THE INFLUENCE OF SPEECH PRODUCTION EXPERIENCE ON THE SIZE AND THE STRUCTURE OF THE SPEECH MOTOR PROGRAM

by

Hyun Seung Kim

B. A., Sogang University, 2004

M. S., Yonsei University, 2006

Submitted to the Graduate Faculty of

School of Health and Rehabilitation Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This dissertation was presented

by

Hyun Seung Kim

It was defended on

July 12, 2018

and approved by

Sheila Pratt, PhD, Professor, Department of Communication Science and Disorders

Susan L. Whitney, DPT, PhD, NCS, ATC, FAPTA, Professor, Department of Physical Therapy

Dissertation Advisors: Malcolm R. McNeil, PhD, Distinguished Service Professor, Emeritus, Department of Communication Science and Disorders (Committee Co-Chair)

Susan Shaiman, PhD, Associate Professor, Department of Communication Science and Disorders (Committee Co-Chair)

THE INFLUENCE OF SPEECH PRODUCTION EXPERIENCE ON THE SIZE AND THE STRUCTURE OF THE SPEECH MOTOR PROGRAM

Hyun Seung Kim, Ph. D.

University of Pittsburgh, 2018

Schema theory (1975) proposed that information about relative timing and force of movements and the order of motor events is stored in Generalized Motor Programs (GMPs). Because some researchers (e. g., Löfqvist, 1991; Max & Caruso, 1997) observed consistent relative timing information in some, but not all, speech rate contexts, this study attempted to provide an alternative explanation for these inconsistent findings of proportional relationships in the trajectories of speech movements. Motivated by Verwey and colleagues (1995, 1996; 1996), who observed changes in production modes from preparing a key press motor response in advance to preparing it in a concurrent manner as the sequence length increased, this study proposes two possible reasons for increased variability in movement trajectories: various motor program sizes and changes in the production mode between advance programming and concurrent programming. The current study hypothesizes that more experienced speakers preserve more proportional relationship information, utilize larger size stored motor programs, and make more flexible switches in their production modes.

Twenty-four native Mandarin and twenty-four non-Mandarin male speakers (19-30 years of age) with normal speech and language functions were recruited. They produced three-syllable, six-syllable, and nine-syllable length Mandarin tone sequences. Interactions between Group and Sequence Length Conditions were investigated in the hierarchical generalized linear model. Several timing, GMP error, and parameter error measurements were examined.

Significant interactions were observed between Group and Sequence Length Condition on the GMP errors per syllable, Hamming distance difference per syllable between slope and parsons' code measurements, and Hamming distance per syllable for parsons' code measurement. In addition, many other significant Group and Sequence Length Condition main or simple main effects were observed.

Results revealed that once motor programs are retrieved, they are executed without being reparameterized. The existence of GMP for lexical tones was supported. Also, it appeared that both native Mandarin and non-Mandarin speakers could switch between advance programming and concurrent programming as the sequence length increased. The timing of this switch occurred later in more-experienced speakers. Furthermore, the attempt to concatenate motor programs appeared to increase variability in movement outcome trajectories, supporting the hypotheses of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I received a lot of love and support while I pursued my Ph. D. program. First, I want to express my special appreciation to my co-advisor, Dr. Malcolm R. McNeil, for his unceasing patience, insight, wisdom, and brilliant guidance throughout my Ph. D. study. I met him in South Korea when he visited my former advisor, Dr. Hyang Hee Kim, and I have a clear memory that I laughed at his jokes and I liked him from that moment. I thank Mick that he accepted me as his advisee and I will not forget all of his help, support, and encouragement. I will miss all the moments of in-depth discussions we had together. You helped me develop the ability to think critically. I am very honored to be your last Ph. D. student who will succeed your academic perspectives and develop from there.

Secondly, I want to thank my co-advisor, Dr. Susan Shaiman, for her generous acceptance of me as her Ph. D. student in 2008 when we first corresponded through emails. Although sometimes you gave me harsh advice, I knew it was like a mother's scolding her child and I knew your warm-hearted intentions, so I appreciated all of your advice, care, and concerns. I will miss all moments I had with you sitting down to discuss research ideas and many aspects of Ph. D study. Because you and Mick had different perspectives and emphasis, the combination of both of you was exactly the kind of help I needed. Without both of your help, I would not have gone this far, so thank you deeply.

I want to thank Dr. Sheila Pratt for her careful thoughts and advice on the aspects I could not see in my dissertation. Sheila helped to complement this dissertation study in many ways. Not only by her thoroughness as a researcher, Sheila inspired me by her charming looks and gentle personality, so you are my role model. Thank you, Sheila.

Fourthly, I want to thank my committee, Dr. Susan Whitney, for her keen and thorough advice. You helped me complete the last puzzle of my dissertation while I developed, wrote, and presented it. So I will say, without your advice, my dissertation would have remained incomplete. I also appreciate your humor and the comfortable atmosphere you created during meetings. Thank you very much.

Not only these four committee members, whenever I met faculty and friends in our school, I was always impressed by their hard work, in-depth conversations which I could have with them, and their openness and good personalities. I regret that I had an introvert personality which sometimes prevented me from developing initimate relationships with them. However, I want to thank all of them because they were the support I had during my Ph. D study.

I especially appreciate all the help and support from Dr. Fassbinder, Dr. Kyoungyuel Lim, Dr. Min Zhang, Dr. Hyunsoo Yoo, Dr. Sang Eun Shin, Dr. Jimin Lee, Dr. Thomas Kovacs, Dr. Kimberly Meigh, Dr. Ruba Almelaifi, Dr. Abel Lei, Dr. Ye Yi (Abby), Dr. Ying Yang, and many other friends, who were once Ph. D students like me, and who proudly received their doctorates over the years, and soon to be Drs. Atsuko Kurosu, Christina Dastolfo-Hromack, Pitchulee Uaypon, Reem Mulla, and Linmin Kang (Carolyn). I appreciate your thoughtful and friendly interactions with me the entire time.

I want to thank all of my research participants and all Chinese friends who helped many apsects of my dissertation. l want to thank faculty members from whom I gained advice and help regarding statistical concerns: Dr. Qi Mi, Dr. Leming Zhou, Dr. Lauren Terhorst, Dr. Jong-Hyeon Jeong, Dr. Satish Iyengar, Dr. Kwonho Jeong, many experts at sas.com, and other stat majoring students who I met through stat-consulting programs. I also want to send my special thanks to Mr. Szuminsky, my ex-roommate Ji Hye Choi, and many other friends, who helped develope computer

programs for this study. I also want to thank Ms. Susan Demo, Ms. Kuhu Tanvir, Ms. Sandy Foster and many other people in the Pitt Writing Center for their help with my writing. With their help, my dissertation could wear a better garment.

On the day I defended, many of my friends and neighbors including three priests prayed for me. I met them at the Korean Catholic Church, Newman Center, St. Paul Church, VA Chapel, and UPMC Chapel. One of the good things about being a catholic is that I can join any church mass whenever I wish to. Through these masses that I tried to attend daily, I met various friends of a wide age range. To me, they were family-like friends who cared about me and prayed for me. They enjoyed my success as if it is theirs. I truly appreciate all of your encouragement and support.

Last, I appreciate my father, Jai Young Kim, my mother, Sung La Oh, my sister, Hee Jung Kim, my brother-in-law, Woo Pan Lee, my nephew, Tae Hwan Lee, my niece, Seo Yeon Lee, and my brother, Kuh San Kim, and all other relatives, for their patience and mental and financial support. They always inspired me as good role-models, and they were my true friends. I love you and I want to be with you forever.

The gospel reading for the day I defended covers Mattheu, 10:7-15, which is the message that Jesus gave to his disciples as he dispatched them. Any honor I experience today has to be raised to heaven because it is God who raised me up and He deserves all glory and praise, not me. In particular, because the gospel said, “Without cost you have received, wihout cost you are to give,” The patients I want to help were my motivation while I pursued this degree. I hope I can continue to aim to help them without cost because that is the message I received on the day I defended. I received so many opportunities and love without even asking. I trust God’s love and plan for me and, in return, I love Jesus and people around me. Thank you all! ♥

1.0 INTRODUCTION

Speech is a coordinated sequence of movements that involves respiratory, phonatory, resonatory and articulatory systems. The act of speaking generates acoustic signals that are eventually perceived by the listeners. Listeners utilize certain aspects of the acoustic signals to recognize speech that have shared meaning among language users. This all occurs while speech is produced by different speakers with different speaking styles (e.g., dialects, different oral structure, etc.) and in different speaking contexts (e.g., various speech rate, pitch[1], loudness, degree of clarity, etc.). Blumstein and Stevens (1979) explained that, in speech, invariance appears in the acoustic signal which corresponds to higher-level representations. They also proposed that this invariance is equivalent to the relative pattern appearing in the acoustic frequency and amplitude spectra. Schmidt (1975) proposed in Schema theory that relative timing or force relationships are the information stored in the Generalized Motor Program (GMP) and this idea also was explored in speech movements by many researchers (e.g., De Jong, Beckman, & Edwards, 1993; Gracco & Abbs, 1986, 1988; Smith et al., 1995; Maas et al., 2008). The invariant information generated by speech movements, or GMP, has been investigated at the muscle activation level using electro-

[1] Pitch and loudness are perceptual correlates of fundamental frequency and intensity respectively. This introduction section will continue to use the terms, pitch and loudness. However, acoustic measures of fundamental frequency and intensity will be made to capture those properties.

myography (EMG), in kinematic trajectories, and in acoustic measures. This study examined the issue of invariance in the frequency and time frame of the acoustic signals during speech production.

The first section of this manuscript begins by introducing Schema theory and the concept of a Generalized Motor Program (GMP) proposed by Schmidt (1975). Information about the order of events, relative timing, and relative force of muscle activation that are consistent across different contexts, has been presumed to be stored in a GMP. In theory, the information stored in the GMP should be constant (or invariant) because the proportional relationship stored in the GMP should be preserved while the context, in which the GMP is embedded, changes. However, several studies were not able to observe this invariant structure consistently across different contexts (e.g., Smith et al. (1995)). Thus, previous studies questioned the existence of the GMP. However, the basic concepts of the GMP still remain valid and useful. People observed reduced variability in movement structure following practice and transfer of learned effects to related but novel movements. Also, the concept of the GMP is effective in explaining rapid speech motor control.

To explain why relative values were inconsistently observed, it is worth examining how people prepare (plan and program) upcoming motor responses in a sequential movement task as they execute a prior motor response. The way in which stored motor programs are concatenated or parameterized may explain the variability observed, which may violate the proportional scaling assumption. Internal model by Wolf and colleagues (1995; 1998) has proposed that movements can be controlled in a concurrent manner. Schmidt and Lee (1999) also revised Schema theory to explain how to control a movement in a concurrent manner. This current study sought the evidence of GMPs for lexical tones when they were produced in a sequence.

Therefore, this current study explored an alternative explanation for previous researchers' failure to find invariant structure. This study investigated whether invariant movement structure can be observed when a large unit of speech GMP is executed without the demands to prepare the next motor response or target speech sound. Because it is presumed that a larger GMP unit develops with practice, this study explored whether speech movement patterns that result in fundamental frequency (f_0) change over time, are determined by the degree of speech production experience of related movement sequences. Additionally, the degree to which concurrent motor control is affected by the size of the GMP unit was explored. The dual route model of speech production proposed by Hankamer (1989) and Varley and colleagues (1999b; 2006), which proposes direct and indirect routes to the GMP, was used to explain two possible manners to control movements. The control of action as a continuum between automatic and conscious controls has been proposed by Norman and Shallice (1980). This study examined whether the evidence of the GMP for speech, which is prepared and executed as a whole, was more evident in 3 syllable phrases as compared to 9 syllable phrases and whether this pattern differs between experienced and novice speakers.

<Key words: Generalized Motor Program (GMP), parameters, Mandarin lexical tone, Dual-route model, motor program size, speech production experience, reaction time, inter-syllable interval>

1.1 SCHEMA THEORY

1.1.1 Schema Theory

Schema theory (Schmidt, 1975) proposes the notions of schema and Generalized Motor Program (GMP). Schema theory suggests that there is a rule that stores an abstract pattern (or a population prototype) for a set of inputs and provides similar inputs with a categorical membership. In this conceptualization, a schema is a rule that describes the relationship between the outcome of a motor program and chosen parameters for those attempts (Schmidt, 2003). Schmidt (1975) proposed that four types of information are stored to form and update the schema after a movement is made. They are 1) the initial state of the muscular and environmental system, 2) the response specification made for that attempt, 3) the sensory consequences of the response produced, and 4) the outcome of the movement, such as knowledge of results (KR) or subjective reinforcements. A schema is therefore a memory trace composed of several interactive pieces of information regarding sensory and motor states. According to Schmidt (2003), a “recall schema” specifies the relationship between response specifications and the outcome of the motor program, and a “recognition schema” denotes the relationship between the past sensory consequences and the outcome of the program.

Schmidt (1975) also proposed an additional concept, called a Generalized Motor Program (GMP) in order to account for discrete and ballistic movement control which does not involve the role of sensory feedback. He suggested the existence of a GMP for a given class of movements. A GMP is “the pre-structured commands for a number of movements if specific response specifications are provided” (Schmidt, 1975, p. 232). Ballard, Maas, and Robin (2007) described GMPs as “abstract representations of movement structures” (p.1196). Schmidt (1975, 2003) and

Schmidt and Lee (1999) proposed that a GMP is a memory structure or program that specifies the order of events, relative timing and relative force (i.e., the amount of force over time) of the movement. The information stored in the GMP is described as "relative" because the proportional relationship between the subcomponents and the whole structure of one GMP has been presumed to remain constant, while the context to which the GMP is scaled and the absolute value of each subcomponent changes (Gentner, 1987).

As addressed previously, the GMP refers to the proportional relationships that remain constant across different contexts, while the response specifications adjusted to meet the contextual needs are considered as parameters. Response specifications, or parameters, allow a GMP to vary in terms of speed and force. Schmidt and Lee (1999) noted that an increase in force is related to an increase in displacement. In other words, it is presumed that a force parameter can be measured indirectly by the magnitude of displacement of an effector. Thus, the information in the GMP is expanded or compressed (Schmidt, 2003) along the parameter dimensions of duration and movement displacement (magnitude of change), although not every movement may be scaled linearly in terms of force dimension depending on the type of movement that is being produced (Schmidt, 2003). For example, the effect of gravity on muscles will change between horizontal and vertical movements. In this case, proportional force scaling across muscles may not achieve the same movement patterns that are produced horizontally and vertically (Schmidt, 2003). The GMP can also be applied across different effectors that involve different muscles (Schmidt & Lee, 1999), so the effector-selection is considered as a parameterization process. The parameters are presumed to define (or specify) absolute timing and force (Schmidt & Lee, 2005) via a recall schema (Shea & Wulf, 2005). The motor planning process, described in the four-level framework by van der Merwe (2009), appears to be an equivalent process to the parameterization process in

Schema theory. According to van der Merwe, a retrieved motor program is adjusted to meet the contextual needs at this stage, and also effectors are selected during this motor planning process. Muscles are selected during the motor programming process. Because Schema theory also suggested that selection of the effectors occurs during the parameterization process, the concept of GMP and parameter align with the motor program and plan notions of van der Merwe's model.

Kent (2015) noted that speech production involves over a 100 muscles that result in 2^{100} or 1 nonillion number of possible movements supposing each muscle has binary states such as contracted vs. relaxed. If every movement had a corresponding motor program, too many motor programs would have to be stored in the brain, and this would cause a storage and retrieval problem (Schmidt, 1975). Additionally, Higgins and Spaeth (1972) noted that no two movement trajectories are the same even when people are producing the same movement. Therefore, proponents of the stored motor program need to explain how a novel movement trajectory is produced when movements are produced using a limited number of stored motor programs. This has been referred to as the novelty problem.

The GMP notion resolves the storage and novelty problems for the same reason the schema notion does. The motor system needs to store only shared information among a given class of movements as information that composes one GMP. This resolves the storage problem. This GMP is utilized with different parameters in various contexts, explaining how the motor system generates novel movements.

Schema theory was originally proposed to explain discrete and ballistic movements that do not require sensory feedback. However, Schmidt (1975, 2003) proposed that sensory feedback is used for an on-going movement control in a slow movement. The theory hypothesized that a rapid action is controlled by recall memory, and the slow action is controlled by the combination of

recall and recognition memories. This is because the recall schema uses previous knowledge of results (KR) to estimate actual outcome and is relatively unaffected by the sensory feedback, while the recognition schema is influenced by the feedback. It was suggested that sensory consequences are used to update both recall and recognition schemata. According to Schmidt (1975), the theory is self-explanatory without including the notion of efference copy. Efference copy is information that the central nervous system maintains to know that the motor commands have been sent to the muscles to make a movement (Schmidt, 1975). Later, Schmidt and Lee (1999) used the concept of efference copy to explain feedforward rapid error correction mechanisms. However, still more work needs to be done in Schema theory in order to better explain how the sensory signals interact with the stored motor programs; in particular, in a movement that involves multiple gestures which are put together in a sequence. This idea of efference copy will be readdressed later when Schema theory is compared to Internal model in section 1.1.4.3.

1.1.2 Speech GMPs

As mentioned above, the notion of the GMP requires the demonstration of temporal invariance. Indeed, temporal invariance has been investigated actively in the speech as well as in the limb literature since 1965 (Kozhevnikov & Chistovich, 1965). It is important to review what has been found regarding speech GMPs and temporal invariance in order to give an idea about possible GMPs for lexical tones, which are the target stimuli of this study. Because temporal invariance will be addressed in the next section 1.1.3, this section will focus on how the GMP relates to linguistic or prosodic units. First, a few researchers have examined GMPs for segments, and different GMP groups were suggested based on distinctive features. Another group of researchers proposed how GMPs would play a role while controlling supra-segmental aspects. Segment

corresponds to a discrete unit, such as consonants and vowels, and supra-segment corresponds to rhythmic features, such as pitch, stress, and duration and is often temporally non-discrete so that it is superimposed over several segments.

For speech movement, Varley, Whiteside, Windsor, and Fisher (2006) and Cholin, Levelt, and Schiller's (2006) studies supported the existence of stored motor programs for frequently used speech units that could possibly vary in length, such as a syllable, a word, or a phrase. Van der Merwe (2009) proposed that the phoneme was the smallest unit of a speech motor program but also acknowledged that program size may be variable. Factors other than habitual f_0, speech rate, intensity, and clarity and related absolute timing and force are considered as parameters (E. Maas, Robin, Hula, et al., 2008) or speech motor plans (van der Merwe, 2009).

To address GMPs for speech segments, such as consonants, Ballard et al. (2007) suggested that voiced and voiceless stops seem to be controlled by different GMPs. These authors proposed this because "rate-dependent variation is permitted as it will not result in the production of different voicing phonemes" (p. 1198). Furthermore, Ballard et al. (2007) suggested that fricatives are likely controlled by a different GMP than stops because when patients with apraxia of speech (AOS) were treated, the treatment effect of the voicing of fricatives did not generalize to the voicing of untrained stops, and vice versa. Similarly, Maas et al. (2008) suggested that the same manner of articulation (e. g., stop) seems to be governed by the same GMP, although the effectors engaged may differ, which are also related to the place of articulation (e. g., lips for /p/ and tongue tip for /t/). On the other hand, different manners of articulation seem to be controlled by different GMPs even when two speech sounds share the same effector (e. g., /t/ vs /s/). This is because the stop requires a full closure and the fricative requires a narrow constriction of the vocal tract.

