encoding, then there are two possible ways that the difference observed between CcT-related and Cc-related conditions can be explained. One is that the process of metrical-frame construction[13] was primed by the CcT-related distractors but was not

[13] Assume that a metrical frame retrieval process is also involved in the present context (i.e., mono-syllabic word production).

by the Cc-related distractors because only the former shared the same tone (i.e., metrical frame in this case) with the target. Another is that the segment-to-frame association process was facilitated by the CcT-related distractors but was not by the Cc-related distractors[14], as it has been shown that the process of segment-to-frame association could be facilitated only if the prime and target shared the same word-initial segment together with an identical prosodic-structure (i.e., a similar tone in this case) (Roelofs & Meyer, 1998). Yet, it is plausible that both of the above reasons contributed to the difference observed. However, all of the above proposed accounts would predict a significant facilitation to be observed in an Onset+Tone-related condition as the distractors in such condition can, in principle, facilitate the processes of metrical-frame construction and segment-to-frame association to the same extent as the CcT-related distractors. However, null effects were obtained in the Onset+Tone-related condition by Wong and Chen (2008), which are inconsistent with the above proposed accounts.

Taken the results of the present and the previous studies together, it appears that the effect of tone can best be captured by a kind of model which assumes an interactive relationship between segmental and tonal processes (e.g., Wong & Chen, 2008; Ye & Connine, 1999), such that although the effect of tone alone is weak and difficult to be observed[15], a significant effect of tone can possibly be found upon a sufficient degree of segmental overlap between target and distractor (e.g., CcT-related vs. Cc-related conditions in Experiment 3, so as the difference between Syllable+Tone-related and Syllable-related conditions in Wong & Chen, 2008). Furthermore, as mentioned in Chapter 2, the Syllable+Tone-related facilitation effect observed in Wong and Chen (2008) might be originated at both lexical and sub-lexical

---

[14] Assume that the present paradigm is also sensitive to the segment-to-frame association process.
[15] The reason for this will be discussed in later sections.

levels (or even the mental syllabary assumed by WEAVER++) because the Syllable+Tone-related distractors were homophonous to the targets. Therefore, it was unclear whether the difference between Syllable+Tone-related and Syllable (without tone) related conditions was due exclusively to the unique effect of tone. Nevertheless, the reliable difference between CcT-related and Cc-related conditions observed in the present study showed clearly that tone has a unique role to play and that interactivity exists between tonal and segmental processes. This is because neither CcT-related nor Cc-related distractors shared the same pronunciation with the target, hence the priming effect observed in these two conditions should only be arising from the early stage of phonological encoding during which sub-lexical units are retrieved.

9.4 Possible language-specific properties

Consistent with the findings from Wong and Chen (2009), significant facilitation effects on naming latencies were obtained in the present study when the target (i.e., monosyllables with a CVC structure) and the distractor shared at least two similar segments, but no reliable effect was found when they shared only an individual segment (see also Table 16). Note that although past studies have already shown that an individual segment such as the onset-consonant does not have a significant impact on Chinese spoken word planning (e.g., Chen et al., 2002; Wong & Chen, 2008), no prior studies have systematically examined the effects of other segments such as vowel and coda. Notwithstanding this, the null effects obtained in the V-related and c-related conditions in Experiment 1 showed clearly that a single segment does not have a salient impact on Cantonese spoken word planning.

The null effect of an individual segment on Chinese spoken word planning observed in this and the previous studies stand in a marked contrast to the findings from languages such as Dutch and English in which significant priming effects on naming latencies were frequently observed even when the prime and target shared only a single segment (see reviews in Chapter 1).

In addition, as discussed in Section 9.2 above, the accumulated evidence from Cantonese PWI experiments suggests that sub-syllabic units such as rhyme have an important role to play in phonological encoding. However, evidence from the speech production studies on Dutch and English indicates that segments are the prominent phonological units. For instance, Schiller (1998, 2000) investigated the effect of visually presented masked primes on word or picture naming latencies. The targets included bi-syllabic words with a CV (e.g., pilot) or a CVC (e.g., picnic) word initial syllable, and two types of prime were included, one shared the same CV components with the target and the other the same CVC components. Schiller found that the size of priming was dependent on the degree of segmental overlap between prime and target, but was not affected by whether the prime shared the same syllable with the target or not. Importantly, these results also suggest that sub-syllabic units such as rhyme do not have a prominent role to play. This is because the syllable-match condition (i.e., a CV-prime with a CV-target, or a CVC-prime with a CVC-target) in Schiller's (1998, 2000) studies can also be considered as a rhyme-match condition. Furthermore, Schiller (2004) investigated the role of onset in word naming using a similar masked priming paradigm and found significant priming when the prime and target shared a similar word initial segment. Importantly, the size of this first-segment priming was not affected by whether the onset of the target was a singleton or a consonant cluster (Schiller, 2004, Experiment 3), indicating that the priming effects were not contingent on sub-syllabic units such as onset. In sum, the findings from languages such as Dutch and English showed clearly that segments, but not sub-syllabic units such as onsets or rhymes, are the prominent units in phonological encoding.

