Gestural Overlap across Word boundaries: Evidence from English and Mandarins Speakers

by

Shan Luo
B.A., Sichuan University, 2007
M.A., Nankai University, 2010

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy

in the Department of Linguistics

Supervisory Committee

Gestural Overlap across Word Boundaries:
Evidence from English and Mandarins Speakers

by

Shan Luo
B.A., Sichuan University, 2007
M.A., Nankai University, 2010

Supervisory Committee

Hua Lin, Linguistics Department, University of Victoria
Supervisor

Sonya Bird, Linguistics Department, University of Victoria
Supervisor

John Esling, Linguistics Department, University of Victoria
Departmental Member

Tsung-Cheng Lin, Department of Pacific and Asian Studies, University of Victoria
Outside Member

Abstract

Supervisory Committee

Hua Lin (Department of Linguistics, University of Victoria)

Supervisor

Sonya Bird (Department of Linguistics, University of Victoria)

Supervisor

John Esling (Department of Linguistics, University of Victoria)

Departmental Member

Tsung-Cheng Lin (Department of Pacific and Asian Studies, University of Victoria)

Outside Member

This research examines how competing factors determine the articulation of English stop-stop sequences across word boundaries in both native (L1) and nonnative (L2) speech. The two general questions that drive this research are 1) how is consonantal coordination implemented across English words? And 2) is this implementation different in L1 versus L2 speech?

A group of 15 native English (NE) speakers and a group of 25 native Mandarin speakers (NM) who use English as a foreign language (ESL) participated in this study. The stimuli employed in this research were designed along four major parameters: 1) place of articulation (i.e., homorganic, front-back, and back-front clusters), 2) lexical frequency (i.e., high frequency vs. low frequency words, and real words vs. nonwords), 3) stress (i.e., word-level stress and sentence-level focus), and 4) speech rate (i.e., stimuli produced in isolation and embedded in conversation). All English stop combinations were considered. The release percentages and closure duration ratios produced by English and Mandarin speakers were measured. Four-way repeated measures ANOVAs analyzed the overall data along four independent factors: *Group, Place of Articulation, Lexical Frequency,* and *Speech Rate*. In addition, two sets of five-way repeated measures ANOVAs analyzed the data at each speech rate separately. At the slow rate, factors considered were: *Group, Place of Articulation, Lexical Frequency, Compound,* and *Voicing*. At the fast rate, factors considered were: *Group, Place of Articulation, Lexical Frequency, Focus,* and *Voicing*.

The results showed that place of articulation had different effects on English and Mandarin speakers in their English stop-stop coarticulation, especially in heterorganic clusters. Specifically, a place order effect (i.e., more releases and more overlap in front-back clusters than in back-front clusters; POE) was only partially supported in native speech but not shown at all in nonnative speech in the current research. The NE group significantly showed the release pattern of the POE while the NM group did not. Although both groups showed slightly more overlap in front-back than in back-front clusters, the temporal pattern predicted by the POE was not significant in either group.

The results also confirmed a gradient lexical frequency effect, finding a significant correlation between self-rated frequency and overlap: both groups tended to have more overlap in more frequently used words than in less frequently words (e.g., *take care* vs. *dock gate*). A group difference was observed in the interaction between the effects of place of articulation and categorical frequency: the NE group significantly exhibited the release pattern predicted by the POE in both real words and nonwords while the NM did not show such a significant release pattern in either lexical context; the NE group also had consistent overlap orders across lexical conditions while the NM group did not. These results suggest that native speakers can apply strategies used in known forms to novel forms, while L2 speakers cannot. Based on this finding, this dissertation supports the idea that L1 speakers have phonologized their strategies of phonetic implementation, while L2 speakers have not: the way L2 speakers coordinate sound patterns is largely dependent upon their familiarity with individual lexical items, implying that L2 sound patterns are registered as discrete chunks.

In addition, the results showed, unexpectedly, a stronger stress effect for the NM group rather than for the NE group. With respect to word-level stress, both groups of participants released significantly more often in compounds (i.e., C1C2, initial stress) than in non-compounds (i.e., C1C2, final stress); but the NM group also had significantly more overlap in compounds than in non-compounds while the NE group had slightly more overlap in non-compounds. With respect to sentential stress, both groups had significantly more releases and less overlap (i.e., longer duration) in sentential focused words (i.e., **F1**F2) than in words without focus (i.e., F1F2). Contrary to the prediction, it was the NM group that showed the overlap order (i.e., more overlap in F1**F2** than in **F1**F2), which was predicted for the NE group; the NE group had more overlap in **F1**F2 than in F1**F2**. Moreover, the NM group had significantly more overlap (i.e., shorter closure duration) under F1F2 (no focus) than under any other focus positions (i.e., **F1**F2, F1**F2**, **F1F2**) in both front-back and back-front clusters while the NE group did not. The unexpected result of Mandarin speakers showing a stronger stress effect than English speakers indicated a trend of over-compensation in L2 speech.

Further analyses showed that increased speech rate did not systematically induce increased temporal overlap, because speakers from both groups varied in their behavior, having either more or less overlap at the fast speech rate than at the slow rate. Lastly, the analyses found no correlation between closure duration ratio and perceived accent in L2 speech. This finding was not predicted, given that timing features had always been considered critical to foreign accent perception.

The current research contributes to the body of literature on consonant sequence realization in three ways. Firstly, it contributes to the sparse literature documenting what determines the surface forms of consonant coarticulation across words in both L1 and L2 speech. Specifically, four factors had been identified to determine the overlap degree in previous research: place of articulation, lexical frequency, stress, and speech rate; this is the first study examining each of these factors concurrently, as well as interactions between these factors, to develop a more comprehensive account of what affects consonantal coarticulation in both L1 and L2 speech. Secondly, this study contributes to our understanding of how phonological representations are differently registered in native versus nonnative processing. This study is among the first, if not the first, to examine how lexical frequency interacts with the POE in both L1 and L2 speech. The findings support a gestural unit representation of speech and suggest that L1 speakers process significantly richer and heavier informational structures than we ever could imagine. In contrast, sound patterns are memorized as acoustic properties of individual lexical items in L2 processing, resulting in their inability to generalize beyond memorized items. Thirdly, this study provides a comprehensive and detailed comparison of L1 and L2 speech features, especially pertaining to Mandarin ESL speakers. This is pedagogically valuable for language instructors and students in that the study pinpoints in which particular aspects the Mandarin speakers, and possibly L2 speakers in general, deviate from English speakers in the production of consonant clusters.

Table of Contents

List of Tables

List of Figures

Acknowledgments

First of all, I would like to take this opportunity to thank my two supervisors, Dr. Hua Lin and Dr. Sonya Bird. I express my deepest gratitude to Dr. Lin for greatly inspiring me in every aspect throughout the years that a student could only hope for. She is such an exceptionally passionate and caring supervisor/mentor and she would never hesitate to share her experience and provide essential help on issues both academic and personal. I also owe a great debt to Dr. Sonya Bird for her valuable guidance, insight, patience and encouragement at every stage of this work and with my other papers. Sonya always attempted to understand the ideas that I could not always phrase transparently, stimulating me to refine my ideas and present greater conceptual details than I might otherwise have done. Thanks are due to my other three committee members, Dr. Alexei Kochetov, Dr. John Esling, and Dr. Tsung-Cheng Lin, for generously dedicating their time and attention despite their terrifically busy schedules. They read (and re-read) every single page of this work, providing detailed, encouraging and constructive commentary through all the aspects of this research project. Also, I wish to thank Dani Byrd and Lisa Davidson who read parts of this dissertation as journal reviewers, providing essential comments and valuable suggestions.

I enormously appreciate all the faculty and staff members in the department of Linguistics at the University of Victoria, especially Leslie Saxon, Li-Shih Huang, Ewa Czaykowska-Higgins, Hossein Nassaji, Sandra Kirkham, Suzanne Urbanczyk, Dave McKercher and Peter Jacobs. Ewa, thank you for being patient and showing me the formalism of phonology. Leslie, thank you for supporting me when I was the CLA student representative, and thank you for saving me when I locked myself out of the lab. I would always remember the days we were the only two people who stayed late in the department. Li-Shih, thank you for your passionate guidance and expertise in applied linguistics; this paper originated from your class, and I have benefited greatly from every conversation we have had. Also, thank you for giving me an opportunity to working with you on the project of assessing students' academic English needs. Jenny and Maureen, thank you for your outstanding administrative support. Thanks also go to Roger Howden at the Continuing Studies of the University of Victoria, who believed in me and invited me to facilitate a series of ESL workshops.

I immensely enjoyed rich, intellectual conversation with other graduate students and friends in and out of Victoria: Tess, Janet, Aki, Jian-Xun, Chris, Carrie, Shu-min, Xiao-Juan, Ning, and Sophia. Tess, thank you for sharing your thoughts on all the random topics and arguing with me like two experienced lawyers in court; those are valuable memories. I will miss you, and I hope our personal friendship continues for many years to come.

I could not have completed this research project without all the generous volunteer participants in this study, especially Corina who also helped with draft proofreading. I thank my dear friend, Buyun, for invaluable friendship and statistical consultation. I thank my parents, my sister, and my boyfriend for unconditional love, and for supporting my choice of linguistics over a more lucrative career. I am deeply grateful for the financial support from both the Linguistics Department at the University of Victoria (scholarship, TA, RA) and the China Scholarship Council, which eventually transformed into enough caffeine to write this paper.

Chapter 1

Introduction

The physical realizations of sounds are often conditioned by the surrounding segmental context in connected speech (e.g., Barry, 1991, 1992; Browman & Goldstein, 1989, 1992; Brown, 1977; Byrd, 1996a, 1996b; Cho 2004; Fowler & Saltzman, 1993; Nolan, 1992; Ohala, 1993; Öhman, 1967; Zsiga, 1994, among many others). For instance, it is typical for coronals to be perceived as assimilated to the following consonant or deleted in English: /t/ in *must be* is not perceived and thus heard as [mʌsbi] (Brown, 1977). Extensive studies within the framework of Articulatory Phonology have provided impressive evidence showing the occurrence of such coarticulation is due to substantial overlap of adjacent gestures in speech, and that a gesture that is not perceptually available is in fact not deleted but hidden by adjacent gestures (Browman & Goldstein, 1989, 1990; Barry, 1991, 1992; Byrd, 1996a, 1996b; Davidson & Stone, 2003; Gafos, 2002; Nolan, 1992; Zsiga, 1994). The implemented degree of overlap affects whether a consonant cluster (CC) is produced with close transition (complete overlap), a release burst, or a transitional schwa (insufficient overlap) (Davidson, 2003; Gafos, 2002; Hall, Hamann & Zygis, 2006). Following this assumption, many nonnative speech deviations are assumed to apply temporal alignment inappropriately in the target language (see Colantoni & Steele, 2008; Davidson, 2003, 2006, 2010; Messing, 2008; Solé, 1997; Zsiga, 2003, 2011).

As part of the ongoing process of understanding speech production, the present study examines how four major factors affect the degree of overlap in both native (L1) and nonnative (L2) speech: place of articulation, lexical frequency, stress, and speech rate. In particular, this study investigates surface forms of English stop-stop sequences spanning words, that is, C1#C2 (# indicates word boundaries). Three important goals motivate the design of the present study. First, extensive evidence has identified four factors that systematically determine the overlap degree in previous research: place of articulation, lexical frequency, stress, and speech rate; this study aims to combine these factors concurrently to develop a more comprehensive account of what affects consonantal coarticulation. Second, only a few studies have dealt with English cross-word cluster coarticulation produced by Mandarin speakers learning English (ESL) (e.g., Chen & Chung 2008; Messing, 2008). It remains poorly understood how coordination is implemented by this community. The current study seeks to provide a detailed account of cluster coarticulation in Mandarin ESL speech. Finally, comparing L1 and L2 speech is expected to be indicative of whether L2 speakers can extract temporal phrasing patterns in detailed phonetic phenomena and generalize beyond lexical items. This promises to inform theories on how gestural overlap patterns are stored in native versus nonnative processing.

The general hypothesis is that Mandarin speakers will deviate from the appropriate gestural coordination used by English speakers. In other words, Mandarin speakers cannot fully exhibit or extend a native-like sound pattern. Overall, four primary research questions concerning the degree of articulatory overlap are raised:

i. Does place of articulation (i.e., homorganic, front-back, and back-front clusters) affect English and Mandarin speakers similarly in their English stop-stop coarticulation? Specifically, previous literature on native English speech has suggested very few internal releases in homorganic clusters and a place order effect in heterorganic clusters (POE, more releases on the one hand and more overlap on the other in front-back clusters than in back-front clusters). Will both L1 and L2 speakers exhibit the same effect of place of articulation, for example, show the POE in a similar fashion?

ii. Does frequency/familiarity affect English and Mandarin speakers similarly in their English stop-stop coarticulation? Specifically, will L1 and L2 speakers coordinate clusters in high frequency words the same way as in low frequency words? And, will L1 and L2 speakers coordinate clusters in real words the same way as in nonwords?

iii. Do stress patterns affect English and Mandarin speakers similarly in their English stop-stop coarticulation? Specifically, English is a stress-timed language in which stressed syllables have longer durations and less coarticulatory effect than unstressed syllables; Mandarin is a syllable-timed language in which every syllable receives relatively equal timing. Does this rhythmic difference affect cluster coordination produced by speakers of the two languages?

iv. Do changes in speech rate affect English and Mandarin speakers similarly in their English stop-stop coarticulation? Will both groups of speakers increase overlap when their speech rate is increased?

In the remainder of this chapter, I outline the approach that this study adopts, discuss the significance of this study, and provide an overview of the body of the dissertation.

1.1.The approach of this study

To answer the research questions above, the stimuli were designed along four major parameters (see Section 3.1): 1) place of articulation (i.e., homorganic, front-back, and back-front clusters); 2) lexical frequency (i.e., high vs. low frequency items in real words, and real words vs. nonwords); 3) stress (i.e., word-level stress and sentential focus); and 4) speech rate (i.e., reading phrases and spontaneous speech). All stop combinations were considered, yielding 36 combinations (e.g., p#p, p#t, p#k). To examine the lexical stress effect, 36 compounds (initial stress) and 36 non-compounds (final stress) were designed with the same C1#C2 pair (e.g., *SOUP pot* vs. *keep PACE*). These 72 real words had varying degrees of frequency (see Section 4.4.1). Another 72 phrases were designed as corresponding nonwords, yielding 144 pairs in total. By corresponding, I mean two conditions were met: 1) their surrounding environments were controlled with the same vowels (e.g., nonsense *peep pate* vs. meaningful *keep pace*), and 2) the initial and final consonants agreed in voicing and manner (e.g., *dak kit* vs. *back kick*).

The research employs a group of native English speakers (NE) and a group of native Mandarin speakers (NM) to complete four tasks. The NE group acts as a baseline for comparison. Task 1 is to rate word familiarity of each syllable pair in order to validate the categorical effect of lexical frequency (real words vs. nonwords). Task 2 is to read a first list that includes 24 baseline items: each stop occurs intervocalically, twice word-initially and twice word-finally (i.e., $V\#C1_bV$, $VC2_b\#V$). This list is designed to elicit the baseline closure duration of each stop. Task 3 is to read the 144 syllable pairs in isolation where two stops are adjacent across words -VC1#C2V- (see Appendix A). Task 4 is to answer questions based on a given dialogue. Specifically, the 144 words (C1#C2) used in Task 3 are embedded in a target position in Task 4 where either the syllable containing C1 or C2 receives sentential focus, or both syllables can be focused, or unfocused (more in Section 3.3.2).

To analyze the realization of clusters, two dependent variables concerning gestural coordination are considered in this study: release percentage and closure duration ratio (see Zsiga, 2003). Release percentage is calculated as the percentage of phrases where C1 has a release burst. The closure duration ratio is calculated as the mean duration of the C1#C2 cluster divided by the sum of the mean closure durations of $C1_b$ and $C2_b$ when $C1_b$ or $C2_b$ occur intervocalically (Task 2). The overlap degree is then calculated as 1 minus the closure duration ratio (see 3.3 Data Analysis).

1.2.The significance of this study

This study examines how English and Mandarin speakers coordinate English stop-stop clusters (C1#C2) across words. Understanding the interaction of word boundaries and articulation is, as Byrd (1996a) states, “a crucial question for both segmental and dynamic phonological theories” (p. 211). This dissertation aims to extend existing findings on cluster overlap at word boundaries by comparing L1 and L2 speakers. Although this acoustic study will not directly measure articulatory contact, it is the first study offering an examination of all possible two-stop combinations. Four factors have been identified to systematically determine the overlap degree in previous research, but this is the first study combining the four main factors in one study: POA, word frequency, stress, and speech rate.

The current research not only examines the phonetic details of sound patterns, but also addresses the ability of transferring phonological knowledge to novel sounds. While much of the current research deals with phonetic descriptions of cluster coordination, a research project to account for patterns of sound distribution and to explain cognitive aspects of this distribution is clearly needed. This study examines the interaction between lexical frequency and sound-related patterns like the POE, across L1 and L2 speech. In doing so, the study contributes to our understanding of how phonological representations are registered in native versus nonnative processing.

Thirdly, this study provides a comprehensive and detailed comparison of L1 and L2 speech features, especially pertaining to Mandarin ESL speakers. This is pedagogically valuable for language instructors and students in that it identifies in which particular aspects of pronunciation among Mandarin speakers, and possibly L2 speakers in general, deviate from native speakers, in terms of consonant clusters.

1.3.Organization

The dissertation is organized as follows. Chapter 2 reviews previous studies and discusses the framework (i.e., Articulatory Phonology), which this study rests upon. Chapter 2 also reviews factors that have been identified to systematically influence C1C2 overlap degree. Following previous studies, research questions are proposed. Chapter 3 details the research methodology. Chapter 4 presents the comprehensive statistical results. Chapter 5 provides an in-depth discussion of results and hypotheses as tested in the current study, discussing the major findings, unexpected results, and their implications. Chapter 6 concludes the dissertation with a synopsis of the main points, outlining limitations and future research directions, and stressing their relevance to understanding language processing.

Chapter 2

Literature Review

The physical realizations of sounds are conditioned by surrounding segmental contexts. Previous studies have used different terms to refer to the process of connecting two adjoining sounds with different terms, such as linking (Brown & Kondo-Brown, 2006; Celce-Murcia, Brinton, & Goodwin, 1996; Grant, 2000; Hieke, 1987; Morley, 1991), blending (Edwards & Strattman, 1996), coarticulation (Browman & Goldstein, 1990; Cho, 2004; Fowler, 1980), sandhi variation (Henrichsen, 1984), co-production (Gafos, 2002), and external sandhi (Zsiga, 2011). Coarticulation, the term that will be used in this dissertation, can occur between a consonant in the coda and a following vowel (C#V) (Gimson, 1970; Prator & Robinett, 1985), between two consonants (C-to-C) (Catford, 1977; Hardcastle & Roach, 1979; Ladefoged, 2001), or between two vowels (V-to-V) (Barb, 2005; Cho, 2004).

This chapter lays the foundation of the dissertation by exploring the articulatory details of stop-stop coarticulation. The data mainly revolve around English clusters; for instance, acoustic evidence shows that adjacent English consonants share a significantly overlapped relationship. This chapter begins with an introduction of language background, including the syllable structures and rhythmic patterns in English and Mandarin – the two languages relevant for the research project, as well as the basic phonetic properties of English stop-stop clusters. In Chapter 2.2, I describe basic ideas of Articulatory Phonology (henceforth, AP), which provides us with a framework to analyze speech gestures. Section 2.3 offers a detailed discussion of previous studies, outlining several factors that have been identified to systematically determine the overlap degree in English clusters. Section 2.4 reviews the literature on cluster production in nonnative speech, including different repair types shown by nonnative speakers, and theoretical accounts that attempt to explain these nonnative speech errors. Having laid out the relevant background literature, Section 2.5 proposes the main research questions and their relative hypotheses.

2.1. Language backgrounds

In order to explore Mandarin speakers' stop-stop coarticulation in English, we need to understand the phonological systems of the two languages, and we must familiarize ourselves with the phonetic properties of English stop-stop clusters. In this section we will discuss the syllable structure and timing patterns in English (Section 2.1.1) and in Mandarin (Section 2.2.2).

2.1.1. English syllable structure and timing patterns

Ladefoged (1999) describes North American English as having 10 simple vowels and five diphthongs, and 24 consonants, including both voiceless (i.e., /p, t, k/) and voiced stops (i.e., /b, d, g/) (Ohala, 1983). In English, 18 consonantal segments can occur in syllable-final position and 21 can occur initially, resulting in 378 possible C1#C2 clusters (Catford, 1977). English, a language well-known for a potential to have very heavy syllables, can have up to three consonants in onset position (e.g., *splash*) and four in coda position (e.g., *sixths*) (Shockey, 2008). Harris (1994, p. 53) formalizes English syllable template as:

(1) $[X^3_0]$Onset $[X^2_1]$Nucleus$[X^4_1]$Coda

This formalization illustrates that possible syllabic constituents in English usually include zero to three consonants as the onset (e.g., *eye*, *pie*, *pry*, *spry*), zero to four consonants as the coda (e.g., *see*, *sick*, *six*, *sixth*, *sixths*), and at least one and at most two vowels at the nuclear part (e.g., *bid* and *boat*).

In an English word that is comprised of more than one syllable, stressed syllables stand out while unstressed ones are weak, for example, "**RE**bel" (the noun) versus "re**BEL**" (the verb) (see Abercrombie, 1967; Beckman, 1986; Faber, 1986; Kreidler, 1989). Phonetically, the "weak" (i.e., unstressed) segments have more coarticulatory effects than strong segments, such as more lenited characteristics (e.g., reduced duration, deletion) (see Avery & Ehrlich, 1992; Celce-Murcia et al., 1996; Crystal, 1969; Cummins & Port, 1998; Öhman, 1967; Weinstein, 2001). For instance, Keating (1984) reported that several subjects increased closure durations of initial voiced stops in stressed syllables. Results in general indicate that the most consistent acoustic correlates underlying lexical stress are an increase in F0, duration, and intensity of stressed vowels in relation to unstressed vowels.

Similar to word level stress, sentence level stress often bears a nuclear pitch accent in English (e.g., "**HE** hit her" vs. "He **HIT** her" vs. "He hit **HER**") (see Cho, 2004; Kent & Netsell, 1971; Ohala & Hirano, 1967). In this study I refer to sentential stress as focus. Speakers are found to produce focused syllables with longer durations than their unfocused counterparts (Cho, 2004; De Jong, Beckman, & Edwards, 1993; Edwards, Beckman, & Fletcher, 1991; Summers, 1987). Less coarticulation also occurs in focused syllables than in unfocused syllables (see Cho, 2004). One example is the study of De

Jong (1995), which discovered that the coarticulation of /t/ into a following /ð/ is reduced when /t/ occurs in focused syllables (e.g., PUT the __ on the table).

Based in part on variation in the duration of syllables and other prosodic units, languages have been traditionally characterized as stress-timed, syllable-timed, or mora-timed (Arvaniti 2009; Abercrombie, 1967; Grabe & Low, 2002; Ladefoged, 2001; Lin & Wang, 2007; Ramus, Nespor, & Mehler, 1999; Roach, 1982). English rhythm is considered to be stress-timed; it has a stream of stressed and unstressed syllables in which stressed syllables stand out. In contrast, other languages such as Spanish and Mandarin, in which every syllable receives the same prominence, are considered syllable-timed (Grabe & Low, 2002; Kreidler, 1989; Lin & Wang, 2007; Ramus et al., 1999). In L2 speech research, rhythmic categorization is often employed to explain why speakers of languages with different rhythmic patterns sound choppy and hard to understand (e.g., Cutler, Mehler, Norris, & Segui, 1986; Lin & Wang, 2007; Ling, Grabe, & Nolan, 2000; Tajima, Port, & Dalby, 1997). The two languages considered here, English (stress-timed) and Mandarin (syllable-timed), provide us with an ideal case to test the timing properties in both L1 and L2 speech, and how these might affect stop-stop coarticulation. In the next subsection, I will introduce Mandarin syllable structure and timing patterns.

2.1.2. Mandarin syllable structure and timing

There are languages that, unlike English, do not allow stops word-finally (e.g., Italian, Tamil, Japanese) (Lisker, 1999), and many languages do not have consonant clusters (e.g., Japanese) (Scarcella & Oxford, 1994). Mandarin is a typical example of these, disallowing both word-final stops and stop-stop clusters.

Mandarin has five vowels (i.e., /i, y, u, ə, a/), four diphthongs (i.e., /əi, əu, ai, au/), and 19 consonants, including unaspirated (/p, t, k/) and aspirated (/p^h, t^h, k^h/) stops (Cheng, 1966; Duanmu, 2007). One major difference between Mandarin and English is that all Mandarin stops are voiceless while English stops may be voiceless or voiced,[1] although the six stops are sometimes transcribed as the same [p, b, t, d, g, k]. Another major difference is that Mandarin has very strong phonotactic constraints on syllable shape. Lin (2001) describes the maximal syllable structure in Mandarin as CGVX (C=consonant, G=glide, V=vowel, X=nasal or glide). Only four syllable types are allowed: vowel (e.g., *é*, "goose"), consonant vowel (e.g., *dǎ*, "hit"), consonant-vowel-nasal/glide (e.g., *dàn* "egg"; *fēng*, "wind"; *tāj* "to stay"), and vowel-nasal/glide (e.g., *ān,* "safe"; *āj,* "to endure") (Cheng, 1966). From these syllable types, it can be seen that Mandarin only allows one consonant in the onset position, and a nasal (/n/, /ŋ/) or glide in coda position.[2] That is, Mandarin prohibits 1) word-final obstruents and 2) consonant clusters in any position.

As mentioned above, Mandarin is considered a syllable-timed language, in which each syllable receives relatively equal timing across a sentence (Grabe & Low, 2002; Lin & Huang, 2009; Lin & Wang, 2007; Mok & Dellwo, 2008). Studies have agreed that Chinese speakers transfer the timing of their L1 in the production of English (e.g., Chen, 2005; Chen & Chung, 2008; Chen, Fan, & Lin, 1996; Lin & Huang, 2009; Lin & Wang, 2007; Tajima et al., 1997). The study by Tajima et al. (1997) examined temporal properties of syllables with a Chinese ESL learner as well as a native English speaker.

---

[1] The Mandarin unaspirated stops [p, t, k] can become voiced [b, d, g] when they occur in an unstressed

[2] Note that the nasal /m/ is not allowed in Mandarin codas. Also, the oral closure of [n, ŋ] in codas is often incomplete (Wang, 1993; Xu, 1986).

Both speakers read a list of approximately 100 short English phrases randomly selected from texts. The researchers then temporally modified the Mandarin speaker's utterances so as to make the duration of acoustic segments match the duration of corresponding segments in the English speaker's productions. The English speaker's utterances were distorted to match the segment durations of the Mandarin speaker. The results showed that the intelligibility improved considerably for the Chinese-accented utterances from 39% to 58% after timing modification, and the native productions declined significantly from 94% to 83%.

Due to the rhythmic differences in English and Mandarin, there have been many acoustic and perceptual studies examining word and sentence level stress in Mandarin speakers' production of English (see Chun, 1982; Chen, Robb, Gilbert, & Lerman, 2001; Wang, 2008; Xu, 1999; Zhang, Nissen, & Francis, 2008).

With respect to word-level stress, Zhang et al. (2008) discovered that both English and Mandarin speakers produced stressed syllables with a higher F0, longer duration and greater intensity than unstressed syllables. In their study, Mandarin speakers produced longer durations than English speakers in both stressed and unstressed syllables (351 ms vs. 329 ms in stressed syllables; 277 vs. 250 ms in unstressed syllables). They concluded that Mandarin speakers seemed to be able to successfully produce native-like patterns of duration in producing stress. Zhang et al. also pointed out that it was difficult to determine whether L2 learners learned to produce the acoustic cues of stress systematically or they had simply already learned these cues for the specific words examined in their study; in other words, they could not exclude a lexical frequency effect.

With respect to sentential stress (i.e., focus), Chun (1982) reported that Mandarin speakers learning English were perceived to produce focused words with abnormally short durations. The study of Chen et al. (2001) found that both American and Mandarin speakers produced focused words with longer durations compared to unfocused words, with Mandarin speakers producing focused words with a shorter duration compared to American speakers (similar results were found for Cantonese speakers, see Ng & Chen, 2011). Further analysis in Chen et al. revealed that Mandarin females produced unfocused words with a significantly longer duration compared to American females; but this difference was not observed between Mandarin and American males. These results were interpreted as variable application of contrastive durations by Mandarin speakers arising from their native language influence.

2.1.3. Summary

In summary, English has voiced stops, and stop clusters are very common in word-initial, word-final, or across-word position; Mandarin only has voiceless stops and categorically prohibits stop-stop clusters. Due to these differences, Mandarin speakers are found to have substantial difficulty in acquiring English clusters (e.g., Anderson, 1987; Broselow & Finer, 1991; Chen & Chung, 2008; Hansen, 2001; Tajima et al., 1997). I will discuss in detail this learning difficulty in Section 2.4.3. Based on previous work, it is expected that Mandarin speakers will have more internal releases than English speakers in producing English consonant clusters, but will be more likely to delete voiced consonants due to learning difficulty. In this study, the deletion will be shown as no release (more in Section 3.4).

In addition, stress affects coarticulation both at the word level and at the sentence level in English: less coarticulatory effect (e.g., longer duration) occurs for lexically and sententially stressed segments than for unstressed segments (Avery & Ehrlich, 1992; Celce-Murcia et al., 1996; Crystal, 1969; Cummins & Port, 1998; Öhman, 1967; Weinstein, 2001). In Mandarin, however, each syllable receives relatively equal prominence throughout a sentence, such that prosodic effects on coarticulation are minimal. A number of works that have measured absolute durations in stressed versus unstressed syllables (Chen et al., 2001; Zhang et al., 2008) seem to suggest that Mandarin speakers acquire native-like durational correlates of stress, while the others find that Mandarin speakers transfer the durational patterns from their native language in ESL speech (e.g., Chen, 2005; Chen et al., 1996; Chen & Chung, 2008; Lin & Huang, 2009; Lin & Wang, 2007; Tajima et al., 1997). What the previous research on L2 speech has not done yet is a) consider the lexical frequency effect in order to establish whether or not the durational patterns of stress are acquired systematically, and b) measure the *relative* duration between stressed versus unstressed syllables, since “strong” and “weak” syllables are relative.

In this study, not only will voicing and stress features be considered, but frequency effects (independent variable) and relative shortening (dependent variable) will be examined in the production of English cluster coordination by L2 speakers (as well as by L1 speakers).

2.2.Theoretical background

In this section, I will present the basic ideas and formalizations of Articulatory Phonology (AP). The primary reason for adopting this framework here is that AP offers an explicit approach to characterizing speech timing, and a large amount of work that I will discuss throughout this chapter and elsewhere in the dissertation is conducted within this framework. To understand how AP developed, I need to first introduce the phonetic properties of English stop clusters (Section 2.2.1) and how they have been analyzed in different phonological traditions (Section 2.2.2). The basic ideas of AP are introduced in Section 2.2.3.

2.2.1. Phonetic properties of English stop clusters

In the production of English consonantal sequences, Catford (1977) refers to three possible transitions in terms of the articulatory relationships between the consonants involved: homorganic, heterorganic, and contiguous. A homorganic sequence is one in which the articulatory location of both consonants is identical, such as [tt, ss, nn]. The sequence is still considered homorganic even if it involves a change of phonation (e.g., [td, zs, bp]), a change of stricture type (e.g., [ts, zd]), a change of oral air-path (e.g., [tl, ld, lz]), or a change from oral to nasal air-paths or vice versa (e.g., [nd, bm, ŋk]). A heterorganic sequence is one in which the articulatory locations of the two consonants are different. The articulators can thus be manipulated relatively independently of each other (e.g., [pk, tk, fx, gv, qʃ]). A contiguous sequence is one in which adjacent parts of the same articulator are used (e.g., [pf, kj, ʃt, θs]). Normally the articulators used in homorganic and contiguous sequences cannot be manipulated independently.

The two kinds of clusters used in this study are homorganic and heterorganic sequences; the latter kind includes front-back and back-front clusters. Front-back clusters are ones in which the articulator of the second consonant is posterior to that of the first consonant: labial-coronal, coronal-dorsal, and labial-dorsal. Back-front clusters have the opposite order: coronal-labial, dorsal-coronal, and dorsal-labial.

Catford (1977) also identifies two possible types of transitions between consonants: "open" and "close". Catford made measurements on high-speed cine film and found that for [pp] in *cop part*, there was no lip movement between the two [p]s; for [p.p] in *cop apart* there was a momentary opening up of a central articulatory channel, with a total duration of 60 ms and reaching a maximum labial channel area of only 20 mm^2. From his measurements, he concluded that in an open transition, there is always a momentary break of articulatory continuity between the successive segments. In a close transition there is no such break. More specifically, there is absolute continuity of the articulatory stricture in homorganic clusters with a close transition. In examples such as *top part*, *that time*, *this sort*, *glove fair* the articulators retain the identical position throughout the cluster even when there is a change in phonation-type. In homorganic clusters with an open transition, the articulatory stricture has a momentary relaxation followed by a renewed tensing into the former position.

In heterorganic clusters with a close transition, the articulatory stricture for the second consonant is formed before the stricture for the first is released such as in [kp] in *back part*. In heterorganic clusters with an open transition, the articulators come together for the second consonant during the release of the first. To further support his claims, Catford also used "continuous palatography" to determine whether the tongue maintained or broke contact with the roof of the mouth in producing clusters. In this method, electrodes are inserted on a false palate, and tongue contact with the palate can be tracked. In the *back part* mentioned above*,* the second articulation of [p] is formed before the first of [k] is released, as shown in Figure 2.1 below.

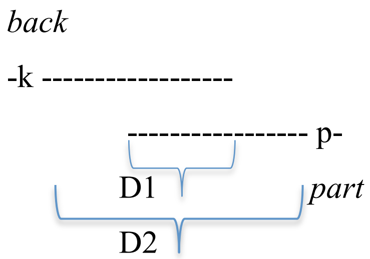


Figure 2.1 An example of a close transition

The results showed that the duration of the overlap in sequences such as [tp, tk, kt, lp, pl, kl] ranged from 29% to 45% of the combined duration of the two consonants (D1/D2*100 – see Figure 2.1). The average overlap was 33.75% in clusters with a close transition. This means that the articulations of the two stops co-occurred during about half of the duration of each of them, and "this amount of overlap seems to be normal" in English heterorganic clusters with close transitions (p. 222). In open transitions, actual contact is avoided so that there is a momentary gap between the two articulations (i.e., D1 ≤ 0, 0% overlap between the consonants).

More recent research findings are congruent with this overlap range and have suggested that English consonant clusters are measured as having approximately 20-60% overlap (see Barry, 1991; Byrd, 1996a; Zsiga, 1994). This is because in most English clusters, the closure for the first stop is not released until the closure for the second is formed; there is "a short period of overlapping or simultaneous closure involving the articulatory organs participating in the production of the stops" (Hardcastle & Roach, 1979, p. 531).

In cases where C1 in C1C2 clusters lacks an internal release (close transitions), it is usually perceived as incomplete, because either the burst is missing (or inaudible) or the closure is indistinguishable from that of C2, or both (see Ghosh & Narayanan, 2009). The burst indicates whether the stop is acoustically released or not (although note that even when perceptual detection fails, the acoustic spectrum can show a visible release of energy). Closure duration is defined as "the time interval between termination of the vowel-formant transition preceding the stop and onset of the transition to the following vowel" (Lisker, 1957, p. 43), which has also been used an index of stop reduction (Zsiga, 2003).

This perception of an incomplete consonant in close transitions is interpreted as either deletion or assimilation. Most previous works use "deletion" to refer to cases in which a segment is inaudible (see Avery & Rice, 1989; Ghosh & Narayanan, 2009; Guy, 1980; Raymond, Dautricourt, & Hume, 2006). For example, /t/ in *perfect memory* is not heard, and thus /pɚfəkt mɛmɚri/ becomes [pɚfək mɛmɚri]. Assimilation, on the other hand, refers to cases in which the place of articulation of the first consonant becomes more perceptually similar to the place of articulation of the second consonant (see Browman & Goldstein, 1989, 1990; Byrd, 1992; Catford, 1977). This is especially evident in clusters across words. For example, *seven plus* (/sɛvn plʌs/) is heard as *sevem plus* ([sɛvm plʌs]).

2.2.2. Phonological approaches to cluster realization

The theoretical framework assumed in the current research is Articulatory Phonology. This theory grew out of shortcomings of previous, feature-based theories of deletion and assimilation. For example, within Autosegmental Phonology (Goldsmith, 1976), two types of formal approaches have been proposed to account for phonetic reductions such as deletion and assimilation. One phonological account assumes "underspecification" of features, especially in the case of coronals, meaning that coronal stops are unspecified for place, whereas labial or velar stops are specified (Archangeli, 1988; Avery & Rice, 1989; Kiparsky, 1985; Pulleyblank, 1988). As a result, words containing coronal stops have a gap in their featural specification and are subject to assimilation and deletion. The second phonological account treats sound reductions as feature-deleting and feature-changing, corresponding to deletion and assimilation, respectively (Clements, 1985; Hayes, 1986; Lass, 1984). For instance, assimilation processes are represented as loss of an existing feature and spreading of an adjacent feature, an operation that permits a single feature to be linked simultaneously to more than one segment (Harris, 1994).

The advantage of analyzing assimilation in Autosegmental Phonology is that it explains processes such as assimilation, their operating contexts, and the output in a non-arbitrary relation.[3] However, a serious problem with an augosegmental account comes from the fact that, if assimilation were a true case of feature deleting/delinking, we would expect to find no trace of the deleted gesture. However, Browman and Goldstein (1989)

---

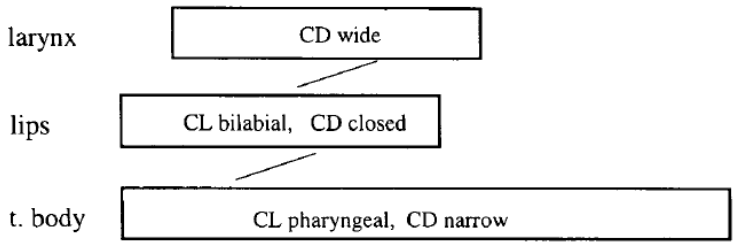
[3] This analysis does require the condition of locality though (Harris, 1994): the segments participating in the spreading operation must be adjacent to one another, which makes long-distance assimilation difficult to account for.

provide X-ray microbeam kinematic data which convincingly show that during the production of *perfect memory*, where deletion seems to occur, there ***is*** a tongue-tip gesture for /t/. This alveolar gesture is even produced with the same magnitude as when the word is produced in isolation. It cannot be heard simply because the gesture is completely overlapped by two other gestures, the preceding velar closure of /k/ and the labial closure of /m/. That is, the alveolar closure gesture is not deleted; all the gestures are present. It is only its temporal relationship to other gestures that has been altered, making it inaudible. Similarly, in the production of *seven plus* (/sɛvn plʌs/), the alveolar closure is still present, in this case with likely reduced magnitude, but it is hidden acoustically by the preceding labial fricative and the following labial stop. This result is in accordance with Catford's (1977) description of close transitions in heterorganic clusters, in which there is articulatory overlap.

To interpret these findings, that a delinked/deleted gesture is in fact produced, Browman and Goldstein proposed an alternative to standard feature based phonology: Articulatory Phonology. Their work led to a rich area of research which confirms and further explores C-to-C coarticulation through the use of various experimental techniques (Browman & Goldstein, 1995; Byrd, 1996a, 1996b; Byrd & Saltzman, 2003; Byrd & Tan, 1996; Davidson, 2003, 2010; Zsiga, 1994). These studies form the framework of Articulatory Phonology (AP).

Following the finding that segments are not deleted even though they are not auditory available, AP argues that a featural representation cannot capture the phonetic details of speech production. Within AP, the primitive units of phonological representations are gestures and their gestural parameters. Speech is thus decomposed into constriction gestures at key locations in the vocal tract: “the lips, tongue tip, tongue dorsum, velum, [and] glottis” (Goldstein, Byrd, & Saltzman, 2006, p. 217). More specifically, each gesture is defined along both spatial and temporal dimensions. The spatial dimensions include vocal tract variables, such as lips, tongue tips, or glottis, which are involved in determining the constriction location (CL) and constriction degree (CD). Under AP, CD includes five values: [closure], [critical], [narrow], [mid], and [wide]. For stops, the CD is [closure], whereas for fricatives, it is [critical]. Combining CL and CD, for example, a [g] would be produced with a tongue body constriction location of velar (CL: TB) and a tongue body constriction degree of closure (CD: TB). Lexical contrast is then defined through different combinations of gestures. For instance, the words *pack* and *tack* contrast with one another in that the former includes a lip gesture and the latter a tongue tip gesture (Goldstein & Fowler, 2003).

The representations reflecting temporal relations among gestures are called *gestural scores*. Figure 2.2 (Gafos, 2002, p. 277) shows the gestural score of the simple CV (consonant-vowel) utterance /p^{h}a/. The larynx box represents the laryngeal opening gesture corresponding to the aspiration (i.e., wide constriction degree); the lips box represents the bilabial gesture of the consonant /p/ (i.e., bilabial constriction location, closed constriction degree); the tongue body box represents the vocalic pharyngeal gesture of /a/ (i.e., pharyngeal constriction location, narrow constriction degree). The length of a box indicates the gestural duration, and the two lines connecting gesture boxes indicate that they are temporarily organized.



(CD=constriction degree; CL=constriction location)

Figure 2.2. Gestural score of /p^ha/

As the gestural score shows, gestures not only have intrinsic space and duration, they also can overlap with each other, since articulators move continuously over time. Increased overlap among gestures can result in different consequences: hiding or blending of gestures. When gestures are on different articulatory tiers (e.g., lip vs. tongue tip), each motion is independent and will be unaffected by the other concurrent gesture. With sufficient overlap, one gesture may completely mask or hide the other acoustically, rendering the latter inaudible. One example of hiding is illustrated through *perfect memory* as discussed previously (see Section 2.1.4): the gesture /t/ is produced with the same magnitude compared to when it is produced in isolation, only hidden by adjacent gestures.

When the same articulator is required for two consecutive gestures that may have different constriction locations, blending is observed to occur (see Recasens, Fontdevila, Pallarès, & Solanas, 1993; Romero, 1996). Because the two gestures are on the same articulatory tier (e.g., tongue tip vs. tongue tip), they are attempting to do different tasks with the identical articulatory structures. They cannot overlap without perturbing each other's tract variable motions (Browman & Goldstein, 1989). In this case, the dynamic parameters for the two overlapping gestures are blended, as shown in *ten themes*.

In the production of sequential stops, the implemented overlap degree varies (c.f. Section 2.2.1). A schematic view of different overlap relations is provided in Gafos (2002), which summarizes consonant cluster coordination into three possible relations. The three relations in Figure 2.3 below are taken from Gafos (2002, p. 271), where 'o' means onset, 't' target, 'cc' c-center, 'r' release, 'roff' r offset.

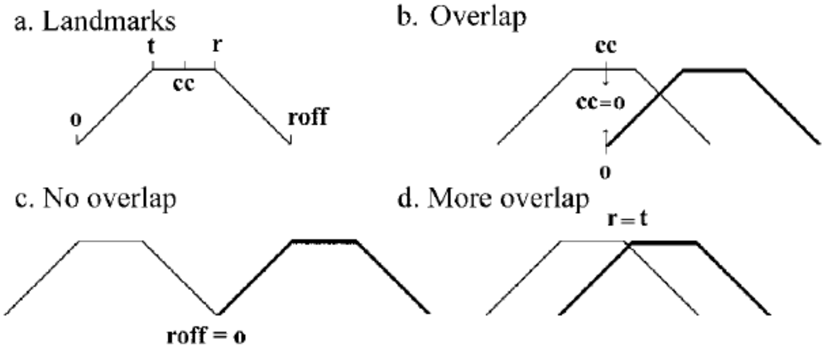


Figure 2.3 Three possible relations of gestural overlap

The coordination relationship in Figure 2.3 (b) indicates an open transition in the sense of Catford (1977): a cluster is produced with an intervening acoustic release (i.e., C^{ə}C). Such a coordination relation is shown in heterorganic sequences in Moroccan Colloquial Arabic (MCA) (see Gafos 2002). Figure 2.3 (c) indicates a relationship requiring the two consonantal gestures to be completely sequenced (i.e., no overlap). This is the relation employed in MCA homorganic clusters to avoid Obligatory Contour Principle (OCP, McCarthy, 1986). Figure 2.3 (d) shows a pattern where the articulatory release of the first gesture and the target of the second gesture occur at the same time, ‘r=t’. This consonant cluster coordination corresponds to Catford’s (1977) close transition.

English clusters, either in onset or coda, are bound in close transition (Figure 2.3d). English speakers “seldom if ever” produce an audible release between the two consonants in a cluster (Zsiga, 2003, p. 400). The close transition results in the lack of an acoustic release in sequences like *robbed*; the pronunciation is [bd] instead of [b$^{\text{ə}}$d] (see Hardcastle & Roach, 1979; Ladefoged, 2001). Even in clusters across words such as */seven plus/*, English speakers start producing the /p/ before the closure for the /n/ has been released. This great amount of overlap often causes the perception of place assimilation (Nolan, 1992; Zsiga, 2011).

Based on previous findings, I propose in this study that in general English speakers will most likely adopt the coordination relationship shown in Figure 2.3(d), while Mandarin speakers will adopt the relation in Figure 2.3(c) in their production of stop clusters across words. More detailed research questions and hypotheses related to this general prediction are outlined in Section 2.5.

2.3.What affects cluster coarticulation

There are several factors contributing to degree of overlap in cluster articulation. Some of these, like vowel context (Hardcastle & Roach, 1979; Lisker, 1999), following pause (Fasold, 1972; Labov, 1969), and gender differences (Neu, 1980; Wolfram, 1969), have shown conflicting results, and are therefore not included in this study. For instance, Hardcastle and Roach (1979) found an effect of vowel context on cluster realization, but Zsiga (2003) did not. Similarly, a pause following a cluster has been claimed to have a dialect-specific impact on cluster realization: it promotes deletion for New Yorkers (Fasold, 1972; Labov, 1969) but inhibits deletion for Philadelphians (Wolfram, 1969).

Other factors have been shown to have relatively systematic effects on gestural overlap; these are the ones that are considered here. The effects of stress on coarticulation were discussed in Section 2.1. The phonetic environment, mainly the place of articulation of C1 and C2, is discussed in Section 2.3.1. The speech rate effect is discussed in Section 2.3.2. Another important factor which has not yet been extensively studied – lexical frequency – is examined in Section 2.3.3. Section 2.3.4 provides a brief summary of the factors known to affect cluster realization.

2.3.1. Place of articulation

A number of factors have been identified to systematically determine the surface realization of English stop-stop (C1C2) sequences. The most consistent result involves place of articulation. In most sequences containing two adjacent English stops, either within words or across words (C1C2 or C1#C2), the closure for C1 is not released until the closure for C2 is formed (Catford, 1977; Hardcastle & Roach, 1979; Ladefoged, 2001). Hardcastle and Roach (1979) reported that in only 32 cases out of 272 was C1 released before onset of the second closure.

This release trend is mitigated by place of articulation of C1 and C2: acoustic releases are produced less often in back-front clusters (e.g., coronal-labial, dorsal-coronal, or dorsal-labial) than in front-back sequences (e.g., labial-coronal, coronal-dorsal, or labial-dorsal) (Byrd, 1996a; Davidson, 2011; Henderson & Repp, 1982; Zsiga, 1994, 2000). Although Ghosh and Narayanan (2009) found no effect of place of articulation on release percentage of C1, other studies seem to suggest otherwise. For example, the study by Henderson and Repp (1982) found that 58% of C1s in [-C1C2-] (e.g., *napkin*, *abdomen*) were released, but release percentages were strongly conditioned by the place of articulation of C2 in the sequence. An average of 16.5% of C1s followed by labial C2s were released, compared to 70% followed by alveolar C2s and 87.5% followed by velar C2s. That is, C1 was released more often when followed by a "back" consonant. Similar results were reported in Zsiga (2000), who discovered that word-final, phrase-medial C1s were released 20% when followed by a labial, 38% when followed by an alveolar, and 39% by a velar. She concluded that in English, "differences in place of articulation seem to account for release patterns in a straightforward way: clusters are more likely to have an audible release if C1 is further forward than C2" (p. 78).

In addition to affecting release patterns, this *place order effect* (POE) determines the degree of overlap, in a seemingly contradictory way (see p. 28 for further discussion); there is more overlap in front-back than in back-front clusters (Byrd, 1996a; Hardcastle & Roach, 1979; Henderson & Repp, 1982; Zsiga, 1994). The acoustic study of Hardcastle and Roach (1979) found that the interval between the onset of the C1 closure and the onset of the C2 closure was significantly shorter for /Vt#Kv/ compared to /Vk#Tv/ sequences, suggesting more overlap in /t#k/. Similarly, Barry (1991) found that [d] in a /d#g/ sequence was overlapped for 87% of its duration, but there was only 53% overlap for [g] in a [g#d] sequence.

Byrd's (1996a) articulatory study measured fricative-stop, stop-stop and stop-fricative sequences, [s#k], [#sk], [sk#], [d#g], [g#d], [gd#], [g#s], and [ks#]. Three indices were calculated to indicate temporal overlap: a) sequence overlap (the percentage of the total sequence duration during which contact occurred in front and back tongue regions), b) C1 overlap (the percentage of C1 duration during which contact for C2 also occurred), and c) C2 overlap (the percentage of C2 duration during which contact for C1 also occurred). The results showed that the sequence overlap and C1 overlap were both much greater for [d#g] than for [g#d] (59% vs. 46% and 87% vs. 53%, respectively). C2 also started much later relative to C1 for [g#d] than for [d#g], suggesting greater latency in back-front clusters. For most speakers, the consonants of [d#g] are "nearly completely overlapped" (p. 218), with contact for [g] often starting synchronously with that for [d]. In general then, the profiles showed tongue tip consonants were more overlapped by a following tongue body consonant than vice versa.

The POE (more releases and more overlap in front-back clusters than in sequences of reversed order) is also evident in other languages such as Georgian, Taiwanese, Korean, and Russian (Chitoran, Goldstein, & Byrd, 2002; Kochetov, Pouplier, & Son, 2007; Peng, 1996). For example, Chitoran et al. (2002) measured timing of the C2 movement onset within the constriction plateau interval of C1 in Georgian, using Electromagnetic Midsagittal Articulometer (EMA). Six combinations were tested in both word-initial and word-internal positions: /bg/, /p^ht^h/, /dg/, /gb/, /t^hb, and /gd/. Three points were measured: movement onset, target achievement (i.e., where constriction is achieved), and target release (where constriction is released). In their study, one speaker showed that in front-back clusters, C2 onset occurred on average after 3% of the C1 constriction interval, while C2 onset occurred after 87% of the interval in back-front clusters, suggesting more delay and thus less overlap in back-front clusters. The second speaker only exhibited a significant POE in word-initial position but not word-medially. Overall, their results pointed towards more gestural overlap in front-back sequences than in back-front sequences. Similar results are found in Kochetov et al. (2007). In their EMA study, degree of overlap was calculated as the time between the C1 release and the achievement of target for C2. Their study showed that the front-back cluster (/pt/) has greater overlap than back-front clusters (/kp/ and /kt/) in both Korean and Russian. Also, an electropalatographic (EPG) study by Peng (1996) considered eight stop-stop sequences in Taiwanese (/t#k, t#k^h, k#t, k#t^h, t#t, t#t^h, k#k, k#k^h/), for example, /pat^{53} taŋ^{55}/, “another year”, /lak^{53} kaŋ^{55}/ “ten days”, and /lak^{53} taŋ^{55}/, “ten years”. She found that “a larger proportion of the dental closing gesture was overlapped by the following gesture than the velar gesture” (p. 60), that is, coronal-dorsal sequence (front-back) was more overlapped than a dorsal-coronal one (back-front).