Speech treatment results support the above ideas about segmental GMPs for consonants. The learning of one type of movement is expected to generalize to other movements of the same GMP category. Raymer and Thompson (1991) reported limited generalization of treatment effects across different voicing groups, supporting the existence of different GMPs for voiced vs. voiceless consonants. In addition, Wambaugh, Martinez, McNeil, and Rogers (1999) observed that the target treatment sound (e.g., /ʃ/) was frequently replaced by a same manner but different place of articulation sound (e.g., /s/) rather than by a different manner sound. Therefore, these findings from treatment studies suggest that speech sounds, which differ in voicing or in the manner of articulation, are governed by different GMPs. On the other hand, Maas et al. (2008) suggested that the phonemes from a different place of articulation seem to belong to the same GMP group as long as they are produced with the same voicing and manner. Thus, it is predicted that transfer will occur across different places of articulation (the same GMP group), but not across different voicing or manners of articulation (different GMP groups) (Ballard et al., 2007; S. N. A. Hula, Robin, Maas, Ballard, & Schmidt, 2008; E. Maas, Robin, Hula, et al., 2008; Wambaugh et al., 1999).

In contrast to the segmental speech sounds discussed above, supra-segmental aspects refer to the rhythmic features, such as pitch, stress, and duration, which are associated with (or overlaid on) segmental units of varying length (Lehiste & Lass, 1976). Lexical stress pattern is a supra-segmental aspect that is assigned to words to distinguish meaning. It is defined in terms of relative prominence of f_0, intensity, and duration of measured acoustic signals (Cooper, Cutler, & Wales, 2002; Van Donselaar, Koster, & Cutler, 2005). Maas et al. (2008) speculated that the movement command for a lexical stress pattern is stored in the GMP and controlled by it. This is because there is a consistent pattern to produce lexical stress despite the variation in overall f_0, degree of clarity, or speech rate (E. Maas, Robin, Hula, et al., 2008). One thing that requires clarification is

that overall f_0, degree of clarity or speech rate are still considered as parameters. These prosodic components, such as pitch (or f_0), loudness (or intensity), degree of clarity, or speech rate may vary every time speech is produced depending on a speaker or context. However, a certain pattern in lexical stress is presumed to stay the same across different speech productions. Thus, it has been proposed that GMPs may exist for lexical stress patterns (E. Maas, Robin, Hula, et al., 2008).

Kim, Shaiman, and McNeil (n.d.) studied lexical stress patterns over four-syllable English nonsense words. The variability around the intra-oral air pressure trajectories was considered to correspond to stress patterns. These authors also speculated that this variability would be equivalent to the GMP error when residual errors around the trajectory are measured after effects of rate and vocal intensity are taken away via proportional scaling along the trajectories. A significant decrease in the residual errors after training was interpreted as support for the existence of GMPs for lexical stress. Additionally, the trained effect of four-syllable lexical stress patterns generalized to several different untrained consonantal contexts when these consonants shared certain aspects of segmental GMPs with the trained consonants. Thus, a GMP for lexical stress may be controlled separately from GMPs for segments, if the segmental GMPs carrying lexical stress have enough similarities among each other (Kim et al., n.d.).

The findings regarding speech GMPs were reviewed in this section to enhance readers' understanding about possible GMPs for speech and to give a rough idea of possible GMPs for lexical tones, which are the target stimuli of this study. Lexical tone is similar to lexical stress in that both distinguish word meaning while they are supra-segmental in nature. However, whether lexical tones have corresponding GMPs has not been explored. The distinction between lexical tone and lexical stress, and the evidence supporting the potential existence of GMPs for lexical tones, will be discussed in section 1.2. Last, although this section reviewed several studies that

proposed possible forms of speech GMP, this current study does not necessarily support the view proposed in this section. The distinction of movement patterns based on distinctive features is a reasonable approach because movement patterns are controlled to achieve linguistic goals. However, this approach limits its unit of interest to a phoneme level, when the actual unit of GMP may exist in variable sizes. Also, motor programs may exist as continuous gestural events rather than as discrete motor commands that correspond to linguistic units (Browman & Goldstein, 1992; Ziegler, Ackermann, & Kappes, 2011). The gesture as a motor program unit will be readdressed in secion 1. 1. 4. 4.

1.1.3 Inconsistent Findings of Invariant Structure

The above studies investigated the generalization context of speech GMPs based on acquisition, retention, and generalization results after training or treatment. The existence of invariant information, which is presumably stored in the GMP, can be derived from the speech acoustic signal as well as from kinematic data. The research that studied the invariant information from speech acoustic and kinematic data examined whether the relative or proportional timing or force relationships are observed. This section describes controversies around the existence of GMPs, which motivated this study to ask about the effect of unit size of motor programs while producing utterances of variable syllable lengths.

It has been presumed that for one GMP, proportional relationships between sub-segments and the whole movement structure remain constant, while the context in which that GMP is embedded may change (Gentner, 1987). Limb studies have reported constant proportional timing (or phasic) relationships among elements of movement across different speeds. For example, constant time to peak velocity, relative to the total movement time was observed in both EMG and

kinematic data of arm movement across various rate conditions (Carter & Shapiro, 1984). Similar patterns were observed in typing (Terzuolo & Viviani, 1979), in piano playing (Shaffer, 1980, 1984), and in overarm throwing movements (Roth, 1988). Other researchers supported the possible existence of GMPs and parameters while they observe the decrease in variability of the relevant measures during training sessions (acquisition, retention and transfer conditions). The results of these limb studies are summarized in section 1.1.4.1.

From a different perspective, Das (1988) suggested that invariance in motor outcome in fact is unimportant in the motor control system. Additionally, a few studies have failed to observe constant proportional relationships among the elements of a movement across different contexts, such as different rate conditions (e. g., Shapiro, 1977; Vu, Isableu, & Berret, 2016). Thus, some limb movement studies have challenged the invariance in movement structure.

Likewise, while the results of some studies are interpreted as support for the idea of invariance by demonstrating preserved proportional timing relationships in kinematic or acoustic data for speech (De Jong, Beckman, & Edwards, 1993; Gracco & Abbs, 1986, 1988; Tuller, Harris, & Kelso, 1981; Tuller, Kelso, & Harris, 1982; Weismer & Fennell, 1985), other researchers have failed to observe consistent proportional relationships in at least one or two conditions in EMG, kinematic, or acoustic data (Kozhevnikov & Chistovich, 1965; Löfqvist, 1991; Max & Caruso, 1997; Munhall, 1985; Smith et al., 1995). Thus, inconsistent findings have weakened the evidence for the existence of the GMP. This section describes several kinematic and acoustic studies' findings that are interpreted as support for or refutation of the consistent proportional timing or force concept during the production of speech.

In kinematic data for speech, Gracco and Abbs (1986) observed that peak velocity timing and sequencing relationships among upper lip, lower lip and jaw movements remained consistent

for lip closure during bilabial production. Because order of motor events is presumably stored in the motor program, this constant phasic relationship among the articulators can be considered as one piece of evidence for the existence of the GMP (Schmidt & Lee, 1999). However, the sequence of gestural events alone may not be adequate evidence to conclude about the existence of the GMP. As noted by Gracco and Abbs (1986), the sequence of gestural events was sometimes evaluated as critical by the speech system and was strictly maintained, but sometimes not. Also, the movements that did not maintain the constant phasic relationships were still frequently perceived as correct speech sounds. For instance, when speech sound /t/ is produced in many different vowel contexts, the movements among articulators change based on the type of vowel that surrounds /t/, but /t/ is perceived as the same speech sound across different contexts (MacNeilage, 1970).

Invariance also has been studied in EMG signals during speech production because muscle activation pattern may reflect the motor commands stored in the GMP (Carter & Shapiro, 1984). When EMG area, peak, and rise time of upper and lower lip muscles were examined during bilabial closure (Gracco, 1988), proportional timing relationships were maintained between fast and slow rate movements. Tuller et al. (1981; 1982) also measured temporal patterns in EMG signals from lip, tongue, and jaw muscles during speech production. They observed that activity of each individual muscle (duration and time of peak magnitude) varied based on the speech rate or stress condition. However, the proportional timing relationship between two landmarks[2] in the EMG signals measured over several articulators remained constant across different speech rate and stress

[2] Ratio of latency (i.e., the time between the onset of muscle activity for one consonant and that of the flanking vowel and vice versa in the /pipap/ sequence) over the period (i.e., the time between the onset of muscle activity for one consonant and that for the next consonant, or the time between the onset of muscle activity for one vowel and that for the next vowel in the /pipap/ sequence) (aka phasic timing relationships among muscles).

conditions (Tuller et al., 1981; Tuller et al., 1982). Kozhevnikov and Chistovich (1965) observed consistent proportional timing patterns at word or phrase-level kinematics as the speech rate increased, although such consistency was not observed at sound- or syllable-level kinematics. In many of the above studies, the proportional timing of certain movement kinematics or EMG signals remained constant across different speech rate conditions.

Constant proportional timing also has been observed in acoustic data for speech. Weismer and Fennell (1985) examined the timing ratio appearing in phrase-level utterances, such as ”Bob hit the big dog,” and observed that the ratios among acoustically defined intervals remained constant between two different speech rate conditions (conversational vs. fast). De Jong (2001) also observed constant proportional timing between two measured points among consonant release, vowel onset, and consonant closure as the speech rate increased in CV structures (e.g., ‘bee’ or ‘pea’), but this pattern did not appear in VC structures (e.g., ‘eeb’ or ‘eep’).

As mentioned earlier, a number of studies have questioned the existence of constant proportional relationships in kinematic and acoustic data for speech (Löfqvist, 1991; Max & Caruso, 1997; Munhall, 1985; Smith et al., 1995). Smith et al. (1995), for example, failed to find constant proportional timing relationships among measured kinematic signals, which tracked lower lip displacement in the superior-inferior dimension, across different speech rate conditions. This is because the relative timing of the middle three lip opening gestures of “Buy Bobby a puppy” in relation to the first and last opening gestures appeared proportionally later as the speech rate slowed. Similarly, Löfqvist (1991) revealed that proportional timing was maintained 67% of the time at the intrasegmental level (the intervals between two gestural landmarks within a consonant (e.g., /t/, /s/)). However, it was preserved only 10% of the time at the intersegmental level (the interval between two gestural landmarks that exist across two different vowels or across a

consonant and a vowel). These two studies provide evidence in kinematic data that questioned the existence of constant proportional relationships.

Furthermore, additional research provided another reason why the idea of constant proportional timing relationship might be challenged. Munhall (1985) observed a high correlation between the first period (the interval between two zero velocity points at the beginning of tongue lowering for the first vowel and for the second vowel, vowel-to-vowel) and the first latency (the interval between the tongue lowering onset for the first vowel and the tongue raising onset for the intervocalic consonant, vowel-to-consonant) in the kinematic tracking signal of tongue dorsum movement. However, Munhall (1985) speculated that in order for this correlation between period and latency to be meaningful, this observed correlation should exceed the baseline part-whole correlation, which comes from comparing two uncorrelated parts that compose the whole period. Because these two parts, which compose the whole, are expected to covary, Munhall (1985) proposed that the observed correlation exceeding this inherent part-whole correlation will be due to relative timing relationship. The period over latency correlation (0.92) of one of their research participants exceeded the 95% confidence interval (0.674-0.818) of the part-whole correlation (0.776). However, the other participant's period over latency correlation (0.88) did not exceed the 95% confidence interval (0.662-0.892) of the part-whole correlation (0.806). In other words, Munhall observed that at least one participant's period over latency correlation may be due to simple part-whole correlation. Therefore, the existence of one invariant relative timing relationship across different rate conditions was questioned.

The existence of the GMP was also challenged when studies failed to find constant proportional relationships in acoustic data. Max and Caruso (1997) failed to find consistent proportional relationships in the acoustic data when they examined five ratios composed of two

pre-determined intervals from the phrase ‘Buy Bobby a puppy.’ The proportional timing relationship changed as the speech rate changed at both syllable- and phrase-level comparisons. This result was consistent with the previous studies that examined acoustic data (Abry, Orliaguet, & Sock, 1990; Gay, 1978).

In summary, these findings suggest that relative timing relationships were not consistently observed in speech EMG, kinematic, and acoustic signal across different rates, and this inconsistency challenges the existence of the GMP in speech.

1.1.4 Limitations of Explanations for the Inconsistency

The above section provided examples of inconsistency in the findings regarding relative timing relationships in the acoustic, EMG, and kinematic data. This section explains several possible reasons for failing to find consistent relative timing relationships.

1.1.4.1 Biomechanical Constraints

One possible explanation for failing to find evidence for a GMP involves biomechanical constraints that change proportional relationships in the acoustic and kinematic signals. Biomechanical constraint refers to the limitations in the range of motion due to musculoskeletal structural organization and stiffness in the muscles (Hu, Murray, & Perreault, 2010). Perrier et al. (2003) noted that “the passive tongue elasticity, the muscle arrangements within the tongue, and the force generation mechanism” are the examples of biomechanical properties in the tongue that may constrain the range of motion. That is, the proportional timing relationship may be influenced by biomechanical constraints imposed by effectors that are involved in a particular movement (Gentner, 1987; Tseng, 1981; Weismer & Fennell, 1985). Understanding of the mechanism for

general limb or hand movement may help to understand speech movement. Thus, a few studies that were intended to explain limb or hand movement will be addressed below before explaining biomechanical constraints for speech.

Gentner (1987) proposed that the relative timing of finger movements for typing was influenced by biomechanical constraints of the effectors involved in depressing two key sequences. When two keys were typed slowly, the duration required was almost constant, regardless of which key combinations were chosen. However, when typing rates increased, the typing duration varied based on the fingers or hands involved, suggesting relative timing relationship was influenced by the biomechanical constraints (Gentner, 1987).

Kelso (1986) also reported that different phasic relationships appeared as the movement rate changed. This is because one phasic relationship among related effectors at one movement rate cannot maintain the biomechanical requirements for the other movement rate. Kelso (1986) suggested this idea because he observed that a horse shifts its locomotion manner from walking to trotting when walking becomes metabolically costly at a certain velocity. Heuer (1988) and Heuer and Schmidt (1988) suggested that the relative relationships found in naturally existing movement trajectories might depend on how natural the trajectory is to the biomechanical system. If the movement requires unnatural (or non-harmonic) relative timing relationship to the biomechanics, it may require additional training. Then, this movement with non-harmonic patterns may be governed by different GMPs if interpreted from Schema theory view. However, Heuer and Schmidt (1988) occasionally observed the occurrence of a transfer between harmonic and non-harmonic movement patterns. If this was the case, their findings were not in line with Schema theory view because Schema theory assumes that the transfer occurs between two movements that are governed by the same GMP.

With respect to speech production, Schema theory does not specify how a single GMP controls several articulators that are engaged in one speech act. It is possible that a single articulator maintains consistent relative timing or force relationships between events, such as the time of peak velocity for consecutive lip opening gestures across speech rate conditions (Smith et al., 1995), while it also makes coordinated movements with other articulators. Each articulator has a potential to move independently from another (Barlow & Bradford, 1992), although the lower lip, tongue, and jaw work together, as do tongue tip, blade, and body.

However, despite the high degrees of freedom in the articulatory systems, Gracco and Abbs (1986) noted that the articulators seemed to covary and to be interdependently controlled to achieve the same speech goal in various ways. Additionally, the degree of coupling of the articulators increased as the speech rate increased (Gracco & Abbs, 1986). For example, Cummins (1998) reported that high variability of the relative ratio was observed at slower rates and that this effect disappeared at the fastest rates. The ratio was obtained by comparing the duration of the final foot to that of the prior foot of a sentence. Cummins speculated that speakers have a restricted coordinative strategy and a fixed rhythmic structure at fast speech rates as compared to slow speech rates. Thus, the degree of biomechanical constraints increases as speech rates increase. In other words, biomechanical constraint may become more influential at a fast speech rate than at a slow speech rate. Thus, manipulations of speaking rate may influence the degree of biomechanical constraints and the observation of invariance in the trajectory.

Similarly, variability in movement trajectories may increase due to biomechanics, because each articulator has different mass, inertia, and stiffness. For example, Müller et al. (1977) and Müller and MacLeod (1982) noted that the inertial characteristics of the lip and jaw are different.

As a result of these biomechanics, each articulator has different timing or coordination requirements from each other.

However, biomechanical constraints, as well as dynamical properties of the articulators, such as mass, stiffness, damping, etc. (Kelso et al., 1986), may become part of the motor program. Gracco and Abbs (1986) observed that movement of the upper lip (UL) consistently preceded the lower lip (LL), and the LL consistently preceded the jaw (J) when lips closed for bilabial stops. However, the constant sequence of articulatory movements disappeared in the lip opening movements. In other words, the articulators did not move in the same sequence during the lip opening movement, as compared to closing movement. Gracco and Abbs (1986) speculated that this finding eliminated the possibility that articulatory movements result from pure biomechanics. Instead, articulatory movements may be controlled by plans and programs. Gracco (1988) also argued that although the inertia and stiffness characteristics are different between UL and LL, the increase in the mass of these two articulators during the computational modeling does not significantly affect the peak EMG velocity when the force profile is maintained. Therefore, Gracco proposed that the differences in the peak velocity timing between UL and LL are caused by active control of the neural-muscular system. Furthermore, Bailly (1998) suggested that the flexibility in the central nervous system (CNS) has the ability to take into account the biomechanics of the articulators. Perkell (2000) also addressed that the speech production system considers the biomechanical constraints when planning movements for a sound sequence.

In summary, there has been the suggestion that inconsistency in the GMP results from biomechanical constraints. In particular, it was observed that proportional relationships were not maintained in some acoustic and kinematic studies and this was attributed to biomechanical constraints at fast rate conditions as compared to normal or slow rate conditions. However, it is

possible that dynamics of the articulators could be taken into account while movements are programmed (Gracco & Abbs, 1986). In conclusion, the biomechanics may be a cause for the variability in the movement trajectory, which violates the invariance assumption of Schema theory. However, it is also possible that the motor system is experienced enough to handle the dynamics in the motor system coming from biomechanical constraints during speech motor planning or programming.

1.1.4.2 Inessential Components (Parameters) Becoming Essential Components (GMPs)

The above section concluded that the central nervous system may be able to account for the biomechanical constraints and dynamical properties of the articulators (or effectors). As an extension of the above concern, a few researchers have proposed that a parameter factor such as speech rate might cause changes to the GMP when one parameter factor changes above a certain threshold. While this explanation is plausible, other scenarios will also be discussed.

Kugler, Kelso, and Turvey (1982) suggested that any variable that changes the topological (or phasic) relationships in the movement structure is an essential variable, and any variable that changes the movement structure but does not change the topological qualities is an inessential variable. When it comes to this point, the concepts of essential and inessential variables look relevant to concepts about GMP and parameter of Schema theory respectively. It is because different phasic relationships are related to different GMPs and any factor that determines different phasic relationship may determine the GMP category as well. The notion of an inessential variable also corresponds to the parameter concept of Schema theory.

Interestingly, Kugler, Kelso, and Turvey (1982) proposed that a continuous change in an inessential variable may eventually result in changes in the phasic relationships in the behavior. Thus, the concepts of essential and inessential variables are not exactly the same as the concepts

of GMP and parameter because they assert that the inessential variable (aka, parameter) has a potential to influence and bring changes to an essential variable (aka, GMP). Likewise, Weismer and Fennell (1985) also suggested that inessential factors, such as speech rate, become essential when speech slows down significantly. They asserted that although rate has been considered as a parameter factor according to Schema theory, if it goes above or below a certain threshold, it may become an essential factor rather than a parameter and may require a different GMP or different relative timing relationship. Fennell and Weismer (1984) and Weismer and Fennell (1985) also addressed that speakers experience difficulty in maintaining existing phonetic plans, equivalent to GMP, when they have to expand them in slow speech, equivalent to parameter change, and speech starts to sound prosodically anomalous when these plans are maintained. Thus, the change of relative temporal structures is necessary in slowed speech as compared to conversational and fast rate speech, supporting Kugler, Kelso, and Turvey's (1982) argument.