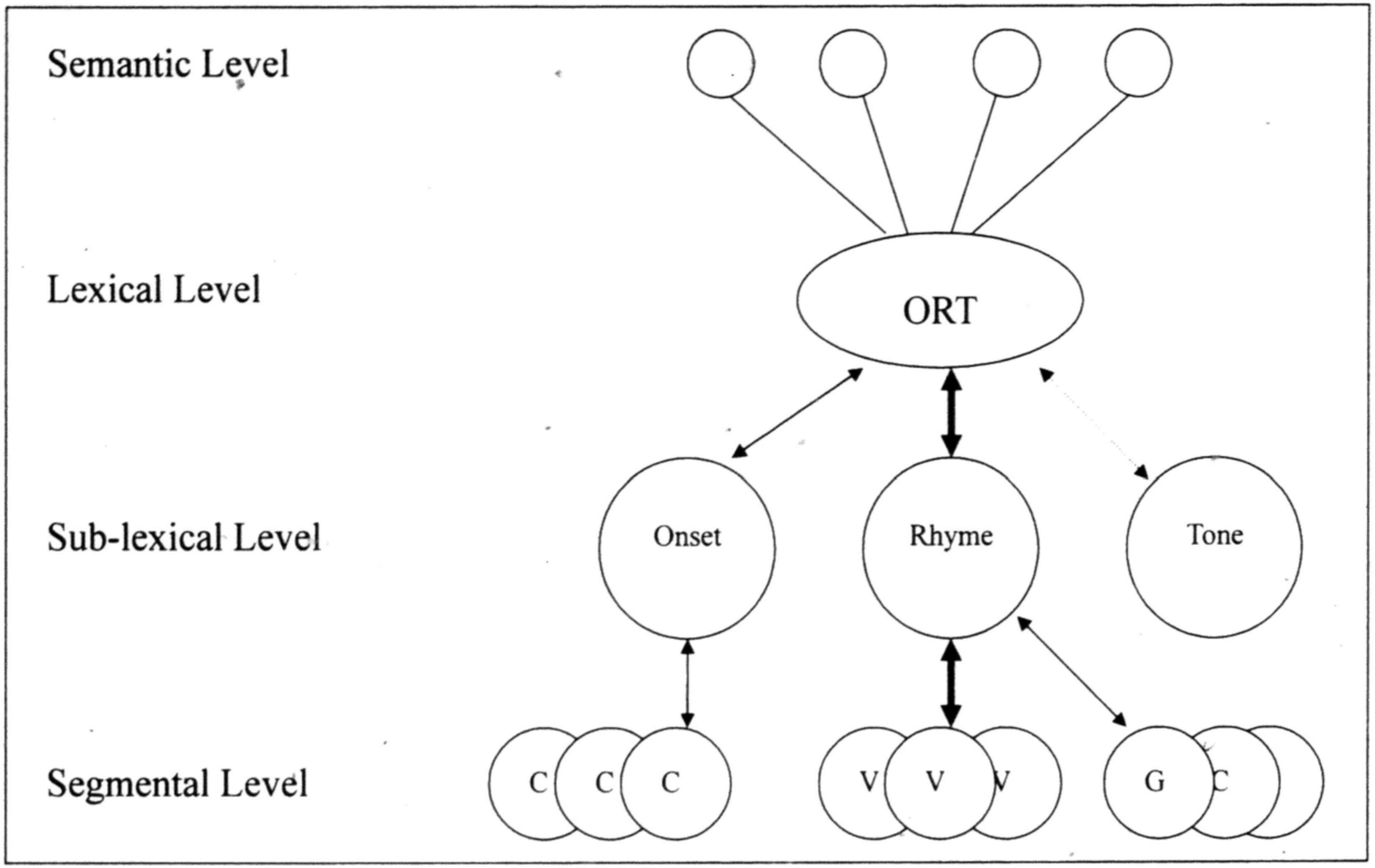
The cross-language discrepancies discussed above might be due in part to the influence of spoken and orthographic experiences. Dissimilar to many non-tonal languages in which consonant-clusters are prevalent, there are no consonant clusters in either Cantonese or Mandarin. In fact, the syllables in Cantonese and Mandarin are relatively simple, and are mostly composed of a CV or CVC structure. In addition, unlike many non-tonal languages using alphabetic scripts, Chinese adopts a logographic writing system in which each orthographic unit maps directly onto a syllable, but not phoneme (Chen, 2006). Consequently, it is plausible that speakers of languages with different phonological complexity and orthographic transparency would show various degrees of sensitivity towards different phonological units.

Indeed, there is evidence showing that both spoken and written language experiences influence the development of phonological awareness (e.g., Bertelson, de Gelder, Tfouni, & Morais, 1989; Caravolas & Bruck, 1993; Cheung, Chen, Lai, Wong, & Hills, 2001). For instance, by comparing the level of phonological awareness among children (including both pre-reading and reading children) coming from Hong Kong (China), Guangzhou (China), and Dunedin (New Zealand), Cheung et al. (2001) found that pre-readers from Dunedin (i.e., children who spoke English, an alphabetic language with a relatively complex phonological structure than Cantonese) outperformed aged-matched pre-readers from either Hong Kong or Guangzhou (i.e., children who acquired Cantonese as their first spoken language), who showed similar performance, on word-matching tasks requiring sub-syllabic analyses. In addition, the authors observed that the reading children from Guangzhou (i.e., children who learnt pinyin, an alphabetic system, along with their learning to read logographic Chinese) performed better than those from Hong Kong (i.e., children who read logographic Chinese without learning the pinyin system) on tasks requiring phonemic analyses. With these results, Cheung et al (2001) argued that the development of phonological awareness is shaped by both spoken and orthographic experiences. As a result, it is plausible that native Chinese speakers are less sensitive to the phonemic level of processing relative to the speakers of alphabetic languages (Cheung & Chen, 2004), and hence reliable effects of an individual segment are much more difficult to be shown among the Chinese speakers relative to the speakers of Dutch or English. Importantly, the present results indicate that such a cross-language difference can also be observed in the context of spoken word production.

9.5 A sketch of a model on Cantonese spoken word production

To account for the available results from Cantonese PWI studies, a tentative model is proposed here (see Figure 2). This model is largely similar to the one proposed by Wong and Chen (2008) except that an additional segmental level is incorporated. The current model posits a four-layer architecture in which activation flows in the following manner during spoken word production. Similar to most existing models of speech production, the first stage is a semantic activation stage during which the semantic representation of the to-be-articulated message (e.g., a picture name in the present context) is retrieved. Then activation flows from the semantic level to the lexical level, a place in where abstract word-form representations are stored. The phonological encoding process starts once activation reached the lexical level and the flow of activation continues to the subsequent sub-lexical level in which sub-syllabic units such as onsets, rhymes, and tones are stored, and finally to a segmental level in which representations for individual segments such as consonants and vowels are specified. Furthermore, similar to Dell's (1986) interactive activation model, in-between levels there are bidirectional connections such that feedbacks are allowed in the current model. Note, however, that one distinct feature entails with the current model is that the strength of connection between lexical and sub-lexical levels is dependent on the type of sub-lexical representations (e.g., onsets, rhymes, and tones) in that the strength of connection between a lexical node and its corresponding rhyme representation is the strongest, followed by the onset representation and then the tone representation (see also Wong & Chen, 2008). As discussed in Wong and Chen (2008), the differential strength of the three types of lexical-to-sub-lexical connections (so as the sub-lexical-to-lexical connections, due to their bi-directional nature) is due to the reason that the three types of sub-lexical components have differential degrees of informational value, and the strength of the different connections is shaped by the long term experience towards Cantonese speech in a way that can optimize speech processing.

Figure 2. A model of phonological encoding in Cantonese spoken word production.



Note: C stands for consonant; V stands for vowel; G stands for glide. The relative strength of each bi-directional connection is indicated by the thickness of the arrow.