Before moving on, it is important to clarify the seemingly contradictory relationship between more releases and more overlap in front-back sequences than in back-front sequences. Indeed, the POE seems to affect releases and overlap in contradictory ways, since one would expect *more* releases to go with *less* overlap, and not vice versa, but front-back sequences are associated with *more* releases and *more* overlap. As shown in both Figure 2.1 and 2.2, overlap (i.e., the simultaneous closure in C1 and C2) should be avoided if there is a momentary gap (release) within a cluster (Catford, 1977; Gafos, 2002). This apparent contradiction is an effect of the metrics used to measure release percentage and degree of overlap. Two kinds of overlap measurements are summarized from the reviewed studies here. One is to use contact profiles and measure how early C2 occurs with respect to C1 offset (Byrd, 1996a; Chitoran et al. 2002; Kochetov et al., 2007); the other is to measure how much the closure duration is shortened (Hardcastle & Roach, 1979; Henderson & Repp, 1982; Zsiga, 1994) in consonants in clusters versus intervocalically. In both cases, it is possible to discuss and compare released clusters with varying degrees of overlap.

As to why the POE emerges across languages, different proposals are briefly reviewed here. Kochetov et al. (2007) interpreted the higher degree of overlap in /pt/ than in /kt, kp/ as resulting from the fact that /p/ and /t/ have different subsets of articulators; even when overlapped, [p] in /pt/ and [t] in /tk/ can be audibly released, because the C1 constrictions (the lips and the tongue tip at the alveolar ridge) are external to the C2 constriction (the dorsum a the velum for [k]), while the same cannot happen with [t] in /tp/ and [k] in /kp/. Hardcastle and Roach (1979) explained that the movement from [t] to [k] in /tk/ (e.g., front-back) involves the contraction of a single intrinsic tongue muscle to raise the back of the tongue while the movement from [k] to [t] requires the use of two muscles including the extrinsic genioglossus to reposition the tongue upwards and forwards. Byrd (1996a), however, contradicted this explanation by finding that the difference in time taken to contract one intrinsic muscle versus one extrinsic and one intrinsic muscle is likely to be vanishingly short. She pointed out that the assumption that using two muscles takes longer than using one is not valid.

More researchers have interpreted the POE from a perceptual account: the weak perceptual cues result from the rapid articulatory characteristics of coronals, given that C1s in most studies involve [t, d]. That is, a tongue tip gesture is more easily hidden by a following velar gesture than vice-versa (e.g., Byrd, 1992; Suprenant & Goldstein, 1998; Winitz, Scheib, & Reeds, 1972). Jun (2004) proposes that tongue tip gestures are rapid and thus “have rapid transition cues”, while tongue dorsum and lip gestures are “more sluggish and thus give rise to long transitions" (p. 10). Similarly, Chitoran et al. (2002) argued from a perceptual principle that a significant amount of overlap between C1 and C2 would impede recovering C1; with this in mind though, back-front clusters are expected to have less overlap because "C1 gestures are more easily hidden by C2 gestures" (p. 424). Silverman (1995, 1997) supports a perceptual explanation of the POE by positing that the phasing of articulatory gestures is governed by the need to recover the perceptual cues of consonants for the listener. Kochetov's (2001) study corroborates the perceptual principle as well. He reports that Russian listeners performed better at identifying the contrast between final /t/ and /tʲ/ before /k/, where the coronal had an audible release burst, than before homorganic /n/ or /s/ where there was no audible release burst.

Table 2.1 below summarizes a few important studies reviewed in this section. In general, previous research points towards a place order effect in native speech: more releases and more overlap allowed in front-back clusters than in back-front clusters (Byrd, 1996a; Chitoran et al., 2002; Hardcastle & Roach, 1979; Henderson & Repp, 1982; Kochetov et al., 2007; Zsiga, 1994). However, previous studies have examined only a subset of stop-stop sequences, such as labials-velars in Davidson (2011), labials-velars/coronals in Zsiga (1994), coronals-velars and velars-coronals in Hardcastle and Roach (1979), Barry (1991) and Byrd (1996a); no studies have yet examined all possible English stop combinations. It thus remains to be tested whether the POE would emerge across all stop-stop combinations.

Language	Stimuli	Measurement	Instruments	Research & Results
English	/t#k/ , /k#t/	C1 by C2 duration	EPG	Hardcastle & Roach (1979): POE
	/d#k/, /d#t/, /d#p/	Vowel transition	Acoustic	Zsiga (1994): POE; no speech rate effect
	/s#k/, /#sk/, /sk#/, /d#g/, /g#d/, /gd#/, /g#s/, and /ks#/.	C1 and C2 contact	EPG	Byrd (1996a) Byrd & Tan (1996): POE; speech rate effect
	/p#p, t, k/, /d#p, t,k/, /k#p, t, k/	Closure duration shortening	Acoustic	Zsiga (2000): POE
Georgian	/#bg(#)/, /#p^{h}t^h(#)/, /#dg(#)/, /#gb(#)/, /#t^{h}b(#)/, /#gd(#)/	How early C2 occurs in C1 offset	EMMA	Chitoran et al. (2002): POE
Russian Korean	/p#k/, /k#p/	How late C2 occurs after C1 offset	EMMA	Kochetov et al. (2007): POE; no speech rate effect
Taiwanese	/t#k, t#k^h, k#t, k#t^h, t#t, t#t^h, k#k, k#k^h/		EPG	Peng (1996): POE

Table 2.1. Summary of previous studies on the place order effect

2.3.2. Syllable and word position

One framework proposed to account for variability in cluster coordination in a principled way is Phase Windows (Byrd, 1996; Keating, 1990; Saltzman & Byrd, 2000). This framework is hypothesized to account for the variation in overlap due to factors such as stress, focus, speech rate, and phrasal boundaries. For instance, phase windows predict a larger variation in phrasal boundaries; gestures get longer, larger, and farther apart at phrase edges, resulting in less overlap. Based on a range of palatalization shown across words in English along with a limited amount of overlap shown in Russian, Zsiga (2000) incorporates the phase window framework into phonetic alignment constraints, and proposes a constraint having possible alignment locations in different clusters as well as in different languages.

Previous studies have suggested that English stop sequences have more timing variation and allow more overlap in codas than in onsets (Browman & Goldstein, 2000; Byrd, 1996a; Byrd & Choi, 2010; Nam & Saltzman, 2003). Also, speakers display a greater amount of overlap across word boundaries than within (Byrd, 1996a; Fougeron & Keating, 1997; Sproat & Fujimura, 1993; Zsiga, 2000). For instance, Byrd's (1996a) EPG study discussed above tested sequences containing stop-stop and stop-fricative sequences across word boundaries ([C1#C2]), in word-initial onsets ([#C1C2]), and in word-final codas ([C1C2#]). Stop-stop sequences in general were found to exhibit more overlap than stop-fricative sequences. Onset clusters (stops in particular) exhibited significantly less overlap and had longer durations than coda clusters or clusters across words. Specifically, the onset clusters had longer absolute and relative latencies, which were also longer in total duration. A series of EPG studies has also shown that consonants are generally produced with greater articulatory magnitude (measured as lingua-palatal contact) in domain-initial positions at each prosodic level than in domain-medial positions at the same level (see Cho & Jun, 2000; Fougeron, 2001; Fougeron & Keating, 1997).

However, more overlap across words than in onsets does not necessarily mean that the place order effect would be stronger in such positions. Davidson's (2011) study suggested that neither English word-final (CC#) nor across word clusters (C1#C2) showed the POE; only the word-internal clusters ([-CC-]) did. She posited that a coordination relationship across words is not stable enough to have a significant place order effect emerge. Rather, speakers were more likely to control the overlapped coordination to “maintain a more cohesive, stable pattern within a lexical item” (-CC-) than across words (p. 1055).

Along a similar vein, Zsiga (1995) reported that palatalization in cross-word positions is gradient and optional while lexical palatalization is categorical. Her EPG results demonstrated that palatalization, as in *miss you* ([mɪʃʊ]), is applied optionally unlike the word-internal process found in words such as *confession*. Specifically, English speakers showed a range of palatalization effects in cross-word clusters from full palatalization (=[ʃ]), to a token that is acoustically similar to [s] at the beginning of the production and closer to [ʃ] at the end (=[sʃ]), to an audible release of the coronal fricative (=[s#j]). Zsiga attributed the different extents of palatalization across words to variations in the amount of overlap of the critical alveolar gesture and the palatal gesture, supporting the idea that coarticulation is less extensive across word boundaries than within words. Comparably, Davidson and Roon (2008) hypothesized that both consonants in C1#C2 are longer than their counterparts in #C1C2 because of the intervening word boundary. Their acoustic analysis confirmed that both monolingual Russian and bilingual Russian-English speakers produced consistent durational differences differentiating C1#C2, #C1C2, and #C1əC2 in connected speech. They discovered that #C1əC2 had the longest duration, followed by C1#C2 sequences, then by #C1C2 sequences, as predicted. The results also found that C2 is the longest in C1#C2, followed by #C1əC2 and then by #C1C2.

The previous findings are likely due to boundary lengthening so that consonants at phrase edges are longer and larger than their non-edge counterparts (Beckman, Edwards, & Fletcher, 1992; Byrd, 2003; Byrd & Saltzman, 1998, 2003). The study by Byrd and Choi (2010) suggested that coda clusters showed less gestural overlap if the word was utterance-final than if it was phrase-final, or than if it was not located at a phrasal boundary at all. Evidence also shows that articulatory gestures spanning a phrase juncture are less overlapped than those phrase medially and that this overlap is further decreased for prosodically stronger phrase boundaries (Browman, 1995; Byrd, Kaun, Narayanan, & Saltzman, 2000; Cho, 2004; Hardcastle, 1985; Holst & Nolan, 1995).

In summary, existing analyses have reported conflicting results with respect to whether cross-word clusters are as significantly overlapped as word-internal clusters (Browman & Goldstein, 2000; Byrd, 1996a; Byrd & Choi, 2010; Davidson, 2011; Davidson & Roon, 2008; Nam & Saltzman, 2003; Zsiga, 1995). It thus remains to be seen whether the POE is strongly shown in cross-word positions.

2.3.3. Speech rate

Saltzman and Byrd (2000) proposed that a change in speech rate is implemented by modifying a coordination parameter. By shifting the coordination landmarks to be earlier in the time course of the first gestural organization in a sequence, more overlap is achieved. A large body of evidence has shown that changes in movement rate provide adequate motivation for changing the coordination dynamics (Byrd & Tan, 1996; Engstrand, 1988; Fosler-Lussier & Morgan, 1999; Gay, 1981; Jurafsky, Bell, Gregory, & Raymond, 2001; Kuehn & Moll, 1976; Munhall & Löfqvist, 1992; Tjaden & Weismer, 1998). The findings from these studies agree that greater fluency is associated with segment shortening (Gay, 1981), reduced velocity of the articulators (Adams, Weismer, & Kent, 1993; Gay, 1981), and reduction in the magnitude of articulatory gestures produced (Byrd & Tan, 1996; Shaiman, 2001). Munhall and Löfqvist (1992), for example, discovered glottal openings were blended into one single movement at fast rate when producing adjacent alveolar fricatives and alveolar stops, but blending did not occur at slow rate.

In a study to determine the effect of speech rate on overlap in clusters, Byrd and Tan (1996) discovered reduced lingua-palatal contact and shortened duration of C1#C2 in association with faster speech rates. Byrd and Tan considered four cross-word sequences (i.e., /s#g/, /g#s/, /g#d/, /d#g/). Five participating speakers were cued to speak each sentence in a “normal”, “medium”, “faster”, and “fastest” fashion successively. Byrd and Tan found that individual consonants were shortened in duration as speaking rate increased regardless of articulation position, place, or manner of the consonants. However, the results suggested that rate had only a minimal effect on [d#g] which “remains almost completely overlapped at all rates” (p. 276). They argued that this presumably was due to a ceiling effect, “whereby [d#g] is so overlapped even at slower rates that only a minimal additional increase due to faster rate is evidenced” (p. 270).

Other studies have reported inconsistent results. The acoustic study of Zsiga (1994) examined three sets of word pairs with stops across word boundaries (C1#C2), in which vowel environment was controlled (only front vowels were used): /d#p/, /d#t/, /d#k/. The results showed that the gestures for C2 began before the closure for the C1 was reached, as predicted. However, comparison of cluster overlap in slow and fast tokens of the same utterance was not indicative of a direct relationship. The interaction of consonant and rate was “rarely significant” (p. 54). Zsiga’s finding has been supported by other work. Kochetov et al. (2007) found slow and fast speech rates do not influence overlap in the production of stop clusters in either Korean or Russian. Similarly, Gao et al. (2011) observed no effect of faster production rates on C-V coarticulation in Taiwanese and in English. Dixit and Flege (1991) found no speech rate effect on the amount of lingua-palatal contact in producing Hindi [t].

The assumption that a speech rate change induces varying amounts of overlap between adjacent gestures has been further discouraged by a number of studies outlining substantial variability among speakers (Ellis & Hardcastle, 1999; Kuehn & Moll, 1976; Ostry & Munhall, 1985; Shaiman, Adams, & Kimelman, 1995). The study of Shaiman et al. (1995) showed that the timing between the jaw closing gesture of a vowel and the upper lip lowering gestures of a following bilabial stop got significantly shorter as speakers moved from normal to fast speech. However, they also found that the direction of the effect was not consistent across speakers. For three of the speakers, as the jaw cycle durations for the vowel decreased, the amount of overlap increased. For two other speakers, the onset of upper lip lowering for the bilabial gesture started earlier in the jaw cycle at the slower rates of speech. They concluded that how speech rate changes were implemented varied across individual speakers.

In all, it is quite possible that the acoustic consequences of overlap are controlled by individual speakers (Barry 1992; Jun, 1995; Lubker & Gay, 1982; Nolan, 1992; Tjaden & Weismer, 1998). Given the inconsistent results and sparse literature examining the speech rate effect, further research incorporating the effect of speech rate is needed.

2.3.4. Word frequency

Another variable that systematically affects coordination is word frequency. In connected speech, more frequent words have shorter durations, and a variety of other phonetically reduced characteristics such as reduced vowels, deleted codas, more tapping, palatalization, and reduced pitch ranges (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Bybee, 2000; Bybee & Hopper, 2001; Fosler-Lussier & Morgan, 1999; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Hooper, 1976; Jurafsky et al., 2001; Munson, 2007; Pluymaekers, Ernestus, & Baayen, 2005). These studies reached the consensus that more reductions occur for words and phrases that are more frequent. For example, Bybee (2000) found that the schwa in frequent words like *memory* is more likely to delete than the schwa in the nearly identical, but infrequent, *mammary*. More importantly, Bybee's data showed a statistically significant relationship of word frequency with rate of /t, d/ deletion. The most frequent word *told* had /d/ deleted in 68% of tokens while the least frequent *meant* never had the /t/ deleted.[4] Comparably, the results in Jurafsky et al. (2001) indicated that the two stops are deleted more often in high-frequency words than in low-frequency words and that, if they are not deleted, their durations are significantly shorter in high-frequency than in low-frequency words. Similar word frequency effects are also shown in the study of Pluymaekers et al. (2005), which found that high-frequency words are more reduced than low-frequency words in Dutch. This could be due to the fact that high-frequency words are more predictable and

---

[4] Note that Bybee's rate relied on the inability of the transcriber to hear the stop on a tape-recording. It is very likely that other cues (e.g., longer vowel duration in *meant*) give the target consonant away.

thus allow more variation without affecting listeners' ability to retrieve them in speech processing (Fosler-Lussier & Morgan, 1999; Gregory et al., 1999).

Previous studies on frequency effects have mostly focused on word final deletion. There is a surprising lack of research on how word frequency affects English cluster coordination that does not involve deletion (see Bybee, 2000; Bush, 2001; Yanagawa, 2006; also see the frequency effect on cluster assimilation in Australian English in Stephenson & Harrington, 2002, and in German in Jaeger & Hoole, 2011). A notable exception is the study of Bush (2001), which showed the palatalization of /tj/ and /dj/ sequences in English conversation occurs only between pairs of words that occur together most frequently in the corpus studied (e.g., *did you*, *don't you*, *would you*, *that you*, *told you, last year*). One question that needs addressing is thus: does frequency affect degree of overlap in consonant sequences, and if so, how? More specifically, one can ask to what extent sound patterns like the POE are evident in low-frequency versus high-frequency words.

One way to answer these kinds of questions is to test speakers' production of real words (ranging from high- to low-frequency) and of nonwords (on the extreme end of low-frequency). This idea has its root in studies that claim that speakers internalize their knowledge of sound patterns and are able to apply this knowledge to new forms (Berko, 1958; Ernestus & Baayen, 2003; Pater & Tessier, 2006; Pycha, Nowak, Shin, & Shosted, 2003; Wilson, 2003, 2006). For example, an early study by Berko (1958) showed that English preschool and first grade children could extend their knowledge of how plural –s is pronounced (*dog[z]* vs. *cat[s]*) and apply this knowledge to form new plurals in wug-words requiring [-s] or [-z]. However, studies have suggested that the application of phonological knowledge may be incomplete (Fleischhacker, 2005; Zhang & Lai, 2010; Zuraw, 2007). The study conducted by Zhang and Lai (2010) examined how native Mandarin speakers produced non-words with two sandhi patterns, one with a strong phonetic motivation (Half Tone 3 sandhi, 214→21), the other with a more opaque pattern (Tone 3 Sandhi, 214→35). Their results showed that participants were able to apply both patterns to nonwords, but applied the pattern induced by the stronger motivation more accurately.

In this study, I will use both real words and nonwords to assess the role of frequency in cluster realization. Specifically, two aspects of lexical frequency are considered: high-frequency words versus low-frequency items in real words (e.g., *take care* vs. *dock gate*), and real words (i.e., high-frequency) versus nonwords (i.e., extremely low-frequency) (e.g., *keep pace* vs. *peep pate*, more in Section 4.4). In doing so, we can make inferences about the ability of speakers to generalize from familiar forms to less familiar forms, and to novel forms.

Although a few studies have examined how lexical frequency affects English cluster coordination (see Bybee, 2000; Bush, 2001; Yanagawa, 2006), it remains unclear how frequency (ranging from high- to low- to extremely low-frequency) affects cluster coordination. Also, previous studies have not examined whether and how the frequency effect might interact with others that have been identified to influence gestural overlap (e.g., place of articulation). Moreover, we know little about whether and how lexical frequency would facilitate specific gestural overlapping patterns (e.g., the POE).

2.3.5. Summary

Existing analyses reviewed in Section 2.3 have identified four factors that systematically influence gestural overlap between English consonantal members in native speech: place of articulation (POE, in particular), syllable positions, speech rate, and word frequency. Together with the stress effect discussed in Section 2.1, these five factors are the main influences that seem to determine the overlap degree in English clusters. No studies that I am aware of have examined these factors concurrently. In addition, previous studies have considered only a subset of stop-stop combinations. Therefore, research examining all of these factors in a research paradigm that includes all possible English stop-stop sequences is much needed, to get a clear picture of how these factors interact in predicting cluster realization in L1 speech. This study examines clusters in cross-word positions with the other four factors manipulated to test each of the effects along with their interactions.

2.4. Non-native production of English clusters

Having discussed the factors that influence the production of stop-stop sequence in L1 speech, I now turn to a discussion of the literature on L2 production, including accounts that attempt to explain L2 speech errors (Section 2.4.1- 2.4.2). I will focus on cluster coordination produced by Mandarin learners of English in Section 2.4.3.

2.4.1. Repair types

Due to the extensive overlap in English cluster coordination, most L2 learners encounter difficulty in perceiving and producing lexical items in connected speech where clusters are present (e.g., Brown & Hilferty, 1986; Henrichsen, 1984; Ito, 2006; Kim, 1995; Kweon, 2000; Rogerson, 2006). In terms of perception, problems arise for L2 listeners because they think that native speakers talk too fast (Gilbert, 1995). In terms of production, L2 speakers sound “choppy” (Celce-Murcia et al. 1996, p. 158), “disconnected and unnatural” (Brown & Kondo-Brown, 2006, p. 5), and “overly careful and stilted” (Zsiga, 2003, p. 401) because they fail to produce appropriate degrees of overlap. The absence of coarticulation at word boundaries among L2 speakers results in a failure to link words together into hierarchical prosodic groupings. These groupings, however, are what native speakers rely on to analyze discourse information (Zsiga, 2003). The following subsections present previous works on English cluster coordination in L2 speech, and discuss several approaches that attempt to account for L2 deviations in their production.

In general, L2 learners often coordinate adjacent gestures in an inappropriate way such that an insufficient overlap pattern arises (e.g., Barlow, 2005; Broselow & Finer, 1991; Colantoni & Steele, 2008; Davidson, 2003, 2006; Major, 1987; Messing, 2008; Solé, 1997; Zsiga, 2003, 2011). The study of Yanagawa (2006) found that even when the native language (i.e., Cantonese) had more overlap, Cantonese speakers produced C1#C2 in English with much less extensive degree. The most salient consequence of this insufficient overlap is schwa insertion in L2 production of sequential consonants when the target language would typically not have a schwa (e.g., Davidson, 2006; Major, 1987; Messing, 2008; Sperbeck, 2000; Tarone, 1987). For example, Tarone (1987) reported that Korean speakers repair English sequences by inserting a schwa (e.g., class → [kəlæs]). Barlow (2005) found Spanish ESL speakers inserted a vowel before word initial /sC/ clusters. Major (1987) found speakers of Brazilian Portuguese (BP) tended to epenthesize

[i] between the two consonants in English final clusters. Inserting schwa to break up consonant clusters is by no means limited to L2 acquisition of English. Davidson's (2006) study examined English speakers' production of #fricative+obstruent and #fricative-nasal clusters in a Slavic language. Subjects were found to have better performance on /f/-initial sequences than /z/-initial sequences. Overall though, L2 learners (in this case, English speakers) were more likely to insert a schwa into Slavic clusters compared to native speakers.

Another acoustic consequence brought about by incorrect gestural coordination is inappropriate assimilation (Cebrian, 2000; Solé, 1997; Weinberger, 1994). A study by Solé (1997) examined the production of English consonant clusters whose members differed in voicing (e.g., [sn-] in *snail*). Simultaneous EPG, electroglottographic (EGG), and acoustic measurements were used to examine the production of two native English speakers and four Catalan-English learners. The Catalan subjects produced one single glottal gesture for two segments (e.g., [sn-] becomes [zn-]) in their production of both English and Catalan words, while this anticipatory effect was not found in the native English speakers. Solé interpreted this result as L1 transfer. This is because in Catalan clusters, C1 and C2 agree in voicing, and take on the voicing of C2. Solé argued that the observed anticipatory effect in the Catalan subjects was derived from difficulty in changing the fossilized articulatory timing habits that govern anticipatory voicing in their L1. Similarly, Rubach (1984) found that Polish subjects also showed an effect of the L1 rule of voicing assimilation in consonant clusters. In Polish, obstruent clusters, both within words and across word boundaries, must agree in voicing. The interference of this rule was evident in examples such as the pronunciation of [z] instead of [s] in *this boy* and *misgiving*. Rubach characterized the types of rules that are likely to transfer as low-level, automatic, context-sensitive rules such as Polish regressive voicing assimilation (also see Hungarian regressive voicing assimilation transfer in Altenberg & Vago, 1983).

However, the cross-word anticipatory voicing in L1 transfer to L2 speech found in Rubach (1984) was not supported in Cebrian (2000). Cebrian found Catalan subjects failed to apply L1 voicing assimilation at word boundaries in English. He discovered that target final obstruents in word pairs such as *Swiss girl* or *smooth ice* did not undergo the L1 rule of anticipatory voicing assimilation. He argued that the low productivity of the L1 voicing rules in L2 speech may result from the prosodic characteristics of the test items. Specifically, because each word in the pair constituted a phonological word, he argued that "the interlanguage constraint blocks the application of regressive voicing assimilation across a word boundary" (p. 19). Based on this finding, Cebrian proposed the Word Integrity Principle, positing a tendency to preserve segments within one word, which implied a failure to coarticulate segments across word boundaries.

2.4.2. Accounts that explain L2 production

In fact, earlier than Cebrian (2000), Weinberger (1994) proposed the Recoverability Principle based on Chinese learners' production of English clusters: L2 speakers prefer to maintain contrastive information even if this means more articulatory effort (e.g., no cross-word coarticulation), in contrast with Rubach (1984). Both the Word Integrity and Recoverability Principles predict that gestures across words produced by L2 speakers would drift away, resulting in less overlap.

This hypothesis is confirmed by Zsiga (2003), who found that both English-Russian learners and Russian-English learners preferred little overlap at word boundaries in producing L2 clusters. Moreover, in her study, the carryover of L1 is shown to be strong as well. Both English and Russian speakers were found to exemplify their native articulation patterns. English speakers produced English consonants with approximately 20% closure duration overlap for all three cluster types (homorganic, back-front, and front-back) at word boundaries, and the sequences rarely had an audible internal release (17.1%). In contrast, Russian speakers had an audible internal release more often when speaking Russian (53.6%). In agreement with the L1 patterns, English speakers produced Russian clusters with fewer C1 releases and greater overlap than did Russian speakers, carrying over from the English patterns; meanwhile, Russian speakers produced English clusters with more C1 releases and smaller overlap, comparable to their articulatory habits in Russian. Zsiga's (2003) finding is in line with previous studies discussing L1 transfer that gives rises to L2 mistiming alignment (Broselow, 1987; Chen et al., 1996; Major, 1987; Solé, 1997; Tajima et al., 1997).

Aside from the Word Integrity and Recoverability Principles, a number of studies have relied on markedness theory to account for L2 speech errors, particularly in three aspects: voicing, cluster size, and sonority sequencing. For example, Japanese, Korean, and Mandarin speakers are found to be more accurate in producing voiceless than voiced English clusters (Broselow & Finer 1991; Broselow, Chen, & Wang, 1998; Major & Faudree, 1996; Hansen, 2001). Studies also have shown that L2 learners acquire onset clusters before coda clusters and acquire smaller size clusters before longer ones (Anderson, 1987; Carlisle, 1998; Hansen, 2001). The study by Anderson (1987) found both Arabic and Chinese speakers' accuracy decreased as final clusters increased from two to three consonants.

Moreover, the markedness relationship is characterized in terms of sonority sequencing when acquiring English clusters (e.g., Broselow & Finer, 1991; Eckman & Iverson, 1993; Cardoso, 2008; Carlisle, 2006). The sonority scale is defined as an abstract phonological property of segments that correlates with acoustic intensity (Clements, 2005; Parker, 2008; Wilson, Davidson, & Martin, 2014). Specifically, glides are the most sonorous, followed by liquids, nasals, fricatives, and stops. The smaller the distance between C1 and C2, the more likely the cluster is to undergo repair (see Berent, Steriade, Lennertz, & Vaknin, 2007). Spanish learners of English in Carlisle (2006) were found to modify /#sn/ more than /#sl/ clusters. This is because /#sn/ is more marked than /#sl/ according to sonority distance. However, the extent to which sonority sequencing can account for L2 consonant cluster processing has been called into question. For example, Davidson, Jusczyk and Smolensky (2004) reported that English speakers have varying accuracy when producing Polish-legal /zm/ and /vn/ clusters (63% vs. 11%), although they have the same sonority distance. In a follow up study, Wilson et al. (2014) discovered that other phonetic details (e.g., voicing) can be equally important for predicting L2 production patterns.

To summarize, the reviewed studies on L2 gestural mistiming of English clusters suggest that a purely articulatorily motivated account is not sufficient. Rather, theories of word integrity, L1 influence, and markedness are all active in explaining L2 cluster production (see Davidson, 2011 for a review). Given the conflicting results found in L2 studies, as well as that the limited number of languages that have been studied, it is important to extend our cross-linguistic analysis. In the next subsection, I will review studies on Mandarin learners' production of English clusters.

2.4.3. Mandarin ESL speakers' production of English clusters

Due to the fact that consonant clusters do not exist in Mandarin, Mandarin speakers have substantial difficulty in acquiring English clusters (see Broselow & Finer, 1991; Broselow et al., 1998; Weinberger, 1987). As mentioned above, the learning difficulty increases as the cluster size increases (Anderson, 1987; Hansen, 2001). For instance, the study of Anderson (1987) showed that Mandarin speakers had significantly more modification in three-member codas than in one-member codas. Also, Mandarin speakers had more difficulties in producing longer consonant clusters in coda position than in onset position.

Another factor affecting learning difficulty is the nature of the consonants in the clusters, specifically voicing. The study of Hansen (2001) found that Mandarin learners had more difficulty acquiring voiced consonants than voiceless ones. The codas with which the participants had the most difficulty, such as /vz/, /vd/, /rd/, /gz/, /rz/, /nt/, /lv/, usually consisted of voiced fricatives and voiced stops. This finding is in line with previous studies, which have shown a devoicing strategy exhibited by Mandarin speakers (e.g., Broselow & Finer, 1991; Broselow et al., 1998). For instance, Weinberger (1987) discovered that devoicing accounts for 66% of modifications in word-final codas produced by Mandarin speakers, followed by deletion (10.2%), and then epenthesis (9.9%). The latter two modifications (deletion and epenthesis) are common strategies to break up consonant clusters in L2 speech, and likely result from Mandarin speakers' preference for bisyllabicity carrying over from their native language (Broselow et al., 1998; Heyer, 1986; Wang, 1995).

A third difficulty among Mandarin learners of English involves timing. Given the timing differences between Mandarin and English, a number of studies have examined Mandarin ESL learners' cluster coarticulation from the timing perspective (e.g., Chen, 2005; Chen et al., 1996; Chen & Chung, 2008; Tajima et al., 1997). They argue that Chinese speakers transfer timing of their L1 in the production of English, resulting in a stream of syllables with relatively equal prominence (also see Section 2.1.2).

A more recent study by Chen and Chung (2008) examined timing patterns of Taiwanese speakers. Three groups of Taiwanese-English learners with different proficiency levels and one control group of American learners read one paragraph in their study. Four acoustic correlates of timing were measured: syllable duration, vowel reduction, C-to-V duration, and C-to-C duration. The results showed that vowel reduction and C-to-V duration were two important indicators for perceiving accent. Consonant deletion was found to be the dominant strategy in reducing clusters for the Taiwanese speakers while American English speakers had higher frequency of reduced syllables with a schwa. The Taiwanese learners also showed a significant deviation from American English speakers in terms of syllable duration and syllable-to-syllable variation. For example, they had much longer consonant-vowel linking duration than that of American speakers (also see Fan, 2011). Moreover, the lower the English proficiency of the learners, the longer the duration of the consonant clusters they produced.

Overall, previous studies have suggested that Mandarin speakers adopt different strategies to break up consonant clusters (e.g., deletion and schwa insertion), and that voiced consonants are particularly difficult to acquire (see Anderson, 1987; Bayley, 1996; Eckman, 1981; Fan, 2011; Hansen, 2001; Wang, 1995; Weinberger, 1987). Given that most studies investigate coda production and only a few studies have dealt with Mandarin speakers' cross-word coarticulation (e.g., Chen & Chung, 2008; Messing, 2008; Tajima et al., 1997), it remains poorly understood how coordination is implemented by the community of Mandarin ESL speakers. Moreover, as discussed in Section 2.1 and 2.3, four factors have been identified to systematically determine overlap degree in consonant clusters in L1 speech: place of articulation, lexical frequency, stress, and speech rate; no studies have yet examined their effects in L2 speech.

2.5.Research hypotheses

As reviewed earlier in this chapter, previous research has identified four factors that determined overlap degree in English clusters coordination: place of articulation, lexical frequency, stress, and speech rate. However, no studies to my knowledge have combined these factors to examine their effects in both L1 and L2 speech. This gap is what essentially drives this research. Also, as outlined in Section 2.2, English and Mandarin differ in a) rhythmic structure (i.e., stressed timed vs. syllable timed) and b) cluster organization (extensive overlap in clusters vs. no clusters categorically). Therefore, comparing native English speakers and Mandarin ESL speakers will provide a much needed window into how these factors play out not only in L1 speech, but also in L2 speech.

Abstracting away from specific influences of place of articulation, frequency, stress, and speech rate, and based on previous works such as Hansen (2001), Messing

(2008), and Wang (1995), Mandarin speakers are hypothesized, in general, to have more internal releases and have less overlap than English speakers.

In terms of specific effects, the following research questions arise. The first research question asks: ***Will place of articulation (i.e., homorganic clusters, front-back clusters, and back-front clusters) have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?*** Specifically, previous research on native English speech suggests very few internal releases in homorganic clusters, and a place order effect (more releases and more overlap in front-back than in back-front clusters; POE) in heterorganic clusters (Byrd, 1996a; Catford, 1977; Henderson & Repp, 1982; Ladefoged, 2001; Zsiga, 1994, 2000). Here are three hypotheses covering the overlap patterns in homorganic clusters (Hypothesis 1) and in heterogenic clusters that vary in place of articulation across the two consonants (Hypothesis 2 & 3):

Hypothesis 1: English speakers will have fewer releases and more overlap than Mandarin speakers in producing homorganic clusters.

Hypothesis 2: English speakers will show the predicted release pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.

Hypothesis 3: English speakers will show the predicted temporal overlap pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.

The second question asks: ***Will lexical frequency have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?*** Two sub-questions are relevant here: a) Will both groups of speakers coordinate clusters in high-frequency words the same way as clusters in low-frequency words (see Bush, 2001; Bybee, 2000; Jurafsky et al., 2000)? b) Will both groups of speakers coordinate clusters in real words the same way as clusters in nonwords (see Berko, 1958; Ernestus & Baayen, 2003; Pater & Tessier, 2006; Pycha et al., 2003; Wilson, 2003, 2006)?

Three hypotheses deriving from research question two are raised: Hypothesis 4 relates to the frequency effect proper (sub-question a) above); Hypotheses 5 and 6 relate to the ability of transferring knowledge of sound patterns from known (i.e., real words) forms to novel forms (i.e., nonwords) (sub-question b) above).

Hypothesis 4: Both English and Mandarin speakers will have fewer releases and more overlap in meaningful high-frequency words than in meaningful low-frequency words.

Hypothesis 5: English speakers will exhibit the release patterns of the POE in both real words and nonwords while Mandarin speakers will not.

Hypothesis 6: English speakers will exhibit the temporal overlap pattern of the POE in both real words and nonwords while Mandarins speakers will not.

The third research question concerns the different stress patterns in English and Mandarin. Previous research seems to suggest that Mandarin speakers can acquire native-like durational patterns in producing English stress (Chen et al., 2001; Zhang et al., 2006), while other works find that Mandarin speakers transfer their L1 timing habits when producing English sentences (e.g., Chen & Chung, 2008; Tajima et al. 1997). Given these inconsistent results, the third research question asks: ***Will stress patterns have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?***

Hypothesis 7: English speakers will release less often in words when they are lexically and sententially unstressed than when they are stressed, while Mandarin speakers will not exhibit such a difference.

Hypothesis 8: English speakers will produce more gestural overlap in words when they are lexically and sententially unstressed than when they are stressed, while Mandarin speakers will not exhibit significant temporal differences.

Finally, the effects of speech rate on gestural overlap are inconsistent as seen from the literature (Byrd & Tan, 1996; Kochetov et al., 2007; Kuehn & Moll, 1976; Munhall & Löfqvist, 1992; Tjaden & Weismer, 1998; Zsiga, 1994). The fourth research question asks: ***Will speech rate have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?***

Hypothesis 9: English and Mandarin speakers will have more gestural overlap when speech rate is increased.

Chapter 3

Experiment

In order to examine how the four factors (place of articulation, lexical frequency, stress, and speech rate) determine overlap degree in English stop clusters spanning words, two studies were designed in which English and Mandarin speakers produced a set of clusters in slow and fast speech rates. In this chapter, I describe the methodology employed in the current research, which was designed to test the hypotheses outlined in the previous chapter.

This chapter begins with the stimuli design (Section 3.1). Section 3.2 provides detailed descriptions of the participants and the experimental tasks. This chapter also discusses data analysis including acoustic and statistical analyses (Section 3.3), in which I lay out the independent and dependent variables for data collected from each study. This chapter concludes with a summary (Section 3.4).

3.1.Stimuli

Numerous studies have investigated the overlap degree of consonantal gestures in L2 speech (Davidson, 2010; Fan, 2011; Messing, 2008; Zsiga, 2003, 2010). A few of these also looked at Mandarin speakers (e.g., Fan, 2011; Messing, 2008), but no previous studies have examined place of articulation and lexical frequency concurrently to investigate overlap degree in L1 and L2 speech. One reason for the lack of such studies may be due to the methodological difficulty of controlling vowel differences between familiar and unfamiliar items. In order to avoid problems associated with inconsistent vowel environments in C1#C2 between familiar and unfamiliar items, words were deliberately designed and chosen for this study.

Prior to the primary study, a pilot study was conducted using different testing materials. In the pilot study, I used an open mid vowel in all nonwords, [ɔ]. For example, the nonsense sequence *kop pos* was used to correspond to the real sequence *keep pace*. This design had one major shortcoming: because the vowel environments differed between the two lexical settings, it is possible that cluster overlap would be affected by the effect of vowel context.

I then designed another corpus in the primary study. The vowel contexts between real words and nonwords were controlled to be **as similar as possible**. For example, the nonword *peep pate* [pip pet] was used to correspond to the real word *keep pace* [kip pes]. The important criteria for building the corpus were: 1) order of place of articulation (i.e., homorganic, front-back, and back-front clusters); 2) lexical frequency (i.e., high vs. low frequency in real words, and real words vs. nonwords); 3) stress positions (i.e., lexical stress and sentential focus); 4) speech rate (reading phrases vs. spontaneous speech). Each factor corresponds to each of the four research questions listed in Chapter 2. Also, because I employed closure duration ratio to infer temporal overlap (more details in Section 3.4), the stimuli in this research included words with single codas and single onsets to establish a baseline closure duration of each stop.

Three lists were designed based on these criteria. Specifically, List 1 contained 24 items, yielding V#C and C#V contexts (see Table 3.1). When C1 was an onset, it was embedded as *saw#C1*. When C1 was a coda, it was embedded as *C1#art* (see Zsiga, 2003). The C1 and C2 in the 24 items were referred as $C1_b$ and $C2_b$ ($_b$ for “baseline”) where no consonant-consonant coarticulation occurs. Differing from the design in Zsiga (2003), each stop appeared as onset and coda ***twice*** to calculate the baseline closure duration in the current research. I believe this could yield a more reliable duration value for each speaker. For instance, the baseline closure duration of coda /g/ was averaged over *leg art* and *big art*; similarly, onset /g/ was averaged over *saw good* and *saw game*. Words containing the target stop in List 1 were randomly chosen from List 2.

Secondly, List 2 included 144 disyllabic items, all in a VC1#C2V environment (see Table 3.2). All possible stop combinations were included to examine Research Question (RQ) 1, which yielded 36 tokens (e.g., p#p, d#k, k#b, etc). To examine the lexical stress effect, that people might release C1 more often in compounds (initial stress) than in non-compounds (final stress), 36 compounds and 36 non-compounds were designed with same C1#C2 pair. For example, *soup pot* and *keep pace* have the same p#p context; *soup pot* is a compound and *keep pace* is not. To further control for lexical stress, all test items were single syllables (with one exception of *diet*). The 72 phrases (compounds and non-compounds) were all meaningful common phrases, which differed in frequency (e.g., *take care* vs. *dock gate*). These 72 items varying in frequency allowed us to address research question 2a. Another 72 phrases were designed as corresponding nonwords. By corresponding, I mean two conditions were met: 1) their surrounding environments were controlled with the same vowels (e.g., nonsense *peep pate* versus meaningful *keep pace*), and 2) their initial consonants and codas agreed in voicing and manner, with the aim of creating consistent segmental contexts. For example, /d/ in *dak kit* agrees with /b/ in *back kick* in terms of voicing (i.e., voiced) and manner (i.e., stops).

These 72 nonwords, along with 72 real words, allow us to examine RQ 2b. The following two tables provide the complete list for experimental stimuli used in the primary study.

Single Coda $C1_b$#V	Single Onset V#$C2_b$
keep art, pep art, job art, rib art, hot art, shut art, look art, rock art, food art, red art, leg art, big art	saw pain, saw pen, saw balm, saw boy, saw tie, saw take, saw care, saw car, saw day, saw down, saw good, saw game

Table 3.1 Testing material: 24 items of /C1#V/ and /V#C2/

		Nonsense compound C1#C2	Meaningful compound C1#C2	Meaningful non-compound C1#C2	Nonsense non-compound C1#C2
Homorganic clusters	p#p	foop pok	soup pot	keep pace	peep pate
	p#b	mip bong	lip balm	deep bite	beep bipe
	t#t	shot tud	hot tub	get tip	det tik
	t#d	zet dep	vet debt	beat down	deat doun
	k#k	dak kit	back kick	take care	pake kare
	k#g	bek gaep	deck gate	look good	mook goob
	b#p	lib peim	rib pain	grab pen	brab pem
	b#b	zob banp	job bank	rub back	mub bap
	d#t	fade tenk	shade tent	need time	meed tine
	d#d	nud dei	mud day	bad day	gad dei
	g#k	grag kar	drag car	smug king	swug kim
	g#g	brug gam	drug gang	big game	dig gane
Front-Back clusters	p#t	kep tot	pep talk	skip test	stip tesp
	p#d	rap dant	lap dance	deep down	beep doun
	p#k	mip kare	lip care	keep calm	teep kalm

	p#g	pape gum	tape gun	top gear	pop geal
	t#g	biet gaib	diet guide	great guy	breat gai
	t#k	doot kamk	boot camp	fat cat	shat kap
	b#t	rab tesp	lab test	glib talk	blib tot
	b#d	prab dis	crab dish	rub down	lub doun
	b#k	rab coak	lab coat	grab cash	brab kaf
	b#g	plub girm	club girl	fab gift	sab gisp
	d#g	fead gane	head game	good game	bood gane
	d#k	sood kork	food court	bad kid	gad kig
Back-front clusters	t#p	fuit panf	suit pants	bright pink	grait pinp
	t#b	srat boi	frat boy	get back	det bap
	k#p	feek paim	cheek pain	look pale	rook peil
	k#t	mek tai	neck tie	black tea	glack tee
	k#b	lock banb	rock band	quick boost	kwik booft
	k#d	bok dape	dock date	kick down	tik doun
	d#p	fead paim	head pain	red pen	led pem
	d#b	ded bud	bed bug	need beer	meed beel
	g#t	gog tael	dog tail	big time	dig tine
	g#d	dag dear	bag deal	big deal	kig deer
	g#p	glog peiz	blog page	hug pug	sug puz
	g#b	reg bone	leg bone	big boy	dig boi

Table 3.2 144 items of /C1#C2/

3.2.Participants

This research project employed two participant groups: one group of 15 native English speakers (NE) and another group of 25 native Mandarin speakers (NM) who use English as a second language (ESL), for a total of 40 participants. For each group, a questionnaire was administered to elicit background demographic information prior to the experiment (see Appendix B and Appendix C). For example, the NM group provided information such as age, gender, total years of English instruction, amount of ESL exposure, knowledge of additional languages, and standardized English test score, etc

(See Appendix D).[5] To ensure within-group homogeneity within the NM group, I selected participants who had intermediate and above ESL proficiency. The criteria were determined by their reported standardized English test scores. Mandarin volunteers were selected only if they scored at/above 90/120 of iBT (an Internet-Based Test of English as a Foreign Language), 550/677 of PBT (a Paper-Based TOEFL), or 6.5/9 of IELTS (International English Language Testing System). There were 14 females and 11 males in the NM group, with an average age of 27.4 years old (ranging from 23 to 40 years old). All Mandarin speakers were students at the University of Victoria, Canada.

The NE group consisted of 15 native English speakers. To control for regional dialect influence, only speakers that were born and raised in British Columbia, Canada, were selected in this study. There were nine females and six males in this group with an average age of 26.3 years old (ranging from 19 to 40 years old). Seven of them reported they had learned French, five had learned Spanish, one Russian, one German, and one Japanese; none of them had reported learning Mandarin.

Participation in this study was completely voluntary. All participants reported no hearing or speech problems. Participants were told that they were taking part in a speech study comparing native versus nonnative production but were given no other details until after the recording session was complete.

---

[5] No Cantonese speakers were used in this study but there were seven Mandarin speakers from Wu dialect areas, which allow glottal stop codas. Although it was unclear whether/how dialects played a role in determining surface forms of stop-stop clusters in this study, it is worth considering in future research.

3.3.Procedure

3.3.1. Pre-test

Prior to the experiment, each participant read and signed a consent form. After providing their signature, the participants received verbal instructions in their L1 (English or Mandarin) on how to complete the tasks. In addition to the self-reported TOEFL/IELTS scores, a **pre-test** was conducted to better assess Mandarin participants' (NM) English proficiency. They were asked to speak about a topic for two to three minutes; the question asked was "What do you feel the most passionate about?", I adopted this topic from Lin and Wang (2008). The accentedness of the monologue was judged by two native English speakers (one participated in the primary study; the other did not) using a Likert scale (see Brill, 2008) from one to seven, where one indicates native-like and seven indicates heavy accent.

Table 3.3 below provides the accent score for each participant from each judge. The table shows that the mean accented score is 3.38 for Judge 1 and 3.06 for Judge 2, 3.22 across both judges. This result suggests that Mandarin speakers in this study were perceived to be moderately accented (3.5 represents the middle point on the scale). No participants received scores higher than 6 (i.e., extremely heavy accent). The Pearson correlation test showed that two judges scores were significantly correlated with one another [$r(15)=.583$, $p<.002$]. Considering the possibility that two judges' scores showed little agreement even though they were correlated, an Intraclass Correlation Coefficient was conducted. The results showed that inter-rater reliability was again confirmed [Cronbach's Alpha=.73, $p<.001$]. Together with their standardized English scores, I am confident that the group homogeneity was validated, and that the NM group in this study represented an ESL group of intermediate to advanced proficiency.

Mandarin Participants	Judge 1	Judge 2
NM1	2	2
NM2	4	5
NM3	4	3.5
NM4	2.5	2.5
NM5	4.5	3
NM6	3.5	2
NM7	2	3
NM8	4	3
NM9	2	2
NM10	4	4
NM11	3.5	3
NM12	2.5	3.5
NM13	3.5	3
NM14	3.5	4
NM15	4	3.5
NM16	3.5	2.5
NM17	3	3
NM18	4	3
NM19	3.5	2.5
NM20	5.5	4.5
NM21	3	2
NM22	2.5	3.5
NM23	3	3
NM24	5	3.5
NM25	2	2
Average	3.38	3.06

Table 3.3 Summary of Foreign Accent Scores by Two Judges

3.3.2. Tasks

The primary experiment was composed of four tasks designed in E-prime 2.0 Professional (Schneider, Eschman, & Zuccolotto, 2012).

To examine the lexical effect on coarticulatory patterns, I needed to first confirm that the participants did in fact make a distinction between a) high frequency and low frequency words, and b) so-called real words and nonwords (see Table 3.2). For instance, real words should be identified by participants as words they were familiar with and had used before. Therefore, **Task 1** asked participants to rate word familiarity on the 144 items used in Task 3 and Task 4 (List 2, Table 3.2) on a seven-point Likert scale, where seven represents the most familiarity. The seven-point Likert scale is very common but is only simply described in previous works (see Imai, Walley, & Flege, 2005). In most cases, participants are only given the instruction such as “7 = extremely familiar, 1 = not familiar at all”. For the aim of this study, I slightly revised the rating scheme, which is described below.

(1). Familiarity Rating Scale

1: “I’ve never heard/seen this phrase before, I don’t know what it means, and I’ve never used it.”

2: “I’ve heard/seen this phrase before, but I don’t know what it means, and I’ve never used it.”

3: “I’ve heard/seen this phrase before, I know what it means, but I’ve never used it.”

4: “I’ve heard/seen this phrase before, I’m not sure what it means, but I’ve used it.”

5: “I’ve heard/seen this phrase before, I know what it means and I’ve used it.”

6: “I’ve heard/seen this phrase frequently, I’m not sure what it means, but I’ve used it.”

7: “I’ve heard/seen this phrase frequently, I know what it means, and I’ve used it frequently.”

Prior to Task 1, participants were reminded to make judgements based on orthography rather than sound because some nonwords might sound like real words in English (e.g., *tik doun*). They were also warned that each word would appear only once. If they pressed the wrong button, they were instructed to inform the researcher (i.e., the author) of the mistake. The researcher would then manually input their real response once participants finished the experiment. The researcher also told participants that they did not have to memorize the seven-point scale because it would appear on every slide as a prompt. Participants were instructed to give their response by pressing a number on a keyboard (1, 2, 3, 4, 5, 6, or 7) to rate each word appearing on the screen.

E-prime collected a total of 5760 familiarity responses from Task 1 (40 participants * 144 stimuli items). I anticipated that for the 72 nonwords the average rating score would be around one; for the 72 real words, the average rating score would be around six or seven. Table 3.4 below summarizes the lexical judgment scores for the two groups. The results showed that for the NE group, the average score for the 72 real words was 5.66, while for the nonwords was 1.06. For the NM group, the average score for the real words was 4.74, while for the nonwords 1.16. Table 3.4 also includes the results of paired sample T-tests (2-tailed), which showed that both groups were able to significantly distinguish real words from nonwords.[6] That is, the categorical lexical effect was validated in terms of differentiating real words and nonwords. The gradient frequency effect will be discussed in Section 4.4.1.

---

[6] In this study, the significance level was set at .05 such that any p-value less than .05 was considered statistically significant.

Group	Real Words (SD)	NonWords (SD)	T-tests
NE (n=15)	M=5.66 (.62)	M=1.06 (.14)	t (14)=30.118, p<.001
NM (n=25)	M=4.74 (.95)	M=1.16 (.38)	t (24)=19.911, p<.001

Table 3.4 Summary of lexical frequency ratings for both groups

Note: N=Number, M=Mean value, SD=Standard Deviation

Task 2 asked participants to read aloud the 24 baseline items contained in List 1 (see Table 3.1). I used closure duration as an index of temporal overlap (see Section 3.4), and Task 2 was designed to provide the necessary baseline closure duration of each stop as coda and onset where no stop-stop coarticulation was involved. I told participants to read each phrase three times ***clearly*** instead of using the term "no coarticulation", so that participants without linguistic training fully understood the instructions. There was no time limit and participants adjusted their own pace by pressing any key when they were ready to produce the next item. Aside from the verbal instructions, instructions written in E-prime reminded participants to repeat each phrase three times as they cued every new item. The middle token was subject to analysis.