However, instead of theorizing that different GMPs are used for different rate conditions, the same GMP may be used for different rate conditions. The concept of a GMP had been first proposed to explain discrete and ballistic movements which are controlled by a single GMP. Although a single GMP for speech may control muscles of several articulators at the same time, it is not clear whether a single GMP will be expanded or compressed linearly across different rate contexts. In other words, the question, how many parameters a single GMP may relate to when this GMP is expanded or compressed, has not been answered.

Also, it has not been experimentally investigated how a retrieved GMP for utterances of variable syllable lengths is parameterized. Motor commands for utterances of multiple syllable lengths may be stored in a single GMP as a whole after repeated practice on the same utterance. This assumption is based on an existing proposal which asserted the existence of motor programs

of varying lengths (Cholin et al., 2006; Guenther, 2006; Guenther, Ghosh, & Tourville, 2006; E. Maas, Robin, Austermann Hula, et al., 2008; Schmidt & Lee, 2005; Varley, Whiteside, Windsor, et al., 2006). Also, no matter how many GMPs are involved in one utterance, another question to explore is how many times each GMP is parameterized as it is executed. For an utterance of variable syllable lengths, the number of GMPs or parameters involved in that utterance is not clear.

Weismer and Fennell (1985) reported that consonants are relatively difficult to compress as compared to vowels. They hypothesized that the effect of the incompressibility of consonants appears to accumulate as speech progresses and this effect becomes larger in the later portion of speech sequences. Fujimura (1987) also suggested that the assumption regarding a constant scaling factor needs to be changed to account for the nonlinearly lengthened intervals in the later portion of an utterance. These changes in the parameters across an utterance, regardless of whether a single GMP or multiple GMPs are involved, may affect the movement trajectories.

Schema theory assumes invariant trajectories. These trajectories are expanded or compressed proportionally when an utterance is examined as a whole. However, if the parameter, such as speech rate, varies as the utterance progresses, constant proportional scaling may not be observed. The changes in the parameters may influence the observation of invariance in the trajectory.

Thus, the variability of movement trajectories may have been increased because various parameters are used while a single GMP is executed, while assuming that the size of a single GMP may vary and it may store motor commands for multiple syllables as a whole. Weismer and Fennell (1985) once proposed that the movement trajectory for a GMP remains intact as long as elongation occurs during pauses for slowed speech. However, the parameterization may occur at various points (vowel, consonant, or inter-syllable interval) as a single GMP is executed, not only in

advance of movement execution or during execution. This may result in the different trajectories for the same utterance every time it is produced.

In addition to the various parameters used, the number of GMPs that are engaged in an utterance becomes unclear. The possible existence of stored motor programs for speech movements of varying lengths, whether shared or not, i. e., GMPs, have been proposed by several researchers (Guenther, 2006; Guenther et al., 2006; E. Maas, Robin, Austermann Hula, et al., 2008; Schmidt & Lee, 2005). In other words, an utterance may be composed of a single GMP or of several GMPs, and how they are put together may affect the trajectory of speech. When the motor control system needs to prepare the upcoming response while executing the current motor response, the execution process may slow down. This may cause temporary changes in movement trajectories, while the relative temporal structures within a single GMP involving multiple gestures do not change. In other words, every concatenation effort may cause changes in the movement trajectory while movement structure of each GMP involved remains the same.

In summary, the explanation that an extreme change in a parameter causes changes in the temporal structure of a GMP may not be satisfactory. It is because the unit size of the GMPs for speech as well as how the retrieved GMPs are concatenated (or put together) or parameterized may provide another explanation. To date, these questions have not been experimentally investigated. In the next section, Internal model addresses the possibility that after retrieving a crude format of a stored motor program, the parameterizing of the motor program will occur in a concurrent manner.

1.1.4.3 Internal Model

Internal model explains the inconsistent findings of relativity in the kinematic and acoustic data by proposing that speech movements are controlled in an on-going, concatenating manner rather

than by means of movement trajectories that are prepared and executed as a whole. Internal model also explains how a movement of longer duration, due to concatenating multiple motor programs, is controlled. This model provides an alternative view to Schema theory.

Researchers have suggested that there are two kinds of internal models: the forward model and the inverse model (Wolpert & Ghahramani, 2000; Wolpert et al., 1998). The forward internal model predicts the consequences of the movement and the environmental states that result from the given motor commands without peripheral information. This model assumes that the central nervous system (CNS) has the ability to predict the future outcome of the motor commands without any contribution from peripheral information. The forward model explains the error correction mechanism of the motor control system as a process comparing the estimated state using efference copies to the desired state.

The inverse internal model estimates motor commands based on the observed transition of the motor system between the states (Wolpert et al., 1995; Wolpert et al., 1998). The inverse internal model does its estimation by considering the state of the motor system and the tasks it is dealing with (Wolpert & Ghahramani, 2000). The inverse internal model converts target displacement into what they called, “motor plans” (Desmurget & Grafton, 2000; Wolpert & Ghahramani, 2000), which can be considered as “motor programs” in this document. In addition, the inverse model reflects the biomechanical state of the motor system, such as inertial and viscous properties of the effector, in its estimation of the outcome of the motor commands (Desmurget & Grafton, 2000).

The forward model assumes that compensatory movements are possible even without the involvement of sensory (visual or proprioceptive) feedback as is evidenced for limb movement executed without visual information (e.g., in the dark) or in surgically deafferented monkeys

(Guthrie, Porter, & Sparks, 1983; Prablanc & Martin, 1992). However, the role of inflow of sensory information (or reafferent signal) also has been emphasized because the feedforward system demonstrated inaccuracy in the outcome of motor commands (Desmurget & Grafton, 2000). Feedback control is necessary to correct for errors in the motor control system (Miall, Weir, Wolpert, & Stein, 1993). Any detected discrepancy between the predicted reafferent signal and the actual reafferent signal is used to correct the motor commands. Also, this feedback is used as the basis for the inverse internal model (Miall et al., 1993). It is presumed that the nervous system learns to estimate the consequences of motor commands, and this forward model feeds the internal feedback loop. Then, there remains an almost negligible delay in the internal feedback loop (Desmurget & Grafton, 2000; Wolpert et al., 1998).

This may apply differently for speech, because the role of sensory feedback often has been proposed while explaining limb movement which involves visual or proprioceptive sensation. When it comes to speech, auditory information may play an important role than visual sensation, and the processing of auditory feedback has been suggested to be slower than other types of feedback. Kawahara (1993) reported participants demonstrated corrective response with around 100-200 ms latency in response to the auditory pitch perturbations. Tourville et al. (2008) suggested around 108-165 ms was required to make a compensatory adjustment when the first formant (F1) value of the /ɛ/ vowels is perturbed. Cai et al. (2011) reported about 150-160 ms until they could observe a compensatory response to an up or down perturbation in the second formant (F2). Thus, even with the help of efference copy, it appears that the auditory feedback is utilized with a bit of latency while controlling speech movement.

The almost negligible delay in the internal feedback loop has been used as evidence of concurrent movement control (Desmurget & Grafton, 2000; Wolpert et al., 1998). The use of

efference copies enables rapid on-going movement control, and this aspect has been the major difference between Internal model and Schema theory. Schema theory has difficulty explaining rapid corrective responses because the information from the sensory feedback loop is presumed to be slow and inefficient to use. However, by implementing the idea of using the efference copies of motor commands, Internal model is better at predicting motor, sensory, and environmental consequences of the motor command and effectively explains how the movement is controlled and adjusted rapidly. Later, Schema theory (Schmidt & Lee, 1999) also adopted the idea of efference copy and explained how a sequence of motor programs are produced.

Whalen's (1990) findings support that this hypothesis of concurrent movement control is applicable to speech. He observed that speech production could be initiated with only partial planning of the whole utterance, and the rest of the utterance was planned/programmed in a concurrent manner. The study participants were asked to produce incomplete stimuli, and as soon as they began their speech, a missing consonant (/p/ or /b/) or vowel (/i/ or /u/) was provided. The stimuli were /a_u/ (consonant unknown condition) and /ab_/(vowel unknown condition) in experiment 1, and /əbu_a/ (consonant unknown condition) or /əb_pa/ (vowel unknown condition) in experiment 2. Anticipatory coarticulation did not appear when the information about the upcoming vowel or consonant was absent, but participants could incorporate coarticulation as soon as the missing information was given as they produced their speech. Longer closure duration than normal for the vowel-unknown condition was evidenced. During this duration, the planning for the upcoming vowel was presumed to occur. The time required to resolve missing segmental information was found to be similar in both consonant and vowel unknown conditions (around 300-360ms). Also, it appeared that the coarticulation plan for the final one or two syllables was recovered after the utterance began rather than it being produced as a result of articulatory

dynamics. This was hypothesized because the effect of the preceding vowel on F2 of the final vowel /a/ appeared in both conditions to an equal degree, regardless of whether the participants knew the final vowel or not at the onset of speech. Thus, Whalen's (1990) study suggested that speech could be initiated with only partial planning of the whole utterance, and the rest could be planned/programmed or reprogrammed in an concurrent or parallel manner.

Although Internal model proposes that movements are controlled in a concurrent manner, the proponents of Internal model still agree with the need for a crude form of a motor program. It has been presumed that the initial motor program is subsequently adjusted by the motor system using internal feedback loops as the movement progresses. Furthermore, although patients with deafferentation can make relatively accurate movements, frequently flawed motor control by these patients (Desmurget & Grafton, 2000; Ghez, Gordon, & Ghilardi, 1995; Sainburg, Poizner, & Ghez, 1993) suggests that the initial motor program is crude and requires adjustments to meet the contextual needs. Subsequent updates are necessary as the movement progresses. Crude motor control has also been observed in speech studies under the condition of delayed auditory feedback (DAF). For example, Borden (1979) reported the increased variabilities in the EMG signals when EMG signals were taken from genioglossus while making tongue fronting and elevation movements for speech under DAF as compared to under no DAF. On the other hand, Perkell (2000) proposed a robust internal model in mature speakers when auditory feedback was not provided due to hearing loss. He presumed that it might be owing to the preserved tactile sensation. Thus, while the sensory feedback may play a role in controlling movements, the degree of influence of the feedback may depend on how robust Internal model became through learning or practice.

Wolpert et al. (1998) and Wolpert and Ghahramani (2000) suggest that motor commands do not control each individual muscle, but rather control "a few motor primitives (Wolpert &

Ghahramani, 2000, p. 1214) [or modules (Wolpert et al., 1998, p. 345)]," which generate patterned muscle activations. Wolpert et al. (1998) propose that multiple pairs of inverse and forward models control these primitives. Additionally, each module can be scaled to numerous contexts (Wolpert & Ghahramani, 2000; Wolpert et al., 1998). Wolpert et al. (1998) mentioned that based on a few stored modules (e.g., 32 inverse models), generation of numerous behaviors (e.g., 2^{32} or 10^{10} behaviors) is possible, which explains every new behavior in our daily lives. These multiple modules are similar to the GMP notion of Schema theory.

Shadmehr and Mussa-Ivaldi (1994) and Wolpert and Ghahramani (2000) propose that the motor system determines the optimal trajectory in which movement is maximally smoothed. For instance, a maximally smoothed trajectory follows a Gaussian function, which tracks the joint velocity of a single joint (Shadmehr & Mussa-Ivaldi, 1994). In addition, Wolpert et al. (1998) and Wolpert and Ghahramani (2000) proposed that motor commands are generated inversely by converting the desired state (along the optimal trajectory) into motor plans. Walker and Hickok (2015), Guenther (1995), and Perkell et al. (2000) proposed that auditory representations drive speech motor control. These speech targets are learned from reafferented feedback via error back-propagation mechanisms (Bailly, 1998). The mapping between auditory targets and motor gestures develop while infants learn to speak (Guenther, 1995; Hickok, 2012). This idea leads to the role of central representation while controlling movement. This is the topic of the next section.

Hickok (2012) adopts this internal model framework and extends it by combining it with the existing hierarchical processing model for speech production. This model is called as Hierarchical State Feedback Control (HSFC) model. In this HSFC model, Hickok (2012) hypothesized that higher-level sensory goals of speech, such as syllables, reside as auditory targets. Lower-level goals of speech, such as phonemic-level targets, reside as somatosensory targets.

Hickok (2012) hypothesized that the auditory system drives higher-level control of cyclic movements, such as cycles and half cycles. Somatosensory system drives lower-level control, such as the end point of vocal tract configuration. This also proposes that the auditory target is independent of the context, whereas the somatosensory goals are flexible and adjust to the phonetic contexts (Hickok, 2012).

Researchers proposed that the posterior parietal cortex (PPC) and the cerebellum are the related areas for internal models (Desmurget & Grafton, 2000; Miall et al., 1993). Likewise, Hickok (2012) proposed that the temporal-parietal boundaries appear to be related to the higher (phonological) level processing. The cerebellum is involved in the sensorimotor integration for lower (phonetic) level processing in the HSFC model.

In the existing internal model, an efference copy is defined as a copy of a completed motor program. Thus, the internal forward prediction only occurs after motor programming is completed. However, Hickok (2012) proposed an input from the lexical level (lemma) that activates both auditory and motor phonological representations and "the activated motor representation sends an inhibitory signal to the auditory target" (Hickok, 2012, p. 141). The motor-induced signals sent to the auditory representation are presumed to be inhibitory. When the inhibited sensory cells match with the reafferent sensory information, no corrective signals are sent to the corresponding motor representation area to correct the motor plans. However, if they do not match with the cells activated by the reafferent sensory information, the corrective signals are sent to the motor cortex to correct the motor commands. This account explains the motor-induced suppression responses, how the sensory area is rapidly suppressed after activation and how it is readied for the next activation. In this case, the efference copy becomes part of the motor planning process, not a result of a completed motor program.

Furthermore, Pickering and Garrod (2013) proposed tighter relationships between comprehension and production systems. They suggested that forward and inverse internal models play a role during conversation, not only when speakers monitor their own speech but also when listeners predict other's connected speech.

In this formulation, Internal model might be an acceptable alternative model to Schema theory. The model is good at explaining how movements are controlled and a movement schema is updated in a concurrent manner. However, the model does not completely exclude the possible existence of a stored motor program, such as a GMP. Also, the type of information stored in the GMP and how stored motor programs are put together still require investigation.

1.1.4.4 Invariance in Central (or Perceptual) Representation

This section describes another hypothesis researchers have proposed as a reason for failing to find consistent relative relationships in the data. It has been proposed that the relative relationship, stored as both sensory and motor information, is centrally represented, and may not necessarily be observed in the movement outcome. Perturbation study results support this idea. Furthermore, in this section, the possibility will be discussed that these relative timing or spatial change relationships may be achieved by the motor system in a concurrent manner rather than at a whole utterance level. For the same reason, examining the motor program unit size becomes important, because it may determine whether the motor program unit of interest contains information regarding proportional timing or force relationships.

Heuer (1988) proposed that while controlling a movement, the motor system might utilize relative timing information that is part of the central level representation. This central representation may not linearly correspond to the relative timing information observed in the peripheral level data (or movement outcome). Heuer (1988) proposed that strict relative timing

relationships might not be observed in the movement trajectories. This is because, while components of the movement trajectories are concatenated and executed, these relationships might be interrupted by motor delays that intervene in this process. Thus, possible concurrent concatenation and parameterization of motor programs may cause changes in the movement trajectories, which, otherwise, would have been invariant. In order to discuss centrally present invariant information, two aspects need consideration. First, whether the central representation contains relative timing information, and second, whether the representations for perceptual and motor systems share information.

Many perturbation studies indirectly addressed whether central representation contained relative timing information. According to Internal model, central representation is important for the control of concurrent movements because this representation is used as a reference state (Bailly, 1998; Shadmehr & Mussa-Ivaldi, 1994). The auditory and somato-sensory perturbation studies observed responses that compensated for the sensory perturbation effect (Abbs & Gracco, 1984; Cai, Ghosh, Guenther, & Perkell, 2010; Shaiman, 1989; Shaiman & Gracco, 2002; Shiller, Sato, Gracco, & Baum, 2009; Villacorta, Perkell, & Guenther, 2007). This fact supports the hypothesis that perceptual representation may be invariant and that speakers try to achieve the relative relationships in the perceptual target (Bailly, 1998). Therefore, it is possible that while motor programs adjust to meet the perceptual representation, it may be an equivalent notion to the central representation which Heuer (1988) suggested, storing relative timing and force relationships.

In addition, with sustained perturbation over several trials, the effect of the perturbed auditory feedback has been reported to last after it is discontinued (Cai et al., 2010; Purcell & Munhall, 2006a; Shiller et al., 2009; Villacorta et al., 2007). Shiller et al. (2009) discovered that

participants demonstrated a boundary shift for /s-ʃ/ sounds after being exposed to an altered feedback condition. This suggests the possibility of an altered or recalibrated acoustic representations (or targets) in which relativeness is maintained after the perturbation phase (Shiller et al., 2009). Both Schema theory and Internal model permit changes in the perceptual and motor representations. This is because Schema theory assumes schemata development after several attempts at the same kind of movement. Internal model also permits the possibility of updating internal models based on sensory feedback.

While perceptual representation is used as a reference state for movement control in these perturbation studies, this perceptual representation may contain information about relative timing and force. The absolute values of the perceptual target may change as long as the relative relationships in the acoustic or kinematic targets are preserved. Ramadoss (2012) reported that Thai tone language speakers often perceived the low tone (L-tone) as the falling tone (F-tone), when the shape of the Thai L-tone trajectory resembled that of an F-tone. This happened even when the absolute fundamental frequency (f_0) values of this L-tone did not reach the values of the F-tone. This finding suggests that changes in the relative f_0 trajectory affect the tone perception more than the absolute f_0 target values do.

The results from a few sensory perturbation studies imply that the perceptual and motor systems attend to the information about relative relationship in the movement outcome. Cai et al. (2010) perturbed the first formant (F1) while speakers produced the Mandarin triphthong /iau$_{55}$/. Here, the triphthong /iau$_{55}$/ indicated that a triphthong /iau/ was combined with two high-flat tones. The number tone marks indicated the height of the fundamental frequency level, 5, thus, being the highest fundamental frequency level and 1 being the lowest. They observed that the relative timing

of the peak of the F1 in the vowel /a/ was maintained across experimental phases with or without perturbation.

Additionally, the compensatory response was observed not only in speech productions that were directly perturbed, but also in unperturbed speech productions. Cai et al. (2010) observed compensatory responses not only in the perturbed triphthong /iau$_{55}$/, but also in the unperturbed vowels, such as /iou$_{55}$/ and /uai$_{55}$/. Interestingly, the generalization effect was weaker when the unperturbed speech sounds were more dissimilar from the perturbed speech sounds. Villacorta et al. (2007) also demonstrated that participants not only quickly adjusted their F1 in the opposite direction when auditory feedback of F1 was perturbed, but also generalized the effect to unperturbed acoustic parameters such as f_0 and F2 (Villacorta et al., 2007). Therefore, participants appeared to maintain the relative distances among the f_0 and formants in the vowel (Villacorta et al., 2007). One might speculate that the whole sound system was recalibrated. In particular, the vowel systems seemed to be recalibrated under the condition of perturbation in such a way as to maintain relative values in the acoustic parameters within and among vowels.

During connected speech, Cai, Ghosh, Guenther, and Perkell (2011) introduced an auditory perturbation to the F2 minimum of a vowel. They observed an immediate compensatory delay in subsequent portions of the utterance, as well as in the F2 of this perturbed vowel. A statistically significant change occurred in timing during the perturbation condition as compared to the no perturbation condition. Thus, it appeared that the participants adjusted their responses in order to maintain the relative timing relationships among speech segments when the speech rate was slowed.

Last, the degree of compensatory response to auditory perturbation may be influenced by how strong the representation is in the perceptual and motor systems. This is supported when

smaller compensatory responses to the auditory perturbations were observed in native Mandarin speakers than in non-native speakers (Ning, Loucks, & Shih, 2015; Ning, Shih, & Loucks, 2014). This result indicates that perceptual representation in more experienced speakers may be so strong that their speech production is less influenced by auditory perturbation than that of novice speakers. This is consistent with Guenther's (1995) proposal that more experienced speakers rely less on the feedback system, but more on the feed-forward system.