The current model can account for the phonological effects obtained in a PWI experiment by either one or both of the following two mechanisms. First, if the distractor shared similar sub-lexical or phonemic components with the target, the target's sub-lexical or phonemic representations would be primed and thus leading to a facilitation of the phonological retrieval process. Second, the source of phonological priming can also come from the feedback activations sending from the sub-lexical to the lexical level due to the bi-directional nature of the connections. If the distractor shared a similar sub-lexical representation (e.g., onset or rhyme) with the target, not only the activation level of the shared sub-lexical representation would be increased, the degree of feedback activation sending from that sub-lexical representation to the target lexical node would also be increased, and consequently, the speed of lexical node selection would be accelerated.

With the above discussed features, the current model can readily account for the available results from Cantonese PWI studies. First, a reliable rhyme (or vowel) related priming was found but a null effect of onset related distractors was observed because the strength of connection between a lexical node and its rhyme representation is stronger than that between the lexical node and its onset representation. As a result, the feedback activation sending from the shared rhyme unit to the target lexical node was stronger than that from a shared onset unit, and hence the rhyme related distractors can more likely facilitate the lexical node selection process to a detectable extent than the onset related distractors. Second, since the vowels and codas of the VC rhymes are also specified in the current model, a distractor which shared with a CVC target a similar vowel or coda can also partially prime the rhyme unit through the feedback from phonemic to sub-lexical levels. As a result, although an onset related distractor was not strong enough to produce a reliable facilitation effect, significant priming effects were observed in both CV related and Cc related conditions. This is because the distractors in the latter two conditions can prime not only the target's onset unit, but can also (partially) prime the target's rhyme component to a certain extent, and the joint effect of these two sources of priming thus lead to the observed phonological facilitation effect. Third, as discussed above, due to the bi-directional nature of the connections between lexical nodes and tonal nodes, as well as those between lexical nodes and sub-syllabic representations such as onsets and rhymes, tonal and segmental processes are operated in an interactive manner based on the current model. Accordingly, although the effect of tone alone is weak and difficult to be observed, a reliable effect of tone can possibly be obtained upon a certain degree of segmental overlap between target and distractor in a PWI task. Indeed, the difference observed between CcT-related and Cc-related conditions, as well as the difference observed between Syllable+Tone related and Syllable related conditions, appear to suggest that tonal and segmental processes are operated in an interactive manner, which can also be explained by the current model on Cantonese spoken word production.

9.6 Other related issues

However, as noted by Starreveld (2000), two processes are involved in a typical PWI experiment. One is word production (i.e., picture name retrieval) and the other is word perception (i.e., distractor word perception). Before making any strong claim on the mechanism of spoken word production based on the results from PWI studies, one needs to consider how much the results would possibly be affected by the mechanism involved in spoken word (e.g., distractor) perception. Otherwise, the conclusion made based on the results from a PWI task might actually reflect the processes of word perception. For instance, it is plausible that there are two phonological lexicons, one for speech production and one for production. And that the reason for the phonological effects observed in Cantonese PWI studies was due to the phonological activations taken place in the lexicon for word perception. Specifically, in a typical PWI task with auditory distractors, when a phonological distractor is presented to the participant, the lexical node for that distractor would be activated within the lexicon for word perception, it is plausible that the activation would spread to other phonologically related lexical nodes in the perception network, including the one corresponding to the target. Furthermore, given that there are direct connections between the corresponding lexical nodes in the two phonological lexicons, it is possible that the phonological facilitation effect obtained was due to the direct influence from the perception lexicon to the target lexical node in production. According to this scenario, the present results might have nothing to do with the priming at the sub-lexical levels within the lexicon for word production.

To verify whether the results obtained from Cantonese PWI studies were due exclusively to the effect of speech perception, one can look into the findings from Cantonese spoken word perception research. For instance, Tsang (2009) investigated the time course of phonological activation in Cantonese spoken word perception using the visual world paradigm. In two experiments which are of most relevant to the present discussion, Tsang (2009) found that the activation level (i.e., as reflected by eye fixation proportions) of a syllable related distractor picture was very robust, and the onset of this syllable effect occurred much earlier than that produced by a rhyme related distractor picture (by some 300 ms). Accordingly, one would expect that in a PWI task with both syllable related and rhyme related distractors, the facilitation effect produced by the syllable related distractors should arise in an SOA condition which is earlier than that by the rhyme related distractors. However, no significant interaction effects involving SOA were observed in Wong and Chen (2008) in which both syllable related and rhyme related distractor conditions, as well as three SOA conditions (i.e., -150 ms, 0 ms, and 150 ms) were included. Furthermore, a CV related distractor picture condition was also included in Tsang (2009), however, no significant effect was found with the CV related distractor picture. On the contrary, significant priming effects were found with CV related distractors in Cantonese PWI experiments regardless of the mode of distractor presentation (e.g., Experiment 3 of the present study; Wong & Chen, 2009). Taken the results from Tsang (2009) and those from Cantonese PWI experiments together, it appears that the influence from speech perception on the results of the present PWI study is minimal at best. Coupled with the results from Experiment 5 of the present study suggesting that the phonological effects obtained were independent from the acoustic features of the distractors, the findings from the present Cantonese PWI study would most likely be reflecting the mechanisms involved in phonological planning.

9.7 Future directions

Nevertheless, the model proposed here is only a sketch of a model for Cantonese spoken word production, much work needs to be done in the future to verify it and to unveil the issues which are still unresolved. For instance, the null effects of single-segment related distractors observed in Cantonese PWI experiments suggest that Cantonese speakers are less sensitive to the manipulation at phonemic processing than the speakers of alphabetic languages, and that individual segments do not have a salient role to play in Cantonese spoken word planning. An interesting but yet unaddressed issue is whether it can be changed through practice. As demonstrated by Cheung (1999), the phonological awareness among a group of Cantonese speakers was improved after a short period of training on phonemic analysis, indicating that although the development of phonological awareness is affected to a large extent by the nature of one's language background, through practice one's level of phonological awareness can possibly be improved. Hence, it would be interesting to see if a reliable phonological facilitation would be seen among a group of Cantonese speakers who have received training in phonemic analysis.

Second, although significant priming effects were found in all distractor conditions which shared two similar segments with the targets (i.e., CV related, Vc related, and Cc related), the lexical status of the primed segments might be differed across different distractor conditions. For instance, CV and Vc overlaps can by themselves be lexical units in many cases but not for the Cc components. Further studies are needed to investigate whether the lexical status of the primed segments would have a significant influence on the phonological effects observed in a PWI experiment.