Task 3 asked participants to read aloud the 144 items contained in List 2 (see Table 3.2) in natural speech. Prior to Task 3, the participants were instructed to produce the words **as naturally as possible**; I also explicitly told participants to pretend all words were English words and to utter them in an English way even if they ran into words they did not know.[7] In Task 3, stimuli were marked by underlining the appropriate lexical

---

[7] I observed how each participant read the stimuli. If they produced items as not anticipated, I simply interrupted them. For instance, one or two participants produced *girm* [gɜrm] as [ʤɜrm]. I then asked them to produce [g] instead.

stress. For example, *soup* was underlined in the compound, *soup pot*, and *pace* was underlined in the non-compound *keep pace*. Similarly, *foop* was underlined in the corresponding nonsense word *foop pok*, and *pate* was underlined in *peep pate*.

To further study effects of prosody and speech rate, a final fourth task was included. Task 4 asked participants to answer questions based on a given dialogue. Specifically, using the 144 disyllabic phrases in List 2 (Table 3.2), List 3 was designed as having 144*4 dialogues. Each dialogue included a question and a given answer. Participants were asked to read the question silently, then produce the given answer. The answers were designed to elicit four sentential stress patterns: focused versus unfocused, unfocused versus focused, focused versus focused, and unfocused versus unfocused. An example from List 3 is given below. The full material is provided in Appendix A.

Example

Focused – Unfocused

Prompt: Did you say "**good** day" just now?

Target: No, I said "**bad** day" just now.

Unfocused – Focused

Prompt: Did you say "bad **boy**" just now?

Target: No, I say "bad **day**" just now.

Focused – Focused

Target: You know what? He said "**bad day"** just now.[8]

---

[8] Following Cho (2004), there is no prompt to elicit a stressed-stressed sequence. The first part, "You know what?", implies focus coming up.

UnFocused – UnFocused

Prompt: Did you know **he** said “bad day” just now?

Target: No, **she** said “bad day” just now

As mentioned in the beginning of this chapter, a pilot study was conducted prior to the primary study. In the pilot study, the researcher participated in the conversation with each participant by reading the prompt sentences. To avoid a researcher effect, Task 4 in the primary study adopted Cho’s (2004) design: the prompt was read silently by each participant. Specifically, participants were instructed to read the question silently and interpret the answer in their own way.

After the participants produced the answer, they could either press “1” to go back and correct themselves if they stuttered, or press “2” to continue. It was assumed that since the speaker produced the target word in a conversation, their production would reflect the properties of fast connected spontaneous speech, in contrast with the more careful speech in Task 3. Participants answered each question once and only clusters embedded in the Target sentences were subject to analysis. Given the considerable length in List 3, I divided Task 4 into eight blocks; after each block, participants were reminded to take a break by instructions provided by E-prime.

All testing materials in the experiment were presented in randomized order generated by E-prime. All participants were recorded in the Speech Research Lab in the Department of Linguistics at the University of Victoria. They were told to take a break any time they needed to. They took an average 1.5 -2 hours to complete all tasks. Recordings were made at 44.1 kHz/16 bits through a condenser microphone linked to M-

Audio Fireware 410 and saved as .wav files. A few participants' sound files contained background noise, which were subsequently subjected to noise removal.

As discussed in the experimental design, one important consideration was to examine the speech rate effect. To do this, I needed to confirm that participants in the current research did indeed have different speech rates at the slow rate (i.e., Task 3) and at the fast rate (i.e., Task 4). The speech rate for each participant was calculated by using a speech rate script developed for Praat by De Jong and Wempe (2009). The script measured the number of syllables produced per second. Figure 3.1 below illustrates how this script measures the total syllables produced at each time point. This sound file was randomly chosen from a native English speaker in this study. The number corresponding to each word indicates the total syllables counted thus far.

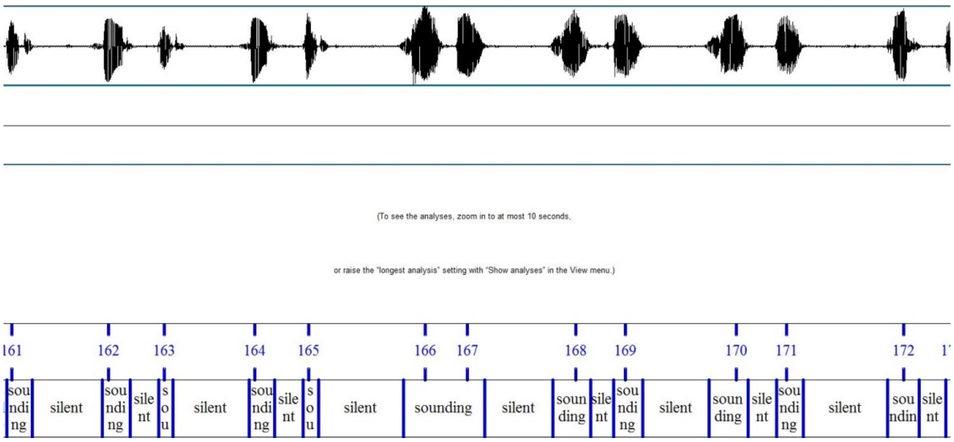


Figure 3.1 An example of Speech Rate Script

Table 3.5 below presents the speech rate data and two-tailed Paired-sample t-tests comparing Task 3 and Task 4 speech rates. The results showed that both groups had significantly faster speech rates at the fast rate (Task 4) than at the slow rate (Task 3). Meanwhile, both groups had high consistency at both speech rates (The NE group: n=15, t=.783, p<.001; The NM group: n=25, t=.704, p<.001). This means that participants' speech rates were significantly correlated as well as distinguished at the same time. In other words, the Speech Rate effect was validated for both groups in this research. Because the *Speech Rate* effect was established, I could be reasonably confident in incorporating it as an independent variable in later analysis.

Group	Word Level	Sentence Level	T-test values
NE (n=15)	M=1.48 (.51)	M=2.22 (1.02)	t (14)= -4.155, p<.001
NM (n=25)	M=1.28 (.4)	M=1.76 (.53)	t (24)= -6.329, p<.001

Table 3.5 Speech Rate summary for both groups

To summarize, this project asks four research questions: how would place of articulation, lexical frequency, stress, and speech rate affect cluster overlap across words in both L1 and L2 speech? To answer these questions, the stimuli incorporated all possible stop combinations to fully examine the effect place of articulation. The stimuli also included both aspects of frequency: a comparison between high versus low frequency clusters in real words (i.e., gradient frequency), and a comparison between real words versus corresponding nonwords (i.e., categorical frequency). In addition, the tasks included both slow and fast speech rates; different stress positions at the word level (i.e., compound vs. non-compound) and at the sentence level (i.e., sentential focus) were considered as well.

3.4.Data analysis

3.4.1. Labelling

A total of 29760 sound files were acoustically analyzed (40 participants * (24 in Task 2 +144 in Task 3 + 576 in Task 4)) using Praat (Boersma & Weenink, 2016). Any phrase in which there was a discernible stumble or pause between C1 and C2 was defined as disfluent. Zsiga (2003) treated disfluent tokens as ones with a period of silence of 350 ms or more between C1 and C2. Given that nonwords were employed in this study, I operationally defined any token with 700 ms or more stumble/pause as disfluent. A total of 58 tokens were excluded based on this criterion to avoid skewing the data (.2% of the total collected).

During the data coding, two measures were of special interest in this research: release percentage and closure duration ratio. For the release percentage, there were three possibilities in this study. Specifically, a cluster of C1#C2 was counted as ***released*** if C1's burst of energy was visible on a spectrogram. A cluster was identified as ***quasi-released*** if an extremely short (usually less than 10ms) and low-intensity event was present for C1, and such an irregular event did not have enough energy throughout the entire frequency range in the spectrogram (see Davidson, 2011). A cluster was identified as having ***no release*** if no burst was present for C1. The following figures demonstrate the three release identifications. In all cases, the highlighted yellow interval also indicates how I labelled closure duration.

A released example is provided in Figure 3.2, where a clear visible burst for /t/ in *boot* can be seen. A quasi-release sample is provided in Figure 3.3, which displays a short and low-intensity burst (less than 10ms) for /b/ in *crab*. A no-release example is provided in Figure 3.4, which shows no burst energy for /b/ in *rub*. Release percentage was then calculated as the amount of tokens that had released C1 divided by the total tokens. For example, Participant 1 (P1) produced 20 released C1s in 144 items in Task 3, therefore the release percentage at the slow rate for P1 is 20/144.[9]

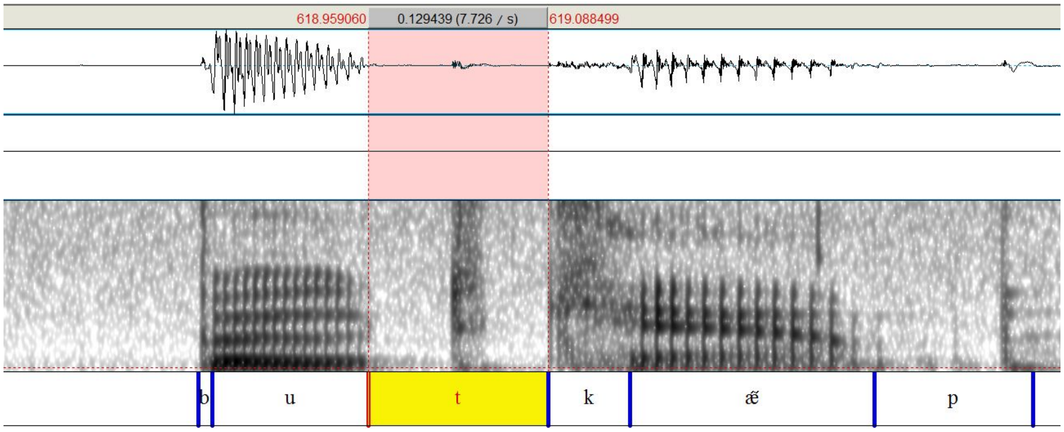


Figure 3.2 Release Example *boot camp*

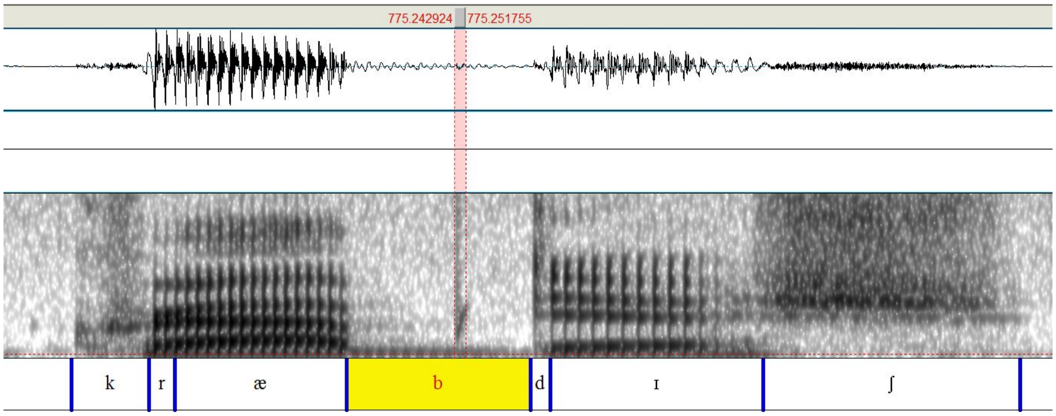


Figure 3.3 Quasi-release Example *crab dish*

---

[9] The calculation of release percentage concerns data in Task 3 and Task 4 since words in Task 2 were employed to infer the standard closure duration.

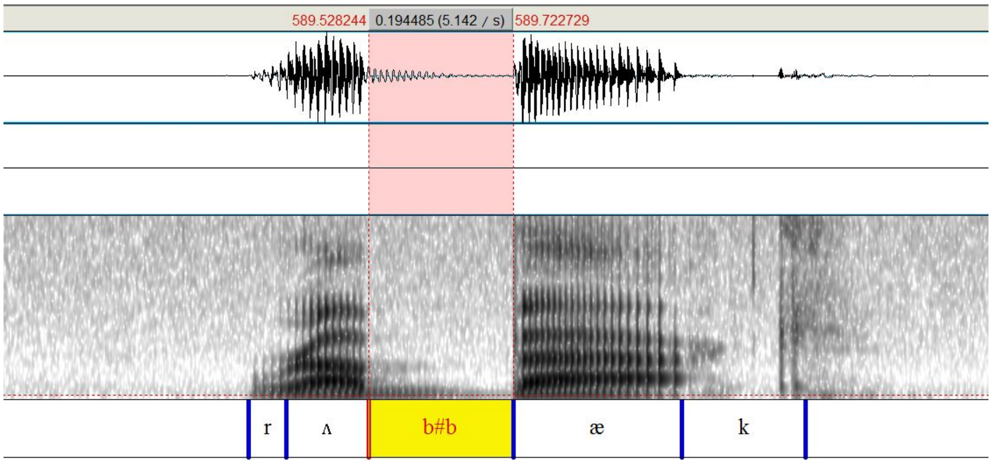


Figure 3.4 No Release Example *rub back*

I need to point out that the acoustic analysis was conducted differently in this research than in previous work: in much of the literature involving acoustic studies, the middle tokens of three repetitions are analyzed. Given that this research project examined articulatory habitual patterns, I also considered whether the middle token was representative of the three repetitions. The middle token was subject to analysis when the release identification agreed with the first or the third token. If the middle token was the only one identified with release while neither the first nor the third token was released, I chose the third token to analyze. That is, I either chose the middle repetition or the third depending on which could represent 2/3 probability in Task 3. All the data were coded by the author. To check reliability of coding, another trained researcher, who was a visiting scholar from China, coded 507 sound files that were randomly chosen from Task 3 (8.8% of the total collected in Task 3). There were 20 tokens that did not agree in release identifications (mainly quasi-release vs. no release), making the disagreement 3.9% (20/507) between judges. The inter-coder reliability was 96.1%, suggesting high consistency.

While the Figures 3.2 – 3.4 provide examples of release coding, the highlighted interval in these figures also indicates how I labelled closure duration. A sample file showing the standard closure duration for an onset stop is provided in Figure 3.5. The time in the highlighted interval represents the closure duration, which is 101 ms for /b/ in *saw boy*. A sample file of labelling coda closure duration is provided in Figure 3.6, suggesting 73 ms for /b/ in *job art*.

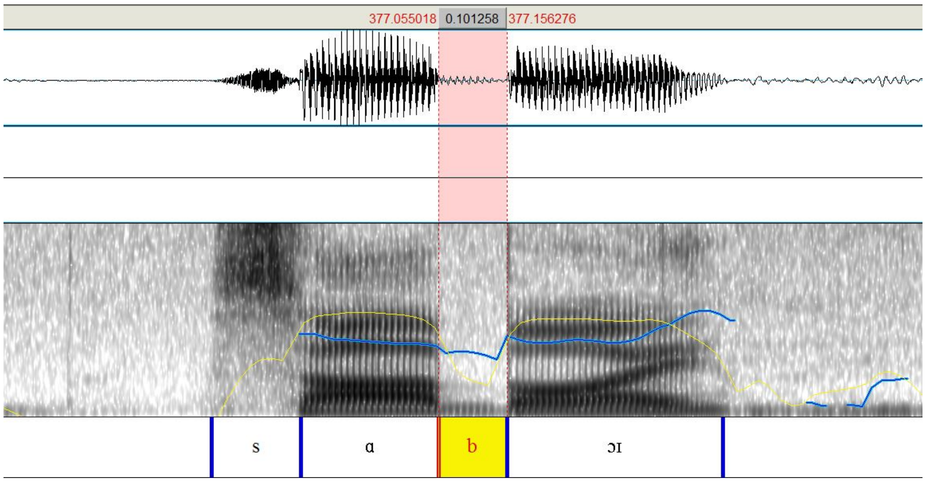


Figure 3.5 Closure duration in onset (V#C1): *saw boy*

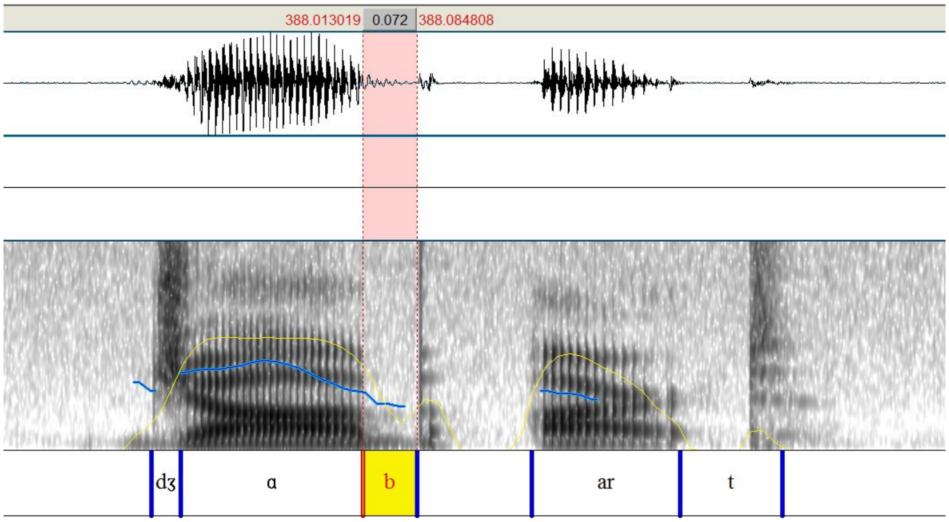


Figure 3.6 Closure duration in coda (C2#V): *job art*

As discussed, I used closure duration ratio as an index of temporal overlap in this study, following Zsiga (2003). The closure duration ratio was measured for each phrase for each speaker with the equation below, which calculates “the mean duration of the C1#C2 cluster divided by the sum of the mean closure durations of $C1_b$ and $C2_b$ occurring intervocalically” (Zsiga, 2003, p. 441). The overlap degree corresponds to 1 – the duration ratio. Therefore, a smaller value indicates a shorter duration of C1#C2, meaning a larger degree of overlap; a bigger value indicates a longer duration of C1#C2, meaning a smaller degree of overlap.

(2) Equation: Duration ratio

$$\frac{M\text{ closure duration C1\#C2}}{(M\text{ closure duration V\#C1}_b) + (M\text{ closure duration C2}_b\text{\#V})}$$

A duration ratio of 1 would indicate that the two consonants in the cluster are exactly sequenced: there is no overlap in the cluster. A ratio of 0.8 would mean that C1 and C2 have 20% overlap.

To sum up, the calculation of duration ratio included the following steps.

1) Segment out stop closure in data from Task 2, Task 3 and Task 4 using Praat
2) Extract the duration of stop closured by a Praat duration script
3) Average the baseline closure duration (BCD) for each stop over two codas and two onsets from words in Task 2
4) Calculate the duration ratio for each phrase by the closure duration of the phrase divided by corresponding BCD.

For example, the closure duration for both /b/s in *rub back* is 194 ms as shown in Figure 3.3 Given BCD of /b/ as onset is 120 ms averaged over *saw boy* and *saw balm*, BCD of /b/ as coda is 80 ms averaged over *rib art* and *job art*, then the duration ratio for /b/s in *rub back* is 0.97, 194 / (120+80). This suggests that onset /b/ and coda /b/ in *rub back* were overlapped by 3% in their intervocalic closure duration for this speaker.

I need to iterate again that in a true articulatory view, actual overlap is avoided if there is a momentary gap between the two consonantal articulations in an open transition. In this study, I follow previous studies (Hardcastle & Roach, 1979; Henderson & Repp, 1982; Zsiga, 2003) and take shortening closure duration to be indicative of overlap. Thus, it is possible to discuss a released cluster that nonetheless has some degree of overlap if the closure durations of C1 and C2 are shortened compared to the baseline. The relationship between release percentage and duration ratio could be a) positively correlated, meaning when speakers have more releases, they tend to have longer durations (i.e., less overlap); b) negatively correlated, meaning when speakers have more releases, they tend to have shorter durations (i.e., more overlap); or c) not correlated, meaning no systematic coordination between release and duration patterns.

3.4.2. Independent and dependent variables

As mentioned in the last subsection, the two dependent variables in this study were release percentage and closure duration ratio. The roles of the independent variables in determining C1#C2 articulatory relationships were established through the two dependent variables. The five independent variables for the data in Task 3 were:

I. ***Group*** (2 levels)

1. NM – Native speakers of Mandarin who also use English as a foreign language
2. NE – Native speakers of English

II. ***Place of Articulation*** (***POA***, 3 levels)

1. Homorganic clusters – Place of articulation of C1 and C2 in C1#C2 is the same (i.e., labial-labial, coronal-coronal, dorsal-dorsal)
2. Front-Back clusters – C1#C2 has a front-preceding-back order of place of articulation (i.e., labial-coronal, coronal-dorsal, or labial-dorsal)
3. Back-Front clusters – C1#C2 has a back-preceding-front order of place of articulation (i.e., coronal-labial, dorsal-coronal, or dorsal-labial)

III. ***Frequency*** (***Categorical Familiarity***, 2 levels)

1. Real words
2. Nonwords

IV. ***Compound*** (2 levels)

1. Compound: Words whose first syllable (i.e., that includes C1) is lexically stressed
2. Non-compound: Words whose second syllable (i.e., that includes C2) is lexically stressed

V. ***Voicing*** (4 levels)

1. VC1: Both C1 and C2 are voiceless
2. VC2: C1 is voiceless and C2 is voiced
3. VC3: C1 is voiced and C2 is voiceless
4. VC4: Both C1 and C2 are voiced

To examine the effect of prosodic position on coarticulatory patterns, the data in Task 4 considers another factor: *Focus*. The independent variables for Task 4 were thus:

I. ***Group*** (2 levels)
   1. NM
   2. NE

II. ***Place of Articulation*** (***POA***, 3 levels)
   1. Homorganic clusters
   2. Front-Back clusters
   3. Back-Front clusters

III. ***Frequency*** (***Categorical Frequency***, 2 levels)
   1. Real words
   2. Nonwords

IV. ***Focus*** (4 levels)
   1. FOCUS 1: Words that includes C1 is focused, **F1**F2
   2. FOCUS 2: Words that includes C2 is focused, F1**F2**
   3. FOCUS 3: Both words that include C1 and C2 are focused, **F1F2**
   4. FOCUS 4: Neither words that include C1 and C2 is focused, F1F2

V. ***Voicing*** (4 levels)
   1. VC1
   2. VC2
   3. VC3
   4. VC4

Three points need to be addressed here. Firstly, since I included all possible English stop-stop combinations, I had access to voicing as a factor to examine, even though previous studies suggested it might not (Ghosh & Narayanan, 2009), and for this reason, it was not included as one of the four primary factors under investigation.

Secondly, frequency was defined as a) a categorical factor with two levels (e.g., real words vs. nonwords), as well as b) a gradient factor (e.g., high vs. low frequency in real words) in this study (see Section 2.3.4). In Section 4.4.1, I will discuss the gradient effect (from highest to lowest frequency including only the real words) to answer the question about frequency effect per se, using correlation tests. For the bulk of the analyses and elsewhere other than Section 4.4.1, categorical frequency was used and examined. This categorical decision here was a simplification strategy, which was done primary due to keep the multi-factor analyses manageable.

Thirdly, note that both *Compound* and *Focus* effects refer to stress positions. The difference is that *Compound* refers to word-level stress in Task 3, while *Focus* refers to sentential stress embedded in conversation in Task 4. I did not include *Compound* as an independent variable during data analysis in Task 4, though I was aware of a possible interaction between *Compound* and *Focus*. This decision was based on previous work indicating that sentential stress would be very likely to override lexical stress. For example, Cho and McQueen (2005) found that in Dutch, accented (i.e., sentential stressed) /d/s always had longer closure durations than unaccented (i.e., no sentential stress) ones, regardless of whether the stop received lexical stress or not. It is reasonable to expect a similar interaction between lexical and sentential stress in English given that both Dutch and English are stress-timed languages. Moreover, the statistical analysis in this study was complicated enough (2*3*2*4*4 design); another subcategory would make the statistics less powerful. Thus, this research examined the *Compound* effect in Task 3 and the *Focus* effect in Task 4.

In an independent analysis, I conducted 4-way Repeated Measures by-subjects ANOVAs to analyze the overall data (i.e., averaged over Task 3 and Task 4). For this test, the independent variables were:

I. ***Group*** (2 levels)

1. NM
2. NE

II. ***Place of Articulation*** (POA, 3 levels)

1. Homorganic clusters
2. Front-Back clusters
3. Back-Front clusters

III. ***Frequency*** (Lexical Familiarity, 2 levels)

1. Real words
2. Nonwords

VI. ***Speech Rate*** (2 levels)

1. Slow rate
2. Fast Rate

3.4.3. Statistical Analysis

As discussed in the last two subsections, Task 3 incorporated five independent variables (*Group, POA, Frequency, Compound,* and *Voicing*) and two dependent variables (release percentage and closure duration ratio); similarly, Task 4 also incorporated five independent variables (*Group, POA, Frequency, Focus,* and *Voicing*) and two dependent variables (release percentage and closure duration ratio). In a later analysis, I conducted 4-way Repeated Measures by-subjects ANOVAs with the independent variables (*Group, POA, Frequency,* and *Speech Rate*) and two dependent variables (release percentage and closure duration ratio).

Table 3.6 and Table 3.7 below summarize these factors and their levels for Task 3 and Task 4, respectively, where the shaded columns indicate the within-subjects factors. Table 3.8 summarizes factors and levels concerning the speech rate.

Variables	Group (between subjects)	Frequency (within-subjects)	POA (within-subjects)	Compound (within-subjects)	Voicing (within-subjects)
Level	Mandarin	Real words (i.e., high frequency)	Homorganic	Compounds	C1 and C2 are voiceless
	English	Nonwords (i.e., low frequency)	Front-Back	Non-compounds	C1 is voiceless and C2 is voiced
			Back-Front		C1 is voiced and C2 is voiceless
					Both C1 and C2 are voiced

Table 3.6 Between- and within-subjects factors for the by-subject analysis in Task 3

Variables	Group (between subjects)	Frequency (within-subjects)	POA (within-subjects)	Focus (within-subjects)	Voicing (within-subjects)
Level	Mandarin	Real words (i.e., high frequency)	Homorganic	Word1 is focused	C1 and C2 are voiceless
			Front-Back	Word 2 is focused	C1 is voiceless and C2 is voiced
	English	Nonwords (i.e., low frequency)	Back-Front	Both words are focused	C1 is voiced and C2 is voiceless
				Neither word is focused	Both C1 and C2 are voiced

Table 3.7 Between- and within-subjects factors for the by-subject analysis in Task 4

Variables	Group (between subjects)	Frequency (within-subjects)	POA (within-subjects)	Speech Rate (within-subjects)
Level	Mandarin	Real words (i.e., high frequency)	Homorganic	Slow
			Front-Back	Fast
	English	Nonwords (i.e., low frequency)	Back-Front	

Table 3.8 Between- and within-subjects factors for the by-subject analysis of the overall data

3.5.Summary

In the examination of how competing factors influence gestural overlap in clusters across words, this chapter has laid out in detail the stimuli choice (Section 3.1) and the experimental design (Section 3.2). This chapter also discusses the data collection and how acoustic and statistical data analyses are conducted (Chapter 3.3). The following Chapter, Chapter 4, reports on the results and analyses of the overall data.

Chapter 4

Results

The previous chapters have outlined and discussed all the necessary research background (i.e., Chapter 2) and methodology upon which the current research rests (i.e., Chapter 3). In this chapter, I will report on the research findings and evaluate the hypotheses laid out in Chapter 2.

Prior to exploring the statistical analyses, I will first briefly review the research questions driving this research (Section 4.1). I will then report on findings from the overall data (Section 4.2). In section 4.3 – 4.6, I evaluate the four primary research questions and their respective hypotheses. In each section, I first discuss data from Task 3 (the slow rate), then data from Task 4 (the fast rate). Section 4.7 reports additional analyses concerning the relationship between overlap degree and accent rating. Finally, Section 4.8 concludes with a general summary, outlining the significant group differences discovered from the data.

4.1. Research Questions

As reviewed earlier in Chapter 2, previous research has identified four factors that determined overlap degree in English clusters coordination: place of articulation, lexical frequency, stress, and speech rate. However, no studies to my knowledge have combined these competing factors to examine their effects in both L1 and L2 speech. This gap is what essentially drives this research. The general assumption is that L2 speakers will not coordinate the adjacent gestures as L1 speakers (e.g., Barlow, 2005; Broselow & Finer, 1991; Colantoni & Steele, 2008; Davidson, 2003, 2006; Solé, 1997; Zsiga, 2003, 2011). English and Mandarin, in particular, differ in rhythmic category (i.e., stressed timed vs. syllable timed) and cluster organization (extensive overlap in clusters vs. no clusters categorically). In general, Mandarin speakers are thus hypothesized to have more internal releases and have less overlap than English speakers (see Hansen, 2001; Wang, 1995; Messing, 2008).

Table 4.1 below outlines the four primary research questions along with their respective hypotheses in this study. Each research question corresponds to one of the four factors identified in previous research that determined overlap degree in English clusters coordination: place of articulation, lexical frequency, stress, and speech rate. I will dedicate the remainder of this chapter to discussing whether each of the hypotheses is supported or not in this study.

Research Questions	Hypotheses
RQ1: Will place of articulation (i.e., homorganic clusters, front-back clusters, and back-front clusters) have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?	H 1: English speakers will have fewer releases and more overlap than Mandarin ESL speakers in producing homorganic clusters.
	H 2: English speakers will show the release pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.
	H 3: English speakers will show the temporal pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.
RQ2: Will lexical/frequency have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?	H 4: Both English and Mandarin speakers will have fewer releases and more overlap in meaningful high-frequency words than in meaningful low-frequency words.

a) Will both groups of speakers coordinate clusters in high-frequency words the same way as clusters in low-frequency words? b) Will both groups of speakers coordinate clusters in real words the same way as clusters in nonwords?	H 5: English speakers will exhibit the release pattern of the POE in both real words and nonwords, while Mandarin speakers will not.
	H 6: English speakers will strongly exhibit the temporal overlap pattern of the POE in both real words and nonwords, while Mandarins speakers will not.
RQ3: Will stress patterns have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?	H 7: English speakers will release less often in words when they are lexically (Task 3) and sententially (Task 4) unstressed than when they are stressed, while Mandarin speakers will not make such a difference.
	H 8: English speakers will produce more gestural overlap in words when they are lexically (Task 3) and sententially (Task 4) unstressed than when they are stressed, while Mandarin speakers will not exhibit significant temporal differences.
RQ4: Will speech rate have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?	H 9: English and Mandarin speakers will have more gestural overlap when speech rate is increased (Task 3 vs. Task 4).

Table 4.1 Outline of the primary research questions and their respective hypotheses

4.2. Descriptive statistics for the overall data

The two dependent variables in this study were release percentage and closure duration ratio. For the overall data, a Pearson correlation showed that the two variables were positively correlated (r=.466, n=40, p<.002). That is, the more often the group released, the greater the closure duration ratio (i.e., the smaller the gestural overlap). This effect was in fact due to the NE group (r=.582, n=15, p<.023). The NM group did not have a significant correlation (r=.369, n=25, p>.07). This result indicated that when the

NE group had more releases, they also tended to have longer closure durations of C1#C2, resulting in less overlap; when the NE group had fewer releases, they also tended to have shorter closure durations, resulting in more overlap. However, this systematicity was not shown in nonnative speech. It seems relatively random how the NM group coordinated English clusters.

Based on the experimental design outlined in Chapter 3, both the release percentage and closure duration ratio were analyzed using factorial repeated measures ANOVAs. For data from Task 3 (the slow rate), 5-way Repeated Measures ANOVAs were conducted with *Group* as between-subjects variable, *Place of Articulation (POA), Frequency* (i.e., categorical frequency, real words vs. nonwords)*, Compound*, and *Voicing* as four independent within-subjects factors. I will discuss the results from Task 3 (the slow rate) in Section 4.2.1. For data from Task 4 (the fast rate), 5-way Repeated Measures ANOVAs were conducted with *Group* as between-subjects variable, *POA*, *Frequency*, *Focus*, and *Voicing* as four independent within-subjects variables. The results will be presented in Section 4.2.2. In a later analysis to examine the overall data, 4-way Repeated Measures ANOVAs were conducted with *Group* as between-subjects variable, *POA*, *Frequency*, and *Speech Rate* as three independent within-subjects variables. The results will be summarized in Section 4.2.3.

4.2.1. Statistical results for the data from Task 3

The following paragraphs provide overall descriptive statistics for the NE and the NM groups in Task 3 (the slow rate). The overall results from Task 3 are provided in Tables 4.2 – 4.4, which summarize the significant effects and interactions for the two dependent variables. Shaded cells contain significant results. From the tables, it can be seen that the interactions are too complex to be fully explained at this point. Also, although the main effect of *Group* is not significant on either release or duration ratio patterns, there are differences across groups within other factors. Multiple comparisons with Bonferroni corrections were conducted to further examine the group differences, and I will elaborate on group effects within each condition and within each significant interaction when examining each research hypothesis (Section 4.3 – 4.6).

Table 4.2 shows that four main effects (*POA*; *Frequency*; *Compound*; *Voicing*) are significant on release percentage but only the *Frequency* effect is significant on closure duration ratio.

Main effects	POA	Frequency	Compound	Voicing
Release Percentage	F(2,74)=54.281, p<.001	F(1,37)=29.851, p<.001	F(1,37)=25.219, p<.001	F(3,111)=27.752, p<.001
Closure Duration Ratio	F(2,74)=.448, p>.626	F(1,37)=33.508, p<.001	F(1,37)=1.473, p>.232	F(3,111)=.191, p>.816

Table 4.2 The significant main effects for the two dependent variables at the slow rate

Table 4.3 shows that the effects of thee 2-way interactions (*POA* * *Voicing*; *Frequency* * *Voicing*; *Compound* * *Voicing*) are significant on release percentage, and effects of three 2-way interactions (*POA* * *Compound*; *POA* * *Voicing*; *Frequency* * *Voicing*) are significant on duration ratio. Note that the effect of *POA* * *Compound* is only significant on closure duration ratio and the effect of *Compound* * *Voicing* is only significant for release percentage.

2-way interactions	POA by Voicing	Frequency by Voicing	Compound by Voicing	POA by Compound
Release Percentage	F (6,222)=8.697, p<.001	F(3,111)=4.542, p<.006	F(3,111)=8.934, p<.001	F(2,74)=1.899, p>.158
Closure Duration Ratio	F(6,222)=2.652, p<.04	F(3,111)=3.288, p<.039	F(3,111)=.543, p>.609	F(2,74)=3.326, p<.045

Table 4.3. The significant 2-way interactions for the two dependent variables at the slow rate

Finally, Table 4.4 displays two significant 3-way interactions (*POA* * *Voicing* * *Group*; *POA* * *Compound* * *Voicing*) for release percentage and one significant 3-way interaction (*POA* * *Frequency* * *Compound*) for closure duration ratio.

3-way interactions	POA by Frequency by Compound	POA by Voicing by Group	POA by Compound by Voicing
Release Percentage	F(2,74)=1.471, p>.242	F(6,222)=4.666, p<.001	F(6,222)=2.424, p<.02
Closure Duration Ratio	F(2,74)=5.085, p<.01	F(6,222)=.844, p>.491	F(6,222)=1.02, p>.403

Table 4.4. The significant 3-way interactions for the two dependent variables at the slow rate

Additionally, there is one 4-way interaction on release percentage (*POA* by *Frequency* by *Compound* by *Voicing*: F(6,228)=2.568, p<.031). Finally, there are one 2-way interaction (*Compound* by *Group*: F(1,38)=3.984, p>.053) and one 3-way interaction involving *Group* (*POA* by *Compound* by *Group*: F(2,76)=3, p>.06) that were marginally significant on closure duration ratio.

These results reflect the complexity of the data at hand, and the extent to which various factors interact with one another. Abstracting away from this complexity, three important points can be made about the data: 1) closure duration ratio is less affected than release percentage, with fewer significant effects and interactions; 2) there is no main group effect, meaning that there were no overall differences between groups that emerged independently of within-group factors; and 3) there were however interactions between group and other factors, suggesting that group differences do exist, but are specific to specific factors. These more subtle group effects are explored starting in Section 4.3.

4.2.2. Statistical results for the data from Task 4

The following paragraphs provide overall descriptive statistics for the NE and the NM groups in Task 4 (the fast rate). Again, I summarize significant effects and interactions below in Table 4.5 – 4.7. Shaded cells represent significant effects/interactions for each dependent variable. Table 4.5 shows that four main effects (*POA*; *Frequency*; *Focus*; *Voicing)* are significant on release percentage, and two main effects (*Frequency*; *Focus*) are significant on closure duration ratio.

Main effects	POA	Frequency	Focus	Voicing
Release Percentage	F (2,76)=65.558, p<.001	F (1,38)=47.087, p<.001	F(3, 114)=36.294, p<.001	F(3,114)=31.416, p<.001
Closure Duration Ratio	F(2,76)=.704, p>.495	F(1,38)=51.97, p<.001	F(3,114)=27.406, p<.001	F(3,114)=.938, p>.393

Table 4.5. The significant main effects for the two dependent variables at the fast rate

Table 4.6 shows that four 2-way interactions (*POA* * *Group*; *POA* * *Focus*; *POA* * *Voicing*; *Frequency* * *Voicing*) are significant on release percentage, and five 2-way interactions are significant on closure duration ratio (*Frequency* * *Group*; *POA* * *Focus*; *Frequency* * *Focus*; *POA* * *Voicing*; *Focus* * *Voicing*).

2-way interactions	POA by Focus	Frequency by Voicing	POA by Voicing	POA by Group	Frequency by Group	Frequency by Focus	Focus by Voicing
Release Percentage	F (6, 228)=4.404, p<.001	F(3,114)=3.673, p<.015	F(6,228)=14.058, p<.001	F (2, 76)=6.63, p<.003	F(1,38)=.021, p>.886	F(3,114)=1.378, p>.254	F(9,342)=1.739, p>.106
Closure Duration Ratio	F(6,228)=3.487, p<.005	F(3,114)=2.602, p>.063	F(6,228)=3.091, p<.04	F(2,76)=1.248, p>.293	F(1,38)=7.543, p<.009	F(3,114)=4.849, p<.008	F(9,342)=1.674, p<.007

Table 4.6 The significant 2-way interactions for the two dependent variables at the fast rate

Also, Table 4.7 shows that five 3-way interactions are significant on release percentage (*POA* * *Voicing* * *Group*; *POA* * *Frequency* * *Voicing*; *Focus** *Voicing* * *Group*; *POA* * *Focus* * *Voicing*; *Frequency* * *Focus* * *Voicing*), and no 3-way interactions affect closure duration ratio.

3-way interactions	**POA by Voicing by Group**	POA by Frequency by Voicing	**Focus by Voicing by Group**	POA by Focus by Voicing	Frequency by Focus by Voicing
Release Percentage	**F(6,228)=7.502, p<.001**	F(6,228)=2.409, p<.042	**F(9,342)=2.626, p<.015**	F(18,684)=2.346, p<.008	F(9,342)=2.763, p<.009
Closure Duration Ratio	F(6,228)=1.603, p>.201	F(6,228)=2.36, p>.058	F(9,342)=1.674, p>.144	F(18,684)=.876, p>.546	F(9,342)=2.2, p>.054

Table 4.7 The significant 3-way interactions for the two dependent variables at the fast rate

Just as with the slow speech rate (Task 3), results indicate that closure duration is not as likely to be affected by various factors as is release ratio. Also, although there is no main effect of group, differences between groups do emerge when one considers interactions with other factors.

4.2.3. Statistical results for the overall data

To analyze the effect of *Speech Rate* specifically (Task 3 vs. Task 4), stress and focus conditions were merged and both the release percentage and duration ratio were analyzed using 4-way Repeated Measures by-subjects ANOVAs with 3*2*2*2 design (*POA*Frequency*Speech Rate*Group*). The following paragraphs provide overall descriptive statistics for the NE and the NM groups. Again, I summarize significant effects and interactions below in Table 4.8 – 4.9. Shaded cells represent significant effects/interactions for each dependent variable. Table 4.8 shows that two main effects are significant on release percentage (*POA*, *Frequency*) and one main effect is significant on duration ratio (*Frequency*). The main effect of *Speech Rate* is not significant on either dependent variable.

Main effects	POA	Frequency	Speech Rate
Release Percentage	F(2,76)=73.606, p<.001	F (1,38)=46.719, p<.001	F(1,38)=.598, p>.444
Closure Duration Ratio	F(2,76)=.551, p>.572	F(1,38)=54.592, p<.001	F (1,38)= .008, p>.928

Table 4.8. The significant main effects for the two dependent variables in analyzing the overall data

Table 4.9 shows that two 2-way interactions are significant on release percentage (*POA* Group*, *Speech Rate* * *Group*) and one 2-way interaction is significant on duration ratio (*Frequency* * *Group*).

2-way interactions	POA by Group	Frequency by Group	Speech Rate by Group
Release Percentage	F (2,76)=4.675, p<.016	F(1,38)=.122, p>.729	F(1,38)=6.085, p<.018
Closure Duration Ratio	F(2,76)= .912, p>.403	F(1,38)=4.825, p<.034	F(1,38)=.046, p>.832

Table 4.9 The significant 2-way interactions for the two dependent variables in analyzing the overall data

Given the complex interactions outlined above, in the following sections I will divide and collapse the data in ways that allow me to answer each of the research questions presented in Chapter 2 and repeated in Section 4.1 above.

4.3. Research Question 1

The first research question (RQ 1) asks whether and how place of articulation (POA) affects English and Mandarins ESL speakers' stop-stop coarticulation. The overall *POA* effect averaged over both speech rates on release patterns for each group is plotted in Figure 4.1 below. The statistical results reported in the last section show that *POA* has significant effect on release percentage at both speech rates. Figure 4.1 shows that *POA* has a similar effect on both groups in terms of the release trend. This trend is visible in that both groups had the highest release percentage in front-back clusters, followed by back-front clusters, then homorganic clusters (i.e., front-back > back-front > homorganic, > indicates more releases).

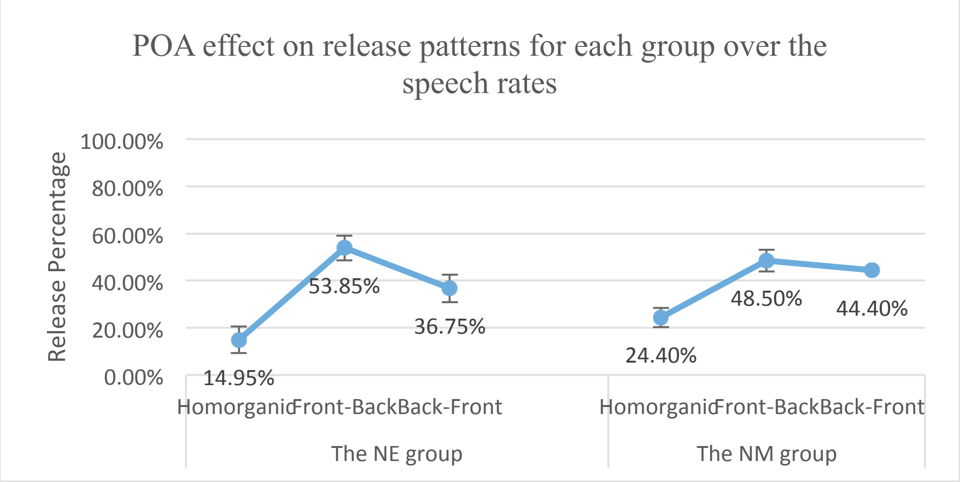


Figure 4.1 Place of Articulation effect on release percentage for both groups

Note: The blue line represents the release percentage, the higher the line, the more often group released;

The overall *POA* effect averaged over both speech rates on duration ratio patterns for each group is plotted in Figure 4.2 below (error bars: 95% CI). Note that the higher the line, the smaller the gestural overlap the group has. The statistical results reported in Section 4.2.2 suggested that *POA* has no significant effect on duration ratio at either speech rate for either group. Figure 4.2 shows that overall, the NE group seemed to have more overlap than the NM group. The NE group had the most amount of overlap in front-back clusters, followed by back-front clusters, then homorganic clusters (i.e., front-back >> back-front >> homorganic, >> indicates more overlap); the NM group did not have much difference in producing the three kinds of clusters.

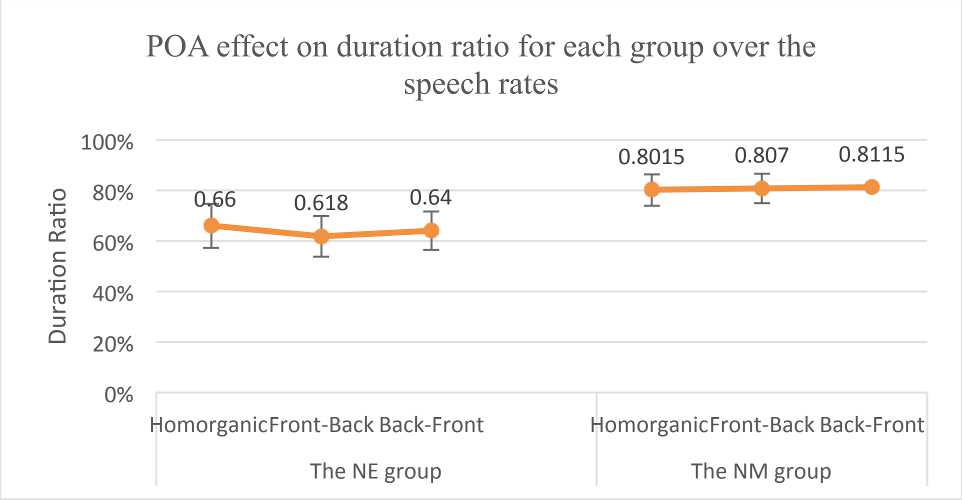


Figure 4.2 Place of Articulation effect on duration ratio for both groups

To address RQ1, I will further break the data and examine each hypothesis in the following subsections.

4.3.1. Hypothesis 1

H1: English speakers will have fewer releases and more overlap than Mandarin speakers in producing homorganic clusters.

To examine H1, release percentage and duration ratio data in Task 3 (48 homorganic phrases * 40 participants) and Task 4 (192 homorganic phrases * 40 participants) were separately analyzed. Given that *Speech Rate* variable was validated (see Chapter 3.3.1), I will refer to Task 3 as the slow speech rate and refer to Task 4 as the fast speech rate.

The overall release data in homorganic clusters are plotted in Figure 4.3 below, including both speech rates. The height of the bars in this graph indicates the mean percentage under the three release identifications: release, no release and quasi-release. The error bars indicate a 95% confidence interval. Figure 4.3 shows that the NM group released more often than the NE group at both speech rates in homorganic clusters. At the slow rate, the NE group released 11.8% and the NM group released 25.7% of homorganic clusters; at the fast rate, the NE group released 18.1% and the NM group released 23.1% of homorganic clusters.

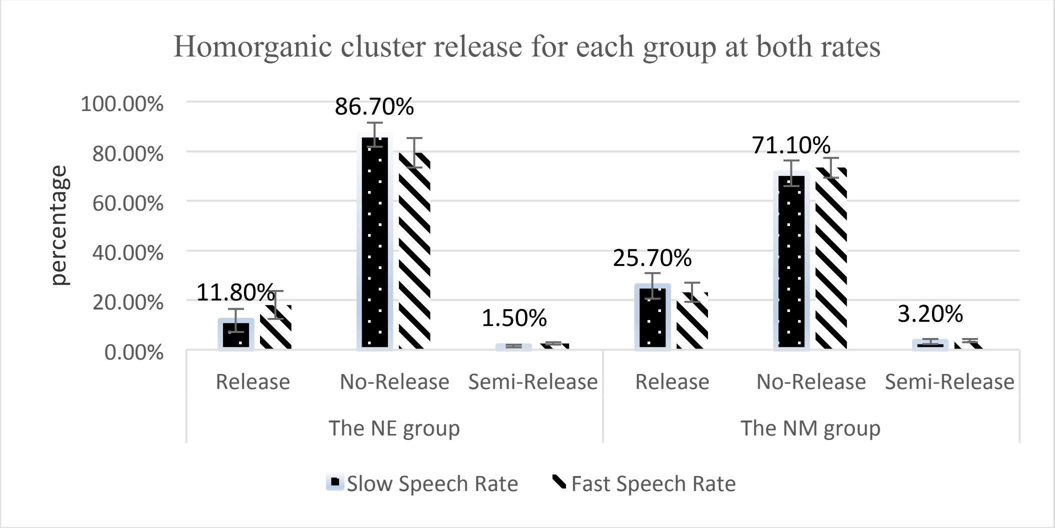


Figure 4.3 Release summary in homorganic cluster for both groups at both speech rates

Given that all research questions and their respective hypotheses revolve around release comparisons, statistical analyses only include release patterns as the dependent variable. The statistical analyses indicate that despite the visible trend, the overall group differences were not statistically significant at the slow rate (MD= -.139, p>.072) or at the fast rate (MD= -.05, p>.457). Although the trend is in the right direction, the statistical analyses do not support Hypothesis 1 (H1). This result means that the two groups did not release homorganic clusters significantly differently from one another. This maybe an effect of lack of power, given that so many factors are involved in this study.

Multiple comparisons with Bonferroni corrections were conducted to further explore possible within- and cross-group differences in release ratios, as a function of other factors included in the study (*POA*, *Frequency*, *Compound, Focus*, and *Voicing*). At the slow rate, the post hoc analyses found significant group differences in one specific voicing condition: VC2 (MD= -.204, p<.026).[10] A closer inspection revealed that the NM group released significantly more often than the NE group in meaningful compound words under VC2 (MD= -.276, p<.008) (e.g., *lip balm*, *vet debt*, and *deck gate*), but the group difference was not significant in meaningful non-compounds such as *beat down* and *look good.* Within the NM group, they had more releases under VC2 than in other conditions in which the first consonant was voiced: VC2 > VC3, VC4 (both p<.033). They also had more significant releases in a voiceless cluster than a voiced cluster: VC1 > VC4 (MD=.097, p<.026).

Similar to the performance at the slow rate, at the fast rate, the NM group released significantly more often in homorganic clusters when C1 was voiceless than when it was voiced: VC1 > VC3, VC4 (both p<.019), VC2 > VC3, VC4 (both p<.001). The NM group's release order in homorganic sequences is VC2 > VC1> VC3 > VC4 (e.g., *lip balm* > *get tip* > *need time* > *bad day*). In contrast, the crucial factor for the NE group was whether the whole cluster was voiced or not: NE speakers released significantly ***less*** often in a voiced homorganic cluster than in other conditions: VC1, VC2, VC3 > VC4 (all p<.015). The NE group's release order in homorganic sequences is VC1 > VC2 > VC3 > VC4 (e.g., *get tip* > *lip balm* > *need time* > *bad day*). That is, the crucial comparison for the NE group was voiced clusters versus other conditions (with *fewer* releases in voiced clusters), whereas for the NM group the crucial comparison was voiceless clusters (and to a lesser extent clusters with a voiceless C1) versus other conditions (with *more* releases in voiceless clusters)

---

[10] Recall VC1= voiceless C1+ voiceless C2; VC2=voiceless C1 + voiced C2; VC 3=voiced C1+voiceless C2; VC4=voiced C1 + voiced C2.

In sum, despite the overall group differences being apparent in Figure 4.3, with the NE group having fewer releases than the NM group in homorganic clusters, H1 is not statistically supported in the release data.