After reviewing these findings from perturbation studies, several assertions may be proposed. First, centrally present perceptual targets may be used as reference states while controlling movements, as proposed by Shadmehr and Mussa-Ivaldi (1994). Second, the auditory perturbation appeared to result in recalibration of the whole sound system when reprogramming after an auditory perturbation. Third, the way perceptual and motor systems compensated for the auditory perturbation was by attending to desired relative relationships within and among the acoustic signals for speech sounds. Last, the degree of compensation seemed to be affected by how robust the representation was, which might be determined by the degree of speech production experience.

On the other hand, perceptual and motor representations may use yet another type of information, other than relative timing or force. This includes the direction and magnitude of change in the acoustic parameters over two consecutive time points, as long as the magnitude of change reaches a certain threshold. This idea also is relevant to the fact that speech is controlled in a concurrent manner.

A speech target has been suggested to exist within a range in acoustic and physical spaces. In the Directions into Velocities of Articulators (DIVA) model, Guenther (1995) suggested that the speech sound target is expressed as a region, not as a point, in the oro-sensory map. This motor

output representation is called a “convex region.” This representation includes spatial movement targets along several dimensions of the relevant effectors (Bailly, 1998; Guenther, 1995). This explanation about the convex region was restricted to the phoneme level in the early DIVA model (Guenther, 1995), but later it was explained as a moving target area along a time axis (Bohland, Bullock, & Guenther, 2010; Guenther, 2006; Guenther et al., 2006).

The speech target seems to exist within a range, but with a clear boundary, because of categorical perception. Some studies report that the speech perception system abruptly identifies two speech sounds as different when the acoustic signals change gradually in a continuous manner (Liberman, Harris, Hoffman, & Griffith, 1957; Lindblom, 1990; Lisker & Abramson, 1970). Stevens (1972) demonstrated that the speech production system is insensitive to any articulatory change when the system produces acoustic signals that fall into the acoustic regions that are perceived as plateau-like. Lisker and Abramson (1970) also reported that participants distinguish voiced vs. voiceless stop consonants based on the length of VOTs when the VOT length changes gradually. This is evidence of abrupt categorical perception. Participants in Liberman et al.’s (1957) study were better at discriminating between two sounds from different phoneme categories when the sounds were near the phoneme boundary, than between two sounds from one phoneme category. Thus, there appears to be a range of oro-sensory areas, in which speech sounds are perceived as identical, and a clear perceptual boundary that is critical for listeners to distinguish one speech sound from another.

Liberman et al. (1957) stated that listeners mainly paid attention to the direction and degree of change of the second formant during the transition period to distinguish the three stop consonants which differed in articulatory positions (/b/, /d/, and /g/). Gandour (1979, 1981) also concluded that Cantonese tones were perceptually distinguished based on information, such as

contour, direction (level, rising, or falling) and height changes of the fundamental frequency. These findings confirm that acoustic information is critical in determining the identity of each speech sound, and that human perception system does not perceive a continuous acoustic signal as continuous, but as ranges with boundaries.

Ramadoss (2012) proposed that the human speech perception system seems to evaluate the direction or magnitude of change, above a certain threshold, while identifying each Thai tone. Ramadoss (2012) collected perceptual judgment data and compared several models to determine the one that explains the data best. During the behavioral experiment, ten native Thai speakers listened to synthesized Thai tones. They made a judgment as to which tone category each tone belonged to and how representative each one sounded (goodness rating). Ramadoss then explored several hypotheses to explain these perceptual judgment results: whether people attend to peaks of the fundamental frequency (f_0) contour in Thai tones, trajectories themselves, or both. She tracked changes in the fundamental frequency (f_0) contour of Thai tones using three different measures: absolute values, velocity values, and Parsons' code values (Parsons, 1975). These measurements tracked the differences between the target (or original) tone trajectory and the synthesized tone trajectory. Parsons' code measurement used three different numbers (-1, 0, +1) to encode the direction and magnitude of change in the fundamental frequency (f_0) values that exceeded a certain threshold between two consecutive time points. The results showed that the largest variance of data was explained when the Parsons' code was used as input to the model, as opposed to velocity or absolute value measurements. This implied that our perceptual system encodes information in a concurrent manner. A judgment is made as to whether incoming information meets the desired direction of change or whether it exceeds a certain threshold. This judgment is only possible when this information is compared to prior information in Parsons' code measurements. In summary,

information about trajectory as well as the direction and magnitude of f_0 change above certain thresholds relative to the prior state are necessary to identify each tone.

If it is true that the perceptual system attends to the direction and magnitude of change instead of the relative timing or force relationships in the f_0 signals across a whole utterance while producing lexical tones, it is likely that the speech production outcome will also contain this type of information. Therefore, the assumption that pre-determined proportional relative values may exist across whole utterance signals will be violated. This violation would then become more apparent when the perceptual and motor systems produce a longer utterance that involves more syllables. Errors that violate rigid proportional relative values across a whole utterance may accumulate as the utterance becomes longer, involving more syllables.

The data from Max and Caruso (1997) supported this expectation when they observed that relative timing relationships were not maintained across different speech rate conditions in both syllable and phrase level utterances. However, the relativeness was maintained in the phrase or word level utternaces but not in the syllable level utterances in Kozhevnikov and Chistovich's (1965) study. Therefore, this inconsistency makes it hard to conclude that concurrent manner of production always produces more variable speech. Speakers may possess information about relative values in the acoustic signals when this information is attended to by the speech production system as a speech target. Furthermore, variances around speech targets might still be perceived as correct speech sounds, when the speech target exists as a range in the acoustic plane as addressed earlier. Overall, although speech motor control systems may attend to a certain direction and magnitude of change in acoustic signals while controlling concurrent speech movements, relative timing or force relationships among the components of an utterance may still appear in the acoustic signals.

Furthermore, the effect of variable sizes of the stored motor program may need more investigation. This is because the concatenation of motor programs may cause more variance in acoustic signals if a movement trajectory is examined at the whole utterance level. If a stored motor program for an utterance of variable syllable lengths exists as a complete unit, the relativeness will be observed for that long utterance. Conversely, if an utterance of variable syllable lengths is not stored as one complete unit, the relativeness will not be observed for that long utterance. This is because the utterance of variable syllable lengths may require concatenation of smaller motor program units and the trajectory will contain more variability due to this concatenation. Thus, the unit size of the stored motor program becomes a critical issue for an utterance of variable syllable lengths.

In particular, if a pre-set movement structure exists, the change of f_0 at a given interval will proportionally expand or compress in a different trial (or context) according to Schema theory. Therefore, the more the f_0 trajectory of one trial fits the proportionally expanded or compressed f_0 trajectories of other trials, the more the f_0 trajectories will generate consistent velocity (or Slope) values along the contour across different trials. Thus, the smaller discrepancy between the velocity measurements that track the two f_0 trajectories will support the GMP notion of Schema theory. In contrast, as explained earlier, the encoding process using Parsons' code requires concurrent judgment to assign -1, 0 or +1 values onto the observed f_0 change. Thus, Parsons' code is expected to capture the concurrent manner of movement control.

Guenther (1995) proposed that four reference frames (phonetic (sounds that speakers wish to produce), acoustic, orosensory (tactile and proprioceptive signals), and motor (articulator movement or muscles controlling the positions of individual articulators)) interact to control the speech production process. Many researchers have proposed sharing information between the

perceptual and motor representations. Turvey (1977) suggested that the perception and action systems seem to share an abstract representation, which drives all sensory perception as well as execution of motor programs. Liberman and Mattingly (1989) also proposed that a phonetic representation are shared by the auditory and motor systems, because the neural loci or biological bases are intimate and are shared by the perception and production systems. Similarly, Studdert-Kennedy (1980) suggested that speech sounds are internally represented; each sound category is perceived from the continuous auditory and articulatory representations using an abstract metric that is common to both auditory and articulatory domains.

However, other researchers proposed that perceptual and motor representations may not share information. Caramazza (1991), Pulvermüller (1996), and Jacquemot, Dupoux, and Bachoud-Lévi (2007) proposed separate auditory and motor representations for speech because patients with conduction aphasia demonstrated relatively intact speech comprehension and production abilities, while the interface between the two systems was damaged. McNeil, Pratt and Fossett (2004) also proposed the separate processes among various levels of the speech production system because impairment at one level of the system did not result in the impairment at another level. Leinen et al. (2015) also reported that, when concurrent visual feedback was present, the movement representation developed in the visual coordinate rather than in the motor coordinate. However, when the performer depended less on concurrent visual feedback, the representation developed in the motor coordinate rather than in the visual sensory coordinate. In addition, motor equivalence, by which the same speech goals are achieved using various articulatory coordinative movements (Abbs, 1986; Gracco & Abbs, 1986; Hughes & Abbs, 1976; Sharkey & Folkins, 1985), provides evidence that invariant speech representation may lie in the auditory coordinates rather than in the motor coordinates. All these suggest that the invariant desired trajectory may be the

information that is contained in both the auditory and motor representations. Additionally, these representations may exist separately.

Thus, it appears that both perceptual and production systems of speech attend to the relative values in the acoustic signals. The relativeness may be crucial information that composes the perceptual representation at first, but may become part of the motor representation as speakers increase their speech production experiences. While relativeness may be important information to both perceptual and production systems, the relativeness may be controlled in an on-going manner. Also, the relativeness at a whole utterance level may be observed more easily if a single complete motor program unit is produced. Thus, the size of the motor program units, which compose an utterance of variable syllable lengths, is important. This study examined this aspect.

1.1.4.5 Only One (Time) Parameter Dimension Examined

Evidence for the existence of a GMP has been inconsistently observed. Two parameter dimensions are discussed that may account for this inconsistency.

Results across the previously mentioned studies failed to find consistent relative values in their acoustic and kinematic data. However, these studies mostly examined the timing dimension of speech production. A movement is a spatial and temporal event and requires simultaneous examination of both parameter dimensions in order to determine the existence of the invariant movement structure. The parameter in schema theory refers to a movement specification or a scaling factor that is used to expand or compress GMP so as to meet the contextual needs. A GMP has been presumed to expand or compress along the time and force parameter dimensions.

There is evidence that two different parameter dimensions are controlled separately. Tuller, Kelso and Harris (1981) reported independently controlled timing and force parameters of movement. According to these researchers, muscles contract differently for changes in speaking

rate and stress. In their study, the duration of genioglossus activity decreased while producing the vowel /i/ as the speaking rate increased, but the peak magnitude of this activity remained unchanged. In contrast, the duration of activity remained unchanged in the lateral pterygoid and the anterior belly of digastric, and the peak magnitude of each muscle's activity increased, while the vowel /a/ was produced and the speaking rate increased. Duration and peak magnitude of these three muscles all increased when a stressed speech sound was produced. The peak magnitude of muscle activation might be related to the force parameter. Thus, these results suggest the potential to control two parameter dimensions separately in a given context.

The two different parameter dimensions that compose movement structure may react differently to knowledge of result (KR) or knowledge of performance (KP). A previous study (Kim et al., n.d.) revealed that the amplitude parameter, which received feedback during the training phase, improved more than the timing parameter, which did not receive feedback. Whether this is true for the opposite case requires more examination. That is, it has not been tested as to whether the time parameter improves when this aspect receives feedback, but the amplitude parameter does not when it does not receive feedback. Two different parameter dimensions that compose the movement trajectory may be controlled separately based on the feedback condition.

Furthermore, although two parameter dimensions may be controlled separately, the interaction of the two may also require consideration. Schmidt and Lee (1999) suggested that the selection of one parameter (e.g., force) affected the selection of the other parameter (e.g., time). Therefore, both parameter dimensions require simultaneous examination while discussing a movement structure.

1.1.4.6 Hierarchical Processes

Another possible explanation for the inconsistent evidence for the GMP is that higher-level processes (semantic, syntactic, and phonological processes, or prosodic processes at whole utterance level) have direct influence on what happens at the lower level processes (phonetic planning) during speech production.

Levelt, Roelofs and Meyer (1999) proposed that speech production processes are composed of four stages: the lexical concept preparation stage, the lemma preparation stage, the morpho-phonological encoding stage, and the phonetic articulatory gesture stage. Spontaneous speech production is initiated by preparing a message at concept preparation stage. At the lexico-syntactic preparation stage, the root forms of words (lemmas) and grammatical markers are prepared. Subsequently, encoding for phonological aspects of each lexical unit and preparation of the phonetic articulatory gestures occur. Preparing a motor program at the phonetic encoding stage would be equivalent to the last stage that prepares articulatory gestures. One thing to consider during this stage is that selection of articulators must be distinguished from the stage that sends commands to muscles. Similarly, van der Merwe (2009) also distinguished the motor planning stage, which specifies articulators, from the motor programming stage, which prepares commands that go to muscles regarding stiffness of the joints, force, direction, distance (or range of motion), velocity, and amount of muscle tension. After this stage, release of these commands to muscles is related to the actual execution of the movement.

According to Levelt (1989), serial process refers to conceptualizing, formulating, and articulating processes that occur strictly serially, one level after the other. On the other hand, it has been hypothesized that message encoding, formulating, and articulating may occur in a parallel manner, but preferred time-relations seem to exist among levels of speech production. The Levelt's

(1989) hypothesis about incremental process combines the serial and parallel processes. Incremental processing refers to a process, in which different components of information are prepared in a parallel manner at all stages, or the order of processing could change (Levelt, 1989).

When there are processing stages, incremental processing of information is possible. Some literature supports the incremental process for speech production (Bohland et al., 2010; Humphreys, Riddoch, & Quinlan, 1988; Kempen & Hoenkamp, 1987; Rapp & Goldrick, 2000). An incremental processing assumes not only a parallel process (Levelt, 1989) but also a bottom-up process. Although the authors did not address it directly, Rogers and Storkel (1998) provide an example of possible bottom-up influence when they report a facilitation in the preparation of the forthcoming articulatory motor programs when the two consecutive speech tasks share the place and manner of articulation. Furthermore, the possible combination of incremental and locally encapsulated serial processes has been proposed for the speech production process, as evidenced by Damian (2003), Meyer (1990), and Roelofs (1999). McNeil, Pratt and Fossett (2004), on the other hand, proposed that degraded performance across different processing stages did not necessarily support the interaction between phonology and motor planning/programming processing stages. Instead, it demonstrated that damage to one level co-occurred with impairment of the other level. According to McNeil, et al. (2004), the interaction across levels of speech production processing is hard to prove.

When it comes to the prosodic structure, similarly to Levelt et al.'s (1999) framework, Pierrehumbert (1980) described a possible hierarchical prosodic structure. The hierarchy of prosodic phonology for English included "the syllable, the foot, the phonological word (also 'prosodic word'), the phonological phrase, the intonational phrase, and the utterance (Gussenhoven, 2004; Hayes, 1989; Nespor & Vogel, 1986; Selkirk, 1978). Later, Beckman and

Pierrehumbert (1986) added the “Intermediate Phrase” level immediately below the intonational phrase to describe English prosodic system. In explaining other languages - such as Korean, Basque and Japanese – “Accentual Phrase” was included immediately above the phonological word level (Pierrehumbert & Beckman, 1988).

How does this hierarchical prosodic structure map onto lexical speech production processes, such as the conceptual, lexico-syntactic, morpho-phonological, and articulatory encoding stages hypothesized by Levelt et al. (1999)? Researchers reported close relationships between the prosodic and non-prosodic units. The prosodic units often correspond to the grouping within an utterance and are important keys to parsing a sentence syntactically (Jun, 2005). It has been proposed that the morpho-phonological module is responsible for computation of the intonation contours, such as the phonological phrase (Van Wijk & Kempen, 1987). These phrases are proposed to frequently correspond to internal pauses within sentences. Similarly, Wijk and Kempen (1985) proposed that the basic intonation patterns (BIP), pitch contours for given linguistic utterances, are computed incrementally as is syntactic structure. Also, the tone-units and tonicity (tonic syllable placement) are highly bonded with syntactic (Crystal, 1979; Duanmu, 1990) or semantic components (Crystal, 1979). Thus, the prosodic units have close relationships with the non-prosodic units in language.

Furthermore, the metrical frame also has been proposed to be critical for phonetic encoding because more integrated phonetic gestures have been found based on metrical frames such as syllabic rhymes or trochaic feet (Browman & Goldstein, 1988; Ziegler, 2013; Ziegler et al., 2011). Browman and Goldstein (1989) suggested that a gesture is a characteristic pattern of the constriction of articulators. This pattern of coordinative constriction of articulators is formed, released, and observed when a given utterance is produced repeatedly. They proposed that lips,

tongue blade, and tongue body, in particular, generate meaningful gestural constrictions for speech. Also, Browman and Goldstein (1989) proposed that the discrete gestures have close relationships with phonological representations. Browman and Goldstein (1992) and Ziegler (2011) proposed that phonetic gestures might be the basic unit of motor planning rather than the discrete phonemes or syllables. Ziegler (2011) suggested that the phonetic gestures might be integrated into a larger motor unit (e. g., metrical foot) and used by the speech motor execution apparatus.

In summary, it appears that linguistic processes are interweaved with the prosodic process. While the question of whether processes are serial or incremental (or cascading) need more exploration, for now, this paper focuses on the fact that the execution of motor programs may be influenced by higher level processes. To reduce any higher-level activation effect on the peripheral-level data, this study did not provide Chinese characters to the speakers.

1.1.4.7 Experience/Practice

Researchers supported the existence of a GMP when they observed the transfer of learned effects to a novel task presumably sharing the same GMP with the trained task. The brief review of the principles of motor learning is included in this section to provide evidence of the existence of a GMP and to justify learning conditions for this study, such as block vs. random practice condition, and the frequency of feedback.

Although the observed inconsistency in the relativeness of movement structure challenges the notion of the GMP, it is supported when GMP errors are reduced during practice sessions and when the trained effect transfers to an untrained task that shares the same GMP. Motor learning occurs when the relationships among several types of information (schemata) develop with practice (Ballard et al., 2007). As explained in the earlier schema theory section, in order to update the schemata, all sources of information are utilized. The experience of a wide range of parameters

accrues and stabilizes the schema (E. Maas, Robin, Wright, & Ballard, 2008). Since the schema enables generation of parameters in novel situations, transfer of learning is assumed to occur in other movements related to the same GMP (E. Maas, Robin, Wright, et al., 2008). GMP learning refers to the acquisition of classes of activities (Schmidt & Lee, 1988; Wulf & Schmidt, 1989), and the learning can be inferred from the retention and/or transfer tests (Knock, Ballard, Robin, & Schmidt, 2000; E. Maas, Robin, Wright, et al., 2008).

A different influence of practice on the development of the GMP and parameters has been suggested (Shea & Wulf, 2005). This suggestion is based on the fact that GMP learning does not affect parameter learning in the same direction. GMPs for complex tasks develop and transfer well when the practice conditions are constant, while variable conditions with random task assignments seem to encourage parameter learning (Lai, Shea, Wulf, & Wright, 2000). A stable GMP was assumed to be pre-requisite for the development of parameter rules (Wulf & Shea, 2002). Thus, early constant practice for GMP learning and later variable condition for parameter learning were suggested as optimal practice conditions (Wulf & Shea, 2002).

Shea, Lai, Wright, Immink, and Black (2001) concluded that the conditions that promote consistency enhance relative timing (GMP) learning, but serial and random practice schedules are more effective for parameter learning, as indexed by absolute timing errors, when examined in transfer tests. When KR was directed to the relative timing, the parameterization was unaffected by the feedback condition, suggesting a dissociation between relative timing (GMP) and parameter errors (Shea et al., 2001).