Third, although the results of Experiment 5 seem to suggest that phonetic factors such as the sonority level and the duration of the primed segment(s) do not have a significant impact on the size of phonological facilitation observed in a PWI task, it is plausible that other acoustic cues may have a role to play in determining the size of priming. For instance, under the context of a PWI paradigm, the phonological similarity between target and distractor can possibly be described in terms of the degree of overlap in their formant structures. However, such a possibility has not yet been examined. To verify this possibility, further studies can be conducted to manipulate the degree of formant overlap between target and distractor, instead of distinguishing a phonologically related from an unrelated condition simply based on segmental overlap (i.e., whether the target and distractor shared similar segments or not). If the formant overlap between target and distractor does have a unique role to play in a PWI task, a reliable difference should be observed between two distractor conditions with different degrees of formant overlap with the targets.

Furthermore, if acoustic information such as the formant pattern does have an effect on the size of phonological priming, it will bring forth a significant insight on how to interpret many of the past PWI results. This is because most of the previous PWI studies adopted an unrelated condition as the baseline in which the distractors did not share any similar segments with the targets, and most of the cross-experiment comparisons made were based largely on the assumption that the baseline (i.e., unrelated condition) across experiments was the same. However, if this is the case that the relatedness between target and distractor should better be described in a continuous manner (e.g., degree of formant similarity), cautions must be taken in comparing the results from different PWI experiments because the degree of "unrelatedness" of the control condition might vary across experiments.

Fourth, past research on speech perception found that the effect of tone alone was difficult to observe in isolated word perception but reliable effects of tone have been reported when a prior context was present, indicating that the effect of tone was sensitive to the availability of context (e.g., Liu & Samuel, 2007; Schirmer et al., 2005). With these findings, it would be interesting to study if a reliable effect of tone alone can be found in sentence or short phase production. In addition, as discussed before, both fundamental frequency level and contour are important features for distinguishing lexical tones in Cantonese (see Figure 1), whether it is the features that were retrieved in spoken word planning, or that the tonal representation as a whole was retrieved are issues still remain open.

Fifth, attempts have been made recently to conduct PWI experiments together with the use of neuro-imaging techniques such as ERP (e.g., Hirschfeld, Jansma, Bolte, & Zwitserlood, 2008) and fMRI (e.g., de Zubicaray & McMahon, 2009), and some fruitful results have been reported. Hence, an interesting follow up study would be to conduct a similar set of Cantonese PWI experiments with the use of such neuro-imaging methods in order to unveil the time course and the neural correlates involved in the processes of spoken word planning.

Sixth, apart from the process of phonological encoding, earlier stages in Cantonese speech production, such as semantic-syntactic processes, are still largely unexplored. It is an open issue of whether syntactic information is necessary in Cantonese spoken word production, and if it is, whether the syntactic mediation model can best explain the case of Cantonese. In addition, it appears that almost all of the existing models of speech production assume that the entire production process starts from the conceptual level (i.e., semantically driven), it remains to be seen how the existing models can account for the possible cases in which the word production process is not initiated from the conceptual level (e.g., mindless speech).

Lastly, in comparison with the speech production research conducted with non-tonal languages such as Dutch or English, research on tonal languages such as Cantonese has just begun. And yet, the model proposed in this paper rests on a representational level of analysis (Marr, 1982), what is lacking is the relative weights for different parameters assumed in the model, the computational algorithms involved, and how such model can be implemented into the neural system. Obviously, more empirical evidence is needed concerning the time course of phonological activation and the neural correlates involved in Cantonese spoken word planning. Only after enough evidence is available, a more concrete model, or a kind of computational model, such as WEAVER++ and the model proposed by Dell (1986), can then be furnished to account for the processes of speech production in Cantonese.

9.8 Conclusion

To conclude, the present findings provide clear evidence that individual segments do not have a prominent role to play in Chinese spoken word preparation, contrary to the robust effects of individual segments observed in the studies on non-tonal languages such as Dutch and English. In addition, the present results indicate that tone is processed or represented differently from segments, and that the effects of tone obtained in this and in the previous work can most readily be explained by a model which assumes an interactive relation between segmental and tonal processes in spoken word planning. Furthermore, the accumulated evidence from Cantonese PWI research suggests that sub-syllabic units such as onsets and rhymes have their own representations, and that the rhyme unit has a relatively more prominent role to play than onset and tone.

Together with the findings from previous Chinese word production research, the present results pose important constraints to the existing models of speech production with respect to three particular areas, namely 1) the salience of individual segments, 2) the nature of sub-syllabic representations, and 3) the relationship between segmental and prosodic processes. The discrepant results observed across languages with distinctive properties are in line with the notion that the cognitive processes involved in speech production are affected by the properties of the language being used. Consequently, the present study has demonstrated the need and significance of cross-language research to the development of a universal theory of language production.

References

Abdel Rahman, R., & Sommer, W. (2003). Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing. *Cognitive Brain Research, 16*, 372-382.

Bertelson, P., de Gelder, B., Tfouni, L. V., & Morais, J. (1989). Metaphonological abilities of adult illiterates: New evidence of heterogeneity. *European Journal of Cognitive Psychology, 1*, 239-250.

Bi, Y., Xu, Y., Caramazza, A. (2009). Orthographic and phonological effects in the picture-word interference paradigm: Evidence from a logographic language. *Applied Psycholinguistics, 30*, 637-658.

Brown, R., & McNeill, D. (1966). The "tip-of-the-tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325-337.

Bruche-Schulz, G. (1997). 'Fuzzy' Chinese: The status of Cantonese in Hong Kong. *Journal of Pragmatics, 27*, 295-314.

Butterworth, B. (1989). Lexical access in speech production. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108-135). London: MIT Press.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177-208.

Caramazza, A. & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence for the "tip-of-the-tongue" phenomenon. *Cognition, 64*, 309-343.

Caramazza, A., Papagno, C., & Ruml, W. (2000). The selective impairment of phonological processing in speech production. *Brain and Language, 75*, 428-450.

Caravolas, M., & Bruck, M. (1993). The effect of oral and written language input on children's phonological awareness: A cross-linguistic study. *Journal of Experimental Child Psychology, 55*, 1-30.

Cuetos, F., Aguado, G., & Caramazza, A. (2000). Dissociation of semantic and phonological errors in naming. *Brain and Language, 75*, 451-460.