H1 also considers the overlap comparison between the two groups. Recall that the closure duration ratio was calculated for each token for each speaker by using the equation in (3) discussed in Chapter 3.3, where a smaller ratio value indicates a larger amount of temporal overlap in C1#C2. For instance, if the closure duration ratio is 0.8, it means the two consonants are overlapped by 20%.

(3) Equation: Duration ratio

$$\frac{M \text{ closure duration C1\#C2}}{(M \text{ closure duration V\#C1}_{b}) + (M \text{ closure duration C2}_{b}\text{\#V})}$$

Table 4.10 below summarizes the descriptive data of the temporal ratios in homorganic clusters produced by the two groups at the slow rate as well as the fast rate. Also included in the table are the standard deviations of these mean ratios (given in parentheses). For example, the table shows that the NE group had 0.65 closure duration ratio at the slow rate and 0.67 at the fast rate. This means that the NE group had 35% overlap at the slow rate and 33% overlap at the fast rate. The NM group had 0.801 closure duration ratio at the slow rate and 0.802 at the fast rate, suggesting 19.9% overlap and 19.8% overlap, respectively. It can be seen that the NE group always had greater overlap than the NM group in producing homorganic C1#C2 at both speech rates.

Duration Ratio	The NE group (SD)	The NM group (SD)
Slow Rate	0.65 (.33)	0.801 (.31)
Fast Rate	0.67 (.47)	0.802 (.27)

Table 4.10 Duration ratio in homorganic clusters for each group at both speech rates

Despite the visible trend, the overall group differences were not significant, either at the slow rate (MD= -.151, p>.163) or at the fast rate (MD= -.133, p>.258). Thus, H1 was not statistically supported by the overall duration ratio data. Further comparisons were conducted to test for significant differences that might emerge under specific conditions. At the slow rate, multiple post hoc comparisons found no within- or cross-group differences. This result suggests that both groups had a similar temporal pattern in coordinating homorganic clusters at the slow rate. At the fast rate, the post hoc analyses found no cross-group differences either, but within-group differences were found in the NM group. They had a significantly greater ratio (i.e., less overlap) under VC2 in homorganic clusters: VC1, VC3, VC4 >> VC2 (all p<.006).

To summarize H1, Figure 4.4 below plots the two groups' release and ratio data in homorganic clusters at both speech rates (error bars: 95% CI). The figure suggests that the NE group had fewer releases and more overlap (i.e., smaller duration ratios) than the NM group regardless of the speech rate. Despite this trend, H1 was not statistically supported. However, significant comparisons were shown in a subset of the data. For instance, the NM group released significantly more often than the NE group in homorganic clusters with a voiceless C1 and a voiced C2 (e.g., *lip balm*). As we shall see in the sections below, this pattern is typical: while overall group differences are not found in the data, they do emerge in more detailed comparisons.

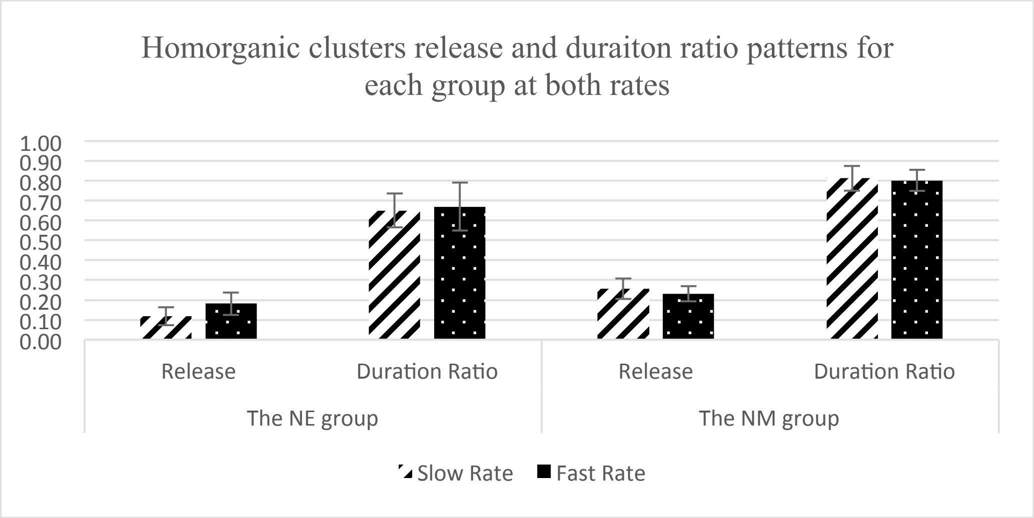


Figure 4.4 Summary of release and duration ratio in homorganic clusters for both groups at both rates (Error bars: 95% CI.)

4.3.2. Hypothesis 2

H2: English speakers will show the predicted release pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.

To examine H2, I compared release patterns in front-back and back-front clusters produced by the NE and the NM groups at both speech rates. The overall release results for the two types of clusters are plotted in Figure 4.5 below (Error bars: 95% CI). The figure illustrates a visible trend that both groups released more often in front-back than in back-front clusters at both speech rates; but the trend is greater in the NE group than in the NM group. The NE group released more often in front-back than in back-front clusters, both at the slow rate (53.8% vs. 33.7%) and at the fast rate (54.5% vs. 39.8%).

The NM group exhibited the same trend but with many smaller differences, both at the slow rate (55% vs. 49.4%) and at the fast rate (42% vs. 39.4%).

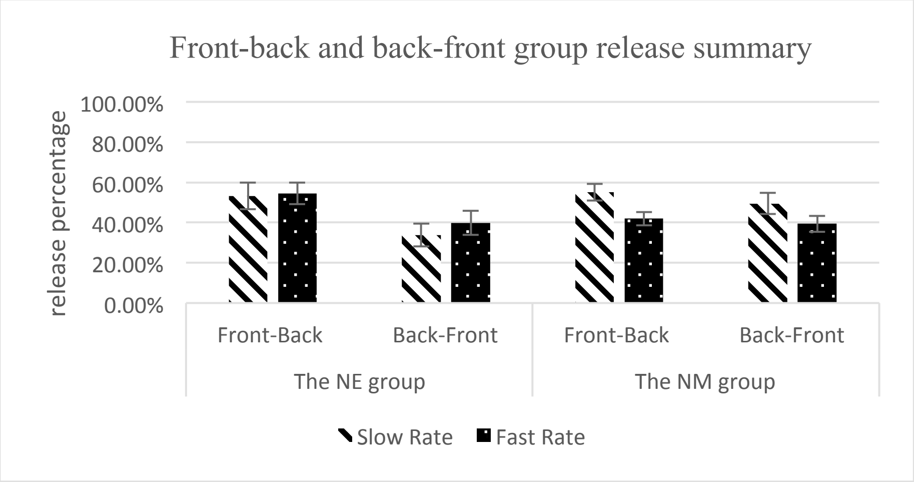


Figure 4.5 Release summary in front-back and back-front clusters for both groups at both speech rates (Error bars: 95% CI0)

The statistical results confirmed that the NE group had significantly more releases in front-back than in back-front clusters at both speech rates (at the slow rate: MD=.194, p<.005; at the fast rate: MD=.147, p<.001) while the NM group did not (at the slow rate: MD=.056, p>.648; at the fast rate: MD=.026, p>.37). This result means that the NE group strongly showed the POE while the NM group did not (although the trend was similar to the NE group). Therefore, H2 is statistically supported.

To further clarify the interaction between the POE and group, within and cross-group differences were examined within each main effect. At the slow rate, the NE group released less often than the NM group in both types of clusters, but overall group differences were not significant in either (in front-back clusters: MD= -.018, p>.806; in back-front clusters: MD= -.157, p>.061). The post hoc analyses found significant group differences in front-back clusters specifically under VC2 (MD= -.159, p<.048), at the slow rate. This result reflects the fact that the NE group had much fewer releases than the NM group in front-back clusters under VC2 (e.g., *deep down*), a pattern similar to the homorganic cluster release pattern. Group differences were also found significant in back-front clusters under VC3 and VC4 (both p<.038). That is, the NE group had significantly fewer releases than the NM group in back-front clusters specifically when C1 was voiced (e.g., *red pen, big boy*). At the fast rate, the NE group released significantly more often than the NM group in front-back clusters (MD=.125, p<.043), but not in back-front clusters (MD=.004, p>.955). The post hoc analyses found significant group differences in front-back clusters under three voicing conditions: VC1, VC3 and VC4 (all p<.047). This means that the NE group released significantly more often than the NM group in front-back clusters except under VC2 (e.g., *deep down*).

At both speech rates, the NE group had the same release order in front-back clusters: VC3 > VC1 > VC4 > VC2 (e.g., *lab coat* > *pep talk* > *good game* > *deep down*). The release percentage was significantly higher when C1 was voiced and C2 was voiceless (i.e., VC3) than when C1 was voiceless and C2 was voiced (VC3 > VC2, p<.014). That is, the NE group released more often in words such as *lab test* than in words such as *deep down*. In back-front clusters, the NE group had the same release order at both speech rates: VC1 > VC2 > VC3 > VC4 (e.g., *suit pants* > *frat boy* > *head pain* > *bed bug*). The back-front clusters with a voiceless C1 had much more releases than the clusters with a voiced C1: VC1 > VC3, VC4 (both p<.001), VC2 > VC3, VC4 (both p<.007). Overall, the NE group had significantly more releases in front-back clusters when C1 was voiced (e.g., *lab test*) than when C1 was voiceless, and had more releases in back-front clusters when C1 was voiceless (e.g., *neck tie*). It was not expected that the NE group would have more releases in front-back clusters when C1 was voiced than when C1 was voiceless. I will discuss the voicing effect in Section 5.3.3.

As for the NM group, they had VC1 > VC2 > VC4 > VC3 (*pep talk* > *deep down* > *good game* > *lab coat*) in front-back clusters at the slow rate and had VC2 >VC1 > VC4 >VC3 (*deep down* > *pep talk* > *good game* > *lab coat*) at the fast rate. The NM group had more releases when C1 was voiceless (e.g., *deep down*) than when C1 was voiced (e.g., *lab test*): VC1 > VC3, and VC2 > VC3, VC4 (all p<.038). In back-front clusters, the NM group exhibited the same release order as the NE group did: VC1 > VC2 > VC3 > VC4, with more releases when C1 was voiceless than when C1 was voiced (VC1 > VC3 / VC4, and VC2 > VC4, all p<.008). This is in accordance with their responses in homorganic clusters, where they produced more internal releases when C1 was voiceless. In sum, the NM group released more often when C1 was voiceless than when C1 was voiced in all three types of clusters (homorganic, front-back, and back-front); the NE group released more often when C1 was voiced than when C1 was voiceless in front-back clusters.

The post hoc analyses also found significant group differences under the interaction of *Focus* * *Voicing* in front-back clusters. The NE group had more releases than the NM group under **FI**F2 and F1**F2** (both p<.048). Specifically, the NE group had more releases than the NM group in releasing front-back clusters under **FI**F2 + VC3 / VC4, F1**F2** + VC3 / VC4 (all p<.043). In sum, the NE group released more often than the NM group in front-back clusters in either **FI**F2 or F1**F2** when C1 was voiced.

To sum up, the analyses indicate that the NE group had significantly more releases in front-back than in back-front clusters at both speech rates while the NM group did not. Thus, H2 was statistically supported. The overall group difference was significant in front-back clusters at the fast rate; other detailed group differences emerged from a subset of data, specifically in relation to *Voicing* and *Focus* factors.

4.3.3. Hypothesis 3

H3: English speakers will show the predicted temporal overlap pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.

To examine H3, I compared duration ratios in front-back and back-front clusters produced by the NE and the NM groups. Table 4.11 provides descriptive data at both speech rates (SD given in parentheses). At the slow rate, the NE group had a duration ration of 0.617 (i.e., 38.3% overlap) in front-back clusters and 0.645 in back-front clusters (i.e., 35.5% overlap); the NM group had 0.806 (i.e., 19.4% overlap) and 0.81 (i.e., 19.0% overlap), respectively. At the fast rate, the NE group had 0.619 closure duration ratio in front-back clusters (i.e., 38.1% overlap) and 0.635 in back-front clusters (i.e., 36.5% overlap); the NM group had 0.808 closure duration ratio in front-back clusters (i.e., 19.2% overlap) and 0.813 in back-front clusters (i.e., 18.7% overlap). Both groups exhibited the overlap trend of the POE at both speech rates.

Group	Front-Back Duration Ratio (SD)		Back-Front Duration Ratio (SD)	
The NE Group	Slow Rate	Fast Rate	Slow Rate	Fast Rate
	0.617 (.29)	0.619 (.39)	0.645 (.28)	0.635 (.35)
The NM Group	Slow Rate	Fast Rate	Slow Rate	Fast Rate
	0.806 (.29)	0.808 (.29)	0.81 (.29)	0.635 (.35)

Table 4.11 Duration ratio for both groups in front-back and back-front clusters at both rates

Unlike release patterns reported in the previous subsection, the NE group did not significantly distinguish temporal overlap patterns significantly between front-back and back-front clusters at the slow rate (MD= -.029, p>.849), nor did the NM group (MD= -.004, p>1). Similarly, neither of the groups had significant timing differences in producing the two types of clusters at the fast rate (The NE group: MD= -.015, p>1; The NM group: MD= -.006, p>1). These analyses suggest that H3 is not statistically supported in the current study, despite some visible trends in the right direction.

At the slow rate, the NE group seemed to have more overlap than the NM group in both cluster types (front-back and back-front), but the overall group differences were not significant in either (front-back clusters: MD= -.189, p>.06[11]; back-front clusters: MD= -.165, p>.097). Post hoc comparisons found that the NE group had significantly more overlap than the NM group in coordinating compound front-back clusters (p<.031, e.g., *lip care*). Group differences were also significant in front-back clusters under VC2 with the NE group having more overlap than the NM group (p<.049, e.g., *deep down*).

---

[11] This is marginally significant, suggesting the group differences were larger in front-back than in back-front clusters.

This result is comparable to their release patterns at the slow rate, where the NM group was found to release more often than the NE group in front-back clusters under VC2.

The NM group produced significantly more overlap in coordinating non-compounds than compounds (MD= -.035, p<.023). These differences came from front-back compounds versus front-back non-compounds (MD=.067, p<.001). Specifically, the NM group produced more overlap in front-back non-compounds such as *bad kid* and *good game* than in front-back compounds such as *lip care* and *head game*.

At the fast speech rate, the NE group had more overlap than the NM group in both front-back and back-front clusters, but again, the overall group differences were not significant in either type (front-back clusters: MD= -.188, p>.083; back-front clusters: MD= -.179, p>.063). The post hoc analysis showed that significant group differences were found in producing nonsense back-front clusters (MD= -.218, p<.033): the NE group had a significantly greater amount of overlap than did the NM group. This significant group difference came from coordinating nonsense back-front clusters under VC2 (MD= -.211, p<.034) (e.g., *lock banb*), with the NE group showing more overlap than the NM group.

Given that neither of the groups showed significantly more overlap in front-back than in back-front clusters at either speech rate, H3 was not statistically supported. Nonetheless, significant group differences did emerge from detailed post hoc comparisons, for instance, in front-back clusters when they were compounds, in nonsense back-front clusters.

4.3.4. Summary for Research Question 1

In this section, I have reported on the results and statistical analyses of the overall data regarding the first research question (RQ 1): Does place of articulation have the same effect for English and Mandarin ESL speakers? Specifically, the POE predicts more releases and more overlap in front-back than in back-front clusters. The general hypothesis was that the POE would be shown in the NE group but not in the NM group.

Table 4.12 below summarizes RQ 1 along with the specific comparisons necessary to answer the question. This table also includes each prediction and indicates whether each result statistically confirmed the research hypothesis or not. The predictions and the results in italics represent statistically significant differences, and those not in italics represent trends. The symbols “>” and “>>” indicate more releases and more overlap, respectively, and the symbol “=” means no differences. The “Yes” or “No” indicates whether each hypothesis is statistically supported (Yes) or not (No). For example, the first prediction was that the NM group would have significantly more releases and less overlap in homorganic clusters than the NE group; the results showed this trend but group differences were not significant. Therefore, H1 was not statistically supported.

	Predictions	Comparisons	Results	Hypotheses
RQ 1: Will place of articulation have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?	*H1* *The NM group > The NE group in homorganic clusters* *The NE group >> The NM group in homorganic clusters*	Both groups' release and duration data in homorganic clusters	The NM group > The NE group The NE group >> The NM group	H1: No (but trend in the right direction)
	H2 *The NE group front-back > back-front* The NM group front-back = back-front	Both groups' release data in front-back and back-front clusters	*The NE group front-back > back-front* The NM group front-back > back-front	H2: Yes
	H3 *The NE group front-back >> back-front* The NM group front-back = back-front	Both groups' duration ratios in front-back and back-front clusters	The NE group front-back >> back-front The NM group front-back >> back-front	H3: No

Table 4.12 Results summary for RQ 1

(Note: > means more releases, and >> means more overlap)

Several points can be made from this table. First, as predicted, the NE group had fewer releases and more overlap than the NM group in coordinating homorganic clusters, but the group differences were not statistically significant. Therefore, H1 is not statistically supported. Second, although both groups had the similar release trend of front-back clusters leading to more releases than back-front clusters, the differences were only significant in the NE group, supporting H2. Third, both groups had slightly more gestural overlap in front-back than in back-front clusters; the differences were not significant in either of the groups, which does not support H3.

In sum, both groups exhibited the release and overlap trends predicted by the POE, but only the NE group significantly showed the release pattern. The temporal overlap pattern was not statistically exhibited in either group. Group differences emerged from detailed post hoc comparisons, I will elaborate on them in Chapter 5.

4.4. Research Question 2

RQ 2 asks whether and how lexical frequency/familiarity determines the cluster realizations: Will English and Mandarin speakers exhibit the same frequency effect on overlap degree? Two sub-questions are a) Will both groups of speakers coordinate clusters in high-frequency words the same way as clusters in low-frequency words? and b) Will both groups of speakers coordinate clusters in real words the same way as clusters in nonwords?

In short, I consider two aspects of lexical frequency/familiarity in this study: the gradient frequency effect, that is, high-frequency words versus low-frequency in real words (e.g., *take care* vs. *dock gate*), and the categorical lexical effect, that is, real words (i.e., high-frequency) versus nonwords (i.e., low-frequency) (e.g., *keep pace* vs. *peep pate*). In this section, I will examine how L1 and L2 speakers coordinate words differing in categorical lexical categories or gradient frequency categories. In particular, the results in the previous section showed that the NE group exhibited a significant POE in terms of release percentage (front-back > back front clusters) in overall data, which included both real words and nonwords. The reasonable follow up questions are: Did speakers exhibit the POE because they were familiar with the lexicon? Did they show the POE in sequences that they were not familiar with at all (i.e., nonwords)?

4.4.1. Hypothesis 4

H4: Both English and Mandarin speakers will have fewer releases and more overlap in meaningful high-frequency words than in meaningful low-frequency words.

Because H4 deals with a comparison between the high-frequency words versus the low-frequency words in real words (e.g., *take care* vs. *dock gate*), I only examined 72 real words in Task 3 and 72*4 items in Task 4 here. To confirm H4, I need to establish a significant correlation between frequency and release percentage, as well as a significant correlation between frequency and duration ratio.

Based on the assigned rating score in Task 1, the 72 real words were ranked from the highest familiarity to the least (i.e., from the highest score to the lowest) for both groups. The means are averaged over speech rates. The results are presented in Table 4.13 below, which includes the means of release percentage and temporal ratio for each of the 72 lexical items produced by each group.

The table shows that the NE group released 30.83% for the 72 items and produced an average duration ratio of 0.609 (i.e., 39.1% overlap); the NM group released 35.59% and produced an average duration ratio of 0.733 (i.e., 26.7% overlap). For instance, both groups ranked *take care* as the most familiar item. In the production of *take care*, the NE group released 10% of the tokens and had 48.9% overlap; the NM group released 12.5% and had 43.6% overlap. In contrast, the NE group ranked *dock gate* as the least familiar item and produced 39.1% overlap; the NM group ranked *hug pug* as the least familiar item, and produced 13.5% overlap.

The NE group				The NM group			
Rank	Lexical score	Release percentage	Closure duration ratio	Rank	lexical score	Release percentage	Closure duration ratio
take care	6.93	10%	0.511	take care	6.96	12.5%	0.564
bad day	6.73	2.5%	0.633	look good	6.96	28.5%	0.763
keep calm	6.73	39.17%	0.49	get back	6.96	30.5%	0.683
great guy	6.73	18.33%	0.515	bad day	6.76	8.5%	0.722
black tea	6.73	52.5%	0.63	keep calm	6.76	40%	0.704
big deal	6.73	13.33%	0.675	big deal	6.68	33.5%	0.698
lip balm	6.67	13.33%	0.584	great guy	6.6	42%	0.672
good game	6.67	51.67%	0.61	black tea	6.6	35%	0.697
get back	6.67	6.67%	0.544	good game	6.56	37%	0.719
hot tub	6.6	5%	0.574	need time	6.48	24%	0.608
look good	6.6	15%	0.64	head pain	6.28	37.5%	0.787
food court	6.6	60.83%	0.58	lab test	6.16	27.5%	0.668
rock band	6.6	56.67%	0.561	deep down	6	54.5%	0.683
lab test	6.53	64.17%	0.611	big game	5.96	18.5%	0.745
deep down	6.53	50%	0.622	big time	5.96	38%	0.748
fat cat	6.53	35.83%	0.53	red pen	5.92	22%	0.715
look pale	6.53	55%	0.663	keep pace	5.88	18%	0.734
soup pot	6.47	17.5%	0.567	bad kid	5.88	49.5%	0.78
need time	6.47	5%	0.623	food court	5.84	79.5%	0.765
bed bug	6.47	5.83%	0.611	look pale	5.8	57.5%	0.825
red pen	6.47	18.33%	0.616	big boy	5.76	20%	0.692
pep talk	6.4	48.33%	0.574	rock band	5.44	46%	0.716
lab coat	6.4	63.33%	0.569	soup pot	5.4	27.5%	0.811
bad kid	6.4	50%	0.636	beat down	5.4	22%	0.737
rub down	6.33	43.33%	0.62	blog page	5.4	27%	0.719
neck tie	6.33	73.33%	0.629	need beer	5.4	50%	0.829

lap dance	6.27	53.33%	0.556	grab pen	5.36	6%	0.683
head pain	6.27	17.5%	0.615	club girl	5.36	35%	0.669
leg bone	6.27	36.67%	0.584	lab coat	5.28	36%	0.776
bright pink	6.27	17.5%	0.557	skip test	5.24	31.5%	0.651
big game	6.2	3.33%	0.648	fat cat	5.24	60.5%	0.773
need beer	6.2	20%	0.711	dog tail	5.24	44%	0.711
big boy	6.2	9.17%	0.597	job bank	5.16	12%	0.645
boot camp	6.13	52.5%	0.527	suit pants	5.16	58.5%	0.701
big time	6.13	19.17%	0.643	bright pink	5.16	39%	0.801
head game	6.07	60.83%	0.658	kick down	5.16	48%	0.65
frat boy	6.07	8.33%	0.493	grab cash	5.08	24%	0.735
crab dish	6	30.83%	0.648	leg bone	5.08	42.5%	0.769
rib pain	5.93	9.17%	0.646	hot tub	5.04	26%	0.662
dog tail	5.87	23.33%	0.603	diet guide	5.04	84.5%	0.828
suit pants	5.8	20%	0.544	get tip	4.84	18.5%	0.704
quick boost	5.8	62.5%	0.618	crab dish	4.8	23.5%	0.671
cheek pain	5.73	67.5%	0.578	lip care	4.68	64%	0.74
keep pace	5.67	13.33%	0.606	top gear	4.64	38.5%	0.689
drug gang	5.6	12.5%	0.639	neck tie	4.6	52.5%	0.677
grab cash	5.6	55%	0.514	quick boost	4.52	57.5%	0.818
beat down	5.53	8.33%	0.585	lap dance	4.48	27%	0.587
blog page	5.53	29.17%	0.562	lip balm	4.36	22.5%	0.859
kick down	5.53	45.83%	0.615	cheek pain	4.36	59.5%	0.806
grab pen	5.47	2.5%	0.567	bed bug	4.12	20.5%	0.811
lip care	5.47	76.67%	0.606	back kick	3.96	26%	0.719
top gear	5.27	49.17%	0.575	boot camp	3.96	73%	0.766
diet guide	5.2	22.5%	0.532	rib pain	3.88	17%	0.737
club girl	5.2	51.67%	0.565	drag car	3.52	19.5%	0.651
skip test	5.2	45%	0.618	mud day	3.48	16.5%	0.72

back kick	4.93	17.5%	0.603	deep bite	3.48	16.5%	0.847
drag car	4.73	10%	0.553	rub down	3.32	38%	0.677
get tip	4.73	10%	0.666	rub back	3.16	12%	0.757
rub back	4.73	10%	0.626	head game	3.16	60%	0.758
shade tent	4.67	15.83%	0.695	shade tent	3.08	27.5%	0.747
job bank	4.6	1.67%	0.555	deck gate	3	32.5%	0.825
deep bite	4.4	13.33%	0.65	tape gun	3	40%	0.798
tape gun	4.33	46.67%	0.6	drug gang	2.96	23%	0.779
deck gate	4.2	15%	0.645	pep talk	2.84	39.5%	0.621
smug king	3.93	22.5%	0.759	smug king	2.64	13%	0.648
glib talk	3.8	52.5%	0.658	bag deal	2.28	35.5%	0.745
fab gift	3.67	60%	0.62	vet debt	2.2	34.5%	0.757
bag deal	3.6	18.33%	0.609	glib talk	2.08	40%	0.82
vet debt	3	12.5%	0.781	frat boy	1.92	36%	0.705
mud day	3	10%	0.808	fab gift	1.88	36.5%	0.857
hug pug	3	47.5%	0.678	dock date	1.84	50.5%	0.764
dock date	2.6	58.33%	0.683	hug pug	1.48	57.5%	0.865
Average	5.66	30.83%	0.609		4.74	35.59%	0.733

Table 4.13. Summary of the lexical rating score and the overlap for both groups

Pearson two-tailed correlation tests were conducted to examine whether there was a correlation between lexical frequency and release percentage, and between lexical frequency and overlap degree. The results showed that for the 72 real words, lexical frequency and release percentage were not correlated in either of the groups (the NE group: r=.024, p>.839; the NM group: r= -.053, p>.657); on the other hand, the results found that for both groups the frequency rating and the closure duration ratio were significantly negatively correlated (the NE group: r= -.511, p< .001; the NM group: r= -.349, p< .003): the more frequent the item was, the shorter the closure duration was (i.e., the more the overlap). H4 was then partially supported: it was supported in terms of a significant correlation between frequency and duration overlap, but not supported in terms of lacking a correlation between frequency and release percentage. In the following section, I will further break the data down and examine other research hypothesis related to the categorical frequency effect.

4.4.2. Hypothesis 5

H5: English speakers will show the predicted release pattern of the POE in both real words and nonwords, while Mandarin speakers will not.

As reported in Section 4.2, the categorical *Frequency* effect (i.e., real words vs. nonwords) was significant on release percentage and closure duration ratio at both speech rates. To give a visual representation, Figure 4.6 below plots the categorical *Frequency* effect on both groups' release patterns over both speech rates. Figure 4.6 suggests that both groups produced more releases in nonwords than in real words (The NE group: 39.5% vs. 30.85%; The NM group: 43% vs. 35.2%). In terms of release percentage, no significant group differences were found in either lexical setting (both p> .061).

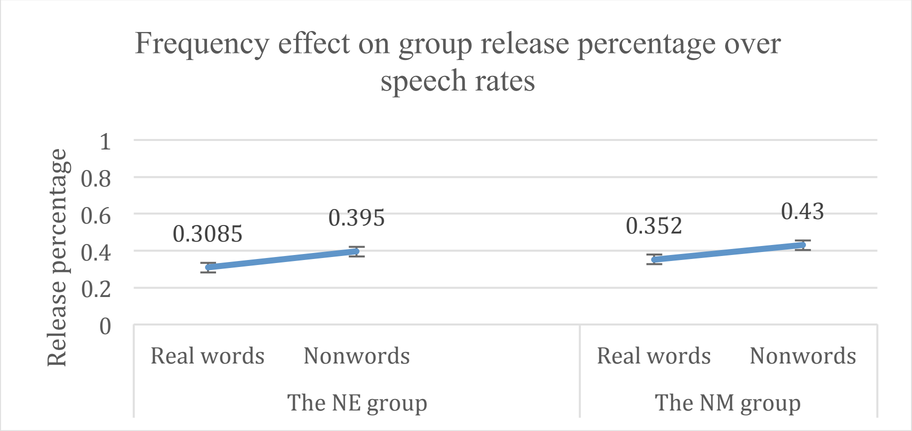


Figure. 4.6. Frequency effect on release percentage for both groups over speech rates

Figure 4.7 plots the categorical *Frequency* effect on both groups' closure duration ratio patterns over both speech rates; the higher the line, the smaller the gestural overlap the group had. Figure 4.7 shows that both groups produced more overlap in real words than in nonwords (The NE group: 39.4% vs. 32.8%; The NM group: 24.8% vs. 12.6%). Group ratio differences were significant in nonwords (F(1.38)=4.121, p<.049) but not in real words (F(1,38)=2.728, p>.107). This suggests that both groups coordinated the real words in the same fashion; however, the NM group failed to coordinate the nonwords in the same way that the NE group did.

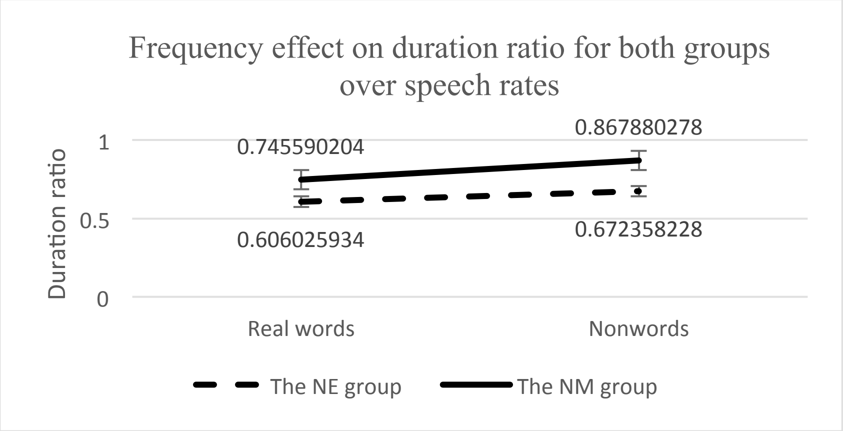


Figure. 4.7. Frequency effect on duration ratio for both groups over speech rates

As reported in Section 4.3, the NE group showed a significant POE in terms of more releases in front-back than in back-front clusters while the NM group did not. To examine H5, I need to confirm that the NE group showed the POE in both real words and nonwords. Figure 4.8 below plots the *Frequency * POA* interaction on both groups' release patterns over both speech rates. The release order showed that both groups had the same trend of front-back > back-front > homorganic in each lexical setting.

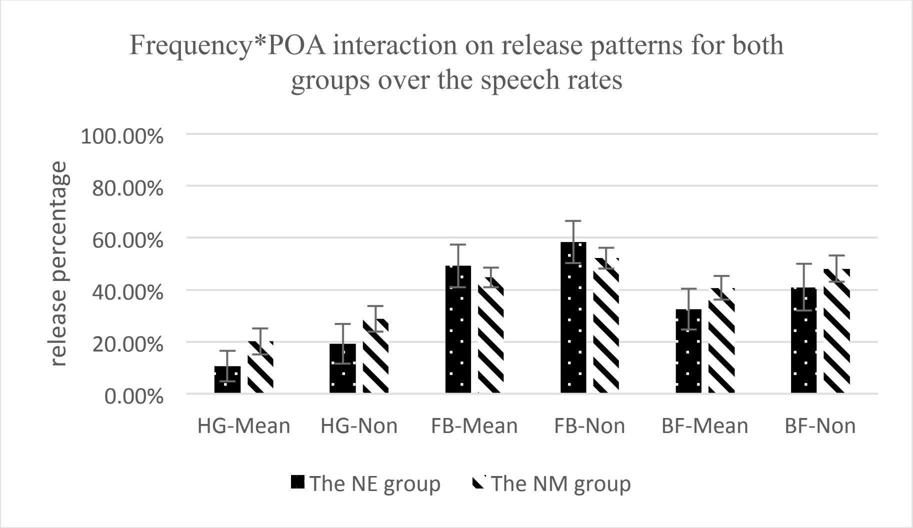


Figure. 4.8. Frequency * POA Interaction on release percentage for both groups

Note: HG=Homorganic. FB=Front-Back. BF=Back-Front. Mean=Real words. Non=Nonwords

To further analyze the data, Table 4.14 below summarizes the mean release percentage of the three cluster types in both real words and nonwords for each group at the slow rate (SD given in parentheses). The statistical analyses confirmed that both groups had significantly ***more*** releases in nonwords than in real words at the slow rate (both p<.001). This suggests that the *Frequency* effect was similar in both groups. Despite a trend that the NM group had more releases than the NE group at the slow rate, the overall group differences were not significant in either real words (MD= -.112, p>.087) or nonwords (MD= -.098, p>.18).

Group	The NE group			The NM group		
Cluster type	Homorganic	Front-Back	Back-Front	Homorganic	Front-Back	Back-Front
Real words	6.9% (.19)	49.2% (.27)	29.2% (.19)	21.7% (.32)	51.5% (.21)	45.7% (.27)
	M=28.4%			M=39.6%		
Nonwords	16.7% (.28)	57.2% (.25)	38.3% (.26)	29.8% (.36)	58.5% (.22)	53.2% (.28)
	M=37.4%			M=47.2%		

Table 4.14 Frequency effect on release percentage for each group at the slow rate

Table 4.14 indicates that in general, both groups had the same release trend of front-back > back-front > homorganic clusters in both real words and nonwords at the slow rate. The statistical results showed that, in both lexical settings, the NE group had significantly more releases in front-back clusters, followed by back-front clusters, then homorganic clusters: front-back > back-front > homorganic (all p<.001). The NM group had a similar pattern with front-back > homorganic and back-front > homorganic in both lexical conditions (all p<.001). However, the NM group did not distinguish front-back and back-front clusters significantly in either lexical environment (in the real words: MD= .58, p>.614; in the nonwords: MD=.053, p>.88). These results indicate that both groups showed the same release order at the slow rate: front-back > back-front > homorganic. The difference was that the NE group statistically distinguished front-back and back-front clusters while the NM group did not. This means that H5 is supported at the slow rate.

The post hoc comparisons found more detailed group differences. The NE group had significantly fewer releases than the NM group under VC2 in real words (MD= -.167, p<.025), but this group difference under VC2 was not significant in nonwords (MD= -.151, p>.082). Also, the NE group had significantly fewer releases than the NM group in producing meaningful homorganic clusters (MD= -.147, p<.041) and meaningful back-front clusters (MD= -.165, p<.047), but the group differences in nonsense homorganic and back-front clusters were not significant (both p>.068). These results further explain why group differences did not extend into the nonwords.

At the fast rate, Table 4.15 below presents the mean release percentage under each cluster type for each group with SD given in parentheses. The statistical analyses confirmed that both groups had significantly ***more*** releases in nonwords than in real words (both p<.001). This suggests a similar *Frequency* effect on both groups, comparable to their responses at the slow rate.

Group	The NE group			The NM group		
Cluster Type	Homorganic	Front-Back	Back-Front	Homorganic	Front-Back	Back-Front
Real words	14.4% (.21)	49.3% (.21)	36% (.19)	18.5% (.18)	38.1% (.17)	35.7% (.18)
	M=33.2%			M=30.8%		
Nonwords	21.7% (.26)	59.7% (.21)	43.5% (.27)	27.7% (.21)	45.8% (.17)	43.1% (.22)
	M=41.6%			M=38.8%		

Table 4.15 Frequency effect on release percentage for each group at the fast rate

At the fast rate, the NE group had the same release pattern in both lexical settings: front-back > back-front > homorganic (all p<.002), similar to their slow rate data. The NM group also had front-back > homorganic and back-front > homorganic in both lexical settings (all p <.001). Again, the NM group did not distinguish front-back and back-front clusters significantly in either lexical setting (both p>.06). Simply put, both groups exhibited the same release pattern at the fast rate: front-back > back-front > homorganic; but the NE group distinguished front-back and back-front clusters significantly while the NM group did not. This means that H5 is statistically supported at the fast rate.

More detailed group differences emerged in post hoc analyses. The NE group had significantly more releases than the NM group in producing ***nonsense*** front-back clusters under VC1, VC3 and VC4 (all p<.038); these group differences were not significant in real words (all p>.07). The NE group also released more often than the NM group in ***meaningful*** front-back clusters under VC3 and VC4 (both p<.024). That is, the NE group released more often in front-back clusters when C1 was voiced (e.g., *lab test, good game*). Again, these results suggest that the group differences were not consistent across lexical settings, suggesting different coordination patterns varying by lexical frequency.

To sum up, both groups had the same release order as front-back> back-front> homorganic clusters in both lexical settings. H5 is supported in the sense that the NE group significantly exhibited the predicted release pattern (i.e., front-back > back-front) of the POE in both lexical settings but the NM group did not in either lexical condition.

4.4.3. Hypothesis 6

H6: English speakers will exhibit the predicted temporal overlap pattern of the POE in both real words and nonwords, while Mandarins speakers will not.

To examine H6, I looked at whether English and Mandarin speakers had significantly more overlap in front-back than in back-front clusters in each of the two lexical settings. Figure 4.9 (Error bars: 95% CI) below plots the *Frequency * POA* interaction on both groups' ratio patterns. The graph illustrates that the NE group had the same pattern in both real words and nonwords. This pattern is visible as front-back clusters having the most amount of overlap, followed by back-front clusters, then homorganic clusters in both lexical environments (i.e., front-back >> back-front >> homorganic). On the other hand, the NM group had different organizations. In real words, the NM group had the most overlap in homorganic clusters, followed by front-back then back-front clusters (i.e., homorganic >> front-back >> back-front). In nonwords, back-front clusters allowed more overlap than front-back clusters (i.e., homorganic >> back-front >>front-back).

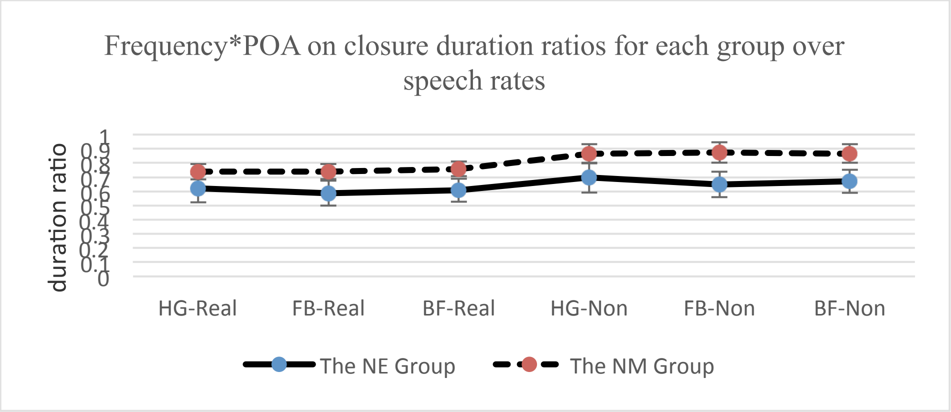


Figure. 4.9 *Frequency * POA* interaction on closure duration ratio for both groups

Note: The higher the line, the smaller the gestural overlap. Each point represents the duration ratio in meaningful homorganic clusters, meaningful front-back clusters, meaningful back-front clusters, nonsense homorganic clusters, nonsense front-back clusters, and nonsense back-front clusters.

As reported in Section 4.3, both groups showed the right trend of the POE, but neither of the group produced significantly more overlap in front-back than in back-front clusters (both p>0.68). In the following paragraphs, I will further analyze the data to examine how *Frequency* affects overlap degree.

At the slow rate, Table 4.16 below provides the descriptive data for the two groups. It includes the mean closure duration ratios in each type of clusters in both lexical settings as well as the standard deviations (given in parentheses) of these numbers. Also included in this table are the overall mean duration ratios in real words and nonwords for each group. For example, the table shows that the NE group had 0.604 duration ratio in producing real words (i.e., 39.6% overlap) and 0.67 in producing nonwords (i.e., 33% overlap) at the slow rate. Meanwhile, the NM group had 0.757 ratio in producing real words (i.e., 24.3% overlap) and 0.854 in nonwords (i.e., 14.6% overlap) at the slow rate. Overall, both groups had significantly more overlap in real words than in nonwords (The

NE group: MD= -.067, p<.005; The NM group: MD= -.097, p<.004). This means the categorical *Frequency* effect was similar on both groups.

Group	Real Words			Nonwords		
	Homorganic	Front-Back	Back-Front	Homorganic	Front-Back	Back-Front
The NE group	0.613 (.34)	0.588 (.31)	0.61 (.29)	0.687 (.33)	0.645 (3)	0.68 (.29)
	M=0.604			M=0.67		
The NM group	0.754 (.33)	0.751 (.31)	0.766 (.3)	0.849 (.37)	0.86 (.38)	0.853 (.38)
	M=0.757			M=0.854		

Table 4.16 Frequency effect on closure duration ratio for both groups at the slow rate

Based on Table 4.16, the NE group seemed to have an overlap order of front-back >> back-front >> homorganic in both real words and nonwords. However, the statistical results showed that the timing differences between front-back and back-front clusters produced by the NE group were not significant in either lexical setting (in the real words: MD= -.022, p>1; in the nonwords: MD= -.035, p>.713). The NM group seemed to have an overlap order of front-back >> homorganic >> back-front clusters in real words and had homorganic >> back-front>> front-back in nonwords. However, the NM group did not show a significant POE either (in the real words: MD= -.014, p>1; in the nonwords: MD= .006, p>1). These results indicated that the overlap pattern predicted by the POE was not statistically shown for either group at the slow rate. This means that H6 cannot be supported at the slow rate.

Worth noting is that the NE group's consistency extended to the nonwords: they had the same overlap order as front-back >> back-front >> homorganic. In contrast, the

NM group had a different order in nonwords than in real words. This result implies one important point: the NE group organized the three types of clusters similarly in both lexical environments while the NM group did not.

Table 4.17 below provides the mean duration ratios in real words and nonwords for each group at the fast rate. It shows that the NE group had 0.608 duration ratio in real words (i.e., 39.2% overlap) and 0.677 ratio in nonwords (i.e., 32.3% overlap) at the fast rate; while the NM group had 0.734 (i.e., 26.6% overlap) and 0.882 (i.e., 11.8% overlap), respectively. Both groups had significantly more overlap in real words than in nonwords (The NE group: MD= -.066, p<.008; The NM group: MD= -.148, p<.001). Similar to performance at the slow rate, the NE group had more overlap than the NM group at the fast rate; but the overall group differences were not significant in either lexical condition (both p >.072).

Group	Real words			Nonwords		
	Homorganic	Front-Back	Back-Front	Homorganic	Front-Back	Back-Front
The NE group	0.632 (.46)	0.585 (.39)	0.608 (.38)	0.701 (.52)	0.654 (.43)	0.661 (.37)
	M=0.608			M=0.674		
The NM group	0.725 (.28)	0.729 (.28)	0.748 (.25)	0.88 (.34)	0.886 (.38)	0.879 (.32)
	M=0.734			M=0.882		

Table 4.17 Frequency effect on closure duration ratio for both groups at the fast rate

Table 4.17 shows the NE group had the overlap order of front-back >> back-front >> homorganic in both real words and nonwords. Similar to their responses at the slow rate, the timing differences between front-back and back-front clusters were not significant in the NE group in either lexical condition (both p>1). On the other hand, the NM group had homorganic >> front-back >> back-front in the real words and had back-front >> homorganic >> front-back clusters in the nonwords. Similar to the NE group, the NM group did not show a significant POE in either lexical condition (both p>1). Therefore, H6 is not supported at the fast rate.

Significant group differences emerged from post hoc analyses. The NE group had significantly more overlap than the NM group in coordinating nonsense back-front clusters at the fast rate (MD= -.218, p<.033). The NE group also had more overlap than the NM group under VC2 in nonwords at both speech rates (both p<.048).

In sum, the analyses showed no significant timing differences between front-back and back-front clusters in either lexical setting for either group. This finding does not support H6. However, the results found that the NE group had a consistent overlap order of front-back >> back-front >> homorganic across lexical conditions regardless of the speech rate. In contrast, the NM group had different orders varying upon lexical and speech rate changes. The post hoc analyses found group differences were significant in cases such as nonsense + VC2 and nonsense + back-front; but these differences did not transfer across lexical settings.

4.4.4. Summary for Research Question 2

The previous subsections have reported on the results and statistical analyses of the overall data with respect to RQ 2. RQ2 asked a) Will both groups of speakers coordinate clusters in high-frequency words the same way as clusters in low-frequency words? and b) Will both groups of speakers coordinate clusters in real words the same way as clusters in nonwords?

In this study, I considered two aspects of frequency effects: the gradient effect (high vs. low frequency real words), and the categorical effect (real words vs. nonwords). To examine RQ2a, I tested whether there was a correlation between gradient frequency and overlap; to examine RQ2b, I examined whether English and Mandarin speakers would exhibit the POE in both real words and nonwords.

Table 4.18 summarizes the predictions, the comparisons as well as the results discussed in this section. The predictions and results in italics indicate significant differences, and those not in italics represent trends. The Yes or No indicate whether the hypothesis is statistically supported (Yes) or not (No). For example, H5 predicts that the NE group would have significantly more releases in front-back than in back-front clusters in both lexical conditions but the NM group would not. The results confirmed that the release differences between the two types of clusters were significant in the NE group, but not in the NM group. Therefore, H5 is statistically supported (Yes).

	Predictions	Comparisons	Results	Hypotheses
Research Question 2: *Will lexical frequency have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?*	H4 *Both groups have a significant correlation between frequency and release percentage, as well as a significant correlation between frequency and duration ratio*	Both groups' release and duration ratios in real words	Neither group has a significant correlation between frequency and release percentage *Both groups have a significant correlation between frequency and duration*	H4: partially Yes
	H5 *a) The NE group:* *FB > BF in both lexical settings* b) The NM group: inconsistent	Both groups' release data in real words and nonwords	*a) The NE group:* *FB > BF in both lexical settings* b) The NM group: FB > BF in both lexical settings	H5: Yes
	H6 *a) The NE group: :* *FB >> BF in both lexical settings* b) The NM group: inconsistent	Both groups' duration ratios in real words and nonwords	a) The NE group: FB>>BF in both lexical settings b1) The NM group: FB>> BF in real words b2) BF>> FB in nonwords	H6: No (in the right trend)

Table 4.18 Results summary for research question two

Table 4.14 also shows that the NE group had a consistent overlap order of front-back >> back-front in both lexical settings; the NM group had inconsistent temporal overlap patterns varying upon lexical environments. However, the timing differences between the two types of clusters were not significant in either group. H6 was thus not statistically supported. These results provide evidence that sound patterns are represented differently in native versus nonnative processing (more in Section 5.2).

4.5. Research Question 3

RQ3 asks: Will English and Mandarin speakers have the same stress effect on stop-stop coarticulation? Specifically, English and Mandarin differ in rhythmic properties. The former is considered a stress-timed language in which stressed words have longer durations and less coarticulatory effects than unstressed words (Avery & Ehrlich, 1992; Celce-Murcia et al., 1996; Cho, 2004; Grabe & Low, 2002; Ramus et al., 1999; Weinstein, 2001). The latter is considered a syllable-timed language where each syllable receives equal timing (Chen et al., 2001; Lin & Wang, 2007; Lin & Huang, 2009; Tajima et al., 1997). The arising question is: Will speakers from the two languages differ in cluster coordination based on their experience with stress placement? To explore this question, the focus in this section is on the different lexical stress positions in Task 3 (C1C2 compound vs. C1C2 non-compound) and four different sentential stress positions embedded in Task 4 (see section 3.3.2): FOCUS 1 refers to when the first word is focused, **FI**F2; FOCUS 2 refers to when the second word is focused, F1**F2**; FOCUS 3 refers to when both words are focused, **FIF2**; FOCUS 4 refers to when neither of the words is focused, F1F2.

4.5.1. Hypothesis 7

H7: English speakers will release less often in words when they are lexically and sententially unstressed than when they are lexically and sententially stressed, while Mandarin speakers will not make such a difference.

To confirm H7, the NE group is expected to have significantly more releases in C1C2 than in C1C2 at the slow rate, and to have significantly more releases in **FI**F2 (the word in which C1 is focused) than in F1**F2** (the word in which C2 is focused) and in F1F2 (no focus) at the fast rate; the NM group is expected not to have significant differences in either pair.

As reported in Section 4.2.1, the *Compound* effect was significant on release percentage. Figure 4.10 below displays the release patterns for each group under compounds versus non-compounds condition. The NE group released 37.5% in compounds and 29.4% in non-compounds; the NM group released 45.1% and 41.7%, respectively. Both groups released more often in compounds (i.e., C1C2) than in non-compounds (i.e., C1C2). The release differences between compounds and non-compounds turned out to be significant in both groups (both p<.02). Group differences were not significant in either compounds or non-compounds (both p>.117). The post hoc comparisons revealed that group differences were significant in back-front clusters: the NM group significantly released more often than the NE group in non-compounds (e.g., *bed bug*, *head pain*) (p<.05).

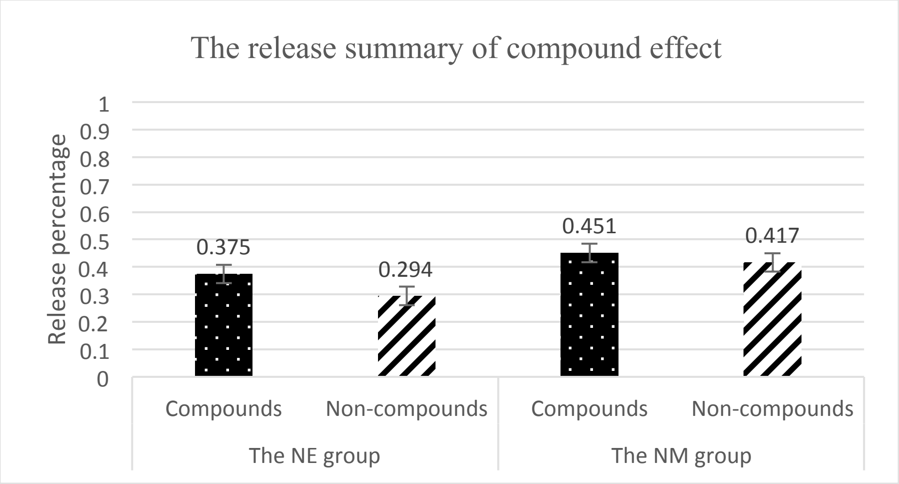


Figure 4.10 Summary of compound effect on release patterns for both groups

As reported in Section 4.2.2, the *Focus* effect was also significant on release percentage. Figure 4.11 below plots both groups' release patterns under each *Focus* condition (error bars: 95% CI). The graph suggests that both groups had the most amount of releases under FOCUS 1 (i.e., **FI**F2) and had the least amount of releases under FOCUS 4 (i.e., F1F2). Both groups had the release order as **FI**F2 > **FIF2** > F1**F2** > F1F2. Figure 4.11 indicates a trend that the NE group released more often than the NM group in all four focus positions. However, the overall group differences were not significant at any focus condition (all p>.447). The statistical results showed that the NE group always released significantly more often under **FI**F2: **FI**F2 > F1**F2**, **FIF2**, and F1F2 (all p<.001). The same focus effect was found in the NM group. The NM group always had significantly more releases in **FI**F2 than in other focus conditions: **FI**F2 > F1**F2**, **FI**F2 > **FIF2**, and **FI**F2 > F1F2 (all p<.001).