Other studies explored the interaction between the effect of feedback frequency and practice conditions. The learning of GMP in variable practice conditions benefited from reduced feedback frequency while parameter learning was unaffected by it. As a reason for this, Schmidt

and Bjork (1992) suggested that "frequent feedback makes performance too variable during practice, preventing the learning of a stabilized representation of the kind necessary to sustain performance" (p. 213). High frequency feedback was somewhat detrimental because of the instability caused by the trial-to-trial correction.

With the guidance hypothesis of KR, more frequent KR has been reported to be more useful early in the acquisition stage because this information can guide performers to learn proper patterns (Salmoni, Schmidt, & Walter, 1984). However, a gradual reduction of KR has been recommended in the later acquisition stage because overreliance on KR may interrupt a learner's own processing of information.

GMPs are viewed as abstract representations (Verwey, 1999), independent of the effectors (Park & Shea, 2002; Shea & Wulf, 2005), and it was proposed that the parameters regarding the effectors are specified before the execution of a movement. Van der Merwe (2009) proposed that specification of the articulators occurs during the motor planning stage. Interestingly, it was observed that effector-independence from the GMP decreased with extended practice as the force parameters became more integrated.

As explained in section 1.1.2, the transfer of the treatment or training effect has been explored to determine speech GMP categories. Ballard et al. (2007) proposed that sounds from different voicing or manner groups (fricatives and stops) seem to be governed by different GMPs, whereas phonemes which have different places of articulation, but the same voicing and manners of articulation, are considered as belonging to the same GMP group (E. Maas, Robin, Austermann Hula, et al., 2008). Maas et al. (2008) proposed that the movement command for a lexical stress pattern is stored in the GMP and controlled by it, rather than by the parameter. This is because there is a consistent pattern to produce lexical stresses despite the variation in overall pitch,

loudness, or speech rate (E. Maas, Robin, Austermann Hula, et al., 2008). Lexical tone is similar to lexical stress in that it distinguishes word meaning while it is supra-segmental in nature. However, few studies appear to have explored whether the lexical tone has a corresponding GMP. Almelaifi (2013) investigated the effect of focus of attention among native English speakers while learning to produce Mandarin tones. The stimuli were presented in a random order with 60% feedback. She observed that learning only occurred at a certain tone and under certain instruction conditions, when these non-Mandarin speakers practiced each tone stimulus for 50 trials. Thus, to ensure participants learn to produce all four Mandarin tones accurately, the current study included a blocked design with increased number of practice trials at 60% feedback frequency.

In summary, the acquisition, retention, and transfer of GMP and parameters have been explored. Differences in the rate of learning of the two in the early learning stage has been proposed (Shea et al., 2001). However, as the learning progresses, the parameters of effectors seem to be integrated with the GMP, and the independence between GMP and parameters has been observed to decrease (Park & Shea, 2002; J. H. Park & C. H. Shea, 2003). In addition, the studies about the frequency or type of KR and structure of the practice conditions revealed that the stable conditions seem to promote GMP learning in the early stage of learning, and variable conditions seem to encourage later parameter learning (Lai & Shea, 1998; Lai et al., 2000; Shea et al., 2001; Shea & Wulf, 2005; Wulf & Schmidt, 1989; Wulf & Shea, 2002). Reduced feedback was beneficial to GMP learning because it encourages performers to process information by themselves (Salmoni et al., 1984; Schmidt & Bjork, 1992). Practice is not only an important factor in the development and existence of the GMP and parameters, but it also affects the unit size of the GMP with extended practice. The development of the size of the motor chunk that is developed with practice will be described in the next section 1.1.5.

1.1.5 Processing Unit

Various sizes of motor program units have been proposed and experimentally examined in limb or hand movement tasks. Diedrichsen and Kornysheva (2015) proposed that an action sequence could be chunked and that an intermediate representation in the hierarchy of motor control system permits this chunking. This intermediate representation is similar to the “module” or “primitive” of Internal model and develops after training. This skilled motor representation reduces the selection process and is usually activated during execution. The execution of the skilled motor response is also uninfluenced by the speed of its execution. These representations facilitate the performance of the transfer tasks that share temporal and spatial features and thus allow efficiency as well as flexibility. Diedrichsen and Kornysheva (2015) speculated that these representations reside in the premotor cortex. They explained that the striatum, basal ganglia, pre-SMA, and SMA may be involved in controlling these intermediate representations. This explanation of intermediate representation (aka modules) resembles schema or GMP notions of Schema theory, except for the fact that the GMP notion was proposed to explain discrete and ballistic movement; not a long movement sequence.

However, Young and Schmidt (1990, 1991) suggested the possibility that more than one program unit may be engaged in controlling a sequential movement. The temporal occurrences of select kinematic landmarks were identified from an arm movement trajectory, including the time of positive acceleration onset, peak positive acceleration, the time that the acceleration returns to the zero baseline (or the peak velocity in the backswing direction), the time of maximum backswing amplitude, the time of peak negative (i.e., in the direction of the target) acceleration, the time of peak negative (toward the target) velocity, and the time that the lever crosses the coincidence-point. Within-subject correlations supported the possibility that more than one

program unit may be engaged in controlling a long sequential movement. This suggestion was made because these researchers observed that the first few kinematic landmarks of a sequential movement correlated more highly with each other than they did with later landmarks. In addition, the last landmark correlated more highly with the adjacent earlier landmarks than with landmarks located farther away. This result suggested that more than one GMP might be involved in longer sequential movements (Young & Schmidt, 1990, 1991). After examining seven wrist twist sequences, Shapiro (1977) also reported that the first and second parts of the sequence seemed to be programmed separately.

Not only might more than one GMP be involved in a movement sequence, but these GMPs might also integrate to form a larger motor program unit. Park and Shea (2005) suggested this possibility because they observed that the motor elements were integrated to form a larger motor unit (subsequence) through practice. They asked participants to move a lever to hit sequential targets. The response time was expected to be slower and more variable before a large motor unit (or subsequence) was loaded and executed than before each element within the large unit. They observed significantly slower but not more variable response times before the 3rd and 6th elements in the 10 element response and before the 3rd, 6th, and 11th elements in the 16 element response. Overall, the decreased number of zero crossings after practice and the smooth transition between adjacent elements, as indicated by "enhanced speed… of response execution" (Park & Shea, 2005, p. 14), proposed the development of motor chunks (or subsequences) through practice.

Interestingly, when the participants practiced the 16 element sequence for four days, a few elements (3, 6, 11, 13) among the 16 elements, which had been slower and more variable on Day 1, were no longer significantly slower or more variable in the retention test. This change suggests that the entire movement sequence is integrated after practice. The data supported the possibility

that the elements of the movement sequence are integrated to form a larger motor program, which can be considered as another GMP. Thus, various sizes of motor programs or GMPs may exist.

Park and Shea (2005) also reported that sequence information may become effector dependent through extended practice. The effector, here, was an arm, but it can be an articulator when it comes to speech. As mentioned earlier, according to Schema theory, the selection of an effector belongs to a parameterization process, and a reduced effector independence after extended practice suggests that the independence between GMP and parameters decreased after extensive practice. In other words, association between GMP and parameters became stronger after extensive practice. Therefore, it appears that retrieval and execution of a stored motor program can become more automatic after extensive practice.

Reaction times, intervals among the elements, or the movement time of each element of the sequence have been examined to infer the unit size of motor programs. Verwey (1995, 1996) and Verwey and Dronkert (1996) reported that a motor chunk developed after participants practiced nine key-pressing tasks. In this study, two structured conditions (333 vs 45) were used, in which fixed response-stimulus intervals (RSIs), such as long (750ms) and zero (0ms) intervals, were used among the nine keys. The nine key sequences were composed of either three repetitions of three key subsequences or a combination of four and five key subsequences with fixed RSIs, such as 750-0-0-750-0-0-750-0-0 in the 333 condition or 750-0-0-0-750-0-0-0-0 in the 45 condition. The participants then performed the tasks in an unstructured condition in which all RSIs were 0ms. It was found that the more participants followed the required timing structure during the acquisition phase, the better they maintained this timing structure in the unstructured condition. It was speculated that motor chunks developed during the structured condition.

A forthcoming response is prepared in a concurrent manner while executing the prior response. This possibility has been supported when the intervals within a sequence (or within-group intervals) were longer in the unstructured condition as compared to the structured condition. As a reason for this, limited capacity shared by the preparation process and the execution process has been proposed by Verwey and Dronkert (1996). Wickens (1976, 2002) proposed the possibility of multiple resources. Participants of Hula and McNeil's (2007) study manually responded to a tone identification task along with a picture-naming task with varying stimulus onset asynchronies (SOAs). The RT for the primary task (the tone identification task) increased as the SOAs became shorter. The RT for the primary task also increased when the secondary task (picture naming task) became more difficult by using low frequency words for picture naming tasks instead of the high frequency words. Because experimental conditions encouraged participants to process both tasks at the same time, the evidence was used to support the central processing of both tasks in a parallel manner. When response selection process (or processing of perceptual stimuli) became more demanding after manipulating stimulus-response compatibility, the delays in RT were not influenced by the length of SOAs. Thus, it appeared that the response selection and central translation between input and output information seemed to require separate resources (McCann & Johnston, 1992). Similarly, Verwey (1995) supported this multiple resource model by proposing that response selection and execution seemed to be controlled by separate memory resources without interfering with each other, while unpacking and executing the loaded element shared a single resource.

Verwey, Shea, and Wright (2015) proposed that three processors are involved in movement production: the perceptual, central and motor levels. These three stages are consistent with psychological refractory period (PRP) model by Pashler (1994) and McNeil and Hula (2008).

According to Verwey et al. (2015), the perceptual processor transmits perceptual representations (or stimulus representations) to the central processor: "[T]he central processor uses this stimulus representation to identify the stimulus, and to construct a new movement representation or to select an existing one" (Verwey, Shea, et al., 2015, p. 15). The central processor stores these low level features in the motor buffer and determines the next movement feature from short-term memory. The motor processor executes the contents in the motor buffer and cycles through the motor loop to assess and execute each movement element. The motor processor includes many feedback loops. In particular, Verwey et al. (2015) suggest that buffer loading is followed by three additional processes: buffer search, unpacking, and execution.

Verwey and Dronkert (1996) reported that the group-start interval was longer in the 3 key group as compared to the 4 and 5 key groups in the structured condition, which contradicted the expectation of the complexity effect view (Sternberg, Monsell, Knoll, & Wright, 1978). In addition, the degree of lengthening of the within-group interval in the unstructured condition as compared to the structured condition was greater in the 3 key group than in the 4 or 5 key group. Verwey (1996) also observed that the lengthening of the within-group response time was greater in the 3 key group than in the 6 key group in the unstructured condition. Therefore, it was speculated that when the preparation time is taken away in the unstructured condition, the preparation of the forthcoming response takes place while executing the current response group and the execution slows down. This slowing down occurred more frequently in the 3 key group than in the 4, 5, and 6 key groups. In other words, in the unstructured condition, the execution of the current response group slows down less by the concurrent preparation of the forthcoming response when the forthcoming response group is longer. It appears that when participants produce a longer

movement sequence, they prepare the responses more in a concurrent manner, rather than preparing the whole sequence at once before executing it.

Furthermore, Verwey (1995, 1996) observed the fastest last key in both structured and unstructured conditions. Verwey et al. (2015) observed a shorter inter-stimulus interval before the last key and a shorter response time for the last key in the 6 key sequence. Verwey (1995, 1996) proposed that the last keypress was faster than earlier keys because this key response involved only the unpacking process. This fact was interpreted as evidence of concurrent processing.

Various sizes of motor programs have also been proposed for speech. A few researchers have focused on the "phoneme" as the input unit to the speech production system, which also could be the processing unit of the GMP for speech (Knock et al., 2000; van der Merwe, 1997). However, Cholin et al. (2006) and many other researchers (Crompton, 1982; Levelt, 1989; Levelt, 1992; Levelt & Wheeldon, 1994) suggested that "frequently used syllables" might be the processing unit of motor programs that are stored in "a repository of articulatory-phonetic syllable programs" called the mental syllabary (Levelt et al., 1999).

On the other hand, a few researchers have proposed flexible GMP sizes for speech based on the human capacity to integrate a series of smaller GMPs to form a single, larger GMP through practice (Park & Shea, 2002; J. H. Park & C. H. Shea, 2003; Schmidt & Lee, 2005). Varley et al. (2006) proposed that the GMPs for speech might store motor commands that correspond to linguistic units, such as a phoneme, a syllable, a word, or a phrase. Similarly, the DIVA or GODIVA models proposed that any highly learned motor behaviors will be stored as motor programs and the size of the motor program may vary from phonemes to syllables to multisyllabic words to phrases (Guenther, 2006; Guenther et al., 2006). Furthermore, Browman and Goldstein

(1992) and Ziegler (2013) proposed that the motor program units are phonetic gestures that may vary in size, which are associated with metrical structures rather than the segmental boundaries.

Although the existence of GMPs in various sizes (Park & Shea, 2002; J. H. Park & C. H. Shea, 2003; Schmidt & Lee, 2005) has been hypothesized, this assumption has not been experimentally tested in speech. Therefore, the current study will examine whether the concatenation of the smaller motor program units is observed more in less experienced speakers, whether more experienced speakers will evidence larger motor program units, and whether any evidence of GMPs is observed in an utterance when it is prepared and produced as a whole.

1.1.5.1 Practice Effect on Time Measurements

The degree of practice may determine whether a motor program corresponding to an utterance of multiple syllables is prepared as a whole unit, or whether more than one program is prepared and produced in a parallel or concurrent manner for the same length of utterance. That is, when the motor programs of varying syllable lengths are stored as whole units after practice, the speech production will become faster, and it becomes unnecessary to retrieve smaller motor program units in order to produce the same length of utterance.

It is true that the inter-stimulus interval (ISI) between two adjacent elements in a movement sequence may not reflect the processing load of the forthcoming element nor signal the transition between two subsequences (or motor chunks) (Verwey, 1995). It is because, as Garcia-Colera and Semjen (1987, 1988) pointed out, if enough time is given, the effect of concurrent preparation may not affect the execution rate of the earlier elements. It becomes hard to infer the concurrent programming process from ISIs. Thus, according to this reasoning, it is important to examine the execution rate in “a paradigm with the highest production rates possible” (Verwey, 1995).

However, it also has been proposed that execution of a prior element takes longer when it concurs with the preparation process of the forthcoming element (Verwey, 1995). This is true especially when the performer does not have time to prepare the forthcoming response before initiating movement, or when the performer is new to the movement sequence. In other words, the effect of concurrent execution and preparation processes may not be detected in the ISI or in the execution rate when participants practice the movement sequence. Verwey (1995) suggested that with practice, performers increased the amount of concurrent programming and became more skilled at selecting forthcoming movement elements, resulting in a shorter ISI.

Verwey (1996) speculated that because there was no advance preparation in the unstructured condition, the existence of a motor chunk became important. Verwey (1996) implied that a motor chunk could be executed more automatically and with less effort as compared to when it did not exist. Verwey (1996) also assumed that motor chunks developed when "the motor buffer [was] consistently loaded with the individual elements included in a single response group" (p. 551).

There are several reasons for proposing the possibility of preparing a response in chunks. A complexity effect (or sequence length effect) that demonstrated a longer group-start interval before the longer sequence than before shorter ones provides support. Support is also provided when the response-stimulus intervals (RSIs) in the unstructured condition resemble that of the structured condition as well as when an error distribution pattern is consistent across responses in the structured condition.

With a limited amount of practice, both the reaction time and within-group response times of the sequence are affected by the complexity of the response (or the demands of response selection) (Sternberg et al., 1978; Verwey, 1995). However, with practice, the complexity effect

decreases in the unstructured condition. Verwey (1996), for example, found that the group-start intervals in the unstructured condition, which is influenced by sequence length, were initially short (3ms), but increased in sessions 3 and 5 (80-90ms), and then decreased in session 22 (22ms) as practice progressed. The fact that the complexity effect appeared and decreased after practice in the unstructured condition, but did not appear in the structured condition, supports the development of motor chunks through practice.

Furthermore, there is additional evidence of concatenating smaller motor chunks to form a larger motor chunk through practice. Verwey (1996) observed no relatively long response time in the sequence toward the end of the practice phase. The relatively fast, last keypress in the early practice phase was no longer the fastest response after practice. In addition, the differences in the within-group intervals disappeared with practice. Also, the ratio of group-start and within-group intervals increased with practice in the structured condition, and the ratio increased more gradually in the unstructured condition. The error distribution demonstrated that the errors slightly increased toward the end of the sequence except at the last key; suggesting the earlier elements in the sequence might have been produced as a chunk. Together these factors support the proposed development of motor chunks. This evidence suggests that the motor system may reduce the frequency of the loading process of motor chunks into the motor buffer after practice with resultant larger but fewer motor chunks loaded to produce the same length of utterance (Verwey, 1996).

Verwey (1996) also reported that the within-group response times of the novel movement sequence, as compared to the practiced sequence, were longer for the 6 key group than the 3 key group. The group-start response time decreased with practice in the unstructured condition, but more slowly in the 6 key group, than in the 3 key group. This was interpreted as evidence that the practiced sequence is loaded as a whole in advance, especially with the shorter (three-key)

sequence. Therefore, the response times for the short sequence seems to be affected less by practice or by the existence of motor chunks, because the response is loaded in the buffer in advance. However, a novel sequence, especially when it is long and exceeds the capacity of the motor buffer, needs to be executed in parts, resulting in a relatively longer within-group interval. Verwey (1996) discovered that the sharing of a subsequence of a whole practice sequence with a novel sequence did not expedite the execution process of the novel sequence. Therefore, it has been speculated that motor chunks are content-specific. Practice seems to have only a limited content-aspecific effect on the performance of movement sequences that share an abstract representation (Chamberlin & Magill, 1992; Schmidt, 1975, 1982; Verwey, 1996).

The possibility of motor chunk development and concurrent movement control does not appear to have been experimentally tested for speech movements. It is possible to formulate two speech-related hypotheses from the experimental results of key-pressing tasks. First, when the unstructured condition is examined, in which no specific response stimulus interval (RSI) is assigned and all RSIs are 0ms, less experienced speakers are expected to demonstrate a longer group-start interval (or reaction time, RT) and a longer within-group intervals (or inter-syllable interval, ISI) than more experienced speakers. It is hypothesized that the RT and ISI will become shorter and its execution rate will become faster with practice.

Second, hypothesizing about the ratio between RT and ISIs becomes more complex. When the complexity effect (or sequence length effect) is assumed and the target sequence develops into a response chunk, the ratio will increase (increased RT relative to average ISI time) after sufficient practice as compared to no practice condition. When the complexity effect is not assumed, the ratio may not increase even after practice because people will produce the movement sequence in a concurrent manner. Following Verwey and Dronker's (1996) argument, this trend of preparing

and executing the motor response as a chunk would be predicted to be greater in a shorter response group. This is because motor chunks may develop through practice and will be prepared in advance of the response initiation for a short response chunk. Movements for a longer response will be controlled in a piecemeal manner (Verwey, 1995, 1996; Verwey & Dronkert, 1996).

Third, due to the complexity effect reported by Verwey (1996) and many others who examined reaction times (more details in section 1.1.5.5), a long response sequence may cause longer RT and ISIs than a short response sequence in the early practice phase. As both long and short response sequences become integrated and develop into motor chunks, both RT and ISIs are expected to decrease with practice. The complexity effect may also decrease.

1.1.5.2 Sensory Feedback

Because the sensory feedback influences the control of the motor response, what becomes important is the size of the motor respose that is executed without the influence of sensory feedback after a motor program releases motor commands to muscles. When people detect a discrepancy between expected sensory feedback and reafferented sensory information, it influences the parameterization process of the current motor program that is under execution, or it influences the retrieval process of the next motor program. The degree of reliance on the auditory feedback may be determined by the production experience as well. Segawa, Tourville, Beal and Guenther (2015) stated that learning a speech movement sequence involves not only the brain area that is responsible for learning a motor sequence, but also the areas that are related to feedback-based speech motor learning. Thus, motor learning seems to involve the development of structural connectivity between the motor and sensory areas in the brain.