Chen, H.-C. (1992). Reading comprehension in Chinese: Implications from character reading times. In H.-C. Chen and O. J. L. Tzeng. (Eds.). *Language Processing in Chinese,* (pp. 175-205). Amsterdam, Netherlands: North Holland.

Chen, H.-C. (2001). Speech processing in Chinese: An introduction. *Journal of Psychology in Chinese Societies, 2,* 155-158.

Chen, H.-C. (2006). Language processing: Implications from the study of Chinese. In Q.C. Jing, M.R. Rosenweig, G. d'Ydewalle, H.C. Zhang, H.-C. Chen, & K. Zhang (Eds.), *Progress in psychological science around the world: Neural,* *cognitive, and developmental issues* (pp. 367-381). London: Psychology Press.

Chen, H.-C., Flores d'Arcais, G.B., & Cheung, S.L. (1995). Orthographic and phonological activation in recognizing Chinese characters. *Psychological Research, 58*, 144-153.

Chen, H.-C. & Juola, J.F. (1982). Dimensions of lexical coding in Chinese and English. *Memory and Cognition, 10*, 216-224.

Chen, H.-C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin & Review, 8*, 511-518.

Chen, H.-C., & Yip, M. C.-W. (2001). Processing syllabic and sub-syllabic information in Cantonese. *Journal of Psychology in Chinese Societies, 2(2)*, 199-210.

Chen, H.-C., & Zhou, X. (1999). Processing East Asian languages: An introduction. *Language and Cognitive Processes, 14*, 425-428.

Chen, J. Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics, 20*, 289-301.

Chen, J. Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia, 43*, 15-26.

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*, 751-781.

Cheung, H. (1999). Improving phonological awareness and word reading in a later learned alphabetic script. *Cognition, 70*, 1-26.

Cheung, H., & Chen, H.-C. (2004). Early orthographic experience modifies both phonological awareness and on-line speech processing. *Language and Cognitive Processes, 19*, 1-28.

Cheung, H., Chen, H.-C., Lai, C.Y., Wong, O.C., & Hills, M. (2001). The development of phonological awareness: Effects of spoken language experience and orthography. *Cognition, 81*, 227-241.

Chinese Character Database: With Word-formations. (2003). Retrieved on August 30, 2008, from http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/

Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99*, 205-235.

Cutler, A. (1980). Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance. Slips of the tongue, ear, pen, and hand* (pp. 67-80). New York: Academic Press.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*, 201-220.

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

Cutler, A. & Clifton, C. (1984). The use of prosodic information in word recognition. *Attention and Performance X: Control of Language Processes,* 183-196.

Cutting, J. C., & Ferreira, V. S. (1999). Semantic and phonological information flow in the production lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 318-344.

Damian, M. F., & Bowers, J. S. (2003). Locus of semantic interference in picture-word interference tasks. *Psychonomic Bulletin and Review, 10*, 111-117.

Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes, 24*, 685-712.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 345-361.

de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/phonological facilitation of naming responses in the picture-word task: An event-related fMRI study using overt vocal responding. *Neuroimage, 16*, 1084-1093.

de Zubicaray, G. I., & McMahon, K. L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective, & Behavioral Neuroscience, 9*, 260-269.

DeFrancis, J. (1984). *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.

Dell, G. S. (1988). The retrieval of phonological forms in production: tests of prediction from a connectionist model. *Journal of Memory and Language, 27*, 124-142.

Dell, G. S. (1990). Effects of frequency and vocabulary types on phonological speech errors. *Language and Cognitive Processes, 5*, 313-349.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*, 801-838.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Fromkin, V. A. (1980). Introduction. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance. Slips of the tongue, ear, pen, and hand.* New York: Academic Press.

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics, 9*, 20-36.

Gandour, J., Petty, S. H., Dardarananda, R. (1988). Perception and production of tone in aphasia. *Brain and Language, 35*, 201-240.

Garner, W. R. (1976). Integration of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8*, 98-123.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation, v. 9* (pp. 133-177), Academic Press, New York.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey, v. III. Biological and psychological aspects of language*. Cambridge, MA: Harvard University Press.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes, 21*, 817-855.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis, 2*, 23-68.

Hirschfeld, G., Jansma, B., Bolte, J., & Zwitserlood, P. (2008). Interference and facilitation in overt speech production investigated with event-related potentials. *Neuroreport, 19*, 1227-1230.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101-144.

Janssen, N., Schirm, W., Mahon, B. Z., & Caramazza, A. (2008). Semantic interference in a delayed naming task: Evidence for the response exclusion hypothesis. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34*, 249-256.

Jerger, S., Martin, R. C., & Damian, M. F. (2002). Semantic and phonological influences on picture naming by children and teenagers. *Journal of Memory and Language, 47*, 229-249.

Jescheniak, J. D., Meyer, A. S., & Levelt, W. J. M. (2003). Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 432-438.

Jescheniak, J. D., & Schriefers, H. (2001). Priming effects from phonologically related distractors in picture-word interference. *The Quarterly Journal of Experimental Psychology, 54A*, 371-382.

Kao, H. S. R., & Hoosain, R. (1986). *Linguistics, psychology, and the Chinese language*. Hong Kong: Centre of Asian Studies, University of Hong Kong.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14*, 185-209.

Ladefoged, P. (2001). *A course in phonetics*. Fort Worth: Harcourt College Publishers.

Law, S.-P., & Or., B. (2001). A case study of acquired dyslexia and dysgraphia in Cantonese: Evidence for non-semantic pathways for reading and writing in Chinese. *Cognitive Neuropsychology, 18*, 729-748.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*, 1-22.

Levelt, W. J. M. (1993). Timing in speech production: With special reference to word form encoding. *Annals of the New York Academy of Sciences, 682*, 283-295.

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Science, 3(6)*, 223-232.

Levelt, W. J. M. (2001). Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 13464-13471.

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). A MEG study of picture naming. *Journal of Cognitive Neuroscience, 10*, 553-567.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. Psychological Review, 98, 122-142.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*, 239-269.

Li, C.N., & Thompson, S.A. (1981). *Mandarin Chinese. A functional reference grammar*. Berkeley, CA: University of California Press.

Liang, J., & Van Heuven, V. J. (2004). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language, 91*, 282-293.

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

Luo, Q., & Weekes, B. (2004). Tonal dyslexia in Chinese. *Brain and Language, 91*, 102-103.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*, 524-545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*, 69-89.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*, 181-211.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 1146-1160.