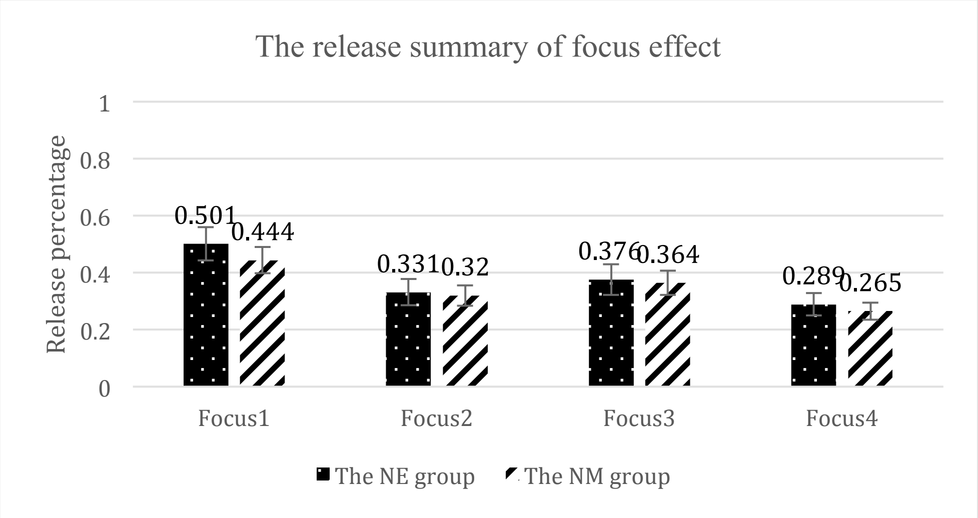


Figure 4.11 Summary of focus effect on release patterns for both groups

Combining results from the word level and the sentence level, H7 was partially supported in that the NE group showed a stress effect in terms of compound (i.e., lexical stress) and focus (i.e., sentential stress), but so did the NM group. Moreover, the differences between F1**F2** and F1F2, and **FIF2** and F1F2 were significant in the NM group (both p <.003) but not in the NE group.

The post hoc analyses found group differences in front-back clusters: the NE group released significantly more often than the NM group in front-back clusters under **FI**F2 and F1**F2** (both p<.048). Also, the NE group released significantly more often than the NM group in nonwords under the combined condition of **FI**F2 and VC3 (MD=.223, p<.019), and F1F2 and VC1 (MD=.168, p<.031). These results suggested that the NM group always released more often as long as there was focus in a cluster, suggesting an even stronger sentential stress effect for the NM group than for the NE group. In the following subsection, I will examine timing differences for both groups under lexical and sentential stress.

4.5.2. Hypothesis 8

H8: English speakers will produce more gestural overlap in words when they are unstressed both lexically and sententially than when they are stressed, while Mandarin speakers will not exhibit significant temporal differences.

To confirm H8, the NE group is expected to have significantly more overlap in non-compounds (i.e., C1C2) than in compounds (i.e., C1C2) at the slow rate, and to have significantly less overlap in **F1**F2 than in F1**F2** and in FIF2 (no focus) at the fast rate; the NM group is expected not to have significant differences in either pair.

As reported in Section 4.2.1, the *Compound* effect was not significant on the duration ratio. Figure 4.12 below shows both groups' duration ratio patterns in compounds versus non-compounds condition (Error bars: 95% CI). The results discovered an opposite effect for the two groups. The NE group showed a trend of having more overlap in compounds than in non-compounds without significant differences (p>.624); the NM group had significantly more overlap in non-compounds than in compounds (p<.013). Group differences were found to be marginally significant in compounds (p>.066): the NE group had more overlap than the NM group in compounds. These results contradicted with H8 in that the NM group showed a strong compound effect but the NE group did not.

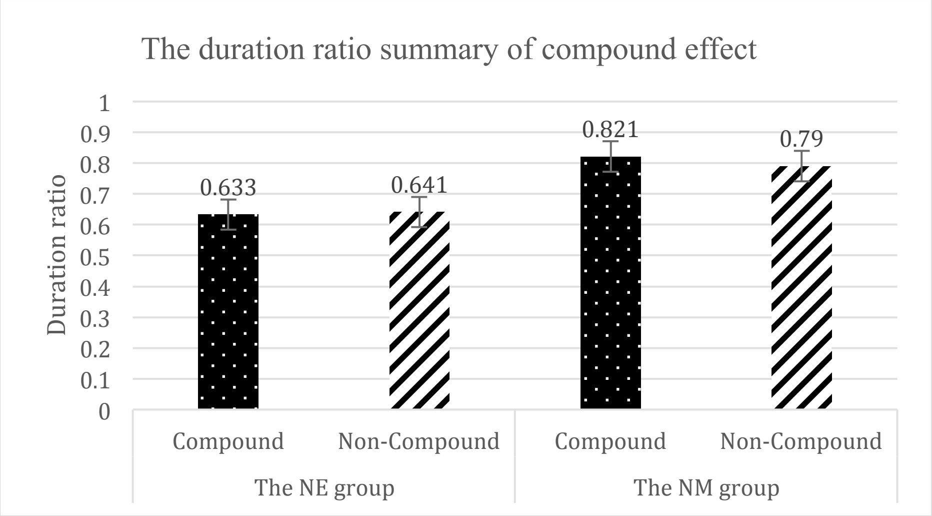


Figure 4.12 Compound effect on closure duration ratio for both groups

As reported in Section 4.2.2, the effect of *Focus* was significant on closure duration ratio. Figure 4.13 below presents a visual representation of both groups' duration ratio patterns under each *Focus* condition (Error bars: 95% CI). The NE group had the overlap order as F1F2 >> **<u>FIF2</u>** >> **<u>FI</u>**F2 >> F1**<u>F2</u>**; the NM group had the overlap order as F1F2 >> **FIF2** >> F1**F2 >> FI**F2. The graph displays that the NE group had a smaller duration ratio than the NM group in all four focus positions. Similar to release patterns, no group differences were found significant in each focus condition (all p>.071).

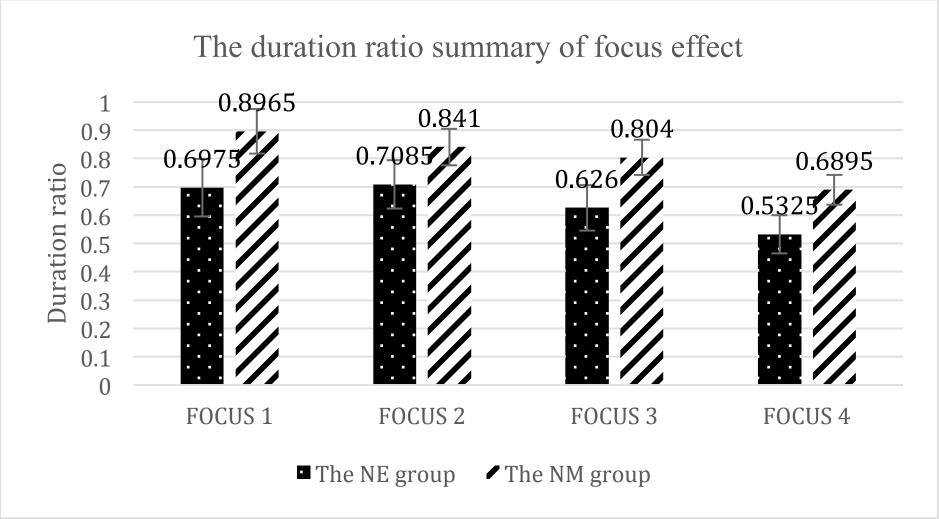


Figure 4.13 Focus effect on closure duration ratio for both groups

Both groups tended to have greater overlap in F1F2 than in other focus conditions. Specifically, the NE group had significantly more overlap in **FIF2** and F1F2, shown as: F1F2 >> **FI**F2 and F1**F2** (both p<.001), **FIF2** >> F1**F2** (p<.001). The same pattern was shown in the NM group: F1F2 >> **FI**F2, F1**F2,** and **FIF2** (all p<.014), **FIF2 >> FI**F2 (p<.014). These results indicate that both groups had more overlap under F1F2 compared to other focus positions (all p<.03). This finding means that both groups significantly shortened closure durations in words without focus than in words with focus. Contrary to the prediction, neither group produced significantly more overlap in F1**F2** than in **F1**F2. Interestingly, the results showed that the NM group was consistent with the trend by having more overlap in F1**F2** than in **F1**F2, while the NE group had slightly more overlap in **F1**F2 than in F1**F2**. Thus, H8 was not supported.

Post hoc comparisons found significant group differences under F1F2. The NE group had more overlap than the NM group in coordinating nonwords under F1F2 (MD= -.199, p<.043). Group differences were also found in front-back and back-front clusters under **FIF2** and F1F2. The NE group had significantly more overlap than the NM group in front-back clusters under **FIF2** (MD= - .212, p<.048). The NE group also had significantly more overlap than the NM group in back-front clusters under **FIF2** and F1F2 (both p<.04).

The difference was that the NM group had significantly more overlap under F1F2 than under **FIF2** (p<.001), but this difference was not significant in the NE group (p>.06). In fact, the NM group had more overlap under F1F2 than under other stress positions in both front-back and back-front clusters (all p<.005), while the NE group did not (both p> .154). These results showed that, similar to the release patterns, the NM group exhibited an even stronger stress effect than the NE group in that the NM group always had much more overlap in clusters without stress.

In all, H8 is not supported in that the NM group, but not the NE group, significantly exhibited the compound effect, and showed an even stronger focus effect than the NE group.

4.5.3. Summary for Research Question 3

RQ 3 asks whether the NE and the NM groups would exhibit the same stress effect. I discussed the stress effect in two aspects, the lexical stress (i.e., compound effect) and the sentential stress (i.e., the focus effect). Table 4.19 below summarizes the predictions and the comparisons along with the results concerning RQ 3, which also includes whether each hypothesis was statistically supported or not. The predictions and results in italics indicate significant differences, and those not in italics represent trends.

Research	Predictions	Comparisons	Results	Hypotheses
Question 3: Will stress patterns have the same effect on English and Mandarin speakers' stop-stop coarticulation?	H7 a) The NE group *C1C2 > C1C2* ***F1**F2 > F1F2* b) The NM group inconsistent	Both groups' release data in different lexical and sentential stress positions	The NE group C1C2 > C1C2 The NM group *C1C2 > C1C2* Both groups: ***F1**F2 > F1F2*	H7: No
	H8 a) The NE group: *C1C2 >> C1C2* *F1**F2** >> **F1**F2* *F1F2 >> **F1**F2* b) The NM group: inconsistent	Both groups' duration ratios in different lexical and sentential stress positions	The NE group C1C2 >> C1C2 The NM group *C1C2 >> C1C2* The NE group **F1**F2 >> F1**F2** The NM group F1**F2** >> **F1**F2 Both groups: *F1F2 >> **F1**F2*	H8:No

Table 4.19 Results summary for research question three

Specifically, the first prediction is that the NE group would have significantly more releases in C1C2 than in C1C2 and more releases under **FI**F2 than under F1**F2** / F1F2 while the NM group would only show a trend. The results discovered that both groups had significantly more releases in C1C2 than in C1C2. With respect to sentential stress, both groups had significantly more releases in **F1**F2 than in F1**F2** and F1F2, while the prediction was that the NM group would not exhibit significant differences under either lexical or sentential stress. Thus, H7 was not supported. Similarly, H8 was not supported in that the NM group significantly showed both the compound and focus effects but the NE group did not. I will discuss these results in Section 5.4.2.

4.6. Research Question 4

The fourth primary research question asks: Will speech rate affect English and Mandarin speakers similarly in their stop-stop coarticulation? The hypothesis is that gestural overlap will increase when speech rate is increased. As I reported in Section 3.3.2, based on the results from the speech rate script, all participants produced more syllables per second in Task 4 (targets embedded in sentences) than in Task 3 (targets in isolation). Two-tailed t-tests showed that both groups had significantly more syllables per second at the fast rate than at the slow rate. The reported statistical analyses so far were conducted separately for Task 3 (slow rate) and for Task 4 (fast rate). Thus, to see if speech rate has an effect on overlap, I need to compare overlap in Task 3 (slow rate) and Task 4 (fast rate).

4.6.1. Speech Rate Effect

Table 4.20 below provides the descriptive statistics of both release percentage and closure duration ratio data, which include the mean release percentage and closure duration ratio for each group at each speech rate (SD given in parentheses). It suggests that the NE group released more often at the fast rate than at the slow rate (37.4% vs.

33%), while the NM group released less often (34.8% vs. 43.4%) at the fast rate than at the slow rate. Also, the NE group had slightly smaller gestural overlap at the fast rate than at the slow rate (35.9% vs. 36.3%) while the NM group had greater overlap at the fast rate (19.2% vs. 18.2%).

Group	The NE group		The NM group	
Variables	Release Percentage	Closure Duration ratio	Release Percentage	Closure Duration ratio
Slow Rate	33% (.36)	0.637 (.3)	43.4% (.38)	0.818 (.29)
Fast Rate	37.4% (.34)	0.641 (.43)	34.8% (.28)	0.808 (.32)
Average	35.2%	0.639	39.3%	0.813

Table 4.20. Summary of release percentage and closure duration ratio for each group at both speech rates

As mentioned in Section 3.4.3, the main effect of *Speech Rate* on release percentage was not significant. The statistical analyses showed that the interaction between *Speech Rate* and *Group* on release percentage was significant (F(1,38)=5.977, p<.019), reflecting the fact that the two groups released differently when the speech rate changed. Specifically, the NE group released less often in a slower speech rate (i.e., at the slow rate), while the NM group released less often in a faster speech rate (i.e., at the fast rate). No significant group differences were found at either slow or fast rate in terms of release percentage (both p>.121).

4.6.2. Hypothesis 9

H9: English and Mandarin speakers will have more gestural overlap when speech rate is increased.

Of special interest among the statistical results was that the main effect of *Speech Rate* was not significant on closure duration ratio (Section 4.2.3). The results showed that the NE group had a smaller ratio (i.e., greater overlap) at the slow rate than at the fast rate, while the NM group had a greater ratio (i.e., less overlap) at the slow rate. The differences between the speech rates were not significant in either group (both p>.805). This means that neither group had significant timing differences when speech rate changed. Meanwhile, the NE group had more overlap than the NM group regardless of the speech rate, but group differences were not significant at either speech rate (at the slow rate, MD= -.18, p>.062; at the fast rate, MD= -.167, p>.114).

In order to further investigate how gestural overlap is implemented to differing degrees by individual speakers, the performance of each participant was examined. Two types of speakers were discerned: those that had more overlap in fast speech (19 speakers, nine NE speakers and 10 NM speakers) (see Figure 4.12), and those that had less overlap in fast speech (21 speakers, six NE speakers and 15 NM speakers) (see Figure 4.13). In Figures 4.14- 4.15, NE stands for native English participants and NM stands for native Mandarin participants.

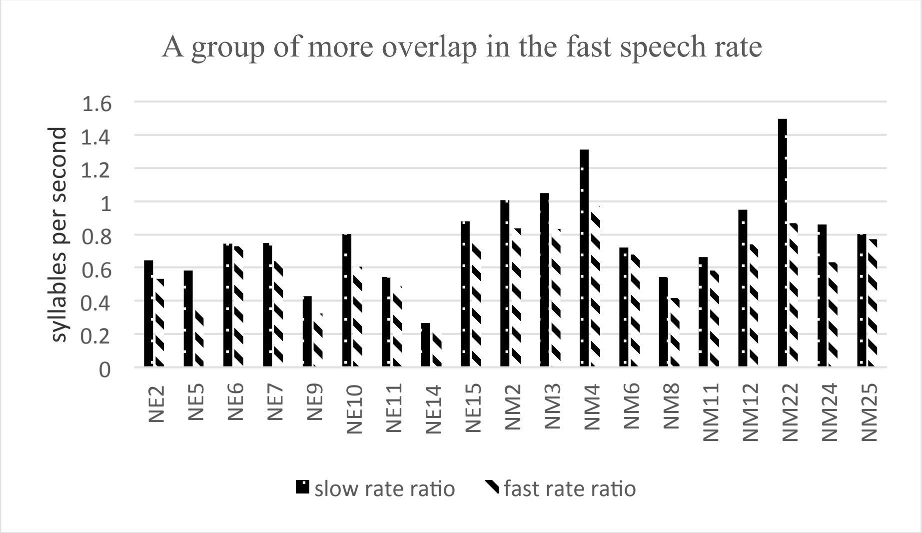


Figure 4.14 Summary of participants who had more overlap in the fast speech rate

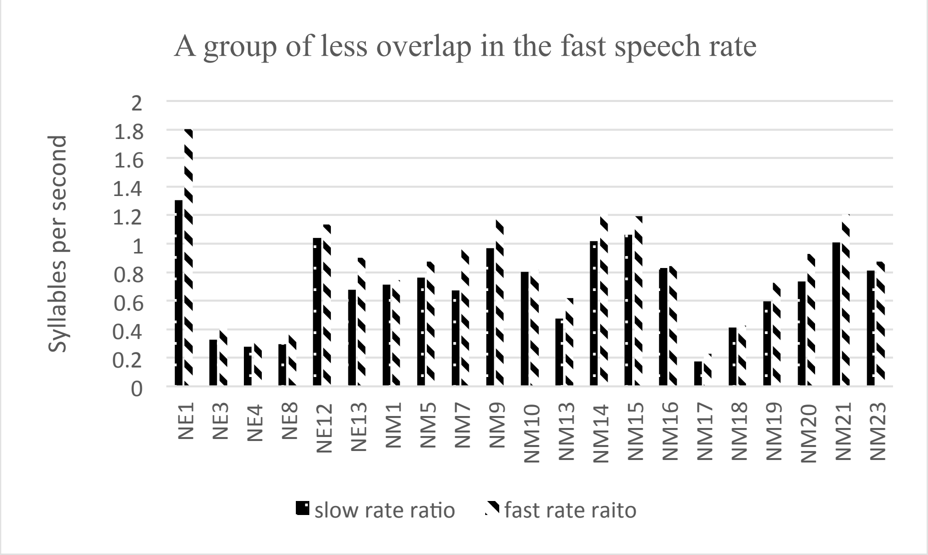


Figure 4.15 Summary of participants who had less overlap in the fast speech rate

There are several reasons why some participants showed less overlap at fast speech rates; these are discussed in detail in the next Chapter. The overall observation is that gestural patterns were not simply accelerated in faster speech rates. For this reason, H9 was not supported. For the first type of speakers, who had more overlap at the fast rate (Figure 4.14), participants' release percentage was correlated with duration ratio during a slow speech rate (r=.555, n=19, p<.014), while no significant correlation was found between release percentage with duration ratio at the fast rate (r=.263, n=19, p>.277). For the second type, who had less overlap during the fast rate, speakers' release percentage was significantly correlated with duration ratio regardless of the speech rate (at the slow rate r=.46, n=21, p<.036; at the fast rate, r=.66, n=21, p<.001).

Neither of the two groups exhibited significantly different overlap patterns among the three positional types of clusters at each speech rate (e.g., overlap in homorganic clusters at the slow rate vs. at the fast rate, all p>.246). This further suggests that both groups had stable timing patterns which were not affected by the speech rate changes.

To sum up, a closer look at individual speakers suggests a large amount of speech variation: speakers could either have more or less gestural overlap at fast rate than at the slow rate. This result indicates that the prediction related to the *Speech Rate* effect is not supported in general; instead, a rather stable overlap pattern is shown and controlled by individual speakers.

4.7. Further analysis

An additional question that data allow us to ask is the following: Is L2 speakers' overlap degree correlated to their perceived accent?[12] As mentioned in Section 3.3.1, the accentedness of the monologue in a pre-task was judged by two native English speakers (one participated in the primary study; the other did not) with a Likert scale from one to seven, where one indicates native-like and seven indicates heavy accent (see Table 3.3).

---

[12] This question had been proposed as a direction for future research in Yanagawa (2006).

To examine the relationship between overlap degree and perceived accent, I averaged perceived accent ratings over two native English-speaking judges and averaged Mandarin participants' closure duration ratio over both speech rates. The results showed that the average accent rating was 3.22 (SD= .77), and the average overlap ratio was 0.82 (SD=.24). Figure 4.16 below plots the relationship between the perceived accent and the overlap degree in the NM group (n=25). The figure does not show a clear pattern. For example, there were three participants who received the lowest accent (2/7), one of them had the closure ratio of 0.2 (80% overlap), another had the closure ratio of 1.2 (no overlap).

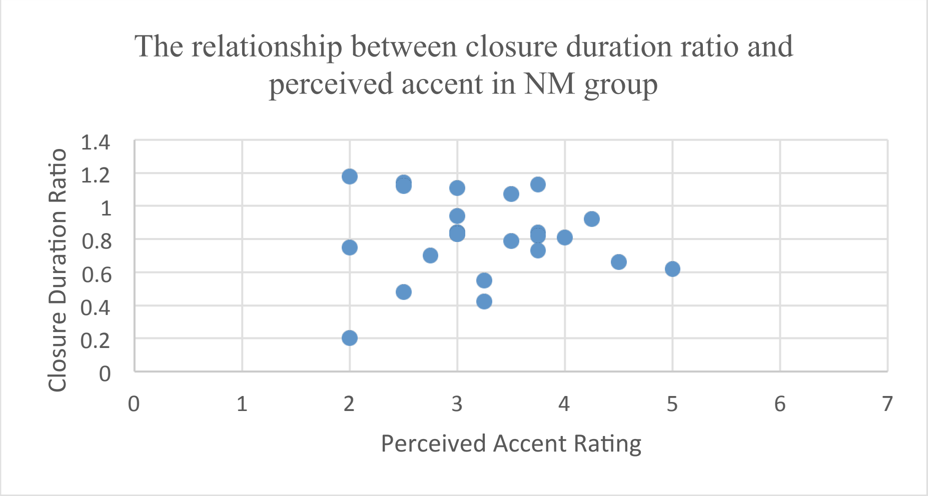


Figure 4.16 The relationship between closure duration ratio and perceived accent

A two-tailed Pearson correlation test was conducted to examine whether there was a correlation between perceived accent rating and temporal ratio. The results showed that they are not significantly correlated (r= -.032, p>.88). This finding was unexpected given the critical importance of the suprasegmental features found in previous literature (e.g., Anderson-Hsieh, Johnson, & Koehler, 1992; Anderson-Hsieh, Riney, & Koehler, 1994; Derwing, Munro, & Wiebe, 1998; Derwing & Munro, 2005). I will discuss the reasons in detail in the next chapter.

4.8. Summary

This chapter has presented and analyzed data in order to examine four primary research questions. These questions concern 1) whether and how *Place of Articulation* affects the gestural coordination for both groups; 2) whether and how *Frequency* affects the coordination for both groups; 3) whether *Stress (i.e., Compound* and *Focus*) has the same effect for both groups; and 4) whether *Speech Rate* has the same effect for both groups.

Firstly, the main effect of *POA* was significant on release percentage but not on closure duration ratio. The overall analyses confirmed the hypotheses that *POA* has different effects on L1 and L2 groups. Specifically, the place order effect (POE) predicts a pattern of more internal releases and more temporal overlap in front-back than in back-front clusters. In terms of the release patterns, both groups showed the trend; however, the NE group's responses significantly confirmed this prediction while the NM group did not. In terms of the timing patterns, both groups tended to have more overlap in front-back than in back-front clusters; but the overlap differences between front-back and back-front clusters were not significant for either of the groups. This means that the overlap pattern predicted by the POE was not statistically supported in the current study. Overall, the POE was only partially supported in native speech but not shown at all in nonnative speech.

Secondly, this study considered two aspects of frequency: gradient and categorical. In terms of the gradient frequency effect, the results discovered that there was no correlation between frequency and release percentage, but a significant correlation between frequency and duration ratio. Both groups had significantly more overlap in words that they were familiar with than in words that they were less familiar with (e.g., *take care* vs. *dock gate*). In terms of the categorical frequency effect, the main effect of *Frequency* was significant on both release and closure duration ratio. Both groups had significantly fewer releases and more overlap in real words than in nonwords (e.g., *keep pace* vs. *peep pate*). A closer inspection revealed that the effect of *Frequency* was different for the two groups. The analyses found that the NE group exhibited the release and temporal trends of the POE consistently in both real words and nonwords. The release pattern (front-back > back-front) was significant across lexical conditions. On the other hand, the NM group exhibited the same release trend but the trend was not significant in either lexical condition. Moreover, they showed different overlap orders in the two lexical environments (i.e., front-back>>back-front in real words, back-front >> front-back in nonwords). These results showed that L1 speakers could transfer their knowledge to new forms while L2 speakers could not, suggesting surface forms are stored differently in native and nonnative processing.

Thirdly, the main effect of *Compound* was significant on release percentage but not on duration ratio. The results discovered that both groups released significantly more often in compounds than in non-compounds. The effect of *Compound* was opposite for the two groups in terms of duration overlap. The NE group showed a trend of having slightly more overlap in compounds than in non-compounds, while the NM group had significantly more overlap in non-compounds than in compounds. The *Focus* effect was significant on both release and closure duration ratio. The analyses showed that despite the different L1 rhythmic patterns, a higher prosodic context (in this case, focus) has the same effect for both groups. The NE and the NM group exhibited the similar sentential stress effect in that they both released significantly less often and had more overlap in F1F2 (no sentential stress) than in **FI**F2 (the first word containing C1 received sentential stress). Unexpectedly, the NM group showed an even stronger stress effect in that they had significantly more overlap in F1**F2** than in **F1**F2 while the NE group had slightly more overlap in **F1**F2 than in F1**F2**; also, the NM group always had significantly fewer releases and more overlap under F1F2 than under any other focus conditions, while this was not the case for the NE group.

Finally, the main effect of *Speech Rate* was not significant on either release or closure duration ratio. The results suggested a non-linear relationship between speech rate and overlap patterns for both groups. Two types of speakers from both groups were evident: a type who had more overlap at faster rates, and another type who had less overlap at faster rates. This result suggested a speaker-dependent effect.

Further analysis was conducted to examine one additional question: whether temporal overlap is correlated with perceived accent. The results showed that temporal overlap is not indicative of perceived accent. NM speakers who received small accent ratings (i.e., very native like), had either 20% overlap or no overlap at all. This unexpected finding suggests that cross-word coarticulation may not be as important as other features (e.g., vowel quality) when judging foreign accent.

Given that one aim of this study was to compare native and nonnative speech, the significant 2-way interactions concerning group differences in this research are summarized again below in Table 4.21 (> indicates more releases, >> indicates more overlap). All comparisons are statistically significant in this table. For example, the table shows that the NM group had significantly more releases than the NE group in lexical homorganic clusters at the slow rate (i.e., NM > NE P1L1). I will elaborate on those results together with other interesting findings in the following chapter.

	At the slow rate	At the fast rate	Averaged over Speech Rate
Release Patterns	NM > NE L1V2 NM > NE P1L1 NM > NE P1V2 NM > NE P2V2 NM > NE P3L1 NM > NE P3V3 NM > NE P3V4	NE > NM P2L2 NE > NM P2F1 NE > NM P2F2 NE > NM P2V1 NE > NM P2V3 NE > NM P2V4	
Temporal Overlap Patterns	NE >>NM L2V2 NE >> NM P2C1 NE >> NM P2V2	NE >> NM L2V2 NE >> NM L2F4 NE >> NM P2F3 NE >> NM P2F4 NE >> NM P3L2 NE >> NM P3F3 NE >> NM P3F4	NE >> NM P2L2 NE >> NM P3L2

Table 4.21. Summary of 2-way interactions on significant group differences

Note: P= Place of articulation. P1, P2, and P3 mean homorganic, front-back and back-front clusters, respectively. L=Lexical. L1 and L2 refer to real words and nonwords. C1 and C2 indicate compounds and non-compounds. V and F refer to Voicing and Focus condition, respectively.

Overall, the results in this chapter point towards three major findings surrounding the three main effects (i.e., *POA, Frequency*, and *Speech Rate*) that determine overlap in

English clusters that cross word boundaries. In addition, the results also suggest two interesting findings that are indicative of modeling sound alternations and L2 speech perception. I will explore and discusses these results in the next Chapter, Chapter 5.

1. The POE was only partially supported in this research. The release pattern was statistically significant in the NE group but not significant in the NM group. The predicted degree of overlap was not significantly shown for either group.
2. The gradient frequency effect was similar for both groups. Both groups exhibited a strong frequency-matching behavior: more frequently used words had more overlap.
3. A group difference was observed in the interaction between categorical frequency * POA. The NE group significantly showed the release pattern of the POE in both real words and nonwords but the NM group did not. The NE group also had consistent temporal overlap orders across lexical conditions, while the NM group had inconsistent overlap orders.
4. Speakers from both groups had either more or less overlap at the fast rate than at the slow rate, suggesting increased speech rate did not always induce increased temporal overlap.
5. The overlap ratio was not correlated with the perceived accent in L2 speech. This unexpected finding suggested that cross-word coarticulation might not be as important as other features (e.g., vowel quality) when judging foreign accent.

Chapter 5

Discussion

Of interest in this study is how gestural overlap is likely to be conditioned by four competing factors: place of articulation, lexical frequency, stress, and speech rate. A number of differences between L1 and L2 speech in producing consonant sequences are predicted on the basis of previous research summarized in Chapter 2. Chapter 3 outlined the procedures of the experiment, which was designed to determine the overlap degree in stop-stop clusters across English words produced by native English (NE) and native Mandarin (NM) speakers. Chapter 4 analyzed and reported on the overall results. The findings raise some important questions about the nature of gestural coordination, how it is implemented, and how it is represented phonologically in both L1 and L2 speech. Many interesting findings and subsequent questions have come to the fore in light of the data analyses.

This chapter discusses each of these findings and questions in depth. To facilitate the discussion, this chapter is divided into five sections. The first section, Section 5.1, provides a review of the research project itself – including the participants, the task, research questions, hypotheses, and the results. Section 5.2 discusses the observed effects of place of articulation and their implications, including homorganic clusters realization (5.2.1), and the place order effect (5.2.2). Section 5.3 concerns the results of lexical frequency, in particular the gradient frequency effect (5.3.1), and its implication in terms of phonological representations in L1 and L2 speech (5.3.2) and language-specificity of overlap patterns (5.3.3). Section 5.4 discusses unexpected results found in this research, including the lack of correlation between overlap and speech rate (5.4.1), the lack of correlation between overlap degree and perceived accent in L2 speech (5.4.2), and the unexpected voicing effect (5.4.3). Finally, section 5.5 concludes the chapter by summarizing the main findings and discussions.

5.1. Research Review

In this section, I will briefly review the research questions and experiment conducted in Chapter 3, along with the results concerning each hypothesis. The overall results of temporal alignment will be briefly discussed as well.

5.1.1. Research motivation

One general question drives this research: what constraints determine the degree of consonantal overlap in L1 and L2 speech? The general assumption is that L2 speakers do not coordinate adjacent gestures the same way that L1 speakers do (e.g., Barlow, 2005; Broselow & Finer, 1991; Colantoni & Steele, 2008; Davidson, 2003, 2006; Solé, 1997; Zsiga, 2003, 2011). In this study, I focused on English and Mandarin, two languages that differ in a) rhythmic category (i.e., stressed timed vs. syllable timed); b) cluster organization (extensive overlap in clusters vs. no clusters categorically); and c) voicing (i.e., both voiceless and voiced stops vs. voiceless stops). Mandarin speakers are hypothesized to have more internal releases and less overlap than English speakers in their ESL productions (e.g., Hansen, 2001; Messing, 2008; Wang, 1995).

Four research questions and their respective hypotheses are as follows. The first research question asks: ***Will place of articulation (i.e., homorganic clusters, front-back***

clusters, and back-front clusters) have the same effect on English and Mandarin speakers' stop-stop coarticulation in English? Specifically, previous studies have shown very few internal releases in homorganic clusters and a place order effect (POE, more releases and more overlap in front-back than in back-front clusters) in native English speech (Byrd, 1996a; Catford, 1977; Henderson & Repp, 1982; Ghosh & Narayanan, 2009; Ladefoged, 2001; Zsiga, 1994, 2000). Will Mandarin speakers exhibit the same effect of place of articulation as English speakers do?

H1: English speakers will have fewer releases and more overlap than Mandarin speakers in producing homorganic clusters.

H2: English speakers will show the predicted release pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.

H3: English speakers will show the predicted temporal overlap pattern of the POE in producing heterorganic clusters, while Mandarin speakers will not.

The second research question concerns whether and how word frequency affects English cluster coordination. It asks: ***Will frequency have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?*** In particular, will both groups of speakers coordinate clusters in high-frequency words the same way as clusters in low-frequency words (see Bush, 2001; Bybee, 2000; Jurafsky et al., 2000)? Moreover, do speakers coordinate clusters in real words the same way as clusters in nonwords (see Berko, 1958; Bush, 2001; Bybee, 2000; Ernestus & Baayen, 2003; Pater & Tessier, 2006; Pycha et al., 2003; Wilson, 2003, 2006)?

H4: Both English and Mandarin speakers will have fewer releases and more overlap in meaningful high-frequency words than in meaningful low-frequency words.

H5: English speakers will exhibit the release pattern of the POE in both real words and nonwords while Mandarin speakers will not.

H6: English speakers will exhibit the temporal overlap pattern of the POE in both real words and nonwords while Mandarins speakers will not.

The third research question concerns the different stress patterns in English and Mandarin. English speakers have shown less coarticulatory effects in stressed segments (e.g., longer duration) than in unstressed segments, while Mandarin speakers are found to transfer the equal syllable prominence from their L1 in ESL speech (e.g., Avery & Ehrlich, 1992; Celce-Murcia et al., 1996; Chen et al., 2001; Cho, 2004; Grabe & Low, 2002; Lin & Huang, 2009; Lin & Wang, 2007; Ramus et al., 1999; Tajima et al., 1997; Weinstein, 2001). The third research question thus asks: ***Will stress patterns have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?***

H7: English speakers will release less often in words when they are unstressed both lexically and sententially than when they are stressed, while Mandarin speakers will not make such a difference.

H8: English speakers will produce more gestural overlap in words when they are unstressed both lexically and sententially than when they are stressed, while Mandarin speakers will not exhibit significant temporal differences.

Given the inconsistent effects of speech rate on gestural overlap seen in the literature (Byrd & Tan, 1996; Kochetov et al., 2007; Kuehn & Moll, 1976; Munhall &

Löfqvist, 1992; Tjaden & Weismer, 1998; Zsiga, 1994), the fourth research question asks: ***Will speech rate have the same effect on English and Mandarin speakers' stop-stop coarticulation in English?***

H9: English and Mandarin speakers will have more gestural overlap when speech rate is increased.

To answer these research questions, a group of 15 native English (NE) speakers and another group of 25 native Mandarin (NM) speakers were recruited in the current study. The stimuli (C1#C2) were created and organized according to three major parameters: place of articulation, word frequency, and stress. For instance, a comparison of homorganic-lexical compound versus homorganic-lexical non-compound will refer to stimuli items like *soup pot* versus *keep pace*. Each sequence was further subcategorized according to voicing conditions, with the stimuli including all possible English stop-stop combinations. The voicing conditions include four combinations: a voiceless C1 + a voiceless C2 (VC1, *keep pace*), a voiceless C1 + a voiced C2 (VC2, *lap dance*), a voiced C1 + a voiceless C2 (VC3, *bad kid*), and a voiced C1 + a voiced C2 (VC4, *good game*).

The experiment consisted of four tasks. Task 1 was to rate lexical familiarity. Task 2 was to produce target words with word-initial and word-final intervocalic stops (V#C_bV and VC_b#V), when they did not occur in a cluster. Task 3 was to read target words with stop-stop sequences (C1#C2) in isolation, and Task 4 was to produce target words in carrier sentences. In Task 4, each target word was embedded in different sentential stress conditions: **FI**F2 (the first word containing C1 received focus), F1**F2** (the second word containing C2 received focus), **FIF2** (both words received focus), and F1F2 (neither of the words received focus).

In sum, this study included all possible stop combinations across words to address research question one (RQ1). A comparison of high frequency versus low frequency words as well as a comparison of real words versus nonwords allow us to examine RQ2. The stimuli design incorporated both compounds (initial stress) versus non-compounds (final stress) at the word level, and different focus positions at the sentence level, to address RQ3. Also, because this study elicited production in both slow (Task 3) and fast speech (Task 4) rates, it examines RQ4.

The two dependent variables were the release percentage and closure duration ratio. Five-factor repeated measures ANOVAs were conducted to analyze the data along five independent factors (*Group, POA, Frequency, Compound,* and *Voicing*) in Task 3 (at the slow rate). Five-factor repeated measures ANOVAs were conducted to analyze the data along five independent factors (*Group, POA, Frequency, Focus,* and *Voicing*) in Task 4 (at the fast rate). After validating the impression that speakers did produce more words per second at the fast rate than at the slow rate, four-factor repeated measures ANOVAs were used to analyze the overall data along four independent factors (*Group, POA, Frequency,* and *Speech Rate)*. In the following section, I will summarize the overall results.

5.1.2. Results overview

As discussed in Section 3.4.1, I follow previous studies (Hardcastle & Roach, 1979; Henderson & Repp, 1982; Zsiga, 2003) in interpreting shortened stop closure durations as indicative of ‘overlap’. Given this interpretation, it is possible to have a released cluster that nevertheless has some degree of overlap, if the release is accompanied by shorter durations of the stop closures (see Figures 5.1 and 5.2), even though from an articulatory point of view, overlap is avoided in open transitions. Thus, there are three possible relationships between release percentage and duration ratio: they could be a) positively correlated, meaning when speakers have more releases, they tend to have longer durations (i.e., less overlap); b) negatively correlated, meaning when speakers have more releases, they tend to have shorter durations (i.e., more overlap); or c) not correlated, meaning there is no systematic relationship between release and duration patterns.

The overall results showed that release percentage and closure duration ratio were significantly positively correlated ($r=.466$, $n=40$, $p<.002$), that is, the greater the release percentage, the greater the closure duration ratio (i.e., the smaller the gestural overlap). This effect was in fact due to the NE group ($r=.582$, $n=15$, $p<.023$). The NM group did not have a significant correlation ($r=.369$, $n=25$, $p>.07$). This result indicates that when the NE group had more releases, they had longer closure durations of C1#C2, resulting in less overlap. This finding is consistent with previous work. For example, Catford (1977) found that there is always a momentary break of articulatory continuity between the successive segments in an open transition (i.e., release). In a close transition there is no such break, meaning shorter closure duration (i.e., more overlap). However, this systematicity was not shown in nonnative speech. It seems fairly random how the NM group coordinated English clusters.

Overall, I found that 35.2% of C1s were released in the NE group and 39% of C1s were released in the NM group, averaged across both speech rates. The English speakers' release percentage falls within the range predicted by previous studies. For instance, Davidson (2011) reported only 25% of C1s were released in native English speech, while Ghosh and Narayanan (2009) found 71% of C1s were released in stop-stop clusters using the TIMIT database. Given that this study included nonwords, it was not surprising that English speakers had a relatively higher release percentage than what was reported in Davidson (2011). The two groups' release orders by places of articulation, averaged over both speech rates, are summarized below:

(1) The NE group release order: front-back (53.85%) > back-front (36.75%) > homorganic (14.94%)

(2) The NM group release order: front-back (48.5%) > back-front (44.4%) > homorganic (24.4%)

Not surprisingly, the NE group's overlap degree was almost twice as the NM group (36.1% vs. 19.3%). The NE group's overlap is consistent with previous research suggesting that English consonant clusters are measured as having approximately 20-60% overlap in articulatory contact (see Barry, 1991; Byrd, 1996a; Zsiga, 1994). The less overlap in NM group here agrees with other studies suggesting no extensive overlap observed in L2 English (see Yanagawa, 2006; Zsiga, 2003). The two groups' overlap orders by places of articulation, averaged over both speech rates, are summarized below from more overlap to less:

(3) The NE group overlap order: front-back (38.2%) >> back-front (36%) >> homorganic (34%)

(4) The NM group overlap order: homorganic (19.85%) >> front-back (19.3%) >> back-front (18.85%)

Worth noting is the overlap order in the two groups. Although the NM group had the most overlap in homorganic clusters compared to heterorganic clusters, the extent was much less significant to that of the NE group. As mentioned above, overlap in this study in fact corresponds to shortening of closure durations (also see Zsiga, 2003). Therefore, it is possible for a cluster to have a released C1 and to have overlap at the same time, although that overlap, in an articulatory view, is avoided due to an articulatory gap between C1 and C2. Figure 5.1 below shows the general relationship between release percentage and closure duration ratio produced by English speakers in this research. The dotted line represents where C1 should begin and where C2 should end in an open transition; the solid line represents where the constriction places actually are in connected speech. In connected speech, English speakers shortened the closure duration of the consonants so that they quickly reached the offset of C1 to recover identities of both consonants as much as possible. This is especially true in front-back clusters (see Chitoran et al., 2002). The overlap in this study thus refers to how much C1 and C2 are shortened (C1 shortening + C2 shortening / C1 + C2), which is 36.1% for the NE group.

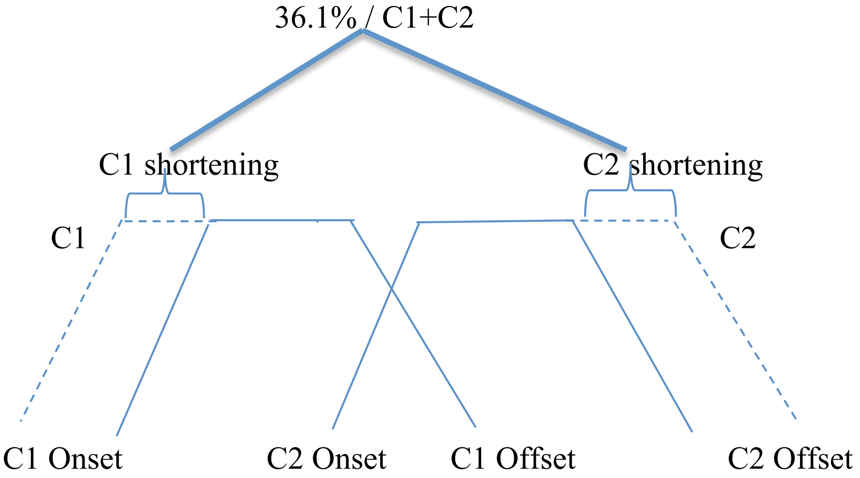


Figure 5.1 The NE group's overlap scheme

Figure 5.2 below shows the relationship between release and closure duration ratio produced by Mandarin speakers in this study. The figure indicates that Mandarin speakers also shortened the closure duration, but the extent (19.3%) was much less compared to English speakers, reflected as a shorter dotted line. Moreover, as reported earlier in this section, the relationship between release percentage and closure duration ratio was significantly correlated in the NE group but not in the NM group. This result suggested the shortening effect was not strongly shown in L2 speech.

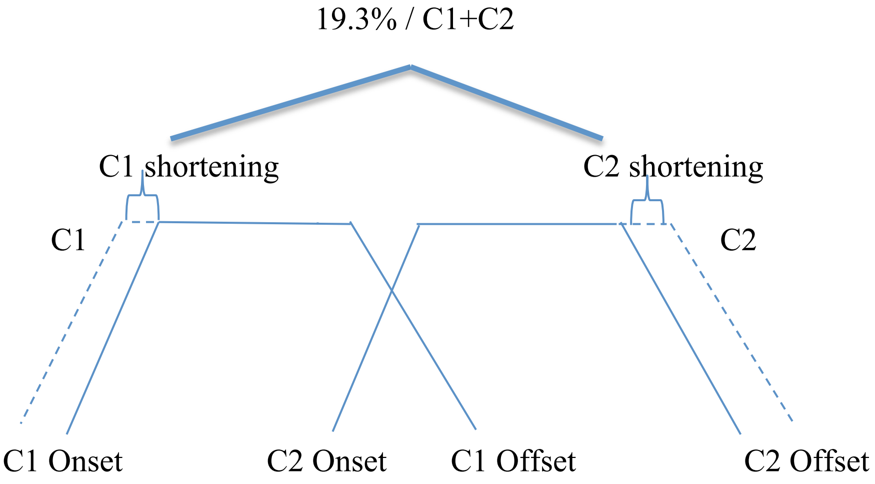


Figure 5.2 The NM group's overlap schema

The following paragraphs provide a brief summary concerning each main effect examined in this study. To answer the first research question, this study included all possible stop-stop combinations across words to examine the place of articulation (POA) effect. The study found that the main effect of *POA* was significant on release percentage but not on closure duration ratio (Section 4.2). First, despite the trend for the NE group to have fewer releases and more overlap than the NM group in coordinating homorganic clusters, the group differences were not statistically significant (i.e., H1 was not supported). Second, while both the NE and NM groups released more often in front-back than in back-front clusters, the NE group had significant differences at both speech rates while the NM group did not at either rate. Thus, H2 was supported. Both groups had slightly more temporal overlap in front-back than in back-front clusters; but the timing differences were not significant in either group. Thus, H3 was not supported. Overall then, the POE was partially supported in the current study: it was supported in terms of the release percentage but not overlap patterns in NE speech. The POE was not shown at all in NM speech.

To answer the second research question, the stimuli in this study incorporated both high and low frequency words as well as nonwords. In terms of the gradient effect (i.e., high vs. low frequency in real words, RQ2a), the results found no significant correlation between frequency and release percentage in either of the groups. However, a significant correlation between frequency and overlap was found for both groups: both groups tended to have more overlap in more frequently used words (e.g., *bad day*, *take care*) than in less frequently used words (e.g., *dock gate*). H4 was thus partially supported (more in Section 5.2).

In terms of the categorical *Frequency* effect (i.e., real words vs. nonwords, RQ2b), it was significant on both release percentage and closure duration ratio: both groups released significantly more often and had less overlap in nonwords than in real words. A group difference was observed in the interaction of *Frequency * POA*. The NE group statistically exhibited the predicted release pattern (i.e., front-back > back-front) of the POE in both real words and nonwords, but the NM group did not in either condition. These results supported H5. As for the predicted temporal pattern (i.e., front-back >> back-front), the NE group exhibited such a trend in both real words and nonwords, but this trend was not significant in either condition. The NM group did not significantly show the predicted degree of overlap in either condition; even worse, they had an order of back-front clusters allowing more overlap than front-back clusters in nonwords. These results did not support H6. The analyses showed that the overlap differences between groups were significant in nonwords, indicating similar performances in coordinating real words but different organizations for nonwords (see in Section 5.3 for further discussion).

To answer the third research question, this study considered stress patterns at the word level (i.e., lexical stress: compounds (initial stress) vs. non-compounds (final stress)) and at the sentence level (i.e., focus). The main effect of *Compound* was significant on release but not on closure duration patterns (Section 4.2.1). The study discovered that both groups released significantly more often in compounds (C1C2) than in non-compounds (C1C2). In terms of duration ratio, the study discovered an opposite effect for the two groups. The NE group showed a trend of having more overlap (i.e., shorter closure duration) in compounds than in non-compounds, while the NM group had significantly more overlap in non-compounds than in compounds. On the other hand, the

Focus effect was significant on both release and duration ratio patterns (Section 4.2.2). Both groups had significantly more releases and less overlap when the first word was focused (i.e., **FI**F2) than when there was no focus (i.e., F1F2). In contrary to the prediction, neither group had significantly more overlap in F1**F2** than in **F1**F2. More surprisingly, it was the NM group that showed such a trend while the NE group had slightly more overlap in **F1**F2 than in F1**F2.** Moreover, the NM group always had significantly more overlap in F1F2 (i.e., without focus) than in other focus conditions, while this was not the case for the NE group. These results showed that the NM group exhibited a stronger stress effect than the NE group sententially. Therefore, H7 and H8 were not supported.

To examine the fourth research question, this study elicited speech from both slow and fast speech rates. The effect of *Speech Rate* was not significant on either release percentage or duration ratio (Section 4.2.3). The results showed that increased speech rate did not induce a greater amount of overlap. Two types of speakers from both groups were evident in this research. One type of speakers had ***more*** temporal overlap accompanied by the faster rate and the other type had ***less*** overlap. This result indicates a large amount of speech freedom, suggesting articulatory patterns are not simply accelerated at fast speech rates. H9 was not supported. Further analyses were conducted to examine whether overlap degree was correlated with L2 speakers' perceived accent. The results suggested that overlap degree was not indicative of perceived accent.

Very few studies have examined the effects of voicing on gestural overlap, finding inconsistent and minor effects (see Byrd, 1993; Ghosh & Narayanan, 2009). For this reason, I did not include the voicing effect as one of the main factors addressed in the research questions (see Chapter 2). However, it became apparent during data analysis that voicing did play an important role in cluster realization. Specifically, the results showed that the main effect of *Voicing* was significant on release percentage but not on duration ratio. The effect was different on each group. The NE group most often released a voiceless C1 in homorganic and back-front clusters, and a voiced C1 in front-back clusters. The NM group most often released a voiceless C1 before a voiced C2 (VC2) in all types of clusters. The NM group also had less overlap under VC2 than under other voicing conditions in general. In terms of group differences, the NE group had more overlap than the NM group in front-back clusters under VC2 at the slow rate. The NE group also had more overlap than the NM group in nonwords under VC2 regardless of speech rates.

In sum, five major findings emerged from this research (taken from Chapter 4). Overall, the hypotheses associated with *POA* and *Frequency* effects were mostly supported while the hypotheses associated with *Stress* and *Speech Rate* effects were not. I will discuss the results of *POA* and *Frequency* in Section 5.2 – 5.3, and will elaborate the results of *Stress* and *Speech Rate* along with other unexpected findings in Section 5.4.

1. The POE was only partially supported in this research. The release pattern was statistically significant in the NE group but not significant in the NM group. The predicted degree of overlap was not significantly shown for either group.
2. The gradient frequency effect was similar for both groups. Both groups exhibited a strong frequency-matching heavier: more frequently used words had more overlap.

3. Group differences were observed in the interaction between the categorical *Frequency* * *POA*. The NE group significantly showed the release pattern of the POE in both real words and nonwords but the NM group did not. The NE group also had consistent temporal overlap orders across lexical conditions, while the NM group had inconsistent overlap orders.
4. Speakers from both groups had either more or less overlap at the fast rate than at the slow rate, suggesting increased speech rate did not always induce increased temporal overlap.
5. Overlap ratio was not correlated with the perceived accent in L2 speech. This interesting finding suggested that cross-word coarticulation might not be as important as other features (e.g., vowel quality) when judging foreign accent.

5.2. Place of articulation

The previous section summarized the results of consonantal overlap gestures examined in this study. In this section, I will elaborate on the effect of *POA*. The results with respect to homorganic clusters are discussed in Section 5.2.1; the front-back and back-front clusters are compared with each other in Section 5.2.2.