Auditory feedback is important for concurrent control of speech movement (Purcell & Munhall, 2006b; Xu, Larson, Bauer, & Hain, 2004). Although some researchers (Keele, 1968;

Russell, 1976) proposed that the motor program is sequentially controlled without access to sensory feedback, Sternberg et al. (1978) suggested that feedback information may influence the onset of the next unit in the sequence. Internal model explains that a rapid use of sensory feedback is possible owing to efference copy (Wolpert et al., 1995; Wolpert et al., 1998). According to this model, the feedforward speech controller monitors incoming acoustic signals (Tourville et al., 2008) and uses auditory feedback to update stored speech motor schemata (Purcell & Munhall, 2006a; Villacorta et al., 2007). The role of auditory feedback for on-going speech movement control also has been proposed. Yates (1963) observed that the speech fluency of well-practiced speakers deteriorated when auditory feedback was delayed by 200 ms in the DAF paradigm.

However, even with the efference copy, the auditory feedback appears to be slow. Kawahara (1993) reported that participants demonstrated corrective responses with about 100-200ms latency after a f_0 perturbation. Tourville et al. (2008) also observed a compensatory adjustment in the F1 value of the /Ɛ/ vowel with about 108-165ms latency when perturbed. Cai et al. (2011) also reported about 150-160ms latency before making any compensatory response to up and down F2 perturbation. This lag was already predicted by Schmidt (1975) because he proposed that running the initial portion of a motor program is necessary in order for the motor system to perceive it and to make any corrective change (Schmidt & Russell, 1972). Thus, some lags seem unavoidable before any auditory feedback can affect concurrent speech motor control.

As addressed in section 1.1.4.4, many auditory and somatosensory perturbation studies reveal that the speech motor control system makes a quick adjustment to motor programs to achieve the same speech sound targets (Abbs & Gracco, 1984; Cai et al., 2010; Golfinopoulos et al., 2011; Ito, Kimura, & Gomi, 2005; Parkinson, Korzyukov, Larson, Litvak, & Robin, 2013; Purcell & Munhall, 2006a; Shaiman, 1989; Shaiman & Gracco, 2002; Shiller et al., 2009; Tourville

et al., 2008; Villacorta et al., 2007; Xu et al., 2004). The motor control system may be subordinate to the acoustic targets, and the compensatory responses to both auditory and somato-sensory perturbation may be made to maintain the relative differences in the acoustic data. However, maintenance of the relative relationships in the acoustic outcome does not guarantee the involvement of the same GMP nor the maintenance of the relative relationships in the kinematic outcomes. The question of whether the compensatory response involves only parameter change, changes in the GMPs, or both, needs more exploration. However, this question is out of scope for this study. The current study focused on the fact that the motor system brings a change to the current GMP and/or its parameters as well as to the retrieval of a GMP for the subsequent movement based on the auditory and somato-sensory feedback.

However, this kind of adjustment to the articulatory movement control may not occur when speakers do not rely heavily on auditory feedback. This was the case for well-practiced speakers in Ning and colleagues' (2015; 2014) studies. They reported that native Mandarin speakers did not demonstrate as great a compensatory response to altered auditory feedback as non-native speakers. Similarly, Perkell (2000) proposed that the production of segmental speech characteristics in the mature system does not require auditory feedback. It only uses auditory feedback to control for the parameters that influences supra-segmental aspects, such as "the average sound level, speaking rate, the degree of prosodically based f_0 and SPL inflection" (Perkell et al., 2000, p. p. 239). Additionally, Perkell mentioned that adults who became deaf still maintained intelligibility to some extent, suggesting the auditory feedback is not essential for mature speech production system. Thus, it is expected that stored motor programs may be executed without any interruption of the auditory feedback when they are highly practiced.

Instead, adult speech production system may rely more on somato-sensory feedback than auditory feedback. This is because the mature speech production system finds a way to circumvent the auditory feedback (Perkell et al., 2000) and because proprioceptive feedback is faster than auditory feedback. This information is more readily available for use. Hickok (2012) asserted different roles between auditory and somato-sensory feedback and suggested that somato-sensory information is used more than auditory to control vocal track configuration, whereas auditory information can be used to monitor whether acoustic output meets the linguistic targets. Interestingly, Feng (2008) argued that the speech production system prioritizes auditory feedback information as compared to somato-sensory feedback when controlling for vowel movement and when two types of sensory information need to be processed simultaneously.

As emphasized earlier, the size of a GMP unit becomes an important issue because it may affect the number of segements that will be executed without the use of sensory feedback. Leinen et al. (2015) reported that performers relied on concurrent visual sensory feedback when it was available. However, if this sensory feedback was not available, the performers developed representations in their motor system. After these representations have developed in the motor system, the motor control system does not rely on concurrent visual sensory feedback to complete the same movement. The same translation of acoustic representation to articulatory representation has been suggested in the DIVA model (Guenther, 1995; Perkell et al., 2000). The auditory feedback may be necessary when learning to produce a speech motor sequence. Once the motor system develops motor sequence representations, it may no longer require concurrent auditory feedback.

In summary, while auditory and somatosensory feedback may cause changes in GMPs, the magnitude of influence by auditory feedback may depend on the size of motor programs and how

strong the motor representations are. It is possible that speakers with more speech production experience may utilize a larger motor program unit and depend less on auditory feedback. Therefore, the GMP information may be preserved and observed more in experienced speakers than in novice speakers. Speakers with less experience may use a smaller motor program unit and depend more on sensory feedback. More variable movement trajectories are expected in less experienced speakers.

1.1.5.3 Individual Differences

Verwey (1996) reported that not all participants in his study seemed to develop the same strategy to reach high performance levels. This was because some participants did not demonstrate the same timing structure as other participants develop motor chunks under the structured conditions. Some participants seemed to have partitioned the unstructured sequences in different ways. Verwey (2015) also reported individual differences in the way participants concatenated motor programs. Thus, individual strategic differences are a potential sources of variability in the speech movement structure.

1.1.5.4 The Dual-Route Model

The concepts of two production modes (direct and indirect routes) are introduced in this section to hypothesize different production modes that experienced speakers and novice speakers may implement.

Varley, Whiteside, and Luff (1999b) and Varley et al. (2006) proposed a dual-route model to explain speech errors produced by patients with AOS. They proposed that speech is controlled by using both a “direct-route” and an “indirect-route.” The direct-route stores phonetic plans for frequently practiced speech sequences and is involved in producing speech sequences in an

automatic manner. In contrast, the indirect-route (or indirect encoding) requires segment-by-segment assembly to produce novel or low frequency forms because these forms are not stored as fully represented programs. This argument is predicated on the finding that the response latency was shorter for high frequency syllables than for low frequency syllables (Levelt & Wheeldon, 1994).

Croot (2001) pointed out that stored phonetic plans enable motor systems to have reduced degrees of freedom and coordinative movements. When a speaker loses this ability, the variability increases as in speech produced by patients with AOS because speakers rely on the indirect route. Whiteside and Varley (1998) also reported variability, altered durational patterns, and limited anticipatory coarticulation in the patients with AOS. Furthermore, Ballard et al. (2007) suggested that patients with AOS tend to produce speech in smaller segments, suggesting use of the indirect route. After reviewing existing behavioral, computational and neuroimaging studies on patients with AOS, Ballard et al. (2014) also concluded that AOS is related to the impairment in the feedforward (or direct route) system and that the feedback (or indirect route) system of this group is relatively intact.

This dual-route theory by Varley et al. (1999b; 2006) has been questioned when evidence of double dissociation has not been found (e. g., Ballard, Barlow, & Robin, 2001; Croot, 2001). Several studies have tried to find evidence of double dissociation but have not been successful. To provide convincing evidence for Varley and Whiteside's theory, the evidence for double dissociation is necessary.

As examples demonstrating impaired direct route and intact indirect route, patients with AOS have been studied. Varley and Whiteside (1999a) asserted that the patients with AOS may utilize the indirect route to produce lexical items because the direct route of these patients is

possibly impaired. However, they later admitted that the brain area that is responsible for the indirect route may lie adjacent to the area for the direct route, and lesions in one route may also affect the other route (Varley & Whiteside, 2001b). Thus, it will be rare to see individuals with AOS who have intact low frequency words, which require indirect route to produce (Varley, Whiteside, Hammill, & Cooper, 2006). This argument was supported when the patients with AOS produced a few intact high frequency words (Varley, Whiteside, Windsor, et al., 2006) and abnormally sounding low frequency words (Miller, 2001). Thus, the existing evidence suggests that patients with AOS may have impairments in both direct and indirect routes. The double dissociation was not supported.

Furthermore, the opposite possibility of intact direct routes and impaired indirect routes was examined to satisfy double dissociation. In this case, patients can produce high frequency words using intact direct routes, but they may show difficulty producing unfamiliar words, non-words, or very-low frequency words due to impaired indirect routes (Ballard et al., 2001; Croot, 2001). Ballard et al. (2001) proposed that patients would need to be reported more in the literature who demonstrate difficulty in repeating or reading aloud unfamiliar words or non-words (Ellis & Young, 2013). However, even if these incidences were reported, still it would be unclear whether these difficulties are caused by impairment in indirect route, assembling segments, or in higher level linguistic processes (Croot, 2001). It is also not clear whether the difficulties are at the phonological level or at the phonetic-motoric level (Ballard et al., 2001). With all these possibilities, evidence for double dissociation was not sufficient to support the existence of direct and indirect routes.

In addition, Rogers and Spencer (2001) questioned the existence of the direct route. They provided evidence that whole syllable substitution or transposition errors did not occur frequently

in normal speakers. This supports the assembly model using sub-syllabic units. Rogers and Storkel (1999) demonstrated slowed response latency before the second word when two successive mono-syllabic words are phonologically similar due to neuronal inhibition after activation. If speakers produce speech as a whole, it becomes hard to explain this phonological similarity effect. Moser et al. (2009) also supported the possible existence of indirect route by proposing possible neural correlates for the indirect route, such as the left inferior frontal gyrus and the left anterior insula. These areas were more active while novel speech was produced. Thus, these results supported the assembly model and questioned the existence of direct route. Overall, the evidence of double dissociation is limited (Croot, 2001). Miller (2001) proposed a possible continuum between extremely direct encoding and extremely indirect encoding mechanisms.

After all these subsequent arguments, Varley and Whiteside (2001a) still suggested the value of the theory and said, "[t]he dual-route hypothesis may allow a unified account of both acquisition of speech control and the operation of the mature system" (p. 84). The dual-route model still has implication in terms of speech motor learning. The existence of two routes is supported when Varley et al. (2006) proposed that an indirect route may assemble speech sounds at sub-lexical units and phonetic plans may be stored in varying sizes once the utterance is produced frequently (Varley & Whiteside, 2001b; Varley, Whiteside, Windsor, et al., 2006). Thus, although evidence is limited for dual-route speech control model, this hypothesis deserves empirical tests (Varley, Whiteside, Windsor, et al., 2006).

Clearly, the existence of separate indirect-routes and direct-routes requires more examination. Despite this lack of evidence, this study borrowed the assumptions about dual-routes to explain two types of production mode: execution of stored large motor program units through the direct-route vs. concatenation of small motor program units through the indirect-route.

Experienced speakers were expected to use both direct and indirect routes, while less-experienced speakers were expected to use mostly the indirect route.

1.1.5.5 Reaction Time

This section reviews reaction time (RT) studies because the RT and inter-response intervals inform researchers of which production mode is in use and what the size of each motor program unit is. In a simple reaction time (RT) paradigm, the target response is identified and programmed in advance of an imperative signal (or "Go" signal). The simple RT refers to the interval between the "Go" signal and the motor response (Klapp, 1995). Since participants can program the response before the "Go" signal, the RT interval has been presumed not to include the programming process (Klapp, 1995).

In the choice RT paradigm, one of the possible alternative target responses is given simultaneously with the imperative signal (Klapp, 1995). Thus, the response selection and programming are made only after the "Go" signal (Klapp, 1995). Because participants cannot program the response before the "Go" signal, the programming time is included in the choice RT. The choice RT tends to be longer than the simple RT in general (Garcia-Colera & Semjen, 1987).

Klapp (1995) suggested that there might be two types of programming: INT and SEQ processes. During the INT process, the motor system composes the internal structure of a motor response. During the SEQ process, motor responses are organized into a sequence (or order). Maas (2006) elaborated these processes by explaining that the INT process is related to forming internal spatiotemporal structures of movement units and loading them into a motor buffer, and the SEQ process sequentially places the movement units in a correct serial order (E. Maas, 2006). Researchers proposed that, in a simple RT paradigm, INT process occurs between the presentation

of the target response and the “Go” signal, and the SEQ process takes place during the RT interval between the “Go” signal and the motor response. On the other hand, during the choice RT paradigm, both the INT and SEQ processes occur during the RT interval. However, if the motor response of the target response is prepared partially in advance of its execution, the remaining INT and SEQ process may occur on-line as this partial motor response is under execution.

In the simple RT paradigm, the target response appears as a precue and a performer executes the response as soon as the “Go” signal appears on the screen. It is presumed that the internal structure of the motor response (INT process) is formed after the precue and before the “Go” signal. On the other hand, the SEQ process is presumed to occur only after the “GO” signal. The time after the “GO” signal is considered as the reaction time (RT). Because the RT latency in the simple RT paradigm increases as the number of chunks increases in the response group, it is speculated that the SEQ process is influenced by the number of chunks in the response group. In the simple RT paradigm, Sternberg et al. (1978) observed that the mean latency of RT increased linearly as the number of syllables per word, the number of words in the sequence, or the number of letters in the typing sequence increased. The authors speculated that the effects of increases in number of syllables per word or words per list were additive. Furthermore, the simple RT seems uninfluenced by the memory load because the latency before the repetition task increased in a similar manner to that of the random speech sequence task. Sternberg et al. (1978) suggested that this lengthening of RT is caused by the unpacking of the supra-ordinate motor program, which seems to happen during the RT and is based on a self-terminating search process.

However, this buffer searching idea was questioned when Garcia-Colera and Semjen (1987) observed a non-linear increase in the inter-tap intervals among eight finger taps in the simple RT paradigm. Also, the existence of a buffer search to find an appropriate sub-program was

unlikely because the response requires repetitive tapping of the same key. Thus, Garcia-Colera and Semjen (1987) speculated that programming of the sequence in advance of the "Go" signal appeared to occur as a whole in the simple RT paradigm.

Garcia-Colera and Semjen (1987) also observed no effect of sequence length on the choice RT. Klapp (1995) proposed that it is because the SEQ and INT processes may operate in a parallel manner during the choice RT and the INT process takes longer than the SEQ process. Thus, the sequence length effect, which affects the SEQ process, may not appear in the choice RT. Rather, it has been proposed that the choice RT is influenced by the response duration (longer choice RT before a long press Morse code "dah," than before a short press Morse code "dit") (Klapp, 1995) or by the response complexity. However, what determines the response complexity has been controversial.

The notion of "chunk" in Klapp's (2003) studies seems to correspond to the notion of "unit" in Sternberg et al.'s (1978) study because both of them are composed of smaller constituents (or elements). For example, a word is a "chunk" or "unit," and a syllable is an "element" or "constituent" of the word unit. Researchers observed that the simple RT increased as the number of elements in the response group increased (Garcia-Colera & Semjen, 1987; Klapp, 1995; Klapp et al., 1979; Sternberg et al., 1978), while the choice RT did not. However, Klapp (2003) observed that choice RT increased as the number of syllables in the response group increased. The simple RT did not increase as a function of number of syllables. Thus, Klapp (2003) concluded that the number of syllables may determine the complexity of the motor response or INT process, while the number of words (or chunks) may influence the SEQ process. Thus, Klapp proposed that the choice RT may increase as the number of "syllables" increases, and the simple RT may increase as the number of "words" increases.

In contrast, Wright et al. (2009) observed that the number of syllables affected only the SEQ process rather than the INT process in the self-selection paradigm. In the self-selection paradigm, participants have study time (ST) to organize the required speech response in advance of the "Go" signal, and they indicate their readiness to respond by pressing a button. Then, a "Go" signal appears after a random interval. The reaction time (RT) is measured as the interval between the time of the "GO" signal and the response onset. Wright et al. (2009) presumed that the INT process would occur during the ST and that the SEQ process would take place during the RT. They reported that the ST increased as a function of syllable complexity (e.g., /ta/ < /stra/)) and sequence complexity (e.g., /ta-ta-ta-ta/ or /stra - stra - stra - stra/ < /ta - ru - stra - ta/ or /ta – stra - ru - ta/). However, the RT increased as a function of number of syllables, demonstrating a longer RT before a longer sequence than before a shorter sequence.

The increase in the number of syllables affected the SEQ process in Wright et al.'s (2009) study, but increase in the number of syllables corresponded to the increase of the complexity effect in Klapp et al.'s (2003) study. Therefore, it is hard to predict how increasing the number of syllables will affect the INT process and the choice RT. If the increase in the number of syllables affects response complexity, it will cause a longer INT process and longer choice RT. However, if it does not, the INT time would not increase as a function of the number of syllables.

Klapp et al. (1979) proposed that a response is programmed in advance of the "Go" signal in the simple RT paradigm. On the other hand, the response can be programmed on a segment-by-segment basis in the choice RT paradigm, and only the initial segment may be programmed before initiating movement during the RT interval. However, Klapp et al. (1995) were aware that this assumption of two different production strategies in the two different RT paradigms (simple and choice) is arbitrary. This is because the interval between the first and second elements in the choice

RT was not longer than that in the simple RT paradigm. If programming of the initial segment was sufficient to initate the movement, the programming of the next segment should lengthen the interval between the first and the second segments in the response group. However, such lengthening of the interval was not observed. Therefore, Klapp et al. (1995) questioned whether production mode (or strategy) changes in response to the RT paradigm used.

Verwey (1995, 1996) proposed that although the motor system has the ability to prepare and execute responses in a concurrent manner, the need to prepare the forthcoming response as the system executes the current response results in lengthened intervals. However, this lengthening may not be observed when more than two elements in the response group are prepared together as a chunk. This possibility may explain why the intervals between the first and second elements in the choice RT were not longer than those in the simple RT paradigm in the study by Klapp et al. (1995).

One question that remains is why the choice RT increases with an increase in the number of syllables in the response group when the performers can initiate a response after the first syllable is prepared in the choice RT paradigm. If the sequence length effect (or complexity effect) occurs in the choice RT, this may be because participants need to prepare some other aspects of a whole response before initiating execution of the first element of the response group or because they choose to prepare the response group as a whole instead of the first element. This also explains why there have been inconsistent reports regarding the length of the choice RT as the response length (or response complexity) increases. The effect of the unit size of the motor chunk and of the production mode (concurrent vs. preparing in advance) requires more examination.

Furthermore, although it has been presumed that all responses are prepared in advance of initiation in the simple RT paradigm, Maslovat et al. (2014) proposed that this might not be true

in every instance. Although the simple RT should be affected by the number of items in the response group, Maslovat et al. (2014) did not observe longer simple RT in all of the three key-pressing conditions as compared to one key-pressing condition. During the post-analysis, participants were divided into two groups based on the length of the simple RT. The group that showed the longer simple RT demonstrated shorter inter-key intervals. The other group that showed a shorter simple RT demonstrated longer inter-key intervals. Thus, the authors concluded that a subset of participants appeared to prepare the response in advance of the "Go" signal, and the rest relied more on concurrent programming as they executed prior motor responses. They speculated that a latter group chose the concurrent production mode because there was a long interval after the initial segment of the three key sequence (e.g., "dit__dit_dit" key-pressing task). The performers who chose a concurrent manner of production used that long interval to prepare the next response.