Moser, D. (1991). *Slips of the tongue and pen in Chinese*. (Sino-Platonic Papers, No. 22). Department of Oriental Studies, University of Pennsylvania, Philadelphia, PA.

Nickels, L., & Howard, D. (1999). Effects of lexical stress on aphasic word production. *Clinical Linguistics and Phonetics, 13*, 269-294.

Nooteboom, S. (1969). The tongue slips into patterns. In: A. G. Sciarone, A. J. Von Essen, & A. A. Van Raad (Eds.). *Nomen: Leyden studies in linguistics and phonetics* (pp. 114-132). The Hague: Kluwer.

Packard, J. L. (1986). Tone production deficits in nonfluent aphasic Chinese speech. *Brain and Language, 29*, 212-223.

Peterson, R. R., & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory & Cognition, 24*, 539-557.

Pike, K. L. (1948). *Tone languages*. Ann Arbor, MI: University of Michigan Press.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191*, 62-88.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*, 460-499.

Rapp, B., & Goldrick, M. (2004). Feedback by any other name is still interactivity: A reply to Roelofs (2004). *Psychological Review, 111*, 573-578.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107-142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249-284.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes, 14*, 173-200.

Roelofs, A. (2000). WEAVER++ and other computational models of lemma retrieval and word-form encoding. In L. R. Wheeldon (Ed.), *Aspects of language production* (pp. 71-114). Philadelphia, PA: Taylor & Francis.

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 922-939.

Roelofs, A., Meyer, A. S., & Levelt, W. J. M. (1996). Interaction between semantic and orthographic factors in conceptually driven naming: Comment on Starreveld and La Heij (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 246-251.

Rosinski, R. R., Michnick-Golinkoff, R., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development, 46*, 247-253.

Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language, 39*, 484-507.

Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 512-528.

Schiller, N. O. (2004). The onset effect in word naming. *Journal of Memory and Language, 50*, 477-490.

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research, 1112*, 201-212.

Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition, 106*, 952-962.

Schiller, N. O., Fikkert, P., & Levelt, C. C. (2004). Stress priming in picture naming: An SOA study. *Brain and Language, 90*, 231-240.

Schiller, N. O., Jansma, B. M., Peters, J., & Levelt, W. J. M. (2006). Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes, 21*, 112-140.

Schiller, N. O., Meyer, A. S., Baayen, H., & Levelt, W. J. M. (1996). A comparison of lexeme and speech syllables in Dutch. *Journal of Quantitative Linguistics, 3*, 8-28.

Schiller, N. O., Meyer, A. S., & Levelt, W. J. M. (1997). The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants. *Language and Speech, 40*, 103-140.

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

Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology, 37*, 473-484.

Schmitt, B. M., Schiltz, K., Zaake, W., Kutas, M., & Münte, T. F. (2001). An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. *Journal of Cognitive Neuroscience, 13*, 510-522.

Schriefers, H., & Meyer, A. S. (1990). Experimental note: Cross-modal, visual-auditory picture-word interference. *Bulletin of the Psychonomic Society, 28*, 418-420.

Schriefers, H., Meyer, A. S. & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86-102.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*, 91-127.

Shattuck-Hufnagel, S. & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*, 41-55.

Starreveld, P. A. (2000). On the interpretation of onsets of auditory context effects in word production. *Journal of Memory and Language, 42*, 497-525.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 686-698.

Starreveld, P. A., & La Heij, W. (1996a). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 869-918.

Starreveld, P. A., & La Heij, W. (1996b). The locus of orthographic-phonological facilitation: Reply to Roelofs, Meyer, and Levelt (1996). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 252-255

Stemberger, J. P. (1985). An interactive activation model of language production. In A. Ellis (Ed.), *Progress in the psychology of language* (pp. 143-186). London: Erlbaum.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: Naming in the picture-word interference task. *Psychology and Aging, 17*, 662-676.

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes, 23*, 689-708.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 39*, 161-181.

Treiman, R. (1995). Errors in short-term memory for speech: A developmental study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1197-1208.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 145-152.

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to syllables, onsets, rimes, and phonemes. *Journal of Experimental Child Psychology, 61*, 193-215.

Tsang, Y.-K. (2009). *Speech perception in Chinese: How are the different levels of ambiguity resolved?* Ph.D. Dissertation, Chinese University of Hong Kong, Hong Kong.

Van Maanen, L., Van Rijn, H., Borst, J. P. (2009). Stroop and picture-word interference are two sides of the same coin. *Psychonomic Bulletin & Review, 16*, 987-999.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 787-806.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science, 280*, 572-574.

Vigliocco, G., Antonioni, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science, 8*, 314-317.

Wong, A. W.-K., & Chen, H.-C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory & Cognition, 34*, 1172-1190.

Wong, A. W.-K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review, 16*, 888-892.

Wu, S.Y. (2001). Modeling speech production: Implications from a Chinese study. Master's Thesis, Chinese University of Hong Kong.

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

Yip, C.-W. (2000). *Probabilistic phonotactics and the segmentation of Cantonese continuous speech.* Ph.D. Dissertation, Chinese University of Hong Kong, Hong Kong.

Zhou, X. & Marslen-Wilson, W. (1999). Phonology, orthography, and semantic activation in reading Chinese. *Journal of Memory and Language, 41*, 579-606.

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1245-1265.

Appendix A: Names of the Pictures and Word Distractors used in Experiments 1 and 2