5.1.1. Temporal blending

In this subsection, I will focus on the release and temporal patterns in homorganic clusters. As mentioned previously, the NE group released 14.95% of the C1s and had 34% closure overlap in producing homorganic clusters averaged over both speech rates; the NM group released 24.4 % and had 19.85% overlap in the clusters. Despite the apparent trend, the group differences did not turn out to be statistically significant in either release or closure duration patterns.

Both groups had fewer releases in homorganic than in heterogenic clusters (see Section 5.1.2). This result is not unexpected considering that existing cross-linguistic findings, which suggest fewer releases in clusters where the adjacent consonants share the same place of articulation than in clusters where the adjacent consonants have different articulation places (Clements, 1985; Elson, 1956; Pouplier, 2003; Zsiga, 2003). For example, in Sierra Popoluca, a Zoquean language spoken in Mexico, consonant clusters are realized with an intervening open transition (i.e., produced with a schwa) if the consonants are heterogenic while homorganic clusters lack such a transition (Clements, 1985; Elson, 1956).

Due to lacking a release between consonants, homorganic clusters are more likely interpreted as a long consonant rather than a sequence of two segments (Catford, 1977). Catford treats homorganic clusters as phonetically geminate sequences, involving "a prolongation of the articulatory posture" (p. 210). Even in L2 speech, this prolongation is strongly evident. Zsiga (2003) discovered that both native (English) and nonnative (Russian) speakers released less often in homorganic clusters than in front-back and back-front clusters. The articulatory rationale behind this, as Pouplier (2003) and Goldstein et al. (2007) argue, is that the consonants in heterorganic clusters, which use different subsets of articulators, are largely independent of each other such that the constriction gestures are compatible and can be produced concurrently. While for the homorganic clusters that share the same constriction gesture, the consonantal gestures are in competition such that they are only manipulated sequentially. The results from both groups in this study suggest a universal grounded articulatory constraint that speakers, regardless of their native language, tend not to release C1 in homorganic clusters.

However, further analysis found significant group differences in the articulation of homorganic clusters at the slow rate (see Table 4.21):

(5) NM > NE P1L1[13]

(6) NM > NE P1V2

The scheme in (5) means that the NM group had significantly more releases than the NE group in meaningful (i.e., L1) homorganic clusters (i.e., P1), such as *keep pace*, *bad day*. Specifically, the results showed that the NE group only released 6.9% of C1s in meaningful homorganic clusters at the slow rate while the NM group released 21.7% (see Section 4.5.1). This finding suggests that Mandarin speakers did not release the homorganic clusters the same way that English speakers did, resulting in more open transitions. The scheme in (6) means that the NM group also had significantly more releases than the NE group in homorganic clusters (i.e., P1) under VC2 (i.e., voiceless C1 + voiced C2) such as *lip balm*, *beat down* and *look good*. This *Voicing* effect was unexpected (i.e., why NM speakers would prefer releasing a voiceless C1 before a voiced C2), and will be discussed in more detail in the Section 5.3.3 below.

An interesting result is the asymmetry between the NE and the NM groups with respect to release patterns in homorganic clusters. Along with fewer releases in homorganic clusters, the NE group also had less overlap in this kind of clusters than in

---

[13] Again, P stands for place of articulation: P1=homorganic clusters, P2=front-back clusters, P3=back-front clusters. L stands for lexical: L1= meaningful words; L2=nonsense words. V stands for voicing conditions: V1=a voiceless C1+ a voiceless C2; V2= a voiceless C1+ a voiced C2; V3= a voiced C1+ a voiceless C2; V4= a voiced C1+ a voiced C2.

front-back and back-front clusters. Specifically, the NE group had 34% closure overlap in homorganic clusters, which is less than the other two types of clusters (38.2% overlap in front-back and 36% in back-front clusters). In contrast, the NM group had slightly more overlap (i.e., shorter durations) in homorganic clusters (19.85%) than in heterogenic clusters (19.3% and 18.85%, respectively).

The NE group's result is comparable to a previous study (Gafos et al., 2010) that showed Moroccan Arabic (MA) speakers had significantly more overlap in heterorganic than in homorganic clusters (both within word-medial and word-final). They found that two homorganic stops are overlapped only across an affixal boundary, otherwise they are sequenced within morphological templates (i.e., no overlap). Gafos et al. interpreted this result as evidence of the Obligatory Contour Principle (OCP) (McCarthy, 1986), which disallows a sequence of two identical units. In order not to violate OCP, a schwa is inserted to break the identical units in homorganic clusters. Such OCP effects are not uncommon cross-linguistically. Suzuki (1998) discussed glottal-dissimilation in the language of Seri, a Hokan language of northwestern Mexico. In Seri, a glottal stop is deleted when preceded by another glottal stop if they are in the same syllable. Note that such constraint-based analyses require adjacent gestures sharing the same place of articulation to reduce overlap as much as possible (also see Hall, 2003). In this sense, the least amount of overlap in homorganic clusters found in this study is in accordance with previous constraint-based analyses.

The findings in this study also provide strong evidence in support of the blending hypothesis of Browman and Goldstein (1989). This hypothesis argues that inter-consonantal gestures that share the same articulatory tier compete with each other, and will show the influence of both consonants. A blending of gestural characteristics “shows itself in spatial changes in one or both of the overlapping gestures” (Browman & Goldstein, 1990, p. 362). In particular, the location of the constriction should not be identical to that of either C1 or C2, but rather should fall somewhere in between.

Browman and Goldstein (1990) discussed *gestural* blending in terms of place compromise: the location of the constriction should fall somewhere in between C1 and C2 when they share the same articulatory tier. Following this idea, I argue for *temporal* blending between adjacent gestures when they share the same articulator, which is reflected as the timing influence of both consonants. Specifically, if homorganic clusters were geminate sequences, they would have a single gesture, and the other gesture would get deleted. However, in the current study, the results showed that the NE group had 0.66 closure duration ratio and the NM group had 0.8015 in producing homorganic clusters. This result meant that the two groups shortened the closure duration of each stop and produced an overlap relation of 34% and 18.85%, respectively. If one gesture were deleted, we would expect a 0.5 closure duration ratio. Given that both groups yielded overlap figures less than 50%, I am confident to conclude that they did not delete a gesture entirely; rather, they shortened each gesture, resulting in less than 50% overlap.

The hypothesis of temporal blending is supported by previous studies such as Browman and Goldstein (1986), Byrd (1995), and Munhall and Löfqvist (1992). Browman and Goldstein (1986) argued for the presence of two overlapping bilabial gestures in Chaga sequences; Byrd (1995) discovered that two lingual gestures are canonically present in geminated sequences (C1#C2), and that the coproduced movement for C1#C2 is longer than the non-coproduced movement for a single gesture.

Overall, both groups tended not to release in homorganic clusters and showed temporal blending instead of geminating homorganic sequences. Despite the fact that the overall group differences were not significant, the NE group had significantly fewer releases in meaningful homorganic clusters than did the NM group, suggesting a trend of an open transition in the nonnative group.

5.1.2. Place order effect

As aforementioned, both native and nonnative speakers in the current research exhibited a trend towards the POE: more releases and more overlap in front-back than in back-front clusters. This result is consistent with studies on the POE across languages (Byrd, 1996a; Chitoran et al., 2002, Davidson, 2011; Hardcastle & Roach, 1979; Henderson & Repp 1982; Ghosh & Narayanan, 2009; Kochetov et al., 2007; Peng, 1996; Zsiga 1994, 2000).

However, findings here were somewhat mixed. The NE group had significantly more releases in front-back clusters than in back-front clusters at both speech rates (at the slow rate: 53.8% vs. 33.7%; at the fast rate 54.5% vs. 39.8%); but the overlap differences were not significant at either speech rate (at the slow rate: 38.3% vs. 35.5%; at the fast rate: 38.1% vs. 36.5%). The NM group, despite showing the same trend as NE speakers, did not show significant release differences in front-back and back-front clusters at either rate (at the slow rate: 55% vs. 49.4%; at the fast rate: 42% vs. 39.4%). Moreover, the NM group did not have significant overlap differences between front-back and back-front clusters at either speech rate (at the slow rate: 19.4% vs. 19.0%; at the fast rate: 19.2% vs.

18.7%). This result means that the POE was partially shown in native speech but not shown at all in nonnative speech.

Here, we have to consider the fact that the consonant clusters used in this study were in boundary positions. In fact, two lines of research make opposite predictions about juncture effects. The first line of research predicts more overlap in clusters across words than in word-initial clusters. Specifically, researchers focusing on perceptual recoverability (e.g., Chitoran et al., 2002, Gafos, 2002; Goldstein et al., 2006) have argued for more limited alignment (i.e., less overlap) in word-initial clusters (#C1C2) than in cross-word clusters (C1#C2). This is because only the formant transition information of C2 to the following vowel is available in word-initial #C1C2 but not during the release phrasing of C1. Due to the limited perceptual cues available for C1, word-initial clusters are protected against the substantial degree of overlap in order to recover both C1 and C2. Following this assumption, cross-word sequences (e.g., VC1#C2V) would allow more overlap than word-initial clusters, considering transitions of both C1 and C2 are available in an intervocalic context, and as such, it is easier to recover both consonants' identities.

Another line of research argues that gestures in cross-word positions stand in a substantially looser relationship than those formed within words, where speakers are likely controlling a more stable pattern (Cebrian, 2000; Davidson, 2011). Goldstein and Fowler (2003) consider that clusters across words are not as tightly bonded as word-initial clusters, which people expect to cohere in speech production and planning. Also, cross-word clusters are as not highly constrained as word-initial clusters in terms of the number and nature of consonant combinations that can occur. For example, tightly bonded units (e.g., word-initial clusters) have sonority hierarchy requirements (with a few exceptions) while cross-word clusters do not. Within this view, word-initial clusters are likely to have a more overlapped relationship than those at word boundaries.

In a similar vein, Cebrian (2000) proposes a Word Integrity Principle, which prefers a more cohesive timing pattern within a syllable and constrains overlap across words. Cebrain interpreted the effect of word integrity as part of the interlanguage grammar, which treats a word as a separate unit and prevents “the articulatory synchronization of sounds belonging to different words” (p. 19). In fact, such a restricting principle is highly active in native grammar, where gestures get longer, larger, and farther apart at phrase edges (Beckman et al., 1992; Byrd, 2000; Byrd & Choi, 2000; Byrd & Saltzman, 1998, 2003; Cho & Keating, 2001; Fougeron & Keating, 1997). As a consequence, the surface outputs at edges have greater splittability (e.g., excrescent vowels) (Zuraw, 2007), longer durations in C1#C2 (Davidson & Roon, 2008), and gradient assimilation (Catford, 1977; Zsiga, 1995).

As reported in Section 5.1, the NE group had 36.1% overall closure overlap, which is in line with the predicted range for close transitions in Catford (1977). However, the 36.1% overlap reported in this study is much less extensive than Byrd (1996b), who found 87% overlap in the front-back cluster (/d#g/) and 53% overlap in the back-front sequence (/g#d/). Byrd (1996b) also reported a wide range of overlap with sequences ranging from 11% to 91% overlapped. In the current study, the individual ratio ranged from 0.242 to 1.554 in the NE group (from 75.8% overlap to a long pause), and from 0.2 to 1.129 in the NM group (from 80% overlap to a short pause). This result is unexpected, especially that the native group showed larger individual differences than the nonnative group.

A closer inspection of the data showed that there were two male native English speakers (2/15) and six Mandarin speakers (6/25) who had a closure duration ratio of more than 1 **even at fast rate**, indicating a noticeable pause between words. The pause (and thus less overlap) was likely due to participants implementing the Word Integrity Principle (Cebrian, 2000) to eliminate contact across words. From a perceptual perspective, pausing at the phrase boundaries in English reflects the loose instantiation of phrase boundaries to help listeners reform the informational structures (Byrd & Saltzman, 2003).

The current study contributes to the debate as whether or not cross-word positions will allow more overlap than word-initial or word-final positions. The findings in this study seemed to show both the POE and juncture effects.

Both groups exhibited the trend of the POE; but the POE was only ***partially*** shown in native speech. The NE group significantly exhibited the predicted release pattern but not the temporal pattern; the NM group did not have significant POE shown in either release or temporal patterns. The significant release pattern of the POE in the NE speech indicates that a constriction transition from front to back articulators is strongly favored and articulatorily grounded even in an English boundary position where gestures are relatively loosely coordinated. Compared to English speakers, Mandarin speakers were more likely to eliminate contact across words, showing a tendency of preserving each word as a unit. What these results suggest is weaker juncture effects in native speech than in nonnative speech. Considering that no significant temporal patterns emerged in either native or nonnative speech, future research needs to explore further the relationship between the POE and juncture effects with different measurements of overlap.

5.2. Frequency

In this section, I will discuss the general effect of *Frequency* on overlap patterns for both groups. The overall frequency results are discussed in Section 5.3.1, and the implications of the results are discussed in Section 5.3.2 – 5.3.3.

5.2.1. Lexical diffusion

As mentioned previously, this study considered two aspects of frequency: the high-frequency words versus the low-frequency phrases in real words (e.g., *take care* vs. *dock gate*), and the real words (i.e., high-frequency) versus the nonwords (i.e., extremely low-frequency) (e.g., *keep pace* vs. *peep pate*).

For the real words, the lexical frequency rating and release percentage were not significantly correlated in either of the groups (also see Yanagawa, 2006). However, the frequency rating and duration ratio were significantly correlated for both groups. Both groups rated real words such as *take care*, *bad day*, and *look good* highly familiar; these reached up to 50% overlap. Words such as *hug pug* and *dock date* were rated as much less familiar, and showed only up to 20% overlap. This result suggests a frequency-matching behavior for both groups: more reductions occur for more frequent words in real words.

These results support existing research on the effect of lexical frequency in sound patterns such as vowel reduction and final obstruent deletion (Bush, 2011; Bybee, 2000, 2007; Gregory et al., 1999). Gregory et al. (1999) reported that tapping of a word-final /t/ or /d/ in an intervocalic context is highly affected by the probabilistic variable "mutual information", which measures the likelihood that two words will occur together. For instance, tapping would be more common in the word pair *lot of* than in *out alone*, which suggests that tapping is a process that "may also preferentially apply internally to highly cohesive pairs" (p. 317). The effect of mutual information on tapping appears to be a gradient effect. As mutual information increases, reflecting a stronger cohesion between the two words, the likelihood of tapping increases.

In this research, as lexical frequency increases, the overlap increases. Here, I adopt the concept of "lexical diffusion" from Bybee (2007), which refers to the way a sound change spreads through the lexicon (also see Wang, 1977). In her discussion of lexical diffusion, Bybee notes that "the more a word is used, the more it is exposed to the reductive effect of articulatory automation" (p. 218). For instance, often-repeated phrases tend to reduce phonetically more than less often-repeated ones: *I will* often sounds like [ajl]. Bybee thus argues that a tightly bonded unit is a construction and accessed as a whole from storage, including "reductions and coarticulation that have accumulated in its representation" (p. 301).

Both groups also showed frequency-matching behavior by essentially differentiating real words from nonwords. The categorical *Frequency* effect (real words vs. nonwords) was found to be significant on both release percentage and closure duration ratio. Both groups had significantly fewer releases and more overlap in clusters in real words than in nonwords (e.g., *keep pace* >> *peep pate*). The NE group had 39.4%

overlap in producing real words and 32.8% in nonwords averaged over both speech rates; the NM group had 24.3% overlap in real words and 14.6% in nonwords.

The results in this study also discovered that the categorical frequency effect was different for the native versus nonnative group in terms of the POE specifically. The NE speakers' treatment of real words versus nonwords exhibited a high degree of consistency (see Section 4.4): in terms of release percentage, the NE group had significantly more releases in front-back than in back-front clusters in both real words and nonwords; in terms of duration ratio, English speakers did not show significantly more overlap in front-back than in back-front clusters in either lexical setting, but the trend was to organize the three cluster types in the same fashion (i.e., front-back >> back-front >> homorganic) in both lexical environments (see Figure 4.9). This result indicated that native speakers maintained a stable timing pattern in new forms (i.e., nonwords) even though they slightly modified the overlap degree in these forms.

Consider the study of Zhang and Lai (2010), which treated phonological alternations motivated by a stronger phonetic motivation as natural, and ones with a weaker phonetic motivation as unnatural. Their Mandarin speaking subjects applied the stronger alternations more accurately in producing Mandarin tone sandhi. They explained that it was due to a stronger phonetic motivation, reflected as a more natural process in the native phonology. Similarly, the stimuli used in this study can be interpreted as a comparison of naturalness versus unnaturalness. The words that speakers are familiar with are considered to be natural (e.g., *keep pace*) while the words that speakers are not familiar with are unnatural (e.g., *peep pate*). The finding that the NE group modified the overlap extent in the unnatural (i.e., nonwords) setting is consistent with Zhang and Lai

(2010), suggesting native processing can extend native phonological knowledge to new forms.

In comparison to the NE group, NM speakers did not significantly exhibit the release pattern (i.e., front-back > back-front) predicted by the POE in either lexical setting, nor did they show consistent temporal orders in organizing real words and nonwords (see Section 4.4). The NM group had more overlap in front-back than in back-front clusters (26% vs. 24.3%) with real words data but had an opposite overlap order with nonwords data (12.7% vs. 13.4%, see Figure 4.9). That is, Mandarin speakers failed to coordinate nonsense front-back and back-front clusters the same way as they did with real words. They also failed to coordinate nonsense front-back and back-front clusters as the NE group would.

To summarize, *Frequency* has similar effects on the two groups in terms of lexical diffusion. Both groups had significantly more overlap in frequent words than in infrequent words (e.g., *take care* vs. *dock gate*); both groups of speakers had significantly more overlap in real words than in nonwords (e.g., *keep pace* vs. *peep pate*). This result is expected since the more often two elements are used together, the more tightly they will be bonded phonologically and semantically, the more likely they are to cohere in speech production and planning (see Goldstein et al. 2007). On the other hand, *Frequency* has different effects on the two groups in terms of the interaction between *Frequency* and the POE. The NE group significantly showed the release pattern of the POE and maintained overlap consistency of the POE across lexical conditions while the NM group did not. This result suggested different phonological representations of sound patterns in native versus nonnative speech, which will be discussed in the following section.

5.2.2. Phonological representation

The reason I incorporated the nonwords into the design of this study was not only to examine lexical diffusion, but also to investigate whether or not speakers would transfer their knowledge of sound patterns (i.e., POE) to words that they are not familiar with. This question is of great importance is because it directly addresses how words along with their surface variations are stored: whether sound patterns are stored as gestural units or discrete outcomes. If it is the former case, gestural units would be available for speakers to reconstruct in unfamiliar forms. As a consequence, such internal knowledge would ensure that speakers' habitual articulatory patterns (measured here as release percentage and closure duration ratio) would remain consistent and would not vary significantly when lexical conditions change. If it is the latter case, sound patterns would be stored as discrete outcomes, and speakers would most likely to fail to generalize their phonetic awareness beyond memorized items. This consequence would be reflected as different coordinating patterns, varying in real words and in nonwords. The prediction is that gestural unit representation would ensure productive extension to nonwords while discrete outcomes would result in partial application.

The consistency shown by the NE group supports previous studies showing native speakers have the ability to identify regularities and systematic alternations, and extend their phonological knowledge to unfamiliar environments (see Berko, 1958; Wilson 2003, 2006; Zhang & Lai, 2010). English speakers' ability to articulate using this knowledge suggests that they have internalized the process and can use it productively. As Wilson (2006) comments, speakers have "detailed knowledge of articulatory and perceptual properties" (p. 946). It is obvious that their grammatical systems can make reference to that knowledge.

More importantly, the extension of native grammar into nonwords suggests that native speakers have an abstract representation of gestural coordination of their own language. The internal knowledge of the coordination of gestural units is reflected as consistent release and overlap patterns across lexical settings. This consistency also implies an extended ability of coding unnatural forms in native processing, which is compatible with "rich memory" as proposed by Port (2007): "speakers simply store the auditory patterns they hear and recognize the word, morpheme or phrasal units ***in whatever sizes*** they were learned in." (emphasis mine; p. 152). The concept of units of varying sizes being memorized requires a large memory load, given the large amount of lexical entries in a language. That is, language is stored in a system that is far richer than linguists have ever considered in the past.

The rich and detailed representation found in NE group also essentially corroborates a fundamental hypothesis of Articulatory Phonology; phonological contrasts are represented in terms of articulatory gestures, not discrete features (Browman & Goldstein, 1989, 1990, 1992; Goldstein et al., 2006). One implication of the gestural unit hypothesis is that this kind of representation is apparently what speakers make use of for their lexicon. Findings in this research suggest a step further: the stored gestural units are available for native speakers to reconstruct in new contexts. If units could not be reorganized, English speakers in this study would coordinate nonwords differently than lexical words, perhaps having a different overlap order among the three cluster types. Yet, they still followed the pattern of front-back >> back-front >> homorganic in non-words.

It seems that native speech not only encodes gestural organizations manifested in lexicon, it also allows for the recombination of new possibilities with the same structure (also see Kita & Özyürek, 2003).

While establishing “rich memory” in the native processing, there is no evidence suggesting Mandarin speakers acquired nonnative phonological knowledge; rather, how to organize L2 clusters seems to be largely dependent on familiarity with lexical knowledge. Especially, the NM group organized nonsense front-back and back-front clusters significantly differently than the NE group. This suggests that have no L2 phonological knowledge of gestural coordination that they can apply in nonwords. The NM group also had different overlap patterns in different lexical conditions. They produced front-back >> back-front in real words, and back-front >> front-back in nonwords. The former pattern of the NM group was consistent with that of the NE group, showing the two groups had similar organizations with familiar lexical items. However, the NM group’s responses for nonwords showed that they could not generalize beyond memorized items. Group differences were found to be significant in nonsense front-back (i.e., (7)) and back-front (i.e., (8)) clusters, with the NE group having more temporal overlap.

(7) NE >> NM P2L2
(8) NE >> NM P3L2

The response of nonwords exhibited by the NM group shows that Mandarin speakers were not sensitive to fine-grained phonetic details in English and thus were unable to reorganize English gestural units that they were familiar with. This implies that L2 sound patterns are stored as undissectable chunks of acoustic consequences. As a result, L2 speakers have a limited repertoire for reconstructing new possibilities within L2 phonotactic constraints. The abovementioned results indicate that the POE is stronger in native speech than in nonnative speech in that it can be applied productively in the former but not in the latter. The NM group failing to exhibit a consistent POE across lexical contexts leads to another question: is the POE universally preferred?

5.2.3. Language specific overlapping rules

Assuming English speakers in the current study represent the norm of how to organize cross-word coarticulation, an apparent trend of front-back clusters allowing more overlap than clusters of the opposite order is shown regardless of lexical status. If such overlapping patterns are universally preferred and considered natural, I predict that the L2 group should exhibit similar surface variations as well. Yet, this is not the case, in particular with respect to the nonwords. The NM group had slightly more overlap in back-front than in front-back clusters in nonwords, failing to exhibit the POE in nonwords.

The fact that Mandarin speakers did not employ nor extend L2 phonetic knowledge provides strong evidence in favor of language-specific sound patterns. This result substantiates previous studies such as Jun (1996), Kochetov and Goldstein (2001), Gobl and Ni Chasaide (1999), Yanagawa (2006), and Zsiga (2000, 2003). These studies agree that coarticulatory patterns are language-specific, “reflecting and being constrained by the phonological and syntactic rules of the language” (Hardcastle, 1982, p. 47). For example, Gobl and Ni Chasaide (1999) have shown that the extent of laryngeal coarticulation associated with laryngeal consonants is language-specific, since the amount of coarticulation is much stronger in Swedish that it is in Italian, while it is rarely present in German and French. In the current study, it could be because of the native language (i.e., Mandarin) categorically disallowing consonant clusters that results in Mandarin speakers' failure to coordinate English clusters in a target-like fashion.

A seemingly redundant question is why the extent of overlap differs among languages? One can simply answer this question by saying that languages are essentially distinctive. Kochetov et al. (2007) argue that language differences in overlap degree "may be related to the propensity of a language to assimilate in consonant clusters" (p. 1364). If one language allows more assimilation than the other, it should also allow more consonantal overlap. This approach accounts for the fact that Korean allows more overlap of consonant clusters than Russian, and more assimilation is found in the former. This claim, however, cannot provide an adequate answer for the puzzle: why do ***some*** languages allow more overlap than the other?

Cross-linguistic studies by Manuel (1990) and Manuel and Krakow (1984) argue that languages with more crowded vowel spaces (and thus smaller vowel convex region targets) show less coarticulation than languages with less crowded vowel spaces. Manuel and Krakow (1984) compared vowel-vowel coarticulation in English, a language with a relatively crowded vowel space (i.e., 10 simple vowels and five diphthongs) and two Bantu languages (Shona and Swahili) for which the vowels are well spread-out (five vowels). Their prediction was that since the English vowels are closer together in the articulatory space, the production range for each English vowel should be relatively small. In contrast, since the vowels of Shona and Swahili are more spread apart, they could tolerate larger ranges of productions without neutralizing phonemic differences. Their results confirmed this prediction. Their studies are consistent with previous analyses that increased vowel dispersion is known to be associated with greater intelligibility and thus clearer speech (e.g., Bradlow, Torretta, & Pisoni, 1996; Moon & Lindblom, 1994; Picheny, Durlach, & Braida, 1985).

Following Manuel and Krakow (1984), more overlap is expected in languages that have fewer vowels (e.g., Taiwanese) than in languages that have more vowels (e.g., English). Gao et al.'s (2011) findings confirmed this hypothesis. They discovered Taiwanese had longer constriction durations for both onset and coda consonants than American English syllables. Their results also found greater overlap of C-C and C-V in Taiwanese both within the same CVC syllable and across syllable boundaries. They argued that greater inter-syllabic overlap may be related to the unreleased coda in Taiwanese: the oral constriction of a preceding consonant is greatly overlapped with the following consonant to achieve a non-audible release. Given the very few studies, the question of why some languages allow more overlap than the other should be considered in future studies.

To summarize, this section has discussed in detail the major findings in this research, including temporal blending in homorganic clusters and the partially shown POE in native speech. I discussed the POE implications and proposed an account for the incomplete POE in terms of juncture effects. I have also discussed lexical effects and their implications: sound patterns are lexically gradient in both groups. Evidence for gestural unit representation was found in native speech: English speakers showed consistent coordination patterns in two lexical settings. In contrast, inappropriate and inconsistent application of sound patterns was confirmed in L2 articulation in this research. I proposed that a set of preferred gestural overlap patterns is language specific, and that this is responsible for the fact that L2 speakers cannot generalize beyond memorized items: they have not acquired native-like patterns.

5.3. Unexpected Results

In Section 5.1 – 5.2, I have discussed the effects of place of articulation, word frequency, and their implications. The effects of speech rate, voicing, and stress, along with other unpredicted findings will be elaborated on in this section.

5.3.1. Overlap as a function of speech rate

This study not only examined how coordination is implemented and how sound patterns are stored in L1 and L2 speech, it also investigated whether changes in speech rate induced greater overlap. Previous studies have suggested that at the faster speaking rates, gestures are coproduced to a greater extent than at the slower rate (Byrd & Tan, 1996; Engstrand, 1988; Fosler-Lussier & Morgan, 1999; Gay, 1981; Jurafsky et al., 2001; Kuehn & Moll, 1976; Munhall & Löfqvist, 1992).

In the current research, the validation analyses established that both groups produced more syllables per second at the fast rate than at the slow rate (see Section 3.3.2); however, the *Speech Rate* effect was not significant on either release or closure duration ratio patterns. The results showed that the NE group released more often when speech rate increased (32.9% at the slow rate vs. 37.4% at the fast rate), and had slightly more overlap at the slow rate (36.3% at the slow rate vs. 35.9% at the fast rate).

Meanwhile, the NM group released less often when speech rate increased (43.4% at the slow rate vs. 34.85% at the fast rate), and had slightly more overlap at the fast rate (18.2% at the slow rate vs. 19.2% at the fast rate). The results suggest that the *Speech Rate* effect was opposite for the two groups.

Further analyses indicated that individual speakers varied in having either ***more*** or ***less*** overlap at the fast rate than at the slow rate, pointing towards a speaker-dependent effect. Nineteen speakers (nine NE speakers and ten NM speakers) had more overlap in fast speech, and 21 speakers (six NE speakers and 15 NM speakers) had less overlap in fast speech. Worth noting is that there was a larger proportion in the NE group (9/15) who had more overlap in faster speech rate than in the NM group (10/25). This finding is consistent with the only study, to my knowledge, that examined the effect of speech rate on overlap produced by Mandarin speakers, Chen and Robb (2004). They found that Mandarin speakers produce English with a slower speaking rate and articulation rate compared to American speakers. Their analyses also indicated that the Mandarin speakers inserted significantly more pauses between syllables compared to American speakers. Comparably, the current study found that the NM group tended to have a longer duration (i.e., smaller overlap) than the NE group even when their general speech rate increased.

The overall results, however, indicated a stable pattern (i.e., overlap was not affected by speech rate changes) along with speaker variations in both slow and fast speech rates, contradicting previous studies suggesting more overlap at faster speech rates (Byrd & Tan, 1996; Engstrand, 1988; Fosler-Lussier & Morgan, 1999; Jurafsky et al., 2001; Kuehn & Moll, 1976; Munhall & Löfqvist, 1992). This finding, however, is consistent with other work showing individual differences in this regard (e.g., Kuehn & Moll, 1976; Ostry & Munhall, 1985; Tsao, Weismer, & Iqbal, 2006; Zsiga, 1994).

I outline three accounts for the lack of correlation between increased overlap and increased speech rate. The first view involves the phonetic nature of stops. As Kessinger and Blumstein (1997) point out, stops have intrinsic limitations in their production. Specifically, stops involve a pressure build-up behind the constriction and a rapid release of the oral closure. As seen from acoustic consequences, stops have a rapid change and an abrupt amplitude increase at their release when preceding a vowel (Stevens, 1980). This abrupt onset is how stops are distinguished from affricates and fricatives, which have more graduate onsets. These physical gestures must occur to implement the obstruent manner of articulation at either slow or fast rate.

Because of their required acoustic properties and consequently the relative invariance with why they must be articulated, stops are less vulnerable than other consonants (e.g., affricates) such that changes in speech rate are less likely to affect them. In fact, studies have shown that speech rate changes affect the duration of glides and affricates but do not affect stops (Miller & Baer, 1983; Shinn, 1984). Also, changes in speech rates do not induce differing voice onset times of English stops (Kessinger & Blumstein 1997; but see different results in Pind, 1995 and Theodore, Miller, & DeSteno, 2007). In addition, it is likely that reduced closure durations in stops are not correlated to speech rate change (see Zsiga, 1994).

A second view is that the rate is simply not quick enough to have exerted an observable influence on gestural coordination. Previous work has shown that the speech organs would rather keep relatively stable during the course of production. For example, the early study by Draper, Ladefoged and Whitteridge (1960) reported that lung pressure is relatively steady in producing a sentence and that the slight variations in pressure do not change breathing pulses corresponding to syllables. Similarly, Lindblom (1983) observed that the speech production system is rarely forced to its physical limits by speakers. Lindblom showed that speakers do not execute extreme displacements or velocities in coarticulation and vowel reduction; they neither hyper-articulate nor constantly whisper or mumble. Instead, speakers prefer behaviors that minimize motoric demands. Given these findings, it is likely that participants' increase in speech rate in Task 4 was not great enough to affect C1C2 coordination. Specifically, studies, that found the speech rate, seemed to use the same task (i.e., reading a sentence), while manipulating speed using a metronome or instructing the participants to speak fast (e.g., Byrd & Tan, 1996). The current study was done differently: reading versus answering questions. Even if the rate of the latter was faster, it was not simply "scaling up" the rate while keeping everything else the same. This is supported by Tjaden and Weismer (1998). They suggested that the increase in speech rate that participants self-produced may not have been great enough to shift coordination patterns in their study. The transition between adjacent gestures cannot be affected if the rate change is not significant enough.

A final view is that people always tacitly find alternative means to reach target values without compromising a steady speech rhythm. For example, Moisik and Esling (2011) and Moisik et al. (2014) discovered that speakers raise their larynx to supplement the raising pitch value supposedly induced by a voiced target without laryngeal constriction. Ellis and Hardcastle (2002) showed that individual speakers may exploit either categorical or gradient assimilation when producing /n#k/. Results from EPG data showed that two speakers never exhibited tongue-palate contact in the velar region during the /n/, whereas four other speakers consistently produced /ŋ/. Two more speakers alternated between complete assimilation and no assimilation. Only two speakers showed a range of tongue-palate contact from full alveolar closure to partial velar contact to total velar closure. Ellis and Hardcastle concluded that speakers' preferred strategies can be fundamentally different. In this research, the fact that speakers had either more or less overlap when speech rate increased suggested gesture coordination was self-controlled; speakers could easily modify the overlap pattern without compromising a steady speech pattern.

5.3.2. Stress effect

This study asked how stress patterns affect English and Mandarin speakers' cluster coordination in English. Based on the literature (see Cho, 2004; Crystal, 1969; Cummins & Port, 1998; De Jong, 1995; Edwards et al., 1991; Keating, 1984; Ohala & Hirano, 1967), it was hypothesized that English speakers would associate more coarticulatory effects such as shorter durations with less stressed segments. On the other hand, Mandarin speakers were predicted either to have significantly longer durations in stressed syllables compared to unstressed syllables (e.g., Chen et al., 2001; Zhang et al., 2008), or not to have significant cluster timing differences in the production of a stream of English syllables (see Chen et al., 2001; Lin & Wang, 2007; Tajima et al., 1997; Wong, 1987). In this study, I followed the latter line and predicated no significant timing differences for Mandarin speakers in their production of stressed versus unstressed clusters in English.

The current study considered two aspects of stress, the lexical stress (compound vs. non-compound) and the sentential stress (focus vs. no focus). To examine the lexical stress effect, both groups of participants produced a set of compounds (initial stress) versus non-compounds (final stress) in isolation; to examine the sentential stress effect, both groups produced target words embedded in four focus conditions: **FI**F2 (the first word containing C1 was focused), F1**F2** (the second word containing C2 was focused), **FIF2** (both words were focused), and F1F2 (neither of the words was focused).

As I stated in Section 3.4.3, being aware of a possible interaction between *Compound* (lexical stress) and *Focus* (sentential stress), I did not include *Compound* as an independent variable during data analysis in Task 4. This research examined the *Compound* effect in Task 3 and the *Focus* effect in Task 4. The main effect of *Compound* was significant on release percentage but not on duration ratio. The main effect of *Focus* was significant on both release and duration ratio patterns.

With respect to the lexical stress, the hypothesis predicted that the NE group would have significantly more releases and less overlap in compounds (i.e., C1C2) than in non-compounds (i.e., C1C2), while the NM group would only show such as a trend. The results discovered a similar effect for both groups in that they released significantly more often in compounds than in non-compounds. Contrary to the prediction, the NE group also showed a trend of having slightly more overlap (i.e., shorter C1#C2 closure duration) in compounds than in non-compounds, while the NM group had significantly less overlap (i.e., longer C1#C2 closure duration) in compounds than in non-compounds. What the results indicated was a stronger lexical stress effect for the NM group rather than for the NE group.

With respect to sentence stress, the hypothesis predicted that the NE group would have significantly more releases and less overlap when the first word received focus (i.e., **F1**F2) than when the second word received focus (i.e., F1**F2**) as well as when the word was without focus (i.e., F1F2), while the NM group would only show such a trend. The overall results discovered that both groups of participants released C1 significantly more often in **FI**F2 than in F1F2, but the release differences between F1**F2** versus F1F2, and **FIF2** versus F1F2 were significant in the NM group but not in the NE group. As for the duration ratio, both groups had significantly less temporal overlap (i.e., longer C1#C2 closure duration) in **FI**F2 than in F1F2. In contrary to the prediction, the NM group had more overlap in F1**F2** than in **F1**F2**,** while the NE group had the opposite overlap order. The NM group also had significantly more overlap under F1F2 than under **FIF2**, but again, this difference was not significant in the NE group. In fact, the NM group had significantly more overlap (i.e., shorter duration) under F1F2 than under any other focus positions in both front-back and back-front clusters while the NE group did not. These facts indicated that the extent of Mandarin speakers' closure duration shortening in words without focus was much more extensive than that in English speakers. These results suggested a stronger sentential stress effect for the NM group rather than for the NE group.

The first consequence of these stress findings is that both groups exhibited the stress lengthening effect at the sentence level. This result generally supports previous studies (e.g., Chen et al., 2001) which found that, similar to English speakers, Mandarin speakers produced English focused words with longer durations compared to unfocused words. Note that previous studies only measured absolute durations. The results in the current study discovered a trend that the NM group had less overlap than the NE group across the four focus positions. This finding indicates that the NM group did not shorten the closure duration as much as the NE group would, suggesting less timing variation in the NM group. This difficulty is likely due to the influence of Mandarin, which has relatively equal syllable timing (see Lin & Wang, 2008; Tajima et al., 1997).

Another interesting consequence of these findings is that, contrary to the prediction, the NE group did not significantly lengthen the closure duration of C1#C2 in compounds as expected, whereas previous studies have suggested that shorter durations are associated with less stressed syllables (see Cho, 2004; Crystal, 1969; Cummins & Port, 1998; De Jong, 1995; Edwards et al., 1991; Keating, 1984; Ohala & Hirano, 1967). Edwards et al. (1991), for instance, argued for less tongue retraction in less stressed syllables because the unstressed syllables are shortened by overlapping the vowel and following consonant gestures. Cases in this study have shown here that stress effects are not always accompanied by durational lengthening in native speech. It is possible that English speakers either lengthened C1 in C1#C2 or C2 in C1#C2, concealing out the overall differences.

Deviating from English speakers, Mandarin speakers in this study significantly shortened the closure duration of C1#C2 in C1C2 than in C1C2. This result, on the one hand, supports Zhang et al. (2008) that Mandarin speakers realized lexical stress through contrastive duration. Such a claim can also be found in Flege and Bohn (1989), for instance, suggesting that L2 learners of English first achieve the durational correlate to produce English lexical stressed versus unstressed syllables. On the other hand, Zhang et al. pointed out a potential lexical frequency effect: it was very likely that the Mandarin participants in their study had simply learned the words examined in their study. The current study included both real words and nonwords, which is why I observed a detailed deviation: a stronger stress effect for the NM group than for the NE group. This deviation from the English norm provides further evidence for ESL speakers' difficulty in acquiring a native-like stress pattern (Chun, 1982; Chen et al., 2001; Ng & Chen, 2011; Yeou. 2004). This stronger stress effect for Mandarin speakers could be due to over-compensation in L2 speech (see Derwing & Munro, 2010; O'Brien, 2004). For example, English learners of German produced L2 intonation patterns with less native-like production than would have been expected due to their over-compensation (O'Brien, 2004).

A stronger stress effect for Mandarin participants than English participants was also reflected at the sentence level: the extent of Mandarin speakers' closure duration shortening in words without focus was much more extensive than that in English speakers. This can be seen from the fact that the NM group had significantly more overlap (i.e., shorter duration) under F1F2 than under any other focus positions in both front-back and back-front clusters while the NE group did not.

In sum, while previous studies suggest that L2 speakers could acquire the durational patterns of English stress (Chen et al., 2001; Flege & Bohn, 1989; Zhang et al., 2008), the results in the current study provide detailed evidence supporting ESL speakers' difficulty in acquiring a native-like stress pattern (Chun, 1982; Chen et al., 2001; Ng & Chen, 2011; Yeou. 2004). What the findings indicate is that Mandarin speakers are affected by their native language having relatively equal syllable timing. The NM group in this research showed an even stronger stress effect than the NE group, suggesting a trend of over-compensation.

5.3.3. Laryngeal specification: voicing

The *Voicing* effect was not explicitly covered by the research questions, but some results are worth noting here. Voicing (VC) 1 indicated that both C1 and C2 were voiceless, VC2 meant a voiceless C1 and a voiced C2, VC3 meant a voiced C1 and a voiceless C2, and VC4 meant that both C1 and C2 were voiced. The statistical results showed that the *Voicing* effect significantly influenced the release percentage but not the closure duration ratio, at both speech rates. This result contradicts Ghosh and Narayanan (2009), which found no voicing effect on release percentage (71% with voiceless C1s vs. 75% with voiced C1s); however, it is consistent with Crystal and House's (1988) study on individual stops, in which voiced stops were released only about half of the time compared with voiceless stops (25.5% vs. 48.5%).

The result showed that the NE group had the same release order at both speech rates of VC1>VC2>VC3>VC4 in homorganic and back-front clusters, and VC3>VC1> VC4>VC2 in front-back clusters. That is, the NE group most often released a voiceless C1 in homorganic and back-front clusters, and released a voiced C1 in front-back clusters. The NM group had the same release order at both speech rates of VC2>VC1> VC3>VC4 in homorganic clusters, and VC1>VC2>VC3>VC4 in back-front clusters. In front-back clusters, the NM group had VC1>VC2>VC4>VC3 at the slow rate and had VC2 >VC1>VC4>VC3 at the fast rate. That is, the NM group most often released a voiceless C1 across the three kinds of clusters.

Literature-based expectations were met in the sense that English speakers tended to release voiceless C1s in C1C2 clusters (see Crystal & House, 1988). However, a preference exhibited by the NE group of releasing voiced C1s in front-back clusters was not predicted. The tendency of releasing voiced C1s in front-back clusters was so strong that the NE group released C1s in frequently used words such as *food court*, *bad kid*, and *good game*. My expectation was that in this context English speakers would have the C1 (e.g., /d/) assimilate to the C2, realizing as /g/ without release and making [fug͡kort] and [gug͡gim]. This hypothesis was not met. I then assumed that releasing a voiced C1 was due to the effect of compounding: speakers would release C1 more often due to the initial lexical stress. In front-back clusters, the results did show that the NE group released more often in compounds (57.2%) than in non-compounds (49.2%) at the slow rate. A closer inspection, however, indicated that the NE group had the release order as VC1 > VC3 > VC2 > VC4 in compounds, and VC4 > VC3 > VC1 > VC2 in non-compounds. It can be seen that, despite the overall higher release percentage in compounds, the NE group released more voiced C1s in non-compounds than in compounds. This result did not support the compounding effect.

An alternative explanation possibly involves the duration differences of voiceless and voiced consonants. It is commonly accepted that the aspirated voiceless segments are usually longer than unaspirated voiced ones (Ladefoged, 1999). As voiced consonants have shorter durations than voiceless ones, releasing a voiceless consonant is incompatible with a significant amount of overlap (i.e., short closure duration) in front-back clusters. An additional explanation is that speakers in this study produced target words in isolation, resulting in more careful pronunciation. However, data of releasing a voiced C1 come from both slow rate and fast rate; I am not certain how much a theory of careful pronunciation is explanatory.

The NM group in this study was found to release a voiceless C1 more often in all three kinds of clusters. Based on the review of the literature (see Anderson, 1987; Hansen, 2001; Messing, 2008; Wang, 1995), it was hypothesized that Mandarin speakers would have more internal releases than would English speakers, given that Mandarin categorically disallows clusters. The preference of releasing a voiceless C1 might reflect the difficulty of acquiring voiced English stops (see Anderson, 1987; Eckman, 1981; Fan, 2011; Hansen, 2001; Wang, 1995). For instance, Flege and Davidian (1984) found that Mandarin speakers devoiced final stops 30% of the time. Broselow et al. (1998) reported Mandarin learners rarely produced voiced stops correctly. It is very likely that Mandarin participants devoiced and/or deleted voiced codas in this research, shown as not released codas. Because I only examined released codas in the current study, the voicing effect was not fully examined. With the data at hand, future research is considered to investigate the devoicing strategy (if any) shown in NM speech, as well as to examine vowel transition to measure deletion.

On the other hand, the *Voicing* effect did not influence closure duration ratio significantly. This was reflected in the fact that both groups had inconsistent overlap orders in terms of the voicing effect. At the slow rate, the NE group had the overlap order of VC1 >> VC4 >>VC2 >> VC3 in homorganic clusters, VC2 >> VC3 >> VC4 >> VC1 in front-back clusters, and VC4 >> VC2 >> VC1 >> VC3 in back-front clusters; the NM group had VC1 >> VC4 >> VC3 >> VC2 in homorganic clusters, VC1 >> VC3 >> VC2 >> VC4 in front-back clusters, and VC3 >> VC4 >> VC2 >> VC1 in back-front clusters. At the fast rate, the NE group had VC1 >> VC2 >> VC3 >> VC4 in homorganic and front-back clusters, and VC2 >> VC1 >> VC3 >> VC4 in back-front clusters; the NM group had VC3 >> VC4 >> VC1 >> VC2 in homorganic clusters, VC2 >> VC1 >> VC4 >> VC3 in front-back clusters, and VC3 >> VC1 >> VC4 >> VC2 in back-front clusters.

One pattern that can be observed is that the NM group always had less overlap under VC2 than under other voicing conditions especially in homorganic and back-front clusters. At the slow rate, the group differences shown in (9) and (10) indicate that the NE group had more overlap than the NM group in front-back clusters under VC2 and had more overlap in nonwords under VC2. At the fast rate, the group differences shown in (11) suggest that the NE group had more overlap than the NE group in nonwords under VC2.

(9) NE >> NM L2V2

(10) NE >> NM P2V2

(11) NE >> NM L2V2

It seems reasonable that the NM group had less overlap under VC2 due to non-agreeing laryngeal specification. Chitoran et al. (2002) argued that the great amount of overlap found in Georgian front-back clusters is due to them agreeing in voicing. As for the non-harmonic (i.e., not agreeing in voicing) back-front clusters, there was “a greater delay between the laryngeal gestures (coordinated with the first consonant) and the release of the second consonant” (p. 443). The study of Hoole, Bombien, Kühnert and

Mooshammer (2009) also suggested that mixed-voicing clusters have little overlap in the examination of German consonant clusters using electromagnetic articulometry (EMA). They discovered that German voiced sequences such as /gl/ and /bl/ (i.e., C1 is phonologically voiced) are produced with more overlap than mixed-voicing clusters such as /kl/ and /pl/ (i.e., C1 is voiceless). [14] Hoole et al. argued for an account based on aerodynamic consideration. In the case of German voiced clusters, low overlap would allow for a release of C1 and thus for a drop of intra-oral pressure. This drop would in turn prevent supra-glottal pressure rising above the threshold where voicing can no longer be maintained for C2.

However, neither account can explain the significant overlap under VC3, a voiced C1 and a voiceless C2, not agreeing in laryngeal specification. In fact, I found the NM group had the most extensive overlap under VC3 in homorganic clusters at the fast rate. The order is as follows, from more overlap to less.

(12) VC3 >> VC4 >> VC1 >> VC2

The overlap order in (12) shows that VC3 has the most overlap and VC2 has the least, while neither agrees in laryngeal specification. Also, I did not expect a voiced cluster to allow more overlap than a voiceless cluster (i.e., VC4 >> VC1) in the NM speech. One possibility comes from the observation that the NM group did not release a voiced C1 due to the learning difficulty of voiced consonants (e.g., Anderson, 1987;

---

[14] In Hoole et al. (2009), French clusters were found to allow little overlap in both voiceless and voiced clusters (also see Bombien & Hoole, 2013).

Broselow & Finer, 1991; Chen & Chung, 2008; Hansen, 2001). In most cases, the NM group released a voiceless C1 before a voiced C2. Thus, a cluster of a voiced C1 without release would have a shorter duration than a cluster of a voiceless C1 with release, which is the case in (12). Lacking comparable findings as support, this preliminary account needs more research to verify.

5.3.4. Overlap degree and perceived accent

A final unpredicted result was the lack of correlation between the overlap degree and the perceived accent for the NM group. Intensive research has established the critical importance of suprasegmental features (i.e., stress, intonation, and rhythm) on comprehensibility and intelligibility of ESL learners (see Anderson-Hsieh et al. 1992; Anderson-Hsieh et al., 1994; Derwing et al., 1998; Derwing & Munro, 2005; Tajima et al., 1997). As one adjustment made in connected speech to promote the regularity of English rhythm, linking is considered the ability to speak English smoothly and to appropriately connect syllables; otherwise, nonnative English speech would present a salient choppy quality (Celce-Murcia et al., 1996; Gilbert, 1984; Wong, 1987). It is hypothesized that more overlap is indicative of linking ability, which would result in clearer speech.

However, in this study, I did not find a significant correlation between overlap degree and perceived accent in the NM group. Possibly, this was because other effects were more salient. Specifically, one of the two judges told me that there were a few Mandarin participants that had near native-like intonation but they did not produce the target vowels accurately. The judge mentioned one female participant with fluent intonation but produced *child* /tʃajld/ as /tʃild/. This observation supports previous research, which suggest the significance of vowel quality in perceiving foreign accent (Chen & Chung, 2008; Flege, 1993; Flege & Bohn, 1989; Hammond, 1986; Lee, Guion, & Harada, 2006; Magen, 1998; Zhang & Francis, 2010). Flege (1993) found that late Mandarin-speaking learners of English had much smaller vowel duration differences than the native speakers. More importantly, the larger the vowel duration difference in word pairs such as "beat", "bead" and "bat" produced by Chinese speakers, the lower the degree of foreign accent perceived by native English listeners.

Findings in this study again raise the importance of vowel acquisition for both instructors and students. The results found that degree of overlap was not correlated with perceived accent, likely due to any existing overlap effect being masked by more salient vowel-related issues.

5.4. Summary

This chapter has discussed the major findings regarding place of articulation, lexical frequency, stress positions, and speech rate. The first major finding here is that place of articulation (POA) had different impacts on English and Mandarin speakers. Specifically, the POE (more releases and more overlap in front-back than in back-front clusters) was only partially supported in native speech but not shown at all in nonnative speech. The NE group exhibited the significant release pattern of the POE but not the temporal pattern; the NM group did not show a significant POE in either release or temporal patterns. I argued for juncture effects weakening the POE, resulting in gestures drifting away.

The second major finding is that sound patterns were lexically gradient with speakers showing frequency-matching behaviors. Both the native and the nonnative groups had more overlap in high frequency words than in low frequency words (e.g., *take care* vs. *dock gate, keep pace* vs. *peep pate*). Group differences were found in the interaction between *POA * Frequency*. The NE group demonstrated the release pattern of the POE and showed the overlap trend of the POE in both real words and nonwords, implying the native grammar can extend to unfamiliar forms. In contrast, NM speakers had inconsistent overlap orders across lexical conditions. The NE group's performance provided strong evidence for a phonological representation of gestural units; further, these gestural units are available for native processing to re-combine and re-construct in new forms. I proposed that a set of gestural overlap patterning was language specific, and that L2 speakers in this study cannot generalize beyond memorized items.