Therefore, presumably when performers have enough time to prepare the forthcoming response in advance of initiation of the movement, they spent less time on execution. When performers do not spend the time to prepare the motor response in advance of the initiation, they spend more time to prepare the forthcoming response as they execute the current response. These results also suggest that there is a time cost when preparation and execution take place simultaneously. Thus, in the case when people choose to prepare the later motor response as they execute the initial part, the effect of sequence length may not appear in the simple RT paradigm as has been presumed.

These inconsistent results suggest that the RTs, inter-element intervals, and execution rates might be determined by the location of the timing measurements analyzed, such as whether they

are measured in the middle of the motor chunk or before it. The duration of inter-element intervals might also be determined by the need to prepare the forthcoming response.

On the other hand, Maslovat et al. (2014) observed that the choice RT did not differ with the number of key presses in the sequence or the timing complexity (isochronous vs. non-isochronous timing structure). Rather, the choice RT was longer than the simple RT in general, and the inter-key intervals were longer in the choice RT paradigm than in the simple RT paradigm. Because advance response programming is generally not possible in the choice RT paradigm, preparation of the motor response might have occurred simultaneously with execution in a concurrent manner in the choice RT paradigm (Maslovat et al., 2014). Spencer and Rogers (2005) and Reilly and Spencer (2013) observed that choice RT and ISIs increased as the number of syllables in the sequence increased while speakers produced the 1 to 5 syllable sequences. However, advance preparation was encouraged in this case because participants were asked to produce each sequence as a whole. Thus, naturally existing subsequences (or motor chunks) in the sequence were not examined. The size of motor program unit and the effect of concurrent programming on RT and ISIs require more investigation.

In summary, the RT and ISI have been used to make inferences about production mode. The more that participants prepare the motor response as a chunk in advance of execution, the longer the RT becomes. However, the more a response is controlled in a concurrent manner, the shorter the RT and the longer the inter-stimulus intervals become. The RT studies also suggest individual differences in the production mode regardless of the RT paradigms used.

Practice Effect

Klapp (1995) speculated that with training, the four chunks were integrated into one chunk as evidenced by reduced simple RT (Klapp, 1995). The simple RT increased as the number of the

syllable in the sequence increased from one to four syllables (Klapp, 1995; Klapp et al., 1979; Sternberg et al., 1978). When asked to produce each sequence as a whole, the choice RT and ISIs increased as the number of syllable in the sequence increased, but both choice RT and ISIs decreased after practice (Reilly & Spencer, 2013; Spencer & Rogers, 2005). It is reasoned that the four-elements in a task are treated as four chunks at first but as a single chunk after practice. Thus, the sequence length effect on the simple and choice RT decreases with practice.

Even with practice, it is expected that the choice RT is generally longer than the simple RT (Klapp, 1995; Maslovat et al., 2014). Additionally, although choice RT decreases with practice, it is speculated that choice RT for the four-element response remains longer than for the one-element response after practice (Klapp, 1995). This is because the four-element response is still more complex than the one-element response.

The above studies observed that practice causes a reduction in simple RT because motor chunks develop with practice. Thus, the effect of the unit size of the motor chunk needs additional examination. Verwey (1999) speculated that the decrease of the sequence length effect on the RT and the execution rate after practice in the simple RT paradigm is evidence of the development of motor chunks. He proposed that the benefit of using motor chunks exists in the faster buffer loading. Also, the concurrent production has been suggested by the prolonged intervals or slowed execution rates.

Additionally, the motor response may start after the whole response is prepared, instead of the first segment of the motor response, even in the choice RT paradigm. This choice of production mode may be determined by the situation (e. g., familiarity with the task) and individual preferences. Thus, the effect of production mode on choice RT and response intervals also requires additional investigation.

Verwey (1996) mentioned that no advance preparation is permitted in the unstructured condition and the existence of a motor chunk becomes important in this condition. The unstructured condition is similar to the choice RT condition. Thus, the choice RT paradigm seems appropriate to examine naturally existing motor program units and the sizes of those units.

1.2 MANDARIN LEXICAL TONES

A tone language is "a language having lexically significant, contrastive, but relative pitch on each syllable" (Pike, 1948, p. 3). Lexical tones are speculated to be similar to lexical stresses. Chen et al. (2002) suggested that lexical tone operates like a metrical frame as lexical stress does, and lexical tone combines with syllable information. It is presumed that lexical tones are prepared during the phonological encoding process. However, the phonetic encoding aspect of lexical tone has not been sufficiently explored in the literature. It is not clear whether a motor program or a GMP for a lexical tone exists.

Speakers may need to learn how to produce each lexical tone, and thus, abstract movement patterns for lexical tones may be stored in the GMPs. Lexical tone has a similarity to lexical stress because both have prosodic properties and are used to distinguish meanings. Both lexical stress and tone have acoustic patterns that are consistently maintained across different contexts (e. g., speaking rate, intensity, etc.). The existence of GMPs for lexical stress has been supported by Kim, Shaiman, and McNeil (n.d.) when they observed the development of GMPs for lexical stress patterns, as evidenced by the reduced GMP errors after practice. If GMPs for lexical stress patterns exist, it is possible that GMPs for lexical tones may also exist and develop with practice. Ning and colleagues (2015; 2014) reported that native Mandarin speakers demonstrated less compensatory

response to auditory perturbation than non-native speakers. Also, the least amount of compensatory response was observed in the level (first learned and presumeably the easiest) tone than in any other tones. This difference between groups appeared even in the easiest tone (level tone), and it might come from different degrees of experience controlling Mandarin tones. The stored motor programs for lexical tones seem robust, particularly when production of those tones becomes automatic after learning. These programs enable native speakers to rely less on auditory feedback. However, we still don't know whether GMPs exist for the lexical tones rather than just motor programs. Thus, it was the focus of this study.

It has been suggested that, in Mandarin, one tone is distinguished from another by the heights and contours of the fundamental frequency (f_0) (Liu, 1924). In particular, speakers and listeners appear to attend to the relative changes rather than absolute changes in the f_0 values to perceptually distinguish lexical tones and to correctly produce them (S. H. Chen, Liu, & Xu, 2007; Duanmu, 1990; Gandour, 1979, 1981; Liu & Xu, 2005). In order to do this, listeners use the f_0 level at the onset of speech as a reference f_0 level and then compute the f_0 range (Duanmu, 1990). Chan (1974) noted that when tones are strung together, the f_0 contour shapes of the tones are maintained in a continuous speech as compared to those in isolated tones, so the frequency height and range, relative frequency level, and contour shape appear to help identify the tones.

The relativeness in the acoustic measurements is used to identify Mandarin tones. Because the relative values of the acoustic signals can be used to infer existence of GMPs for speech, the importance of relativeness in lexical tones make them desirable speech targets above all other speech tasks when studying for GMPs.

The auditory information will be available even when people produce a single Mandarin syllable. This evidence was observed in the auditory perturbation studies, in which the acoustic

signals of the auditory feedback were manipulated and changed, and research participants made rapid adjustments in response to these perturbations. The reported syllable duration for Mandarin is as follows: 270ms for average CV production ranging 214~349ms (Xu, 1997) and 180ms for the average non-focused syllables, such as /mao, mi, mo, na, mai, tao/ with many tone combinations (Xu, 1999). Also, researchers reported that the compensatory response to an auditory perturbation began between 100ms and 200ms after the onset of perturbation; the median response latency was 143ms and the mean was 164ms (Xu et al., 2004). This latency was within CVC syllable duration. Thus, these results suggested that while the motor program for a syllable is being executed, the auditory feedback is available for use by the speakers. In other words, the speakers can monitor their own productions on-line and make an adjustment to their motor plan/program processes if necessary.

Native Mandarin speakers, unlike non-native speakers, are expected to have stored motor programs for each lexical tone. They are also expected to be skillful in concatenating tones to form tone sequences. Auditory feedback may contribute to speech motor control while producing lexical tone sequences that are longer than one syllable. When more than one-syllable tone sequences are produced, auditory feedback will naturally be available after a short latency. However, because of abundant production experience, the native speakers will utilize stored motor programs to produce lexical tone sequences and rely less on auditory feedback than the non-native speakers. This, in turn, will cause less variability of the f_0 trajectories for the native Mandarin speakers' productions.

1.3 RESEARCH QUESTION

1.3.1 Background Summary and Research Questions

Schmidt's (1975) Schema theory has proposed that there is a Generalized Motor Program (GMP), which contains invariant movement structure (kinematic or acoustic) in which information about order of events, relative timing, and relative force relationships in the movement representations are stored. The GMP is a pre-set motor command in which information is expanded or compressed as the context in which the GMP is produced changes. Therefore, before GMPs are executed, it has been proposed that the response specifications regarding absolute timing, absolute force, and effectors have to be determined through a parameterization as a part of the motor planning process. It is assumed that a proportional relationship of the sub-segment compared to the total-movement structure of one GMP remains the same while the context changes.

Yet, many studies of speech production have failed to observe consistent proportional relationships in the kinematic or acoustic speech data (Kozhevnikov & Chistovich, 1965; Löfqvist, 1991; Max & Caruso, 1997; Smith et al., 1995). To explain why they could not, the current study focused on the fact that speech motor programs can be controlled concurrently. The variability in movement trajectories in previous studies may have been caused by utterances that contained one or more GMPs (Young & Schmidt, 1990, 1991) or multiple parameterization processes (Fujimura, 1987).

Also, what GMPs are retrieved, executed, and concatenated may be determined by higher level processing, such as semantic, syntactic, phonologic, or prosodic planning processes, or by the individual differences in the manner of concatenating the GMPs (Verwey, Shea, et al., 2015). Also, the speech motor programs may be monitored and corrected based on the auditory and

somato-sensory feedback. All these with sensory and higher level processes may affect the movement trajectory for an utterance of multiple syllables and prevent this trajectory from demonstrating proportional timing or force relationships. However, the effect of higher level processes was eliminated by using the same language across speaker groups and by manipulating the target speech motor responses. Also, the degree of effect of auditory feedback was considered related to the degree of practice experience in the current study.

On the other hand, Schmidt (1975) has proposed that after several attempts at the same kind of movement, the relationship of four sources of information (the initial states of the muscular system and of the environment preceding the movement, the specified parameters, the outcome of the movement, and the sensory consequences of these motor commands) leads to the construction of an abstract form of schema and to the development of a GMP. Wulf and Schmidt (1989), Lai et al. (2000), Shea et al. (2001), Wulf and Shea (2002), and many other researchers have corroborated this idea of the existence and development of GMPs by providing evidence of decreased GMP and parameter measurement errors in the movement trajectory after practice. This idea was also supported when the trained effect of one GMP transferred to novel tasks that shared the same proportional timing or force relationships with the trained tasks.

Additionally, previous studies suggested that various sizes of GMPs seem to exist (Guenther, 2006; Guenther et al., 2006; E. Maas, Robin, Austermann Hula, et al., 2008; Schmidt & Lee, 2005) and performers seem to develop larger GMPs (or motor chunks) after practice (Park & Shea, 2002; J. H. Park & C. H. Shea, 2003; Park & Shea, 2005; Verwey, 1995, 1996, 1999; Verwey & Dronkert, 1996). Thus, it appears that variation in the movement structure may be determined by the size of the motor program and amount of practice with the motor program.

Varley, Whiteside, and Luff (1999b) and Varley et al. (2006) proposed two different routes to access motor representations, a direct route and an indirect route. The direct-route accesses phonetic plans for frequently practiced speech sequences and produces them in an automatic manner. The indirect-route accesses smaller segments that have to be assembled to produce a larger speech unit. This dual-route theory has been challenged because a double dissociation has not been observed (e. g.,Ballard et al., 2001; Croot, 2001). However, the dual-route theory has been later proposed as a continuum of controlling and producing movements. Miller (2001) proposed a continuum of the automaticity that have two extremely different encoding ends (direct vs. indirect encodings). The theory has been useful in explaining acquisition of motor programs and the operation of the mature motor control system (Varley & Whiteside, 2001a; Varley et al., 1999a; Varley, Whiteside, Windsor, et al., 2006). From this continuum, it was hypothesized that the degree of language experience would affect the size of the GMPs and the manner in which the stored motor programs would be utilized. Speakers with more speech production experience was expected to employ a larger motor program unit in an automatized manner and produce it at once without interacting with sensory feedback. This may guarantee observation of the relative timing and force relationships in the speech movement outcomes. In contrast, speakers with less production experience may need to concatenate smaller motor program units in a concurrent manner.

The current study was motivated by studies that examined key-pressing tasks and reported a change in the production mode from preparing the motor response in advance to preparing the motor response in a concurrent manner when the motor response becomes longer (Verwey, 1995, 1996; Verwey & Dronkert, 1996). The effect of sequence length on the production mode was investigated in this study.

Although the existence of GMPs in various sizes has been hypothesized based on key pressing tasks or limb movement studies, this assumption has not been experimentally tested for speech. It is expected that the acoustic signals obtained from speech motor chunks will contain relative timing or force relationships. Therefore, the current study examined the existence of different sizes of the motor program for speech and whether the experience with one language would affect the size of the motor program or the manner by which stored motor programs are produced (producing a larger unit all at once vs. concatenating smaller units). Therefore, this study recruited native Mandarin speakers and non-Mandarin speakers in order to determine whether the native Mandarin speakers tend to produce larger motor chunks of tone sequences with more preserved relative timing or force relationships than the non-Mandarin speakers.

Because previous literature examined reaction time (RT) and inter-stimulus intervals (ISIs) to determine the development of motor chunks after practice, this study also examined the RTs and ISIs in a choice RT paradigm to determine whether different sizes of speech motor programs were used. Additionally, the variability in the f_0 trajectory of Mandarin tone sequences was examined to discuss evidence for the existence of GMPs.

Experimental Questions: Did native speakers of Mandarin utilize larger motor chunks than non-Mandarin speakers as indicated by a larger ratio of RT over average ISIs in three and six syllable conditions? Did the speakers change their production mode from retrieving and executing a stored motor program all at once to controlling speech in a concurrent manner as the tone sequence became longer as indicated by shorter RTs and longer ISIs in the longer sequence than in the shorter sequence conditions? Did the non-Mandarin speakers produce novel tone sequences concurrently regardless of the sequence length as indicated by similar values of RT, ISI and ratio of RT over average ISIs in all sequence length conditions? Also, this conclusion would be made

when the Hamming distance differences between Slope and Parson's code measurements stay the same across different sequence length conditions. Did the tone sequence, which is produced using larger motor chunks, preserve relative timing or force relationships in the acoustic signals more than speech produced utilizing smaller motor chunks, when variability of the fundamental frequency trajectory is examined? Would the GMP be parameterized more than once while producing Mandarin tone sequences of multiple syllable lengths?

In order to answer the above questions, two factors were manipulated: speaker group (native Mandarin speakers and non-Mandarin speakers) and response length (3, 6, and 9 syllables). Then, the reaction time (RT), inter-syllable intervals (ISIs), ratio of RT over average ISI (RT/ average ISI), parameter variability of the f_0 trajectory, and errors of the f_0 trajectory coded in three different ways (Euclidean distance errors, Slope measurement, and Parsons' code measurement) were analyzed (The detailes for each dependent variable will be explained in section 2.7.3).

1.3.2 Assumptions for Research Questions

The overall purpose of this study was to determine if larger motor programs develop and whether relative timing or force relationships are maintained in the acoustic signals of native speakers. That is, do speakers prepare and execute a larger motor program unit all at once or choose to concatenate smaller units in a concurrent manner when speakers produce syllables of varying numbers? In order to pursue this line of inquiry, several prior assumptions were necessary. They included:

1) Speech production experience affects motor program size. Thus, speakers with more experience will utilize larger motor programs (Park & Shea, 2005).
2) The loading of complex motor responses results in longer choice RTs.

3) After a motor chunk is loaded into the motor buffer, it is executed without a delay. Thus, the ISIs within the motor chunk will decrease while speech production system executes this motor chunk.
4) In the current study, higher level processes or attentional demands were not considered as the cause of variability in the measurements because this study utilized the same stimuli and tasks across groups and conditions.
5) The residual variability observed in the f_0 trajectories of target speech will reflect the GMP errors because speech rate and vocal intensity effects on the acoustic signals will be accounted for before examining the invariant structure in the speech outcome.

1.3.3 Research Questions and Hypotheses

The research questions below are organized by dependent variable. For each dependent variable, the first question relates to the interaction between the two independent variables, group (native Mandarin and non-Mandarin speakers) and sequence length (3, 6 and 9 syllables). Significant interactions were explored with post-hoc analyses of the simple main effects to determine which specific contrasts are significantly different. Simple main effect questions asked if each independent variable is significantly different across one level of the other independent variable (e.g., if the sequence length effect is significantly different for native speakers).

1.3.3.1 Parameter Variability

RQ1: Is there a significant group by sequence length interaction for the parameter variability measurement?

Hypothesis1: There will be a significant interaction between group and sequence length condition (see Figure 1).

RQ2: Is there a significant difference in the parameter variability between native Mandarin and non-Mandarin speaker groups at each sequence length condition?

Hypothesis2: The native Mandarin speaker group will demonstrate significantly lower parameter variability than the non-Mandarin speaker group at all sequence length conditions (see Figure 1).

RQ3: Is there a significant difference in the parameter variability among 3, 6, and 9 syllable conditions for each speaker group?

Hypothesis3: The parameter variability in the native Mandarin group will be significantly higher in the 9 syllable length condition than in the 3 and 6 syllable length conditions. The parameter variability of non-Mandarin speaker group will not be significantly different across 3, 6, and 9 syllable length conditions (see Figure 1).

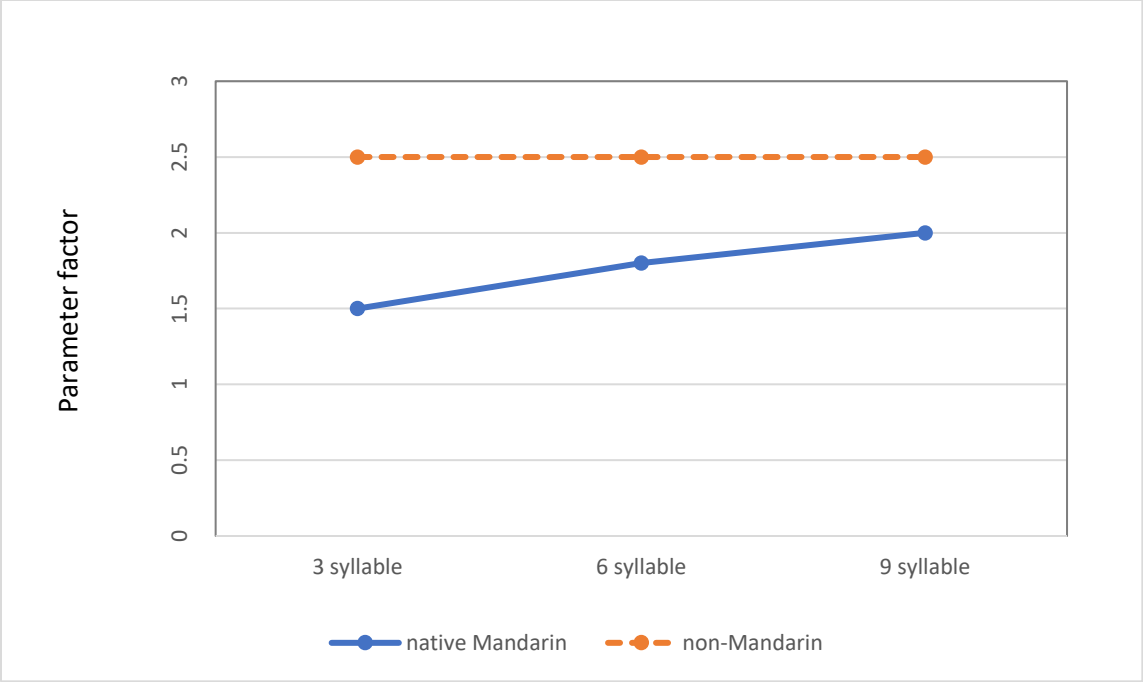

Figure 1 Hypothesized plot of parameter variability results

1.3.3.2 Average GMP Errors per Syllable

RQ4: Is there a significant group by sequence length interaction for the average GMP errors per syllable?