Picture name		VC	V	C	VCT	VT	CT	Control
橙 (caang2)	*orange*	冷 (laang5)	踏 (daap6)	幫 (bong1)	省 (saang2)	減 (gaam2)	榜 (bong2)	兼 (gim1)
箭 (zin3)	*arrow*	淺 (cin2)	點 (dim2)	班 (baan1)	建 (gin3)	店 (dim3)	盼 (paan3)	及 (kap6)
象 (zoeng6)	*elephant*	良 (loeng4)	桌 (coek3)	肯 (hang2)	尚 (soeng6)	略 (loek6)	幸 (hang6)	答 (daap3)
鷹 (jing1)	*eagle*	成 (sing4)	杰 (git6)	病 (beng6)	精 (zing1)	尖 (zim1)	東 (dung1)	發 (fat3)
田 (tin4)	*farmland*	建 (gin3)	協 (hip3)	趕 (gon2)	然 (jin4)	潛 (cim4)	門 (mun4)	律 (leot6)
星 (sing1)	*star*	境 (ging2)	店 (dim3)	孔 (hung2)	京 (ging1)	必 (bit1)	風 (fung1)	各 (gok3)
心 (sam1)	*heart*	禁 (gam3)	能 (nang4)	險 (him2)	音 (jam1)	根 (gan1)	添 (tim1)	喝 (hot3)
藥 (joek6)	*medicine*	卻 (koek3)	槍 (coeng1)	谷 (guk1)	略 (loek6)	量 (loeng6)	目 (muk6)	明 (ming4)
琴 (kam4)	*piano*	甚 (sam6)	品 (ban2)	尖 (zim1)	沉 (cam4)	宏 (wang4)	廉 (lim4)	清 (cing1)
鐘 (zung1)	*clock*	凍 (dung3)	沒 (mut6)	放 (fong3)	充 (cung1)	督 (duk1)	方 (fong1)	別 (bit6)
燈 (dang1)	*lamp*	肯 (hang2)	入 (jap6)	項 (hong6)	增 (zang1)	音 (jam1)	康 (hong1)	沒 (mut6)
冰 (bing1)	*ice cubes*	徑 (ging3)	接 (zip3)	盟 (mang4)	清 (cing1)	色 (sik1)	亨 (hang1)	尺 (cek3)
衫 (saam1)	*shirt*	斬 (zaam2)	犯 (faan6)	嚴 (jim4)	貪 (taam1)	耕 (gaang1)	簽 (cim1)	敵 (dik6)
葉 (jip6)	*leave*	妾 (cip3)	跌 (dit3)	答 (daap3)	獵 (lip6)	敵 (dik6)	立 (lap6)	革 (gaak3)
菜 (coi3)	*vegie*	代 (doi6)	保 (bou2)	每 (mui5)	再 (zoi3)	布 (bou3)	悔 (fui3)	巧 (haau2)
牛 (ngau4)	*bull*	秋 (cau1)	閉 (bai3)	超 (ciu1)	求 (kau4)	迷 (mai4)	潮 (ciu4)	紀 (gei2)
豹 (paau3)	*leopard*	交 (gaau1)	猜 (caai1)	操 (cou1)	教 (gaau3)	泰 (taai3)	告 (gou3)	採 (coi2)
梯 (tai1)	*ladder*	迷 (mai4)	究 (gau3)	配 (pui3)	西 (sai1)	秋 (cau1)	灰 (fui1)	保 (bou2)
鑊 (wok6)	*frying pan*	托 (tok3)	割 (got3)	革 (gaak3)	莫 (mok6)	望 (mong6)	麥 (mak6)	涉 (sip3)
繩 (sing2)	*rope*	政 (zing3)	協 (hip3)	層 (cang4)	境 (ging2)	檢 (gim2)	訪 (fong2)	客 (haak3)

Note: Picture names and word distractors are shown in traditional Chinese (VC = vowel and coda related only; V = vowel related only; C = coda related only; VCT = vowel, coda, and tone related only; VT = vowel and tone related only; CT = coda and tone related only; Control = unrelated control condition). English translations for the picture names are shown in italics. Pronunciations of the Chinese characters are given in parentheses, and all syllable markings are adopted from the Chinese Character Database: With Word-formations (2003). The number besides each syllable marking denotes the lexical tone.

Appendix B: Names of the Pictures and Word Distractors used in Experiment 3

Picture name	CCT	CC	CV	Control
橙 (caang2) *orange*	寵 (cung2)	場 (coeng4)	蠶 (caam4)	兼 (gim1)
箭 (zin3) *arrow*	振 (zan3)	真 (zan1)	漸 (zim6)	及 (kap6)
象 (zoeng6) *elephant*	贈 (zang6)	政 (zing3)	著 (zoek3)	答 (daap3)
鷹 (jing1) *eagle*	央 (joeng1)	養 (joeng5)	染 (jim5)	發 (faat3)
田 (tin4) *farmland*	團 (tyun4)	盾 (teon5)	貼 (tip3)	律 (leot6)
星 (sing1) *star*	商 (soeng1)	相 (soeng3)	食 (sik6)	各 (gok3)
藥 (joek6) *medicine*	育 (juk6)	沃 (juk1)	陽 (joeng4)	明 (ming4)
鐘 (zung1) *clock*	張 (zoeng1)	井 (zeng2)	族 (zuk6)	別 (bit6)
燈 (dang1) *lamp*	東 (dung1)	洞 (dung6)	特 (dak6)	沒 (mut6)
冰 (bing1) *ice cubes*	崩 (bang1)	餅 (beng2)	扁 (bin2)	尺 (cek3)
葉 (jip6) *leave*	入 (jap6)	泣 (jap1)	厭 (jim3)	革 (gaak3)
菜 (coi3) *vegie*	砌 (cai4)	猜 (caai1)	操 (cou1)	巧 (haau2)
豹 (paau3) *leopard*	票 (piu3)	葡 (pou4)	排 (paai4)	採 (coi2)
梯 (tai1) *ladder*	推 (teoi1)	退 (teoi3)	頭 (tau4)	保 (bou2)
繩 (sing2) *rope*	爽 (song2)	送 (sung3)	禪 (sim4)	客 (haak3)
床 (cong4) *bed*	情 (cing4)	充 (cung1)	錯 (cok3)	給 (kap1)
桶 (tung2) *bucket*	躺 (tong2)	停 (ting4)	禿 (tuk1)	塞 (sak1)
船 (syun4) *ship*	純 (seon4)	信 (seon3)	說 (syut3)	店 (dim3)
窗 (coeng1) *windows*	清 (cing1)	寵 (cung2)	桌 (coek3)	極 (gik6)
碗 (wun2) *bowl*	穩 (wan2)	溫 (wan1)	活 (wut6)	尋 (cam4)

Note: Picture names and word distractors are shown in traditional Chinese (CCT = onset-consonant, coda, and tone related only; CC = onset-consonant and coda related only; CV = onset-consonant and vowel related only; Control = unrelated control condition). English translations for the picture names are shown in italics. Pronunciations of the Chinese characters are given in parentheses, and all syllable markings are adopted from the Chinese Character Database: With Word-formations (2003). The number besides each syllable marking denotes the lexical tone.