This chapter also outlined and discussed unexpected findings from the data analyses. First, no correlation between overlap and speech rate was observed in either groups but a speaker-dependent effect was observed. Second, a trend of over-compensation was observed in the NM group: the NM group showed an even stronger stress effect than the NE group. The over-compensation comes from the influence of their native language, which has relatively equal syllable timing. Thirdly, NM speakers most often released a voiceless C1 before a voiced C2, and had less overlap in such a voicing condition. Previous research has argued for an effect of non-agreeing laryngeal specification. However, this account cannot explain Mandarin speakers also had significant overlap when C1 was voiced and C2 was voiceless, which does not agree in voicing either. I proposed it is because a cluster with a voiced C1 without release would have a shorter duration than a cluster with a voiceless C1 with release. However, I have not encountered other studies concerning this specific finding and this preliminary account calls for more examination. Lastly, the overlap degree produced by Mandarin speakers was not indicative of their perceived accent. This result was contrary to previous research, which suggests that coarticulation across words is critical to comprehensibility of spontaneously produced speech

Chapter 6

Conclusion

This final chapter completes this dissertation with three closing sections. The first section summarizes the research, from the research questions and hypotheses to the conclusions. The second section discusses the significance and the contributions this research makes to the growing body of literature on coarticulation. The final section concludes this dissertation by outlining the limitations of the research and suggesting potential future research endeavours.

6.1. Summary

This dissertation has outlined several linguistic constraints or factors that, together, can determine the overlap degree allowed in consonant clusters across English words. I have offered a detailed acoustic study, investigating the surface forms of stop-stop sequences. This research mainly sought to:

i. confirm that place of articulation affects L1 and L2 speakers differently
ii. determine whether and how lexical frequency affects L1 and L2 speakers
iii. determine whether and how stress patterns affect L1 and L2 speakers
iv. examine whether increased speech rate induces more overlap
v. ascertain whether the L2 speakers' overlap degree is correlated to their perceived foreign accent.

The research employed a group of native English speakers and a group of native Mandarin speakers with intermediate to advanced English proficiency. The stimuli were deliberately created and organized according to four major parameters: place of articulation, lexical frequency, stress, and speech rate. The general hypothesis was that the NM group would deviate from the NE group in terms of gestural coordination.

The data were analyzed according to release percentage and closure duration ratio in order to investigate the overlap degree and manner. The results showed that the place order effect (more releases and more overlap in front-back than in back-front clusters; POE) was only partially supported in native speech while not shown at all in nonnative speech: the NE group significantly exhibited the release pattern of the POE; but neither of the groups significantly showed the predicted temporal pattern. I argue that this was due to juncture effects where gestures get larger and longer at edges, weakening the POE. Nonetheless, the NE group extended the release patterns of the POE into nonwords and had consistent overlap orders across lexical settings; in contrast, the NM group deviated significantly from the native group especially in coordinating nonsense front-back and back-front clusters, suggesting they cannot extend their L2 phonetic awareness (if any) into unfamiliar phonological production. The consistency of the NE group showed that the POE is learnable in native processing, indicating sound patterns are registered in memory along with their phonetic details. In addition, both groups exhibited similar stress effects, which was even stronger in the NM group. Further analyses indicated that a faster speech rate did not consistently induce more overlap in either group. Lastly but surprisingly, there was no correlation between the overlap amount and the perceived accent in the NM group.

6.2. Significance

This study contributes to our understanding of gestural overlap in both L1 and L2 speech, and by extension, to L2 pedagogy.

6.2.1. Theoretical implications

First, this study offered a comprehensive and systematic approach to examining what factors contribute to overlap degree. Three factors have been identified to systematically determine the overlap degree in previous research; but this is the first study that combined *four* main factors in one study: place of articulation (POA), lexical frequency, stress, and speech rate. An exhaustive list of stop-stop combinations including over 30,000 tokens was examined in this research. For instance, corresponding nonwords were designed with similar vowel environments to correspond to real words, to eliminate possible vowel context influences. Also, all native English speakers in this study were born and raised in British Columbia, and all native Mandarin speakers were learners of above intermediate proficiency, assuring group homogeneity.

Second, this study contributes to the existing literature on what occurs with respect to phonetic timing in both native and nonnative speech, especially in relation to Mandarin speakers. Most existing studies have focused on word-initial or word-final clusters produced by Mandarin speakers (e.g., Bayley, 1996; Broselow et al., 1998; Fan, 2011; Hansen, 2001; Wang, 1995) while only a few have dealt with cross-word coarticulation (e.g., Chen & Chung, 2008; Messing, 2008; Tajima et al., 1997). When previous analyses mainly showed schwa insertion by Mandarin speakers in the production of English clusters, our findings suggested more detailed deviations from the native norm. The detailed findings in this study complement previous studies suggesting L2 speakers have less overlap and cannot apply appropriate speech coordination in the L2 (Sperbeck 2000; Davidson 2006; Zsiga 2003, 2011). The L2 findings in this study also provide us with a better understanding of acquiring coarticulation under different prosodic contexts (e.g., stressed versus unstressed position), which reflects the rhythmic effect on L2 speech development. This study suggests that NM speakers are affected by the rhythm of their native language, shown here as a trend towards over-compensation.

Third, this study expands our understanding of how phonological representations are differently registered in memory. This study is among the first, if not the first, examining the interaction between lexical frequency and the POE in both L1 and L2 speech. The results found that the native speakers could extend phonological knowledge to nonwords while the nonnative speakers could not. This finding essentially extends the fundamental hypothesis of Articulatory Phonology, indicating not only a gestural unit representation, but also an account of re-combinable gestures. Moreover, this result provides answers to the question of how sound patterns are registered in cognitive processing. Tightly bonded units are accessed as a whole along with their phonetic variations. This result points towards rich memory workload, suggesting native speakers process and store significantly richer and heavier informational structures than we ever could imagine (see Bybee, 2007; Kita & Özyürek, 2003; Port, 2007). In contrast, the way L2 speakers coordinate sound patterns is largely dependent upon their familiarity with individual lexical items, implying that L2 sound patterns are registered as discrete chunks.

6.2.2. Pedagogical implications

This study provides a detailed comparison of native speech versus nonnative speech. This is pedagogically valuable for second language instructors in that it identifies which particular articulatory patterns among Mandarin speakers, and L2 speakers in general, deviate from native speakers. For example, this research demonstrates that Mandarin speakers would most likely release a voiceless C1 before a voiced C2 (e.g., *lap dance*) while native English speakers would not. Another example is that Mandarin speakers have the most significant overlap in homorganic clusters, compared to the other two kinds of clusters while English speakers have the least overlap in homorganic clusters. These findings could raise instructors' awareness to pinpoint where pronunciation differences exist.

Another important implication comes from the finding of no correlation between overlap amount and the perceived accent in L2 speech. Previous studies show that suprasegmental instruction has a significant effect on listeners' impressions of the comprehensibility of spontaneously produced speech (e.g., Celce-Murcia et al., 1996; Derwing et al., 1998; Morley, 1991; Gilbert, 1984). Of course, having a near-native intonation pattern is beneficial to building first impressions. Two native English-speaking judges mentioned that Mandarin ESL learners who had good English intonation "sound confident". What hinders developing clearer speech, however, is that vowel pronunciations are not "full" or complete. Specifically, one participant in this study was found to have near native-like intonation but pronounced *child* /tʃajld/ as /tʃild/. This finding suggests that vowel quality may provide a better indicator of perceived accent. Benefitting from this finding, instructors could develop teaching strategies focusing on target like vowels, especially diphthongs, for L2 speakers. Future research is encouraged to investigate further on this aspect, which I believe would be beneficial for learners who have fossilized articulatory habits.

6.3. Limitations and future research

6.3.1. Limitations

While every methodological decision was carefully considered in this study, it is not without its limitations. One limitation concerns the elicitation tasks, in particular, Task 4. To elicit natural speech at fast speech rate, participants read the answer in each given dialogue (Task 4). This design was necessary to incorporate sentential stress. However, it is likely that participants' speech could be more natural if I did a picture description task rather than a reading task. Also, I elicited baseline closure durations from C1#V and V#C2 sequences in the current research; participants could either over-articulate or under-articulate while given the instruction of "producing the words as clearly as possible".

Another type of limitation concerns the measurement, which was purely dependent on acoustic features. No direct articulatory contact was measured. Following Zsiga (2003) and Ghosh and Narayanan (2009), I used closure duration to indicate incomplete stops and overlap amount in this study. Ghosh and Narayanan (2009) argue that closure duration can be used as an appropriate feature to detect incomplete stops for speech recognition with around 70% accuracy. One rationale is that given the considerable amount of stimuli employed in this study, it seems implausible to detect contact using Ultrasound or EMMA. On the other hand, no measurement technique is perfect. Ultrasound, for example, is slow for measuring raising and lowering tongue tip movements (Perkell, 1969). Another reason is, as discussed in the introduction, is that this study aims not only to examine the phonetic details of sound patterns (e.g., the POE), but also to seek the cognitive aspects of how sound patterns are registered in memory. In this view, acoustic consequences serve well for this purpose.

On the other hand, lacking direct articulatory contact evidence is not without limitations. As Browman (1995) comments, despite “a lawful relation” between the articulation and the acoustic signal, “this relation is not always linear” (p. 336). Coarticulation is a simultaneous modification of articulations (Catford, 1977). For example, there may be many positions of the tongue that could correspond to a single set of acoustic values. Thus, there are many directions in which this research can be extended in the future.

6.3.2. Future research

There are two directions of future research that could develop from the current study. The first line is to investigate further the large amount of rich data collected in the current research. In particular, I would like to examine the voicing effect, the correlation between frequency and release percentage, as well the correlation between perceived accent and release percentage. Specifically, I did not examine whether Mandarin speakers devoiced or deleted voiced stops (C1s) in this study, although these two common strategies were largely documented in previous literature. If they did use these strategies, this would likely affect the overall results. It would also be worth considering the effect of L1 dialect for the NM speakers. In particular, Wu dialect speakers have glottal stop as a possible coda consonant. This may have led to differences within NM speakers, based on dialect, which could be further explored. Moreover, I did not find a significant correlation between frequency and release percentage in this study. Analyses of frequency could have been done separately for each condition of voicing and place of articulation, and may have shown more subtle effects. For example, Table 4.13 (see Section 4.4.1) seems to suggest that homorganic clusters agreeing in voicing have fewer releases in high frequency items than heterorganic clusters agreeing in voicing. Different statistical analyses could be conducted to better assess how the frequency effect interacts with other factors. Finally, I did not evaluate whether there was a significant correlation between perceived accent and release percentage for the NM group; this was because duration ratio was more stable and less affected than release percentage, with fewer significant effects (see Section 4.2). It is possible, however, that L2 speakers' perceived accent is correlated with release percentage; this should be further explored.

The second line of future research is to complement the findings in this research with follow-studies involving additional data collection. Given that acoustic consequences are only partially indicative of coordination patterns, investigations of using various experimental instruments should be conducted in order to determine more precisely the manner of coarticulation and the degree of overlap. One direction is to use EMA, EPG or other techniques to examine gestural overlap in stop-stop sequences spanning word boundaries. The other direction is to examine other sound patterns in both L1 and L2 speech, and compare the results with the POE examined in this study. Research is especially encouraged on nasal assimilation produced by Mandarin speakers. This is because although Mandarin does not allow oral stops in coda position, it does allow nasal stops in this position, making us wonder whether Mandarin speakers will coordinate nasal-stop sequences more similarly to English speakers than stop-stop sequences.

Bibliography

Abercrombie, D. (1967). Elements of general phonetics (Vol. 203). Edinburgh: Edinburgh University Press.

Adams, S. G., Weismer, G., & Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech, Language, and Hearing Research*, *36*(1), 41-54.

Altenberg, E. P., & Vago, R. M. (1983). Theoretical implications of an error analysis of second language phonology production. *Language learning*, *33*(4), 427-447.

Anderson, J. (1987). The markedness differential hypothesis and syllable structure difficulty. In G. Ioup and S.H.Weinberger (Eds.) *Interlanguage phonology: The Acquisition of a Second Language Sound System,* pp, 279-291. Cambridge, MA: Newbury House.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning,* 42,529-555.

Anderson-Hsieh, J., Riney, T., & Koehler, K. (1994). Connected speech modifications in the English of Japanese learners. *Issues and Developments in English and Applied Linguistics,* 7, 31-52.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5(2), 183-207.

Avery, P., & Ehrlich, S. (1992). *Teaching American English Pronunciation: A Textbook* *and Reference Manual on Teaching the Pronunciation of North American English, Written Specifically for Teachers of English as a Second Language (ESL)*. RoutledgeFalmer (UK).

Avery, P., & Rice, K. (1989). Segment structure and coronal underspecification. *Phonology*, *6*(02), 179-200.

Barb, C. (2005). Suprasegmantals and comprehensibility: A comparative study in accent modification. Ph.D. dissertation, Wichita State University.

Barry, M. (1991). Temporal modelling of gestures in articulatory assimilation. In *Proceedings of the XXIIth congress of phonetic sciences* (Vol. 4, p. 14).

Barry, M. C. (1992). Palatalisation, assimilation and gestural weakening in connected speech. Speech communication, 11(4), 393-400.

Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. *Second Language Acquisition and Linguistic Variation*, 97-12.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. *Papers in Laboratory Phonology II*, 68-86.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92-111.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*,*104*(3), 591-630.

Berko, J. (1958). *The child's learning of English morphology*. Doctoral dissertation, Radcliffe College.

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press.

Bloomfield, L. (1933). *Language*. New York: Henry Holt.

Boersma, Paul & Weenink, David (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.10, retrieved 8 January 2016 from http://www.praat.org/

Bombien, L., Mooshammer, C., Hoole, P., Kühnert, B., & Schneeberg, J. (2006). An EPG study of initial/kl/clusters in varying prosodic conditions in German. In *Proc. 7th ISSP*. Ubatuba, Brasil.

Bombien, L., & Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. *The Journal of the Acoustical Society of America*, *134*(1), 539-550.

Bradley, T. G. (2007). Constraints on the metathesis of sonorant consonants in Judeo-Spanish. *Probus*, 19(2), 171-207.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3), 255-272.

Brill, J. (2008). Likert scale. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods.* (pp. 428-430). Thousand Oaks, CA: SAGE Publications,

Broselow, E. (1987). An investigation of transfer in second language phonology. *Interlanguage Phonology*, 261-278.

Broselow, E., Chen, S. I., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*(2), 261-28.

Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research*, *7*(1), 35-59.

Browman, C. P. (1995). Assimilation as gestural overlap: comments on Holst and Nolan. In Bruce Connell and Amalia Arvaniti (eds.), *Phonology and Phonetic Evidence, Papers in Laboratory Phonology IV*, pp. 334-342. Cambridge University Press.

Browman, C, & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3. 219-252.

Browman, C, & Goldstein, L. (1989). Articulatory gestures as phonology units. *Phonology*, 6, 201-251.

Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. In John Kingston and Mary Beckman (Eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, pp. 152-178. Cambridge University Press. Cambridge, England.

Browman, C, & Goldstein, L. (1992). Articulatory Phonology: An overview. *Phonetica,* 49, 155-18.

Browman, C, & Goldstein, L. (1995). Dynamics and articulatory phonology. *Mind as Motion*, 175-193.

Browman, C. P., & Goldstein, L. (1995). Gestural syllable position effects in American English. *Producing speech: Contemporary issues*, 19-33.

Browman, C, & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP,*Bulletin de la Communication Parlee, 5, 25-34.

Brown, G. (1977). *Listening to spoken English*. London: Longman.

Brown, J. D., & Kondo-Brown, K. (2006). Introducing connected speech. In J.D.Brown & K.Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, pp. 1-15. Honolulu, HI: University of Hawai'i. National Foreign Language Resource Center.

Brown, J. D., & Hilferty, A. (1986). The effectiveness of teaching reduced forms of listening comprehension. *RELC Journal*, *17*(2), 59-7.

Browne, S., & Huckin, T. (1987). Pronunciation tutorials for nonnative technical professionals: A program description. *Current Perspectives on Pronunciation: Practices Anchored in Theory*, 41-57.

Bush, N. (2001). Palatalization in English. *Frequency and the emergence of linguistic structure*, *45*, 255.

Butcher, A., & Tabain, M. (2004). On the back of the tongue: Dorsal sounds in Australian languages. *Phonetica*, *61*(1), 22-52.

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. *Usage-Based Models of Language*, 65-85.

Bybee, J. (2007). *Frequency of Use and the Organization of Language.* Oxford: Oxford University Press.

Bybee, J. L., & Hopper, P. J. (Eds.). (2001). *Frequency and the Emergence of Linguistic Structure* (Vol. 45). John Benjamins Publishing.

Byrd, D. (1992). Perception of Assimilation in Consonants Clusters: A Gestural Model. *Phonetica*, *49*(1), 1-24.

Byrd, D. (1993). 54,000 American stops. *UCLA Working Papers in Phonetics*, *83*, 97-116.

Byrd, D. (1995). Articulatory characteristics of single and blended lingual gestures. In *Proceedings of the XIIIth international congress of phonetic sciences* (pp. 438-441).

Byrd, D. (1996a). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, *24*(2), 209-244.

Byrd, D. (1996b). A phase window framework for articulatory timing. *Phonology*,*13*, 139-17.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, *57*(1), 3-16.

Byrd, D., & Choi, S. (2010). At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Laboratory Phonology*, *10*, 31-59.

Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In Michael B.Broe and Janet B. Pierrehumbert (Eds.), *Papers in laboratory phonology V*, pp. 70-87. Cambridge: Cambridge University Press.

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, *26*(2), 173-199.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, *31*(2), 149-18.

Byrd, D., & Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics, 24*(2), 263-282.

Cardoso, W. (2008). The development of sC onset clusters in interlanguage: Markedness vs. frequency effects. In R. Slabakova et al. (Eds.), *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference*, pp. 15-29 Somerville, MA (Cascadilla Proceedings Project).

Carlisle, R. (2006). The sonority cycle and the acquisition of complex onsets. In Barbara Baptista and Michael Watkins (Eds.) *English with a Latin Beat: Studies in Portuguese/Spanish-English Interphonology,* pp. 105-137.Amsterdam: John Benjamins.

Catford, J. C. (1977). *Fundamental Problems in Phonetics.* Midland Books.

Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition*, *22*(1), 1-26.

Celce-Murcia, Brinton., D., & Goodwin, J.M. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages.* Cambridge: Cambridge University Press.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese* . Berkeley, CA: University of California Press.

Chen, C., Fan, C., & Lin, H. (1996). A new perspective on teaching English pronunciation: rhythm. In *The proceedings of the 4th international symposium on English teaching* (pp. 24-41). Taipei: Crane, Inc.

Chen, H. C. (2005). The acquisition of American English rhythm by Taiwanese EFL learners. *English Teaching and Learning*, *29*(3), 1-23.

Chen, H. C., & Chung, R. F. (2008). Interlanguage analysis of phonetic timing patterns by Taiwanese learners. *Concentric: Studies in Linguistics*, *34*(1), 79-108.

Chen, Y., & Robb, M. P. (2004). A study of speaking rate in Mandarin speakers of

American English. *Asia Pacific Journal of Speech, Language and Hearing*, 9(3), 173-188.

Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). A study of sentence stress production in Mandarin speakers of American English. *J. Acoustic Soc. Am,* 109(4), 1681-90.

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*(02), 135-158.

Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. *Laboratory Phonology*, *7*, 419-447.

Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, *32*(2), 141-176.

Cho, T., & Jun, S. A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA working papers in phonetics*, 57-70.

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155-190.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207-229.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics, 33*(2), 121-157.

Chun, D. (1982). A contrastive study of the suprasegmental pitch in modern German, American English and Mandarin Chinese. Unpublished doctoral dissertation, University of California, Berkeley.

Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change*, *7*(02), 209-228.

Clements, G. N. (1985). The geometry of phonological features. *Phonology*, 2(01), 225-252.

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*(3), 489.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America*, *83*(4), 1553-1573.

Crystal, T. H., & House, A. S. (1990). Articulation rate and the duration of syllables and stress groups in connected speech. *The Journal of the Acoustical Society of America*, 88(1), 101-112.

Cutler, A. (1984). Stress and accent in language production and understanding. *Intonation, accent and rhythm: Studies in discourse phonology*, 8, 76-90.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of memory and language*, 25(4), 385-400.

Davidson, L. (2003). The atoms of phonological representation: Gestures, coordination and perceptual features in consonant cluster phonotactics. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore.

Davidson, L. (2006). Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. *Cognitive Science*, *30*(5), 837-862.

Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, *38*(2), 272-288.

Davidson, L. (2011). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, *53*(8), 1042-1058.

Davidson, L., Jusczyk, P., & Smolensky, P. (2004). 10 The initial and final states: theoretical implications and experimental explorations of Richness of the Base. *Constraints in Phonological Acquisition*, 321.

Davidson, L., & Roon, K. (2008). Durational correlates for differentiating consonant sequences in Russian. *Journal of the International Phonetic Association*, *38*(02), 137-165.

Davidson, L., & Stone, M. (2003). Epenthesis versus gestural mistiming in consonant cluster production: an ultrasound study. In *Proc. of the West Coast Conference on Formal Linguistics* (Vol. 22, pp. 165-178).

De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of the Acoustical Society of America*, *97*(1), 491-504.

De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics*, 32(4), 493-516.

De Jong, K., Beckman, M. E., & Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and speech*, *36*(2-3), 197-212.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385-390.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient forΔ C. *Language* *and language-processing,* 231-241.

Dinnsen, D., & Garcia-Zamor, M. (1971). The three degrees of vowel length in German. *Research on Language & Social Interaction*, *4*(1), 111-126.

Dixit, R. P., & Flege, J. E. (1991). Vowel context, rate and loudness effects of linguopalatal contact patterns in Hindi retroflex. *Journal of Phonetics*, *19*(2), 213-229.

Draper, M. H., Ladefoged, P., & Whitteridge, D. (1960). Expiratory pressures and air flow during speech. *British Medical Journal* 1(5189):1837-1843.

Duanmu, S. (2007). *The Phonology of Standard Chinese* (second edition). Oxford: Oxford University Press.

Eckman, F. R. (1981). On the naturalness of interlanguage phonological rules. *Language Learning*, *31*(1), 195-216.

Eckman, F. R., & Iverson, G. K. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research*, *9*(3), 234-252.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America*, 89(1), 369-382.

Edwards, H.T., & Strattman, K. H. (1996). *Accent Modification Manual: Materials and Activities.* San Diego, CA: Singular Publishing Group.

Ellis, L., & Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: evidence from EPG and EMA data. *Journal of Phonetics*, *30*(3), 373-396.

Elson, Benjamin. (1956). Sierra Popoluca syllable structure. *International Journal of American Linguistics.* 13, 13-17.

Engstrand, O. 1988. Articulatory correlates of stress and speaking rate in Swedish VCV utterances. *The journal of the Acoustical society of America*, *83*(5), 1863-1875.

Ernestus, M., & Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, 5-38.

Faber, D. (1986). Teaching the rhythms of English: a new theoretical base. *IRAL-International Review of Applied Linguistics in Language Teaching,* 24(1-4), 205-216.

Fan, Y. N, (2011). *Articulatory Timing of English Consonant Clusters in the Coda Positions: A Study of Chinese-English Interlanguage.* Master Thesis. University of Victoria.

Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589-1608.

Flege, J. E., & Bohn, O. S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in second language acquisition*, 11(01), 35-62.

Flege, J. E., & Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics*, *5*(04), 323-347.

Fourakis, M., & Iverson, G. K. (1984). On the 'incomplete neutralization'of German final obstruents. *Phonetica*, *41*(3), 140-149.

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, *8*(1), 113-133.

Fowler, C. A., & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech*, *36*(2-3), 171-195.

Gafos, A. (2002). A Grammar of Gestural Coordination. *Natural Language and Linguistic Theory*, 20, 269-337.

Gafos, A., Hoole, P., Roon, K., & Zeroual, C. (2010). Variation in overlap and phonological grammar in Moroccan Arabic clusters. *Laboratory Phonology X, Mouton de Gruyter, Berlin/New York*.

Gao, M., Mooshammer, C., Hagedorn, C., Nam, H., Tiede, M., Chang, Y. C., & Goldstein, L. (2011). Intra-and inter-syllabic coordination: An articulatory study of Taiwanese and English. In *Proceedings of XVI International Congress of Phonetic Sciences* (pp. 723-736).

Ghosh, P. K., & Narayanan, S. S. (2009). Closure duration analysis of incomplete stop consonants due to stop-stop interaction. *The Journal of the Acoustical Society of America*, *126*(1), EL1-EL7.

Gick, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology* 16, 29-54.

Gick, B., Wilson, I., & Derrick, D. (2012). *Articulatory Phonetics*. John Wiley & Sons.

Gilbert, J.B. (1984). *Clear Speech: Pronunciation and Listening Comprehension in American English.* (1st ed.) New York: Cambridge University Press.

Gilbert, J. (1995). Pronunciation practice as an aid to listening comprehension. *A Guide for the Teaching of Second Language Learning*, 97-111.

Gimson, A.C. (1970). *An introduction to the pronunciation of English* (2nd ed.) London: Arnold.

Gobl, C., & Ní Chasaide, A. (1999). Voice source variation in the vowel as a function of consonantal context. *Coarticulation: Theory, Data and Techniques*, 122-143.

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

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. *Action to language via the mirror neuron system*, 215-249.

Goldstein, L., & Fowler, C. A. (2003). Articulatory phonology: A phonology for public language use. *Phonetics and phonology in language comprehension and production: Differences and similarities*, 159-207.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology*, 7(515-546).

Grant, L. (2000). *Well Said: Advanced English Pronunciation*. Boston: Heinle & Heinle.

Gregory, M., Raymond, W., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). "The effects of collocational strength and contextual predictability in lexical production," *Chicago Ling. Soc.* **35**, 151–166.

Guy, G. (1980). Variation in the group and the individual: the case of final stop deletion. In W. Labov (Ed.), *Locating Language in Time and Space*, pp. 1-36. New York: Academic.

Hall, N. E. (2003). *Gestures and segments: Vowel intrusion as overlap*. Doctoral Dissertation, University of Massachusetts.

Hall, T. A., Hamann, S., and Zygis, M. (2006). The phonetic motivation for phonological stop assibilation. *Journal of the International Phonetic Association*, *36*(1), 59-81.

Hardcastle, W. J., & Roach, P. J. (1979). An instrumental investigation of coarticulation in stop consonant sequences. *Current Issues in the Phonetic Sciences*, *1*, 531-54.

Hardcastle, W. J. (1982). Constraints on coarticulatory processes. In D. Crystal (Ed.) *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour of FR Palmer*, pp. 33-49. Edward Arnold, London.

Hardcastle, W. J. (1985). Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, *4*(1), 247-263.

Hansen, J. G. (2001). Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, *22*(3), 338-365.

Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell.

Hayes, B. (1986). Assimilation as spreading in Toba Batak. *Linguistic Inquiry*, 467-499.

Hayes, B., Siptár, P., Zuraw, K., & Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, *85*(4), 822-863.

Henrichsen, L. E. (1984). Sandhi-variation: A filter of input for learners of ESL. *Language Learning*, *34*(3), 103-123.

Heyer, S. (1986). English final consonants and the Chinese learner. Unpublished master's thesis, Southern Illinois University, Edwardsville

Hieke, A. (1987). Absorption and fluency in native and non-native casual English speech. *Sound Patterns in Second Language Acquisition*, 41-58.

Hoequist Jr, C. (1983). Syllable duration in stress-, syllable-and mora-timed languages. *Phonetica*, 40(3), 203-237.

Holst, T., & Nolan, F. (1995). The influence of syntactic structure on [s] to [S] assimilation. *Papers in Laboratory Phonology, IV*, 315-333.

Hoole, P., Bombien, L., Kühnert, B., & Mooshammer, C. (2009). Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. *Frontiers in Phonetics and Speech Science*, 275-287.

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, 117(2), 896-907.

Ito, Y. (2006). The significance of reduced forms in L2 pedagogy. . In J. D. Brown & K. Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, pp. 17-25. Honolulu, HI: University of Hawai'i. National Foreign Language Resource Center.

Jaeger, M., & Hoole, P. (2011). Articulatory factors influencing regressive place assimilation across word boundaries in German. *Journal of Phonetics*, *39*(3), 413-428.

Jannedy, S. (1994). Rate effects on German unstressed syllables. *OSU Working Papers in Linguistics* 44, 105-124.

Jun, J. (1995). *Perceptual and articulatory factors in place assimilation: An Optimality Theoretic approach*. Ph.D dissertation, UCLA.

Jun, S. A. (1996). *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure*. Taylor & Francis.

Jun, J. (2004). Place assimilation. In Bruce Hayes, Robert Kirchner & Donca Steriade (Eds.) *Phonetically Based Phonology*, pp. 58-86. Cambridge: Cambridge University Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typological Studies in Language*, *45*, 229-254.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 286-319.

Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. *Papers in laboratory phonology* I, 26, 451-470.

Kent, R. D., & Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica*, 24(1), 23-44.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25(2), 143-168.

Kim, H. (1995). Intake from the speech stream: Speech elements that L2 learners attend to. *Attention and Awareness in Foreign Language Learning*, 65-83.

Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonology*, 2(01), 85-138.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research,* 18(4), 686-706.

Kratochvil, P. (1968). *The Chinese Language Today*. London: Hutchinson University Press.

Kreidler, C. W. (1989). *The Pronunciation of English: A Course Book in Phonology*. Oxford, England: Blackwell.

Kochetov, A. (2001). *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization.* PhD Dissertation. University of Toronto

Kochetov, A., & Goldstein, L. (2001). Competing recoverability factors and inter-gestural phasing in Russian stop clusters. In *Annual Meeting of Linguistic Society of America*.

Kochetov, A., & Pouplier, M. (2008). Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology*, *25*(3), 399-431.

Kochetov, A., Pouplier, M., & Son, M. (2007). Cross-language differences in overlap and assimilation patterns in Korean and Russian. In *Proceedings of the 16th International Congress of Phonetics Sciences (ICPhS),* pp. 1361-1364, *Saarbrücken, Germany*.

Kohler, K.J. (1992). Gestural reorganization in connected speech: a functional viewpoint on Articulatory Phonology. *Phonetica* 49, 205-211.

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, 715-762.

Ladd, D. R., & Scobbie, J. M. (2003). External sandhi as gestural overlap? Counterevidence from Sardinian. *Phonetic interpretation: Papers in laboratory phonology VI*, 164-182.

Ladefoged, P. (1999). American English. *Handbook of the International Phonetic Association,* 4144.

Ladefoged, P. (2001). *A Course in Phonetics*. San Diego, CA: Harcourt Brace.

Lass, R. (1984). *Phonology: An Introduction to Basic Concepts*. Cambridge University Press.

Lin, H. (2001). *A Grammar of Mandarin Chinese*. Vol. 344. Muenchen: Lincom Europa.

Lin, H., & Wang, Q. (2007). Mandarin rhythm: An acoustic study. *Journal of Chinese Linguistics and Computing, 17 (3),* 127-140.

Lin, H, & Huang, S. (2009). Rhythm in the Chinese dialects. *Paper presented at the Annual Canadian Linguistics Association Conference*, Ottawa.

Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (Ed.), *The Production of Speech*, pp. 217-245. Springer-Verlag, New York.

Ling, L. E., Grabe, E., & Nolan, F. (2000). Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English. *Language and speech*, *43*(4), 377-401.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language*, *33* (1), 42-49.

Lisker, L. (1999). Perceiving final voiceless stops without release: effects of preceding monophthongs versus nonmonophthongs. *Phonetica*, *56* (1-2), 44-55.

Luo, S. (2013). Explaining Mandarin T3 Sandhi within an Articulatory Phonology Framework. Paper presented in the *Annual Conference of the Canadian Linguistics Association*, Victoria, Canada.

Luo, C., & Wang, J. (1981). *Putong Yuyinxue Gangyao [Outline of General Phonetics].* Beijing:Shangwu Yinshuguan.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge university press.

Major, R. C. (1987). Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition*, *9*(01), 63-82.

Major, R. C., & Faudree, M. C. (1996). Markedness universals and the acquisition of voicing contrasts by Korean speakers of English. *Studies in Second Language Acquisition*, *18*(01), 69-9.

Marchal, A. (1988). Coproduction: Evidence from EPG data. *Speech Communication*, 7(3), 287-295.

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic inquiry*, 207-263.

McMahon, A., Foulkes, P., & Tollfree, L. (1994). Gestural representation and lexical phonology. *Phonology*, *11*(02), 277-316.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *The Journal of the Acoustical Society of America*, *88*(3), 1286-1298.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research*, *77*(78), 69-78.

Messing, C.L. (2008). *Intrusive Segments as Instances of Gestural Mistiming: Mandarin L1 English Learners' Production of English Consonant Clusters.* Master Thesis. University of California, Davis.

Moisik, S., & Esling, J. H. (2011). The 'whole larynx' approach to laryngeal features. *Proceedings of ICPhS, Hong Kong*, 1406-1409.

Moisik, S. R., Lin, H., & Esling, J. H. (2010). An investigation of laryngeal behavior during Mandarin tone production using simultaneous laryngoscopy and laryngeal ultrasound. In *Proceedings of the 9th Phonetics Conference of China*.

Moisik, S., Lin, H., & Esling, J. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). Journal of the International Phonetic Association, 44 (01), pp. 21 – 58.

Mok, P. P., & Dellwo, V. (2008). Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *In Proc. Speech Prosody* (pp. 423-426).

Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical society of America*, *96*(1), 40-55.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *Tesol Quarterly*, 481-520.

Nakatani, L. H., O'Connor, K. D., & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *The Journal of the Acoustical Society of America*, 69(F1), S82-S82.

Ng, M. L., & Chen, Y. (2011). Proficiency in English sentence stress production by Cantonese speakers who speak English as a second language (ESL). *International journal of speech-language pathology,* 13(6), 526-535.

Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. *Papers in laboratory phonology II: Gesture, segment, prosody*, 261-28.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In Peter MacNeilage (Ed.) *The Production of Speech*, pp. 189-216. Springer New York.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik and Ernst Håkon Jahr (Eds.), *Language change: Contributions to the Study of its Causes*, pp, 173-198. Berlin: Mouton de Gruyter.

Ohala, J. J. (1993). Coarticulation and phonology. *Language and Speech*, *36*(2-3), 155-17.

Ohala, J., & Hirano, M. (1967). An experimental investigation of pitch change in speech.

The Journal of the Acoustical Society of America, 42(5), 1208-1209.

Öhman, S. (1967). Word and sentence intonation: A quantitative model. Speech Transmission Laboratory, Department of Speech Communication, Royal Institute of Technology.

Peng, Shu-Hui (1996). Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi. Ph.D dissertation, Ohio State University.

Peng, S. H. (2000). Lexical versus 'phonological'representations of Mandarin Sandhi tones. *Papers in laboratory phonology V: Acquisition and the lexicon*, 152-167.

Perkell, J. S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study* (No. 53). MIT Press.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*(4), 2561-2569.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking Clearly for the Hard of Hearing IIntelligibility Differences between Clear and Conversational Speech. *Journal of Speech, Language, and Hearing Research*, *28*(1), 96-103.

Pind, J. (1995). Constancy and normalization in the perception of voice offset time as a cue for preaspiration. *Acta psychologica*, *89*(1), 53-81.

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, *25*(2), 143-170.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, *17*(4), 257-282.

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German.

Journal of Phonetics.13, 455-471

Prator, C. H., & Robinett, B. W. (1985). *Manual of American English pronunciation*. Holt, Rinehart and Winston.

Pulleyblank, D. (1988). Vocalic underspecification in Yoruba*. Linguistic Inquiry*, 233-270.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In *Proceedings of the 22nd west coast conference on formal linguistics* (Vol. 22, pp. 101-114). Somerville, MA: Cascadilla Press.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73(3), 265-292.

Raymond., W.D., Dautricourt, R., & Hume, E. (2006).Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18, 43-55.

Recasens, D., Fontdevila, J., Pallarès, M. D., & Solanas, A. (1993). An electropalatographic study of stop consonant clusters. *Speech Communication,* 12(4), 335-355.

Rubach, J. (1984). Rule typology and phonological interference. *Theoretical Issues in Contrastive Phonology. Heidelberg: Julius Groos Verlag*, 37-50.

Rogers, H. (2000). *The Sounds of Language: An Introduction to Phonetics*. Essex, UK: Pearson Education Limited.

Rogerson, M. (2006). Don’cha know? A survey of ESL teachers’ perspectives on reduced forms instruction. In J.D.Brown & K.Kondo-Brown (Eds.), *Perspectives on teaching connected speech to second language speakers*, pp. 85-97. Honolulu, HI: University of Hawai'i. National Foreign Language Resource Center.

Romero, Joaquín. (1996). Articulatory blending of lingual gestures. *Journal of Phonetics* 24: 99–111.

Sato, C. (1987). Phonological processes in second language acquisition: another look at interlanguage syllable structure. In G. Ioup & S. Weinberger (Eds.), *Interlanguage Phonology: the Acquisition of a Second Language Sound System*, pp. 248–6. Cambridge, MA: Newbury House

Scarcella, R. C., & Oxford, R. L. (1994). Second language pronunciation: State of the art in instruction. *System*, *22*(2), 221-23.

Schneider, W., Eschman, A., and Zuccolotto, A. (2012). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Scobbie, M. (2007) Interface and overlap in phonetics and phonology. In Gillian Ramchand & Charles Reiss (Eds.) *The Oxford Handbook of Linguistic Interfaces*. pp. 17-52. Oxford: Oxford University Press.

Shaiman, S. (2001). Kinematics of compensatory vowel shortening: the effect of speaking rate and coda composition on intra-and inter-articulatory timing. *Journal of Phonetics*, *29*(1), 89-107.

Shaiman, S., Adams, S. G., & Kimelman, M. D. (1995). Timing relationships of the upper lip and jawacross changes in speaking rate. *Journal of Phonetics*, *23*(1), 119-128.

Shockey, L. (2008). *Sound Patterns of Spoken English*. John Wiley & Sons.

Solé, M. J. (1997). Timing patterns in the production of a foreign language. In *Views on the Acquisition and Use of a Second Language: Proceedings of the European Second Language Association 7th annual conference* (pp. 539-551).

Stephenson, L., & Harrington, J. (2002). Assimilation of place of articulation: Evidence from English and Japanese. In *Proceedings of the 9th Australian international conference on speech science and technology* (pp. 592-597).

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *The Journal of the Acoustical Society of America*, *68*(3), 836-842.

Stevens, K. N. (2000). Diverse acoustic cues at consonantal landmarks. *Phonetica*, *57*(2-4

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America*, 82(3), 847-863.), 139-151.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*(1), 1-24.

Tarone, E. (1987). Some influences on the syllable structure of interlanguage phonology. *Interlanguage phonology: The acquisition of a second language sound system.* Cambridge: Newbury House Publishers.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2007). The effect of speaking rate on voice-onset-time is talker-specific. In *Proceedings of ICPHs* (Vol. 16, pp. 473-476).

Tjaden, K., & Weismer, G. (1998). Speaking-rate-induced variability in F2 trajectories. *Journal of Speech, Language, and Hearing Research*, *41*(5), 976-989.

Tsao, Y. C., Weismer, G., & Iqbal, K. (2006). Interspeaker variation in habitual speaking rate: Additional evidence. *Journal of Speech, Language, and Hearing Research* 49(5):1156-1164.

Tseng, C. Y. (1990). *An Acoustic Phonetic Study on Tones in Mandarin Chinese.* Nankang, Taipei: Institute of History & Philology, Academia Sinica.

Wang, C. (1995). The acquisition of English word-final obstruents by Chinese speakers. PhD dissertation, State University of New York at Stony Brook.

Wang, W. S. (Ed.). (1977). The lexicon in phonological change (Vol. 5). Walter de Gruyter.

Wang, G., & Kong, J. (2010). The relation between larynx height and F0 during the four tones of Mandarin in X-ray movie. In Chinese Spoken Language Processing (ISCSLP), Taiwan. pp. 335-338.

Wang, W. S., & Li, K. P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, *10*(3), 629.

Warner, N., Jongman, A., Cutler, A., & Mücke, D. (2001). The phonological status of Dutch epenthetic schwa. *Phonology*, *18*(03), 387-420.

Weinberger, S. (1987). The influence of linguistic context on syllable simplification. In G. Ioup & S.Weinberger (Eds.) *Interlanguage phonology: The Acquisition of a Second Language Sound System,* pp, 401-417. Cambridge, MA: Newbury House.

Weinberger, S. (1994). Functional and phonetic constraints on second language phonology. In MehmetYavaş (Ed*.), First and Second Language Phonology*, pp. 283-302. San Diego, CA: Singular.

Weinstein, N. J. (2001). *Whaddaya say?: Guided practice in relaxed speech*. Longman.

Wilson, C. (2003). Experimental investigation of phonological naturalness. In *Proceedings of the 22nd west coast conference on formal linguistics* (Vol. 22, pp. 533-546). Somerville, MA: Cascadilla Press.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945-982

Wilson, C., Davidson, L., & Martin, S. (2014). Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language*, *77*, 1-24.

Yanagawa, M. (2006). Articulatory timing in first and second language: a cross-linguistic study. PhD dissertation. Yale University.

Yeou, M. (2004). The realization of accentual focus in Moroccan learners of English. In Speech Prosody 2004, International Conference.

Zeroual, C., Hoole, P., Gafos, A. I., & Esling, J. H. (2014). Gestural overlap within word medial stop-stop sequences in Moroccan Arabic. 10th ISSP Cologne.

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics*, 38(2), 260-271.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*(01), 153-201.

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, *123*(6), 4498-4513.

Zsiga, E. C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics*, *22*(2), 121-14.

Zsiga, E. C. (1995). An acoustic and electropalatographic study of lexical and postlexical palatalization in American English.. In Bruce Connell and Amalia Arvaniti (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, *4*, 282.

Zsiga, E. C. (1997). Features, gestures, and Igbo vowels: An approach to the phonology-phonetics interface. *Language*, 227-274.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of phonetics*, *28*(1), 69-102

Zsiga, E. C. (2003). Articulatory timing in a second language. *Studies in Second Language Acquisition*, *25*(03), 399-432.

Zsiga, E. C. (2011). External Sandhi in a second language: the phonetics and phonology of obstruent nasalization in Korean-accented English. *Language*,*87*(2), 289-345.

Zue, V. W. (1976). *Acoustic Characteristics of Stop Consonants: A Controlled Study*. Sc.D. Thesis. MIT, Cambridge, MA.

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language*, *83*(2), 277-316.

Appendix A: Testing Material

Pre-test: Please elaborate on a topic of "what do you feel most passionate about?" for about 5 minutes.

List 1: 24 Items of /C1#V/ and /V#C2/

Single Coda C1$_b$#V	Single Onset V#C2$_b$
keep art, pep art, job art, rib art, hot art, shut art, look art, rock art, food art, red art, leg art, big art	saw pain, saw pen, saw balm, saw boy, saw tie, saw take, saw care, saw car, saw day, saw down, saw good, ,saw game

List 2: 144 Items of /C1#C2/

		Nonwords C1#C2	Real words C1#C2	Real words C1#C2	Nonwords C1#C2
homorganic clusters	p#p	foop pok	soup pot	keep pace	peep pate
	p#b	mip bong	lip balm	deep bite	beep bipe
	t#t	shot tud	hot tub	get tip	det tik
	t#d	zet dep	vet debt	beat down	deat doun
	k#k	dak kit	back kick	take care	pake kare
	k#g	bek gaep	deck gate	look good	mook goob
	b#p	lib peim	rib pain	grab pen	brab pem
	b#b	zob banp	job bank	rub back	mub bap
	d#t	fade tenk	shade tent	need time	meed tine
	d#d	nud dei	mud day	bad day	gad dei
	g#k	grag kar	drag car	smug king	swug kim
	g#g	brug gam	drug gang	big game	dig gane
front-to-back clusters	p#t	kep tot	pep talk	skip test	stip tesp
	p#d	rap dant	lap dance	deep down	beep doun
	p#k	mip kare	lip care	keep calm	teep kalm
	p#g	pape gum	tape gun	top gear	pop geal
	t#g	biet gaib	diet guide	great guy	breat gai
	t#k	doot kamk	boot camp	fat cat	shat kap

	b#t	rab tesp	lab test	glib talk	blib tot
	b#d	prab dis	crab dish	rub down	lub doun
	b#k	rab coak	lab coat	grab cash	brab kaf
	b#g	plub girm	club girl	fab gift	sab gisp
	d#g	fead gane	head game	good game	bood gane
	d#k	sood kork	food court	bad kid	gad kig
back-to-front clusters	t#p	fuit panf	suit pants	bright pink	grait pinp
	t#b	srat boi	frat boy	get back	det bap
	k#p	feek paim	cheek pain	look pale	rook peil
	k#t	mek tai	neck tie	black tea	glack tee
	k#b	lock banb	rock band	quick boost	kwik booft
	k#d	bok dape	dock date	kick down	tik doun
	d#p	fead paim	head pain	red pen	led pem
	d#b	ded bud	bed bug	need beer	meed beel
	g#t	gog tael	dog tail	big time	dig tine
	g#d	dag dear	bag deal	big deal	kig deer
	g#p	glog peiz	blog page	hug pug	sug puz
	g#b	reg bome	leg bone	big boy	dig boi

List 3: Engage in a dialogue.

1. keep pace

a). Prompt: Did you say “**take** pace” just now? Target: No, I said “**keep** pace” just now.
b). Prompt: Did you say “keep **baby**” just now? Target: No, I said “keep **pace”** just now.
c). Target: You know what? He said “**keep pace”** just now.
d). Prompt: Did **he** say “keep pace” just now? Target: No, **she** said “keep pace” just now.

2. peep pate
a). Prompt: Did you say “**pob** pate” just now? Target: No, I said “**peep** pate” just now.
b). Prompt: Did you say “peep **bot**” just now? Target: No, I said “peep **pate**” just now.
c). Target: You know what? He said “**peep pate”** just now.
d). Prompt: Did **he** say “peep pate” just now? Target: No, **she** said “peep pate” just now.

3. bed bug
a). Prompt: Did you say “**frat** bug” just now? Target: No, I said “**bed** bug” just now.
b). Prompt: Did you say “bed **pants**” just now? Target: No, I said “bed **bug”** just now.
c). Target: You know what? He said “**bed bug”** just now.
d). Prompt: Did **he** say “bed bug” just now? Target: No, **she** said “bed bug” just now.

4. ked bud
a). Prompt: Did you say “**rot** bud” just now? Target: No, I said “**ked** bud” just now.
b). Prompt: Did you say “ked **pot**” just now? Target: No, I said “ked b**ud**” just now.
c). Target: You know what? He said “**ked bud”** just now.
d). Prompt: Did **he** say “ked bud” just now? Target: No, **she** said “ked bud” just now.

5. rock band
a). Prompt: Did you say “**cab** band” just now? Target: No, I said “**rock** band” just now.
b). Prompt: Did you say “rock **pants**” just now? Target: No, I said “rock **band”** just now.
c). Target: You know what? He said “**rock band”** just now.
d). Prompt: Did **he** say “rock band” just now? Target: No, **she** said “rock band” just now.

6. lock banb
a). Prompt: Did you say “**dog** bot” just now? Target: No, I said “**dock** bant” just now.
b). Prompt: Did you say “dock **pot**” just now? Target: No, I said “dock **bant**” just now.
c). Target: You know what? He said “**dock bant”** just now.
d). Prompt: Did **he** say “dock bant” just now? Target: No, **she** said “dock bant” just now.

7. leg bone
a). Prompt: Did you say “**cab** bone” just now? Target: No, I said “**leg** bone” just now.
b). Prompt: Did you say “leg **pants**” just now? Target: No, I said “leg **bone”** just now.
c). Target: You know what? He said “**leg bone”** just now.
d). Prompt: Did **he** say “leg bone” just now? Target: No, **she** said “leg bone” just now.

8. reg bome
a). Prompt: Did you say “**dock** boun” just now? Target: No, I said “**reg** boun” just now.
b). Prompt: Did you say “reg **pot**” just now? Target: No, I said “reg **boun**” just now.
c). Target: You know what? He said “**reg boun”** just now.
d). Prompt: Did **he** say “reg boun” just now? Target: No, **she** said “reg boun” just now.

9. pep talk
a). Prompt: Did you say “**cab** talk” just now? Target: No, I said “**pep** talk” just now.
b). Prompt: Did you say “pep **door**” just now? Target: No, I said “pep **talk”** just now.
c). Target: You know what? He said “**pep talk”** just now.
d). Prompt: Did **he** say “pep talk” just now? Target: No, **she** said “pep talk” just now.

10. kep tot
a). Prompt: Did you say “**mob** tot” just now? Target: No, I said “**kep** tot” just now.
b). Prompt: Did you say “kep **dot**” just now? Target: No, I said “kep **tot**” just now.
c). Target: You know what? He said “**kep tot”** just now.
d). Prompt: Did **he** say “kep tot” just now? Target: No, **she** said “kep tot” just now.

11. lab test
a). Prompt: Did you say “**get** test” just now? Target: No, I said “**lab** test” just now.
b). Prompt: Did you say “lab **people**” just now? Target: No, I said “lab **test”** just now.

c). Target: You know what? He said “**lab test”** just now.
d). Prompt: Did **he** say “lab test” just now? Target: No, **she** said “lab test” just now.

12. rab tesp
a). Prompt: Did you say “**pop** tesp” just now? Target: No, I said “**rab** tesp” just now.
b). Prompt: Did you say “rab **dot**” just now? Target: No, I said “rab **tesp**” just now.
c). Target: You know what? He said “**rab tesp”** just now.
d). Prompt: Did **he** say “rab tesp” just now? Target: No, **she** said “rab tesp” just now.

13. hot tub
a). Prompt: Did you say “**cab** tub” just now? Target: No, I said “**hot** tub” just now.
b). Prompt: Did you say “hot **pants**” just now? Target: No, I said “hot **tub”** just now.
c). Target: You know what? He said “**hot tub”** just now.
d). Prompt: Did **he** say “hot tub” just now? Target: No, **she** said “hot tub” just now.

14. shot tud
a). Prompt: Did you say “**cod** tud” just now? Target: No, I said “**shot** tud” just now.
b). Prompt: Did you say “shot **dot**” just now? Target: No, I said “shot **tud**” just now.
c). Target: You know what? He said “**shot tud”** just now.
d). Prompt: Did **he** say “shot tud” just now? Target: No, **she** said “shot tud” just now.

15. good tip
a). Prompt: Did you say “**big** tip” just now? Target: No, I said “**good** tip” just now.
b). Prompt: Did you say “good **pants**” just now? Target: No, I said “good **tip”** just now.
c). Target: You know what? He said “**good tip”** just now.
d). Prompt: Did **he** say “good tip” just now? Target: No, **she** said “good tip” just now.

16. dood tik
a). Prompt: Did you say “**rot** tik” just now? Target: No, I said “**dood** tik” just now.
b). Prompt: Did you say “dood **dot**” just now? Target: No, I said “dood **tik**” just now.
c). Target: You know what? He said “**dood tik”** just now.
d). Prompt: Did **he** say “dood tik” just now? Target: No, **she** said “dood tik” just now.

17. neck tie
a). Prompt: Did you say “**cab** tie” just now? Target: No, I said “**neck** tie” just now.
b). Prompt: Did you say “neck **baby**” just now? Target: No, I said “neck **tie”** just now.
c). Target: You know what? He said “**neck tie”** just now.
d). Prompt: Did **he** say “neck tie” just now? Target: No, **she** said “neck tie” just now.

18. mek tai
a). Prompt: Did you say “**dog** tai” just now? Target: No, I said “**mek** tai” just now.
b). Prompt: Did you say “mek **dot**” just now? Target: No, I said “mek **tai**” just now.
c). Target: You know what? He said “**mek tai”** just now.
d). Prompt: Did **he** say “mek tai” just now? Target: No, **she** said “mek tai” just now.