Appendix C: Names of the Pictures and Word Distractors used in Experiment 4

	Picture name		OR	OR Control	O	O Control
SR	蛇 (se4)	*snake*	些 (se1)	注 (zyu3)	市 (si5)	禍 (wo6)
	豬 (zyu1)	*pig*	注 (zyu3)	邪 (ce4)	借 (ze3)	市 (si5)
	畫 (waa2)	*painting*	娃 (waa1)	姑 (gu1)	禍 (wo6)	夫 (fu1)
	褲 (fu3)	*trousers*	乎 (fu4)	些 (se1)	火 (fo2)	查 (caa4)
	車 (ce1)	*car*	邪 (ce4)	乎 (fu4)	查 (caa4)	賀 (ho6)
	梳 (so1)	*comb*	所 (so2)	霞 (haa4)	射 (se6)	價 (gaa3)
	蝦 (haa1)	*shrimp*	霞 (haa4)	所 (so2)	賀 (ho6)	火 (fo2)
	鼓 (gu2)	*drum*	姑 (gu1)	貨 (fo3)	價 (gaa3)	傻 (so4)
	樹 (syu6)	*tree*	殊 (syu4)	娃 (waa1)	傻 (so4)	借 (ze3)
	火 (fo2)	*fire*	貨 (fo3)	殊 (syu4)	夫 (fu1)	射 (se6)
CR	橙 (caang2)	*lamp*	燦 (caang3)	氈 (zin1)	簽 (cim1)	佔 (zim3)
	箭 (zin3)	*arrow*	氈 (zin1)	堂 (tong4)	站 (zaam6)	飲 (jam2)
	象 (zoeng6)	*elephant*	障 (zoeng3)	天 (tin1)	佔 (zim3)	該 (goi1)
	鷹 (jing1)	*eagle*	形 (jing4)	械 (haai6)	飲 (jam2)	站 (zaam6)
	田 (tin4)	*farmland*	天 (tin1)	燦 (caang3)	貪 (taam1)	審 (sam2)
	鞋 (haai4)	*shoes*	械 (haai6)	障 (zoeng3)	毫 (hou4)	禿 (tuk1)
	星 (sing1)	*star*	成 (sing4)	究 (gau3)	審 (sam2)	毫 (hou4)
	糖 (tong2)	*sugar*	堂 (tong4)	約 (joek3)	禿 (tuk1)	因 (yan1)
	狗 (gau2)	*dog*	究 (gau3)	形 (jing4)	該 (goi1)	簽 (cim1)
	藥 (joek6)	*medicine*	約 (joek3)	成 (sing4)	因 (yan1)	貪 (taam1)

Note: Picture names and word distractors are shown in traditional Chinese (OR = onset and rhyme related only; OR Control = unrelated baseline for the OR condition; O = onset related only; O Control = unrelated baseline for the O condition). English translations for the picture names are shown in italics. Pronunciations of the Chinese characters are given in parentheses, and all syllable markings are adopted from the Chinese Character Database: With Word-formations (2003). The number besides each syllable marking denotes the lexical tone. SR = simple rhyme condition; CR = complex rhyme condition.

Appendix D: Names of the Pictures and Word Distractors used in Experiment 5

Experiment 5A

Rhyme Sonority	Picture name			R		R Control	
High	遮	(ze1)	*umbrella*	騎	(ke4)	查	(caa4)
	河	(ho4)	*river*	初	(co1)	車	(ce1)
	鎖	(so2)	*lock*	羅	(lo4)	騎	(ke4)
	蝦	(haa1)	*shrimp*	查	(caa4)	羅	(lo4)
	梳	(so1)	*comb*	賀	(ho6)	拿	(naa4)
	馬	(maa3)	*horse*	拿	(naa4)	苛	(ho1)
	畫	(waa2)	*painting*	沙	(saa1)	初	(co1)
	蛇	(se4)	*snake*	車	(ce1)	賀	(ho6)
	火	(fo2)	*fire*	苛	(ho1)	沙	(saa1)
Low	耳	(yi5)	*ear*	試	(si3)	處	(cyu3)
	鼓	(gu2)	*drum*	污	(wu1)	試	(si3)
	紙	(zi2)	*paper*	似	(ci5)	預	(jyu6)
	褲	(fu3)	*trousers*	箍	(ku1)	於	(jyu1)
	豬	(zyu1)	*pig*	預	(jyu6)	婦	(fu5)
	魚	(jyu2)	*fish*	處	(cyu3)	箍	(ku1)
	書	(syu1)	*book*	住	(zyu6)	似	(ci5)
	湖	(wu4)	*lake*	婦	(fu5)	住	(zyu6)
	樹	(syu6)	*tree*	於	(jyu1)	污	(wu1)

Experiment 5B

Rhyme Duration	Picture name			R		R Control	
Short	木	(muk6)	*wood*	哭	(huk1)	屈	(wat1)
	竹	(zuk1)	*bamboo*	僕	(buk6)	奪	(dyut6)
	尺	(cek3)	*ruler*	石	(sek6)	實	(sat6)
	盒	(haap2)	*box*	甲	(gaap3)	托	(tok3)
	筆	(bat1)	*pencil*	實	(sat6)	石	(sek6)
	葉	(jip6)	*leave*	貼	(tip3)	哭	(huk1)
	襪	(mat6)	*sock*	屈	(wat1)	啄	(doek3)
	雪	(syut3)	*snow*	奪	(dyut6)	僕	(buk6)
	藥	(joek6)	*medicine*	啄	(doek3)	甲	(gaap3)
	鑊	(wok6)	*frying pan*	托	(tok3)	貼	(tip3)
Long	衫	(saam1)	*shirt*	淡	(daam6)	挺	(ting5)
	碗	(wun2)	*bowl*	悶	(mun6)	常	(soeng4)
	田	(tin4)	*farmland*	先	(sin1)	贈	(zang6)

橙	(caang2)	*orange*	冷	(laang5)	先	(sin1)
床	(cong4)	*bed*	項	(hong6)	淡	(daam6)
星	(sing1)	*star*	挺	(ting5)	悶	(mun6)
窗	(coeng1)	*windows*	常	(soeng4)	扁	(bin2)
箭	(zin3)	*arrow*	扁	(bin2)	項	(hong6)
燈	(dang1)	*lamp*	贈	(zang6)	寸	(cyun3)
船	(syun4)	*ship*	寸	(cyun3)	冷	(laang5)

Note: Picture names and word distractors are shown in traditional Chinese (R = rhyme related only; R Control = unrelated baseline for the R condition). English translations for the picture names are shown in italics. Pronunciations of the Chinese characters are given in parentheses, and all syllable markings are adopted from the Chinese Character Database: With Word-formations (2003). The number besides each syllable marking denotes the lexical tone.