19. dog tail
a). Prompt: Did you say "**cab** tail" just now? Target: No, I said "**dog** tail" just now.
b). Prompt: Did you say "dog **baby**" just now? Target: No, I said "dog **tail"** just now.
c). Target: You know what? He said "**dog tail"** just now.
d). Prompt: Did **he** say "dog tail" just now? Target: No, **she** said "dog tail" just now.

20. kog teir
a). Prompt: Did you say "**dock** teir" just now? Target: No, I said "**kog** teir" just now.
b). Prompt: Did you say "kog **dot**" just now? Target: No, I said "kog **teir**" just now.
c). Target: You know what? He said "**kog teir"** just now.
d). Prompt: Did **he** say "kog teir" just now? Target: No, **she** said "kog teir" just now.

21. lap dance
a). Prompt: Did you say "**club** dance" just now? Target: No, I said "**lap** dance" just now.
b). Prompt: Did you say "lap **pants**" just now? Target: No, I said "lap **dance"** just now.
c). Target: You know what? He said "**lap dance"** just now.
d). Prompt: Did **he** say "lap dance" just now? Target: No, **she** said "lap dance" just now.

22. rap dant
a). Prompt: Did you say "**mob** dante" just now? Target: No, I said "**rap** dante" just now.
b). Prompt: Did you say "rap **kant**" just now? Target: No, I said "rap **dante**" just now.
c). Target: You know what? He said "**rap dante"** just now.
d). Prompt: Did **he** say "rap dante" just now? Target: No, **she** said "rap dante" just now.

23. rib pain
a). Prompt: Did you say "**neck** pain" just now? Target: No, I said "**rib** pain" just now.
b). Prompt: Did you say "rib **roast**" just now? Target: No, I said "rib **pain"** just now.
c). Target: You know what? He said "**rib pain"** just now.
d). Prompt: Did **he** say "rib pain" just now? Target: No, **she** said "rib pain" just now.

24. lib peim
a). Prompt: Did you say "**peep** peim" just now? Target: No, I said "**lib** peim" just now.
b). Prompt: Did you say "lib **bot**" just now? Target: No, I said "lib **peim**" just now.
c). Target: You know what? He said "**lib peim"** just now.
d). Prompt: Did **he** say "lib peim" just now? Target: No, **she** said "lib peim" just now.

25. crab dish
a). Prompt: Did you say "**pork** dish" just now? Target: No, I said "**crab** dish" just now.
b). Prompt: Did you say "crab **pants**" just now? Target: No, I said "crab **dish"** just now.
c). Target: You know what? He said "**crab dish"** just now.
d). Prompt: Did **he** say "crab dish" just now? Target: No, **she** said "crab dish" just now.

26. prab dis
a). Prompt: Did you say "**pop** dis" just now? Target: No, I said "**prab** dis" just now.
b). Prompt: Did you say "prab **tot**" just now? Target: No, I said "prab **dis**" just now.

c). Target: You know what? He said “**prab dis”** just now.
d). Prompt: Did **he** say “prab dis” just now? Target: No, **she** said “prab dis” just now.

27. beat down
a). Prompt: Did you say “**cab** down” just now? Target: No, I said “**beat** down” just now.
b). Prompt: Did you say “beat **him**” just now? Target: No, I said “beat **down”** just now.
c). Target: You know what? He said “**beat down”** just now.
d). Prompt: Did **he** say “beat down” just now? Target: No, **she** said “beat down” just now.

27. deat doun
a). Prompt: Did you say “**cod** doun” just now? Target: No, I said “**deat** doun” just now.
b). Prompt: Did you say “deat **tot**” just now? Target: No, I said “deat **doun**” just now.
c). Target: You know what? He said “**deat doun”** just now.
d). Prompt: Did **he** say “deat doun” just now? Target: No, **she** said “deat doun” just now.

29. bad day
a). Prompt: Did you say “**crab** day” just now? Target: No, I said “**bad** day” just now.
b). Prompt: Did you say “bad **pants**” just now? Target: No, I said “bad **day”** just now.
c). Target: You know what? He said “**bad day”** just now.
d). Prompt: Did **he** say “bad day” just now? Target: No, **she** said “bad day” just now.

30. gad dei
a). Prompt: Did you say “**rot** dei” just now? Target: No, I said “**gad** dei” just now.
b). Prompt: Did you say “gad **tot**” just now? Target: No, I said “gad **dei**” just now.
c). Target: You know what? He said “**gad dei”** just now.
d). Prompt: Did **he** say “gad dei” just now? Target: No, **she** said “gad dei” just now.

31. back door
a). Prompt: Did you say “**front** door” just now? Target: No, I said “**back** door” just now.
b). Prompt: Did you say “back **baby**” just now? Target: No, I said “back **door”** just now.
c). Target: You know what? He said “**back door”** just now.
d). Prompt: Did **he** say “back door” just now? Target: No, **she** said “back door” just now.

32. dak dol
a). Prompt: Did you say “**dog** dol” just now? Target: No, I said “**dak** dol” just now.
b). Prompt: Did you say “dak **tot**” just now? Target: No, I said “dak **dol**” just now.
c). Target: You know what? He said “**dak dol”** just now.
d). Prompt: Did **he** say “dak dol” just now? Target: No, **she** said “dak dol” just now.

33. big deal
a). Prompt: Did you say “**cab** deal” just now? Target: No, I said “**big** deal” just now.
b). Prompt: Did you say “big **pants**” just now? Target: No, I said “big **deal”** just now.
c). Target: You know what? He said “**big deal”** just now.
d). Prompt: Did **he** say “big deal” just now? Target: No, **she** said “big deal” just now.

34. kig deer
a). Prompt: Did you say “**dock** deer” just now? Target: No, I said “**kig** deer” just now.
b). Prompt: Did you say “kig **tot**” just now? Target: No, I said “kig **deer**” just now.
c). Target: You know what? He said “**kig deer”** just now.
d). Prompt: Did **he** say “kig deer” just now? Target: No, **she** said “kig deer” just now

35. keep calm
a). Prompt: Did you say “**stay** clam” just now? Target: No, I said “**keep** calm” just now.
b). Prompt: Did you say “keep **pace**” just now? Target: No, I said “keep **calm”** just now.
c). Target: You know what? He said “**keep calm”** just now.
d). Prompt: Did **he** say “keep calm” just now? Target: No, **she** said “keep calm” just now.

36. teep kalm
a). Prompt: Did you say “**mob** kalm” just now? Target: No, I said “**teep** kalm” just now.
b). Prompt: Did you say “teep **gop**” just now? Target: No, I said “teep **kalm**” just now.
c). Target: You know what? He said “**teep kalm”** just now.
d). Prompt: Did **he** say “teep kalm” just now? Target: No, **she** said “teep kalm” just now.

37. bright pink
a). Prompt: Did you say “**red** pink” just now? Target: No, I said “**bright** pink” just now.
b). Prompt: Did you say “bright **red**” just now? Target: No, I said “bright **pink**” just now.
c). Target: You know what? He said “**bright pink”** just now.
d). Prompt: Did **he** say “bright pink” just now? Target: No, **she** said “bright pink” just now.

38. grait pinp
a). Prompt: Did you say “**mob** pinp” just now? Target: No, I said “**grait** pinp” just now.
b). Prompt: Did you say “grait **pinp**” just now? Target: No, I said “grait **pinp**” just now.
c). Target: You know what? He said “**grait pinp”** just now.
d). Prompt: Did **he** say “grait pinp” just now? Target: No, **she** said “grait pinp” just now.

39. boot camp
a). Prompt: Did you say “**food** camp” just now? Target: No, I said “**boot** camp” just now.
b). Prompt: Did you say “boot **pants**” just now? Target: No, I said “boot **camp”** just now.
c). Target: You know what? He said “**boot camp”** just now.
d). Prompt: Did **he** say “boot camp” just now? Target: No, **she** said “boot camp” just now.

40. doot kamk
a). Prompt: Did you say “**cod** kamk” just now? Target: No, I said “**doot** kamk” just now.
b). Prompt: Did you say “doot **gop**” just now? Target: No, I said “doot **kamk**” just now.
c). Target: You know what? He said “**doot kamk”** just now.
d). Prompt: Did **he** say “doot kamk” just now? Target: No, **she** said “doot kamk” just now.

41. food court
a). Prompt: Did you say “**sports** court” just now? Target: No, I said “**food** court” just now.

b). Prompt: Did you say “food **pants**” just now? Target: No, I said “food **court”** just now.
c). Target: You know what? He said “**food court”** just now.
d). Prompt: Did **he** say “food court” just now? Target: No, **she** said “food court” just now.

42. sood kork
a). Prompt: Did you say “**rot** kork” just now? Target: No, I said “**sood** kork” just now.
b). Prompt: Did you say “sood **gop**” just now? Target: No, I said “sood **kork**” just now.
c). Target: You know what? He said “**sood kork”** just now.
d). Prompt: Did **he** say “sood kork” just now? Target: No, **she** said “sood kork” just now.

43. take care
a). Prompt: Did you say “**day** care” just now? Target: No, I said “**take** care” just now.
b). Prompt: Did you say “take **pants**” just now? Target: No, I said “take **care”** just now.
c). Target: You know what? He said “**take care”** just now.
d). Prompt: Did **he** say “take care” just now? Target: No, **she** said “take care” just now.

44. peik ker
a). Prompt: Did you say “**dog** ker” just now? Target: No, I said “**peik** ker” just now.
b). Prompt: Did you say “peik **gop**” just now? Target: No, I said “peik **ker**” just now.
c). Target: You know what? He said “**peik ker”** just now.
d). Prompt: Did **he** say “peik ker” just now? Target: No, **she** said “peik ker” just now.

45. cheek pain
a). Prompt: Did you say “ri**b** pain” just now? Target: No, I said “**cheek** pain” just now.
b). Prompt: Did you say “cheek **pain**” just now? Target: No, I said “cheek **pain**” just now.
c). Target: You know what? He said “**cheek pain”** just now.
d). Prompt: Did **he** say “cheek pain” just now? Target: No, **she** said “cheek pain” just now.

46. feek paim
a). Prompt: Did you say “ri**b** paim” just now? Target: No, I said “**feek** paim” just now.
b). Prompt: Did you say “feek **pain**” just now? Target: No, I said “feek **paim**” just now.
c). Target: You know what? He said “**feek paim”** just now.
d). Prompt: Did **he** say “feek paim” just now? Target: No, **she** said “feek paim” just now.

47. drag car
a). Prompt: Did you say “**racing** car” just now? Target: No, I said “**drag** car” just now.
b). Prompt: Did you say “drag **flag**” just now? Target: No, I said “drag **car**” just now.
c). Target: You know what? He said “**drag car”** just now.
d). Prompt: Did **he** say “drag car” just now? Target: No, **she** said “drag car” just now.

48. grag kar
a). Prompt: Did you say “**dock** kar” just now? Target: No, I said “**grag** kar” just now.
b). Prompt: Did you say “grag **gop**” just now? Target: No, I said “grag **kar**” just now.
c). Target: You know what? He said “**grag kar”** just now.

d). Prompt: Did **he** say “grag kar” just now? Target: No, **she** said “grag kar” just now.

49. top gear
a). Prompt: Did you say “**good** gear” just now? Target: No, I said “**top** gear” just now.
b). Prompt: Did you say “top **gun**” just now? Target: No, I said “top **gear”** just now.
c). Target: You know what? He said “**top gear”** just now.
d). Prompt: Did **he** say “top gear” just now? Target: No, **she** said “top gear” just now.

50. pop geal
a). Prompt: Did you say “**mob** geal” just now? Target: No, I said “**pop** geal” just now.
b). Prompt: Did you say “pop **kop**” just now? Target: No, I said “pop **geal**” just now.
c). Target: You know what? He said “**pop geal”** just now.
d). Prompt: Did **he** say “pop geal” just now? Target: No, **she** said “pop geal” just now.

51. club girl
a). Prompt: Did you say “**big** girl” just now? Target: No, I said “**club** girl” just now.
b). Prompt: Did you say “club **pants**” just now? Target: No, I said “club **girl”** just now.
c). Target: You know what? He said “**club girl”** just now.
d). Prompt: Did **he** say “club girl” just now? Target: No, **she** said “club girl” just now.

52. plub girm
a). Prompt: Did you say “**pop** girm” just now? Target: No, I said “**plub** girm” just now.
b). Prompt: Did you say “plub **kop**” just now? Target: No, I said “plub **girm**” just now.
c). Target: You know what? He said “**plub girm”** just now.
d). Prompt: Did **he** say “plub girm” just now? Target: No, **she** said “plub girm” just now.

53. great guy
a). Prompt: Did you say “**good** guy” just now? Target: No, I said “**great** guy” just now.
b). Prompt: Did you say “great **pants**” just now? Target: No, I said “great **guy”** just now.
c). Target: You know what? He said “**great guy”** just now.
d). Prompt: Did **he** say “great guy” just now? Target: No, **she** said “great guy” just now.

54. breat gai
a). Prompt: Did you say “**cod** gai” just now? Target: No, I said “**breat** gai” just now.
b). Prompt: Did you say “breat **kop**” just now? Target: No, I said “breat **gai**” just now.
c). Target: You know what? He said “**breat gai”** just now.
d). Prompt: Did **he** say “breat gai” just now? Target: No, **she** said “breat gai” just now.

55. good game
a). Prompt: Did you say “**video** game” just now? Target: No, I said “**good** game” just now.
b). Prompt: Did you say “good **bed**” just now? Target: No, I said “good **game”** just now.
c).Target: You know what? He said “**good game”** just now.
d). Prompt: Did **he** say “good game” just now? Target: No, **she** said “good game” just now.

56. dood gane

a). Prompt: Did you say “**mob** gane” just now? Target: No, I said “**dood** gane” just now.
b). Prompt: Did you say “dood **kop**” just now? Target: No, I said “dood **gane**” just now.
c). Target: You know what? He said “**dood gane”** just now.
d). Prompt: Did **he** say “dood gane” just now? Target: No, **she** said “dood gane” just now.

57. look good
a). Prompt: Did you say “**dog** good” just now? Target: No, I said “**look** good” just now.
b). Prompt: Did you say “look **baby**” just now? Target: No, I said “look **good”** just now.
c). Target: You know what? He said “**look good”** just now.
d). Prompt: Did **he** say “look good” just now? Target: No, **she** said “look good” just now.

58. rook goog
a). Prompt: Did you say “**dog** goog” just now? Target: No, I said “**rook** goog” just now.
b). Prompt: Did you say “rook **kop**” just now? Target: No, I said “rook **goog**” just now.
c). Target: You know what? He said “**rook goog”** just now.
d). Prompt: Did **he** say “rook goog” just now? Target: No, **she** said “rook goog” just now.

59. big game
a). Prompt: Did you say “**good** game” just now? Target: No, I said “**big** game” just now.
b). Prompt: Did you say “big **baby**” just now? Target: No, I said “big **game”** just now.
c). Target: You know what? He said “**big game”** just now.
d). Prompt: Did **he** say “big game” just now? Target: No, **she** said “big game” just now.

60. dig gane
a). Prompt: Did you say “**dock** gane” just now? Target: No, I said “**dig** gane” just now.
b). Prompt: Did you say “dig **kop**” just now? Target: No, I said “dig **gane”** just now.
c). Target: You know what? He said “**dig gane”** just now.
d). Prompt: Did **he** say “dig gane” just now? Target: No, **she** said “dig gane” just now.

61. red pen
a). Prompt: Did you say “**blue** pen” just now? Target: No, I said “**red** pen” just now.
b). Prompt: Did you say “red **bill**” just now? Target: No, I said “red **pen”** just now.
c). Target: You know what? He said “**red pen”** just now.
d). Prompt: Did **he** say “red pen” just now? Target: No, **she** said “red pen” just now.

62. led pem
a). Prompt: Did you say “**rot** pem” just now? Target: No, I said “**led** pem” just now.
b). Prompt: Did you say “led **bot**” just now? Target: No, I said “led **pem**” just now.
c). Target: You know what? He said “**led pem”** just now.
d). Prompt: Did **he** say “led pem” just now? Target: No, **she** said “led pem” just now.

63. back pack
a). Prompt: Did you say “**cab** pack” just now? Target: No, I said “**back** pack” just now.
b). Prompt: Did you say “back **baby**” just now? Target: No, I said “back **pack”** just now.
c). Target: You know what? He said “**back pack”** just now.

d). Prompt: Did **he** say “back pack” just now?   Target: No, **she** said “back pack” just now.

64. dak pak
a). Prompt: Did you say “**dog** pak” just now?   Target: No, I said “**dak** pak” just now.
b). Prompt: Did you say “dock **bot**” just now?   Target: No, I said “dak **pak**” just now.
c). Target: You know what? He said “**dak pak”** just now.
d). Prompt: Did **he** say “dak pak” just now?   Target: No, **she** said “dak pak” just now.

65. blog page
a). Prompt: Did you say “**web** page” just now?   Target: No, I said “**blog** page” just now.
b). Prompt: Did you say “blog **music**” just now?   Target: No, I said “blog **page”** just now.
c). Target: You know what? He said “**blog page”** just now.
d). Prompt: Did **he** say “blog page” just now?   Target: No, **she** said “blog page” just now.

66. glog paze
a). Prompt: Did you say “**dock** paze” just now?   Target: No, I said “**glog** paze” just now.
b). Prompt: Did you say “glog **bot**” just now?   Target: No, I said “glog **paze**” just now.
c). Target: You know what? He said “**glog paze”** just now.
d). Prompt: Did **he** say “glog paze” just now?   Target: No, **she** said “glog paze” just now.

67. lip balm
a). Prompt: Did you say “**cab** balm” just now?   Target: No, I said “**lip** balm” just now.
b). Prompt: Did you say “lip **pants**” just now?   Target: No, I said “lip **balm”** just now.
c). Target: You know what? He said “**lip balm”** just now.
d). Prompt: Did **he** say “lip balm” just now?   Target: No, **she** said “lip balm” just now.

68. mip bong
a). Prompt: Did you say “**mob** bong” just now?   Target: No, I said “**mip** bong” just now.
b). Prompt: Did you say “mip **pate**” just now?   Target: No, I said “mip **bong**” just now.
c). Target: You know what? He said “**mip bong”** just now.
d). Prompt: Did **he** say “mip bong” just now?   Target: No, **she** said “mip bong” just now.

69.job bank
a). Prompt: Did you say “**keep** bank” just now?   Target: No, I said “**job** bank” just now.
b). Prompt: Did you say “job **pants**” just now?   Target: No, I said “job **bank”** just now.
c). Target: You know what? He said “**job bank”** just now.
d). Prompt: Did **he** say “job bank” just now?   Target: No, **she** said “job bank” just now.

70. zob bant
a). Prompt: Did you say “**pop** bant” just now?   Target: No, I said “**zob** bant” just now.
b). Prompt: Did you say “zob **pot**” just now?   Target: No, I said “zob **bant**” just now.
c). Target: You know what? He said “**zob bant”** just now.
d). Prompt: Did **he** say “zob bant” just now?   Target: No, **she** said “zob bant” just now.

71. frat boy

a). Prompt: Did you say “**bad** boy” just now? Target: No, I said “**frat** boy” just now.
b). Prompt: Did you say “frat **pants**” just now? Target: No, I said “frat **boy”** just now.
c). Target: You know what? He said “**frat boy”** just now.
d). Prompt: Did **he** say “frat boy” just now? Target: No, **she** said “frat boy” just now.

72. srat boi
a). Prompt: Did you say “**cod** boi” just now? Target: No, I said “**srat** boi” just now.
b). Prompt: Did you say “srat **pot**” just now? Target: No, I said “srat **boi**” just now.
c). Target: You know what? He said “**srat boi”** just now.
d). Prompt: Did **he** say “srat boi” just now? Target: No, **she** said “srat bot” just now.

73. soup pot
a). Prompt: Did you say “**big** pot” just now? Target: No, I said “**soup** pot” just now.
b). Prompt: Did you say “soup **bowel**” just now? Target: No, I said “soup **pot**” just now.
c). Target: You know what? He said “**soup pot”** just now.
d). Prompt: Did **he** say “soup pot” just now? Target: No, **she** said “soup pot” just now.

74. foop pok
a). Prompt: Did you say “**food** pok” just now? Target: No, I said “**foop** pok” just now.
b). Prompt: Did you say “foop **bowel**” just now? Target: No, I said “foop **pok**” just now.
c). Target: You know what? He said “**foop pok”** just now.
d). Prompt: Did **he** say “foop pok” just now? Target: No, **she** said “foop pok” just now.

75. grab pen
a). Prompt: Did you say “**red** pen” just now? Target: No, I said “**grab** pen” just now.
b). Prompt: Did you say “grab **bowel**” just now? Target: No, I said “grab **pen**” just now.
c). Target: You know what? He said “**grab pen”** just now.
d). Prompt: Did **he** say “grab pen” just now? Target: No, **she** said “grab pen” just now.

76. brab pem
a). Prompt: Did you say “**big** pem” just now? Target: No, I said “**brab** pem” just now.
b). Prompt: Did you say “brab **bowel**” just now? Target: No, I said “brab **pem**” just now.
c). Target: You know what? He said “**brab pem”** just now.
d). Prompt: Did **he** say “brab pem” just now? Target: No, **she** said “brab pem” just now.

77. deep bite
a). Prompt: Did you say “**big** bite” just now? Target: No, I said “**deep** bite” just now.
b). Prompt: Did you say “deep **down**” just now? Target: No, I said “deep **bite**” just now.
c). Target: You know what? He said “**deep bite”** just now.
d). Prompt: Did **he** say “deep bite” just now? Target: No, **she** said “deep bite” just now.

78. beep bipe
a). Prompt: Did you say “**big** bipe” just now? Target: No, I said “**beep** bipe” just now.
b). Prompt: Did you say “beep **bowel**” just now? Target: No, I said “beep **bipe**” just now.
c). Target: You know what? He said “**beep bipe”** just now.

d). Prompt: Did **he** say “beep bipe” just now? Target: No, **she** said “beep bipe” just now.

79. rub back
a). Prompt: Did you say “**kick** back” just now? Target: No, I said “**rub** back” just now.
b). Prompt: Did you say “rub **head**” just now? Target: No, I said “rub **back**” just now.
c). Target: You know what? He said “**rub back”** just now.
d). Prompt: Did **he** say “rub back” just now? Target: No, **she** said “rub back” just now.

80. mub bait
a). Prompt: Did you say “**big** bait” just now? Target: No, I said “**mub** bait” just now.
b). Prompt: Did you say “mub **pot**” just now? Target: No, I said “mub **bait**” just now.
c). Target: You know what? He said “**mub bait”** just now.
d). Prompt: Did **he** say “mub bait” just now? Target: No, **she** said “mub bait” just now.

81. get tip
a). Prompt: Did you say “**big** tip” just now? Target: No, I said “**get** tip” just now.
b). Prompt: Did you say “get **done**” just now? Target: No, I said “get **tip**” just now.
c). Target: You know what? He said “**get tip”** just now.
d). Prompt: Did **he** say “get tip” just now? Target: No, **she** said “get tip” just now.

82. det tik
a). Prompt: Did you say “**big** tik” just now? Target: No, I said “**det** tik” just now.
b). Prompt: Did you say “det **tik**” just now? Target: No, I said “det **tik**” just now.
c). Target: You know what? He said “**det tik”** just now.
d). Prompt: Did **he** say “det tik” just now? Target: No, **she** said “det tik” just now.

83. shade tent
a). Prompt: Did you say “**big** tent” just now? Target: No, I said “**shade** tent” just now.
b). Prompt: Did you say “shade **plants**” just now? Target: No, I said “shade **tent**” just now.
c). Target: You know what? He said “**shade tent”** just now.
d). Prompt: Did **he** say “shade tent” just now? Target: No, **she** said “shade tent” just now.

84. fade tenk
a). Prompt: Did you say “**big** tenk” just now? Target: No, I said “**fade** tenk” just now.
b). Prompt: Did you say “fade **bowel**” just now? Target: No, I said “fade **tenk**” just now.
c). Target: You know what? He said “**fade tenk”** just now.
d). Prompt: Did **he** say “fade tenk” just now? Target: No, **she** said “fade tenk” just now.

85. vet debt
a). Prompt: Did you say “**big** debt” just now? Target: No, I said “**vet** debt” just now.
b). Prompt: Did you say “vet **clinic**” just now? Target: No, I said “vet **debt**” just now.
c). Target: You know what? He said “**vet debt”** just now.
d). Prompt: Did **he** say “vet debt” just now? Target: No, **she** said “vet debt” just now.

86. zet dep

a). Prompt: Did you say “**big** dep” just now? Target: No, I said “**zet** dep” just now.
b). Prompt: Did you say “zet **gala**” just now? Target: No, I said “zet **dep**” just now.
c). Target: You know what? He said “**zet dep”** just now.
d). Prompt: Did **he** say “zet dep” just now? Target: No, **she** said “zet dep” just now.

87. mud day
a). Prompt: Did you say “**big** day” just now? Target: No, I said “**mud** day” just now.
b). Prompt: Did you say “mud **car**” just now? Target: No, I said “mud **day**” just now.
c). Target: You know what? He said “**mud day”** just now.
d). Prompt: Did **he** say “mud day” just now? Target: No, **she** said “mud day” just now.

88. nud dei
a). Prompt: Did you say “**big** dei” just now? Target: No, I said “**nud** dei” just now.
b). Prompt: Did you say “nud **bowel**” just now? Target: No, I said “nud **dei**” just now.
c). Target: You know what? He said “**nud dei”** just now.
d). Prompt: Did **he** say “nud dei” just now? Target: No, **she** said “nud dei” just now.

89. back kick
a). Prompt: Did you say “**big** kick” just now? Target: No, I said “**back** kick” just now.
b). Prompt: Did you say “back **pack**” just now? Target: No, I said “back **kick**” just now.
c). Target: You know what? He said “**back kick”** just now.
d). Prompt: Did **he** say “back kick” just now? Target: No, **she** said “back kick” just now.

90. dak kit
a). Prompt: Did you say “**big** kit” just now? Target: No, I said “**dak** kit” just now.
b). Prompt: Did you say “dak **bowel**” just now? Target: No, I said “dak **kit**” just now.
c). Target: You know what? He said “**dak kit”** just now.
d). Prompt: Did **he** say “dak kit” just now? Target: No, **she** said “dak kit” just now.

91. smug king
a). Prompt: Did you say “**big** king” just now? Target: No, I said “**smug** king” just now.
b). Prompt: Did you say “smug **girl**” just now? Target: No, I said “smug **king**” just now.
c). Target: You know what? He said “**smug king”** just now.
d). Prompt: Did **he** say “smug king” just now? Target: No, **she** said “smug king” just now.

92. swag kim
a). Prompt: Did you say “**big** kim” just now? Target: No, I said “**swag** kim” just now.
b). Prompt: Did you say “swag **bowel**” just now? Target: No, I said “swag **kim**” just now.
c). Target: You know what? He said “**swag kim”** just now.
d). Prompt: Did **he** say “swag kim” just now? Target: No, **she** said “swag kim” just now.

93. deck gate
a). Prompt: Did you say “**big** gate” just now? Target: No, I said “**deck** gate” just now.
b). Prompt: Did you say “deck **plans**” just now? Target: No, I said “deck **gate**” just now.
c). Target: You know what? He said “**deck gate”** just now.

d). Prompt: Did **he** say “deck gate” just now? Target: No, **she** said “deck gate” just now.

94. bek geit
a). Prompt: Did you say “**big** geit” just now? Target: No, I said “**bek** geit” just now.
b). Prompt: Did you say “bek **bowel**” just now? Target: No, I said “bek **geit**” just now.
c). Target: You know what? He said “**bek geit”** just now.
d). Prompt: Did **he** say “bek geit” just now? Target: No, **she** said “bek geit” just now.

95. drug gang
a). Prompt: Did you say “**mob** gang” just now? Target: No, I said “**drug** gang” just now.
b). Prompt: Did you say “drug **head**” just now? Target: No, I said “drug **gang**” just now.
c). Target: You know what? He said “**drug gang”** just now.
d). Prompt: Did **he** say “drug gang” just now? Target: No, **she** said “drug gang” just now.

96. brug gam
a). Prompt: Did you say “**big** gam” just now? Target: No, I said “**brug** gam” just now.
b). Prompt: Did you say “brug **bowel**” just now? Target: No, I said “brug **gam**” just now.
c). Target: You know what? He said “**brug gam”** just now.
d). Prompt: Did **he** say “brug gam” just now? Target: No, **she** said “brug gam” just now.

97. skip test
a). Prompt: Did you say “**good** test” just now? Target: No, I said “**skip** test” just now.
b). Prompt: Did you say “skip **clothes**” just now? Target: No, I said “skip **test**” just now.
c). Target: You know what? He said “**skip test”** just now.
d). Prompt: Did **he** say “skip test” just now? Target: No, **she** said “skip test” just now.

98. stip tesp
a). Prompt: Did you say “**big** tesp” just now? Target: No, I said “**stip** tesp” just now.
b). Prompt: Did you say “stip **bowel**” just now? Target: No, I said “stip **tesp**” just now.
c). Target: You know what? He said “**stip tesp”** just now.
d). Prompt: Did **he** say “stip tesp” just now? Target: No, **she** said “stip tesp” just now.

99. lap dance
a). Prompt: Did you say “**good** dance” just now? Target: No, I said “**lap** dance” just now.
b). Prompt: Did you say “lap **desk**” just now? Target: No, I said “lap **dance**” just now.
c). Target: You know what? He said “**lap dance”** just now.
d). Prompt: Did **he** say “lap dance” just now? Target: No, **she** said “lap dance” just now.

100. rap dant
a). Prompt: Did you say “**big** dant” just now? Target: No, I said “**rap** dant” just now.
b). Prompt: Did you say “rap **dant**” just now? Target: No, I said “rap **dant**” just now.
c). Target: You know what? He said “**rap dant”** just now.
d). Prompt: Did **he** say “rap dant” just now? Target: No, **she** said “rap dant” just now.

101.glib talk

a). Prompt: Did you say “**big** talk” just now? Target: No, I said “**glib** talk” just now.
b). Prompt: Did you say “glib **person**” just now? Target: No, I said “glib **talk**” just now.
c). Target: You know what? He said “**glib talk”** just now.
d). Prompt: Did **he** say “glib talk” just now? Target: No, **she** said “glib talk” just now.

102. blib tot
a). Prompt: Did you say “**big** tot” just now? Target: No, I said “**blib** tot” just now.
b). Prompt: Did you say “blib **bowel**” just now? Target: No, I said “blib **tot**” just now.
c). Target: You know what? He said “**blib tot”** just now.
d). Prompt: Did **he** say “blib tot” just now? Target: No, **she** said “blib tot” just now.

103. rub down
a). Prompt: Did you say “**get** down” just now? Target: No, I said “**rub** down” just now.
b). Prompt: Did you say “rub **back**” just now? Target: No, I said “rub **down**” just now.
c). Target: You know what? He said “**rub down”** just now.
d). Prompt: Did **he** say “rub down” just now? Target: No, **she** said “rub down” just now.

104. lub doun
a). Prompt: Did you say “**get** doun” just now? Target: No, I said “**lub** doun” just now.
b). Prompt: Did you say “lub **bowel**” just now? Target: No, I said “lub **doun**” just now.
c). Target: You know what? He said “**lub doun”** just now.
d). Prompt: Did **he** say “lub doun” just now? Target: No, **she** said “lub doun” just now.

105. diet guide
a). Prompt: Did you say “**tour** guide” just now? Target: No, I said “**diet** guide” just now.
b). Prompt: Did you say “diet **plan**” just now? Target: No, I said “diet **guide**” just now.
c). Target: You know what? He said “**diet guide”** just now.
d). Prompt: Did **he** say “diet guide” just now? Target: No, **she** said “diet guide” just now.

106. biet gaib
a). Prompt: Did you say “**big** gaib” just now? Target: No, I said “**biet** gaib” just now.
b). Prompt: Did you say “biet **bowel**” just now? Target: No, I said “biet **gaib**” just now.
c). Target: You know what? He said “**biet gaib”** just now.
d). Prompt: Did **he** say “biet gaib” just now? Target: No, **she** said “biet gaib” just now.

107. fat cat
a). Prompt: Did you say “**small** cat” just now? Target: No, I said “**fat** cat” just now.
b). Prompt: Did you say “fat **girl**” just now? Target: No, I said “fat **cat**” just now.
c). Target: You know what? He said “**fat cat”** just now.
d). Prompt: Did **he** say “fat cat” just now? Target: No, **she** said “fat cat” just now.

108. shat kap
a). Prompt: Did you say “**big** kap” just now? Target: No, I said “**shat** kap” just now.
b). Prompt: Did you say “shat **kap**” just now? Target: No, I said “shat **kap**” just now.
c). Target: You know what? He said “**shat kap”** just now.

d). Prompt: Did **he** say “shat kap” just now? Target: No, **she** said “shat kap” just now.

109. head game
a). Prompt: Did you say “**big** game” just now? Target: No, I said “**head** game” just now.
b). Prompt: Did you say “head **pain**” just now? Target: No, I said “head **game**” just now.
c). Target: You know what? He said “**head game”** just now.
d). Prompt: Did **he** say “head game” just now? Target: No, **she** said “head game” just now.

110. fead gane
a). Prompt: Did you say “**big** gane” just now? Target: No, I said “**fead** gane” just now.
b). Prompt: Did you say “fead **bowel**” just now? Target: No, I said “fead **gane**” just now.
c). Target: You know what? He said “**fead gane”** just now.
d). Prompt: Did **he** say “fead gane” just now? Target: No, **she** said “fead gane” just now.

111. bad kid
a). Prompt: Did you say “**big** kid” just now? Target: No, I said “**bad** kid” just now.
b). Prompt: Did you say “bad **girl**” just now? Target: No, I said “bad **kid**” just now.
c). Target: You know what? He said “**bad kid”** just now.
d). Prompt: Did **he** say “bad kid” just now? Target: No, **she** said “bad kid” just now.

112. gad kig
a). Prompt: Did you say “**big** kig” just now? Target: No, I said “**gad** kig” just now.
b). Prompt: Did you say “gad **kig**” just now? Target: No, I said “gad **kig**” just now.
c). Target: You know what? He said “**gad kig”** just now.
d). Prompt: Did **he** say “gad kig” just now? Target: No, **she** said “gad kig” just now.

113. lip care
a). Prompt: Did you say “**skin** care” just now? Target: No, I said “**lip** care” just now.
b). Prompt: Did you say “lip **kiss**” just now? Target: No, I said “lip **care**” just now.
c). Target: You know what? He said “**lip care”** just now.
d). Prompt: Did **he** say “lip care” just now? Target: No, **she** said “lip care” just now.

114. mip kare
a). Prompt: Did you say “**big** kare” just now? Target: No, I said “**mip** kare” just now.
b). Prompt: Did you say “mip **bowel**” just now? Target: No, I said “mip **kare**” just now.
c). Target: You know what? He said “**mip kare”** just now.
d). Prompt: Did **he** say “mip kare” just now? Target: No, **she** said “mip kare” just now.

115. tape gun
a). Prompt: Did you say “**big** gun” just now? Target: No, I said “**tape** gun” just now.
b). Prompt: Did you say “tape **player**” just now? Target: No, I said “tape **gun**” just now.
c). Target: You know what? He said “**tape gun”** just now.
d). Prompt: Did **he** say “tape gun” just now? Target: No, **she** said “tape gun” just now.

116. pape gum

a). Prompt: Did you say “**big** gum” just now? Target: No, I said “**pape** gum” just now.
b). Prompt: Did you say “pape **bowel**” just now? Target: No, I said “pape **gum**” just now.
c). Target: You know what? He said “**pape gum”** just now.
d). Prompt: Did **he** say “pape gum” just now? Target: No, **she** said “pape gum” just now.

117. grab cash
a). Prompt: Did you say “**big** cash” just now? Target: No, I said “**grab** cash” just now.
b). Prompt: Did you say “grab **bowel**” just now? Target: No, I said “grab **cash**” just now.
c). Target: You know what? He said “**grab cash”** just now.
d). Prompt: Did **he** say “grab cash” just now? Target: No, **she** said “grab cash” just now.

118. brab kaf
a). Prompt: Did you say “**big** kaf” just now? Target: No, I said “**brab** kaf” just now.
b). Prompt: Did you say “brab **bowel**” just now? Target: No, I said “brab **kaf**” just now.
c). Target: You know what? He said “**brab kaf”** just now.
d). Prompt: Did **he** say “brab kaf” just now? Target: No, **she** said “brab kaf” just now.

119. fab gift
a). Prompt: Did you say “**big** gift” just now? Target: No, I said “**fab** gift” just now.
b). Prompt: Did you say “fab **card**” just now? Target: No, I said “fab **gift**” just now.
c). Target: You know what? He said “**fab gift”** just now.
d). Prompt: Did **he** say “fab gift” just now? Target: No, **she** said “fab gift” just now.

120. suit pants
a). Prompt: Did you say “**big** pants” just now? Target: No, I said “**suit** pants” just now.
b). Prompt: Did you say “suit **tie**” just now? Target: No, I said “suit **pants**” just now.
c). Target: You know what? He said “**suit pants”** just now.
d). Prompt: Did **he** say “suit pants” just now? Target: No, **she** said “suit pants” just now.

121.fuit panf
a). Prompt: Did you say “**big** panf” just now? Target: No, I said “**fuit** panf” just now.
b). Prompt: Did you say “fuit **bowel**” just now? Target: No, I said “fuit **panf**” just now.
c). Target: You know what? He said “**fuit panf”** just now.
d). Prompt: Did **he** say “fuit panf” just now? Target: No, **she** said “fuit panf” just now.

122. head pain
a). Prompt: Did you say “**back** pain” just now? Target: No, I said “**head** pain” just now.
b). Prompt: Did you say “head **band**” just now? Target: No, I said “head **pain**” just now.
c). Target: You know what? He said “**head pain”** just now.
d). Prompt: Did **he** say “head pain” just now? Target: No, **she** said “head pain” just now.

123. fead paim
a). Prompt: Did you say “**big** paim” just now? Target: No, I said “**fead** paim” just now.
b). Prompt: Did you say “fead **bowel**” just now? Target: No, I said “fead **paim**” just now.
c). Target: You know what? He said “**fead paim”** just now.

d). Prompt: Did **he** say “fead paim” just now? Target: No, **she** said “fead paim” just now.

124. get back
a). Prompt: Did you say “**rub** back” just now? Target: No, I said “**get** back” just now.
b). Prompt: Did you say “get **done**” just now? Target: No, I said “get **back”** just now.
c). Target: You know what? He said “**get back”** just now.
d). Prompt: Did **he** say “get back” just now? Target: No, **she** said “get back” just now.

125. bet bait
a). Prompt: Did you say “**big** bait” just now? Target: No, I said “**bet** bait” just now.
b). Prompt: Did you say “bet **bowel**” just now? Target: No, I said “bet **bait**” just now.
c). Target: You know what? He said “**bet bait”** just now.
d). Prompt: Did **he** say “bet bait” just now? Target: No, **she** said “bet bait” just now.

126. need beer
a). Prompt: Did you say “**big** beer” just now? Target: No, I said “**need** beer” just now.
b). Prompt: Did you say “need **soup**” just now? Target: No, I said “need **beer**” just now.
c). Target: You know what? He said “**need beer”** just now.
d). Prompt: Did **he** say “need beer” just now? Target: No, **she** said “need beer” just now.

127. meed beel
a). Prompt: Did you say “**big** beel” just now? Target: No, I said “**meed** beel” just now.
b). Prompt: Did you say “meed **bowel**” just now? Target: No, I said “meed **beel**” just now.
c). Target: You know what? He said “**meed beel”** just now.
d). Prompt: Did **he** say “meed beel” just now? Target: No, **she** said “meed beel” just now.

128.dog tail
a). Prompt: Did you say “**big** tail” just now? Target: No, I said “**dog** tail” just now.
b). Prompt: Did you say “dog **bark**” just now? Target: No, I said “dog **tail**” just now.
c). Target: You know what? He said “**dog tail”** just now.
d). Prompt: Did **he** say “dog tail” just now? Target: No, **she** said “dog tail” just now.

129. gog teir
a). Prompt: Did you say “**big** teir” just now? Target: No, I said “**gog** teir” just now.
b). Prompt: Did you say “gog **bowel**” just now? Target: No, I said “gog **teir**” just now.
c). Target: You know what? He said “**gog teir”** just now.
d). Prompt: Did **he** say “gog teir” just now? Target: No, **she** said “gog teir” just now.

130. look pale
a). Prompt: Did you say “**big** pale” just now? Target: No, I said “**look** pale” just now.
b). Prompt: Did you say “look **red**” just now? Target: No, I said “look **pale**” just now.
c). Target: You know what? He said “**look pale”** just now.
d). Prompt: Did **he** say “look pale” just now? Target: No, **she** said “look pale” just now.

131. rook peil

a). Prompt: Did you say “**big** peil” just now? Target: No, I said “**rook** peil” just now.
b). Prompt: Did you say “rook **bowel**” just now? Target: No, I said “rook **peil**” just now.
c). Target: You know what? He said “**rook peil”** just now.
d). Prompt: Did **he** say “rook peil” just now? Target: No, **she** said “rook peil” just now.

132. black tea
a). Prompt: Did you say “**green** tea” just now? Target: No, I said “**black** tea” just now.
b). Prompt: Did you say “black **bowel**” just now? Target: No, I said “black **tea**” just now.
c). Target: You know what? He said “**black tea”** just now.
d). Prompt: Did **he** say “black tea” just now? Target: No, **she** said “black tea” just now.

133. glaik tee
a). Prompt: Did you say “**big** tee” just now? Target: No, I said “**glaik** tee” just now.
b). Prompt: Did you say “glaik **bowel**” just now? Target: No, I said “glaik **tee**” just now.
c). Target: You know what? He said “**glaik tee”** just now.
d). Prompt: Did **he** say “glaik tee” just now? Target: No, **she** said “glaik tee” just now.

134. bag deal
a). Prompt: Did you say “**good** deal” just now? Target: No, I said “**bag** deal” just now.
b). Prompt: Did you say “bag **price**” just now? Target: No, I said “bag **deal**” just now.
c). Target: You know what? He said “**bag deal”** just now.
d). Prompt: Did **he** say “bag deal” just now? Target: No, **she** said “bag deal” just now.

135. dag dear
a). Prompt: Did you say “**hi** dear” just now? Target: No, I said “**dag** dear” just now.
b). Prompt: Did you say “dag **bowel**” just now? Target: No, I said “dag **dear**” just now.
c). Target: You know what? He said “**dag dear”** just now.
d). Prompt: Did **he** say “dag dear” just now? Target: No, **she** said “dag dear” just now.

136. dock date
a). Prompt: Did you say “**big** date” just now? Target: No, I said “**dock** date” just now.
b). Prompt: Did you say “dock **plan**” just now? Target: No, I said “dock **date**” just now.
c). Target: You know what? He said “**dock date”** just now.
d). Prompt: Did **he** say “dock date” just now? Target: No, **she** said “dock date” just now.

137. bok deit
a). Prompt: Did you say “**big** deit” just now? Target: No, I said “**bok** deit” just now.
b). Prompt: Did you say “bok **bowel**” just now? Target: No, I said “bok **deit**” just now.
c). Target: You know what? He said “**bok deit”** just now.
d). Prompt: Did **he** say “bok deit” just now? Target: No, **she** said “bok deit” just now.

138. hug pug
a). Prompt: Did you say “**big** pug” just now? Target: No, I said “**hug** pug” just now.
b). Prompt: Did you say “hug **people**” just now? Target: No, I said “hug **pug**” just now.
c). Target: You know what? He said “**hug pug”** just now.

d). Prompt: Did **he** say “hug pug” just now? Target: No, **she** said “hug pug” just now.

139. sug puz
a). Prompt: Did you say “**big** puz” just now? Target: No, I said “**sug** puz” just now.
b). Prompt: Did you say “sug **bowel**” just now? Target: No, I said “sug **puz**” just now.
c). Target: You know what? He said “**sug puz”** just now.
d). Prompt: Did **he** say “sug puz” just now? Target: No, **she** said “sug puz” just now.

140. quick boost
a). Prompt: Did you say “**big** boost” just now? Target: No, I said “**quick** boost” just now.
b). Prompt: Did you say “quick **meal**” just now? Target: No, I said “quick **boost**” just now.
c). Target: You know what? He said “**quick boost”** just now.
d). Prompt: Did **he** say “quick boost” just now? Target: No, **she** said “quick boost” just now.

141. kwik booft
a). Prompt: Did you say “**big** booft” just now? Target: No, I said “**kwik** booft” just now.
b). Prompt: Did you say “kwik **bowel**” just now? Target: No, I said “kwik **booft**” just now.
c). Target: You know what? He said “**kwik booft”** just now.
d). Prompt: Did **he** say “kwik booft” just now? Target: No, **she** said “kwik booft” just now.

142. big boy
a). Prompt: Did you say “**bad** boy” just now? Target: No, I said “**big** boy” just now.
b). Prompt: Did you say “big **bowel**” just now? Target: No, I said “big **boy**” just now.
c). Target: You know what? He said “**big boy”** just now.
d). Prompt: Did **he** say “big boy” just now? Target: No, **she** said “big boy” just now.

143. dig boi
a). Prompt: Did you say “**big** boi” just now? Target: No, I said “**dig** boi” just now.
b). Prompt: Did you say “dig **bowel**” just now? Target: No, I said “dig **boi**” just now.
c). Target: You know what? He said “**dig boi”** just now.
d). Prompt: Did **he** say “dig boi” just now? Target: No, **she** said “dig boi” just now.

144. sab gisp
a). Prompt: Did you say “**big** gisp” just now? Target: No, I said “**sab** gisp” just now.
b). Prompt: Did you say “sab **card**” just now? Target: No, I said “sab **gisp**” just now.
c). Target: You know what? He said “**sab gisp”** just now.
d). Prompt: Did **he** say “sab gisp” just now? Target: No, **she** said “sab gisp” just now.

Appendix B: Questionnaire for the NM group

1. a) Gender: M ___ F_____ b) Age ______
2. Your Program (e.g., Engineering, PhD) ___________
3. How long have you been studying English?

   ____

4. a) How long have you lived in Canada?

   ____

   b) Have you ever been to other English-speaking countries? If yes, how long did you stay there each time?

   ____

5. Where are you originally from, i.e., Beijing, Shanghai, and Chengdu? What is your dialect?

   ____

6. Do you know a language other than Mandarin and English? Please specify all other languages and the level of proficiency in each.

   ____

7. On average how much time do you spend speaking English on a daily basis?

   ____

8. Which language/dialect is more frequently used on a daily basis, if not English?

   ____

9. What is your TOEFL/IELTS test score? When did you take the test?

   TOEFL____

   IELTS____

10. Have you received any musical training? If yes, when and for how long?

    ____

Appendix C: Questionnaire for the NE group

1. a) Gender : M ___ F_____ b) Age _______
2. Your Program (i.e., LING, Undergraduate) _________
3. Where are you originally from, i.e., Victoria, Edmonton, and Waterloo?

   ____

4. Do you know a language other than English? Please specify all other languages and your level of proficiency in each.

   ____

5. Which language/dialect is more frequently used on a daily basis, if not English?

   ____

6. Have you received any musical training? If yes, when and for how long?

   ____

Appendix D: Background of the NM group

Participants	Gender	Age	TOEFL/IELTS (Speaking)	Hometown	Dialect[15]	YOA (year of arrival)	Daily English Use
P1 JL	M	26	110 (23)	Jiangsu	N/A	5 Y	1 Hour
P2 ZMY	M	23	108 (23)	Sichuan	Northern	1 Y	5 Hours
P3 CYD	M	26	98 (20)	JiLin	Northern	4 Y	1 -2 Hours
P4 DJ	F	28	101 (25)	Shanghai	Wu	5 Y	4 Hours
P5 YWY	M	30	100 (19)	ZheJiang	Wu	4 Y	3 -8 Hours
P6 DDX	F	25	8 (8.5)	Shanghai	Wu	2.5 Y	6 Hours
P7 XBY	M	30	113 (23)	Shanghai	Wu	3.5 Y	8 Hours
P8 LY	M	28	580	Shandong	N/A	7 Y	1 -2 Hours
P9 YW	F	25	7	Shanghai	N/A	1.5 Y	5 Hours
P10 MBX	M	29	91 (23)	Taiyuan	Northern	3 Y	3 mins
P11 JKY	F	29	7.5 (8)	Jinan	Northern	1 week	2 Hours
P12 Kl	F	27	7 (7)	Nanjing	Wu	3 Y	1 Hour
P13 LS	F	29	6.5 (6.5)	Wuhan	Northern	3 Y	20 mins
P14 ZF	F	23	6.5 (6)	Anhui	N/A	1 Y	10 mins
P15 TK	F	33	7.5 (7)	Hunan	Xiang	1 month	Only in class
P16 CR	F	23	6.5 (6.5)	Zhejiang	Wu	3 months	2 -3 Hours
P17 WXW	F	25	7 (7.5)	Beijing	Northern	2 Y	12 Hours
P18 SW	M	24	6.5 (6.5)	Shanghai	Wu	2 Y	2 Hours
P19 MX	M	25	103 (22)	Gansu	Northern	3 Y	2 Hours
P20 QXJ	F	35	7.5 (8)	Xi’an	Northern	5 Y	10 mins
P21 ZYM	M	27	7.5 (6.5)	Ha’erbin	N/A	2 Y	N/A
P22 PCX	F	25	102 (19)	Sichuan	Northern	2 weeks	2 Hours
P23 MY	F	40	7 (7)	Shanxi	Northern	10 Y	5 Hours
P24 YZJ	M	25	96	Xi’an	Northern	2 Y	5 Hours
P25 YSJ	F	24	6.5 (7)	Qingdao	Northern	1.5 Y	2-3 Hours

---

[15] N/A means no dialect used in growing up.

Appendix E: Background of the NE group

Participants	Gender	Age	Occupation	Hometown	Another Language (Level)
P1 Al	F	23	Undergraduate	Victoria, BC	French (advanced), Italian (intermediate)
P2 As	M	26	Worker	Victoria, BC	N/A
P3 Bo	M	29	Worker	Salmon Arm, BC	French (intermediate)
P4 Ca	F	19	Undergraduate	Maple Ridge, BC	French (fluent), Japanese (intermediate)
P5 Co	F	28	Undergraduate	Victoria, BC	Spanish (intermediate), Arabic (beginner)
P6 Ev	M	39	Worker	Victoria, BC	French (intermediate), Russian (beginner)
P7 Ha	F	23	Undergraduate	Vancouver, BC	French (beginner)
P8 He	F	26	Graduate	Kaslo, BC	French, German (intermediate)
P9 Jo	F	26	Worker	Victoria, BC	N/A
P10 Man	F	20	Undergraduate	Victoria, BC	Spanish (beginner)
P11 Ma	M	28	Self-employed	Langley, BC	N/A
P12 Ro	M	23	Undergraduate	Courtney, BC	Spanish (beginner)
P13 St	F	22	Undergraduate	Maple Ridge, BC	Spanish, French (intermediate)
P14 Te	F	23	Worker	Pemberton, BC	Spanish (intermediate)
P15 To	M	40	Graduate	Squamish, BC	Japanese (advanced), French (beginner)