Hypothesis4: There will be a significant group and sequence length interaction for the average GMP errors per syllable (see Figure 2).

RQ5: Is there a significant difference in the average GMP errors per syllable between native Mandarin and non-Mandarin speaker groups at each sequence length?

Hypothesis5: The native speaker group will demonstrate lower average GMP errors per syllable than the non-Mandarin speaker group in all sequence length conditions (see Figure 2).

RQ6: Is there a significant difference in the average GMP errors per syllable among 3, 6 and 9 syllable conditions for each speaker group?

Hypothesis6: The native speaker group will demonstrate higher average GMP errors per syllable in the 9 syllable condition than in the 3 and 6 syllable conditions, whereas the non-Mandarin speaker group will demonstrate non-significant differences in the average GMP errors per syllable among three length conditions (see Figure 2).

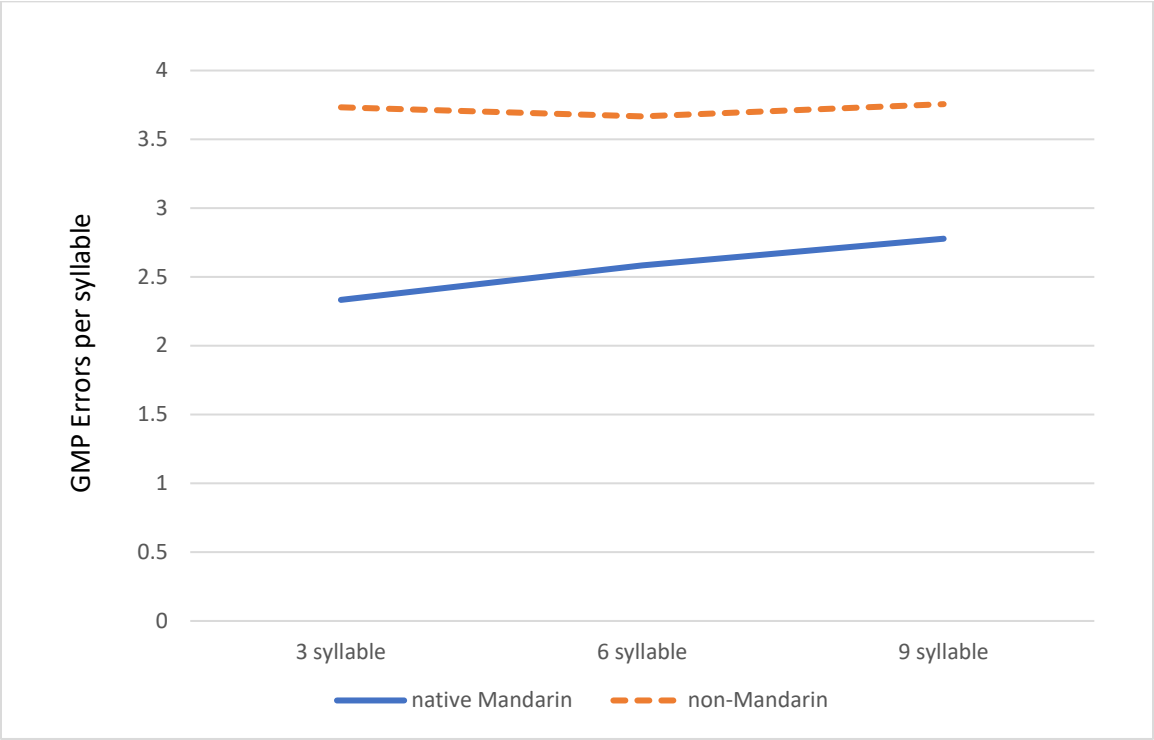

Figure 2 Hypothesized plot of average GMP errors per syllable

1.3.3.3 Average GMP Errors in Tone 2 Obtained from Second and Eighth Syllable Positions of the Target Sequence (Tone 4-2-1-3-4-1-4-2-1)

RQ7: Is there a significant interaction between group and syllable position (second and eighth syllable positions) for the average GMP errors in Tone 2 in the target tone sequence (Tone 4-2-1-3-4-1-4-2-1)?

Hypothesis7: There will be a significant interaction between groups and syllable positions for the average GMP errors in Tone 2 (see Figure 3).

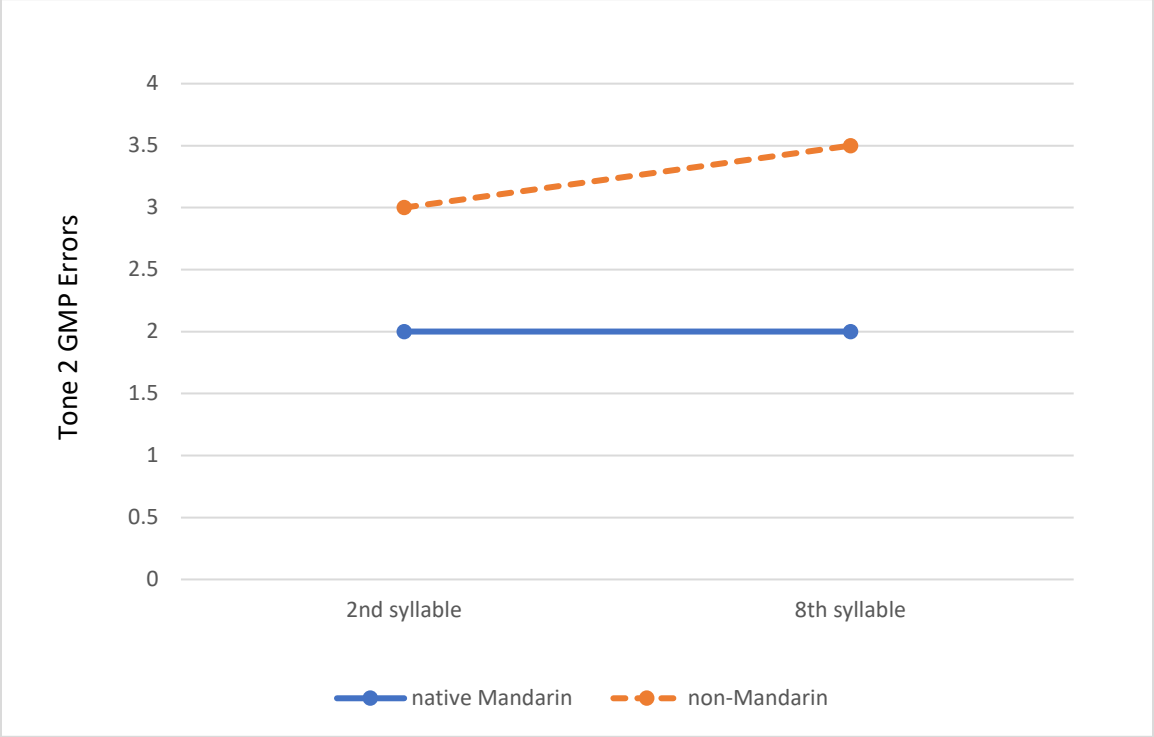

Figure 3 Hypothesized plot of interaction between Syllable Positions and Groups in the average GMP errors in Tone 2

RQ8: Is there a significant difference in the average GMP errors in Tone 2 between native Mandarin and non-Mandarin speaker groups at each syllable position?

Hypothesis8: The native Mandarin speaker group will demonstrate a smaller average GMP error in Tone 2 than the non-Mandarin speaker group at both syllable positions (see Figure 3).

RQ9: Is there a significant difference in the average GMP errors in Tone 2 between the second and eighth syllable positions for each speaker group?

Hypothesis9: The average GMP errors in Tone 2 will not be significantly different between the second and eighth syllable positions in the native Mandarin speaker group. The average GMP

errors in Tone 2 will be significantly different between the two syllable positions in the non-Mandarin speakers. The direction of change will not be relevant because the variability in the GMP errors of non-Mandarin speakers is expected to occur at random (see Figure 3).

1.3.3.4 Hamming Distance Difference per Syllable between Slope Measurement and Parsons' Code Measurement

RQ10: Is there a significant interaction between syllable length by group in the Hamming Distance difference per syllable (between Slope measurement and Parsons' code measurement)?

Hypothesis10: There will be a significant interaction between group and sequence length for the Hamming Distance difference per syllable (see Figure 4).

RQ11: Is there a significant difference in the Hamming Distance difference per syllable between native Mandarin and non-Mandarin speaker groups at each sequence length?

Hypothesis11: The non-Mandarin speaker group will demonstrate significantly greater Hamming Distance difference per syllable between Slope measurement and Parsons' code measurement than the native Mandarin speaker group at all syllable length conditions (see Figure 4).

RQ12: Is there a significant difference in the Hamming Distance difference per syllable among 3, 6, and 9 syllable conditions for each speaker group?

Hypothesis12: The native Mandarin speaker group will demonstrate significantly greater Hamming Distance difference per syllable in the 9 syllable condition than in the 3 syllable condition, but non-Mandarin speaker group will not demonstrate a significant difference among the syllable length conditions (see Figure 4).

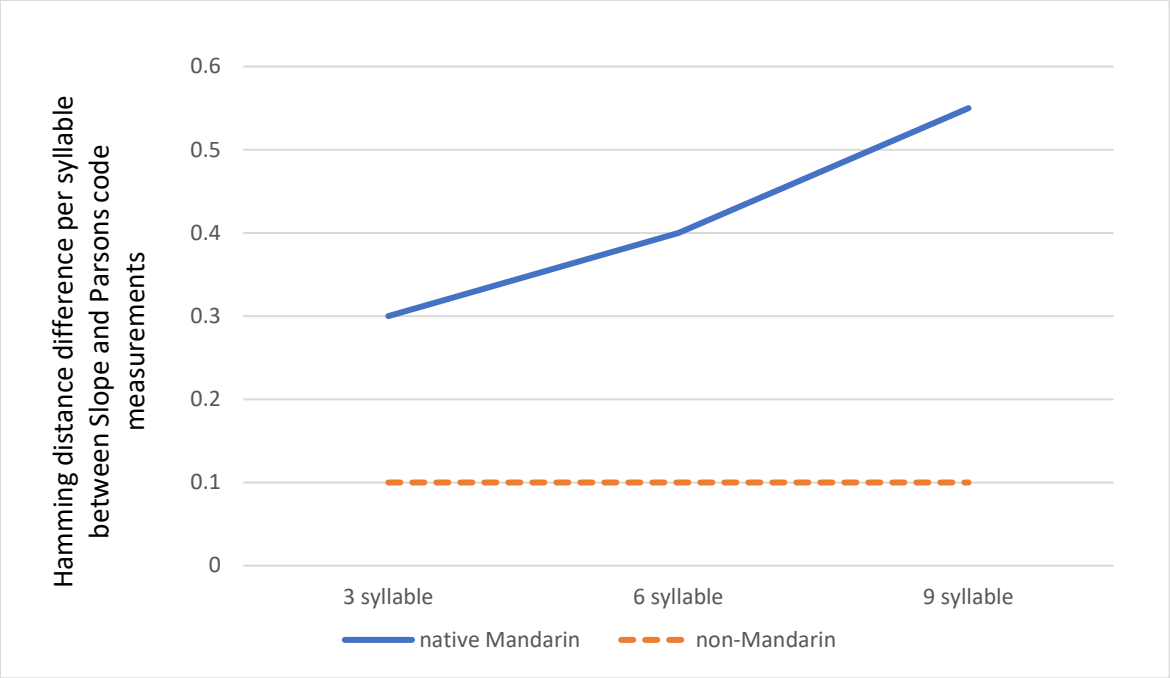

Figure 4 Hypothesized plot of the Hamming Distance difference per syllable between Slope and Parsons' code measurements

1.3.3.5 Reaction Time (RT), Average ISI, Ratio of RT/average ISI

The following questions relate to complexity effect, which refers to a longer RT before longer sequences. However, results in the literature inconsistently observe this effect. Therfore, the research questions and hypotheses about these timing measures (dependent variables) are organized in two different ways: 1) the complexity effect is not assumed; and 2) the complexity effect is assumed.

Complexity Effect Not Assumed

Reaction Time (RT)

RQ13: Is there a significant interaction between group and sequence length for the RT?

Hypothesis13: There will be significant interaction between group and sequence length conditions for the RT (see Figure 5).

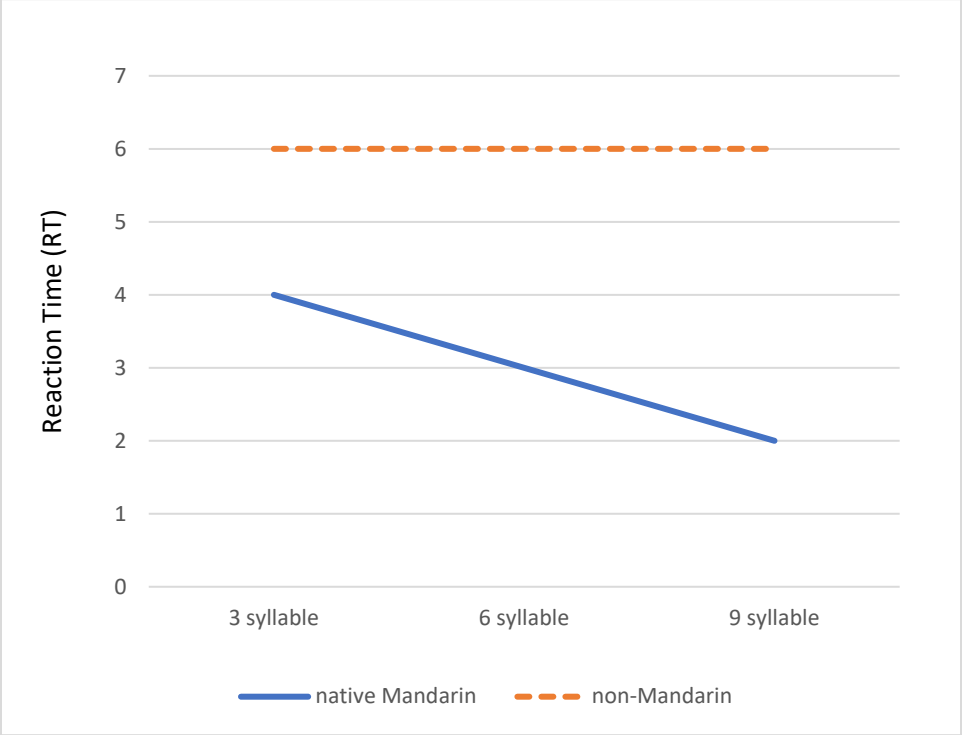

Figure 5 Hypothesized plot of the interaction between Group and Sequence Length on the RT

RQ14: Is there a significant difference on the RT between native Mandarin and non-Mandarin groups at each sequence length condition?

Hypothesis14: There will be significantly shorter RTs for the native speakers than for the non-Mandarin speakers in all sequence length conditions (see Figure 5).

RQ15: Is there a significant difference on the RT among sequence length conditions at each speaker group?

Hypothesis15: There will be significantly longer RT in the short sequence condition than in the long sequence condition in the native Mandarin group. There will not be significant difference on the RT among sequence length conditions in the non-Mandarin group (see Figure 5).

Average ISI

RQ16: Is there a significant interaction between group and sequence length conditions in the average ISI?

Hypothesis16: There will be a significant interaction between group and sequence length conditions in the average ISI (see Figure 6).

RQ17: Is there a significant difference in the average ISI between native Mandarin speakers and non-Mandarin speakers at each sequence length condition?

Hypothesis17: There will be a significantly shorter average ISI for the native Mandarin speakers than for the non-Mandarin speakers at all sequence length conditions (see Figure 6).

RQ18: Is there a significant difference in the average ISI among sequence length conditions for each speaker group?

Hypothesis18: There will be a significantly longer average ISI in the long sequence length condition than in the short sequence length condition for the native Mandarin speaker, but no difference in the average ISI among sequence length conditions for the non-Mandarin speaker group (see Figure 6).

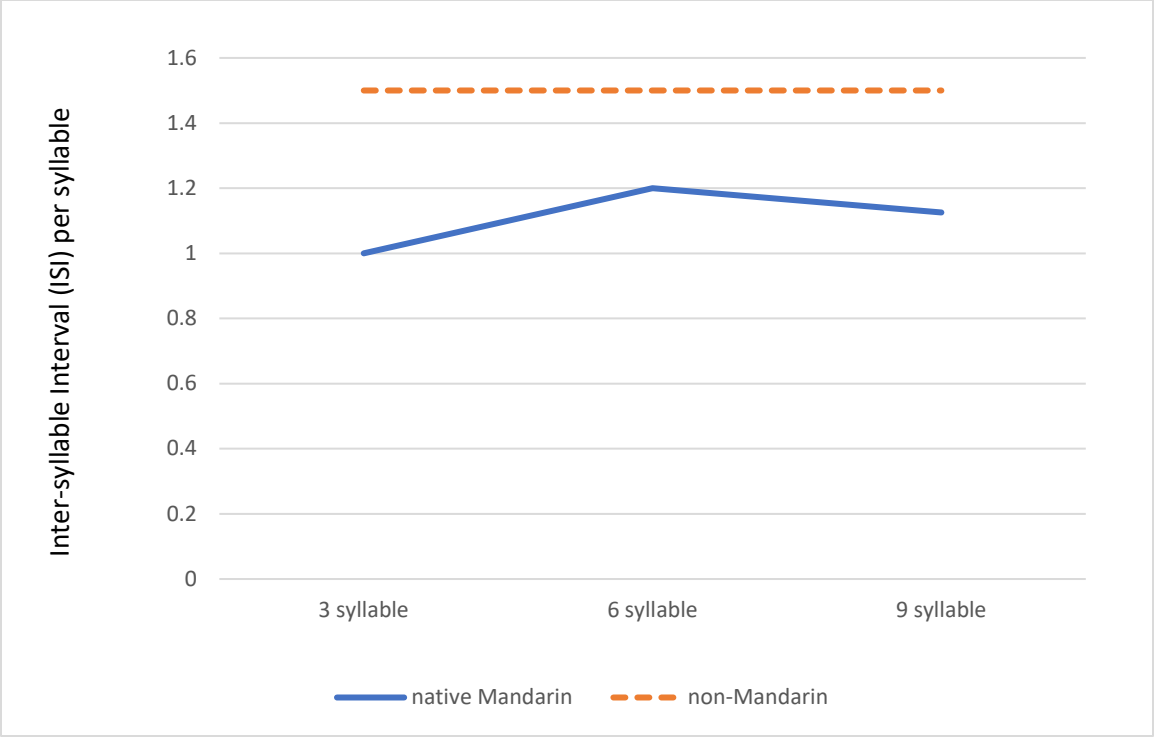

Figure 6 Hypothesized plot of the interaction between Group and Sequence Length on the average ISI

Ratio of RT/Average ISI

RQ19: Is there a significant interaction between group and sequence length conditions for the ratio of RT/average ISI?

Hypothesis19: There will be a significant interaction between group and sequence length conditions for the ratio of RT/average ISI (see Figure 7).

RQ20: Is there a significant difference in the ratio of RT/average ISI between native Mandarin speakers and non-Mandarin speakers at each sequence length?

Hypothesis20: There will be a significantly higher ratio of RT/average ISI for the native Mandarin group than for the non-Mandarin group at the short sequence length condition and a

lower ratio for the native Mandarin group than for the non-Mandarin group at the long sequence length condition (see Figure 7).

RQ21: Is there a significant difference in the ratio of RT/average ISI among sequence length conditions for each speaker group?

Hypothesis21: There will be a significantly different ratio of RT/average ISI among sequence length conditions for the native Mandarin speaker group (larger ratio in the short sequence length than in the long sequence length), but the ratio among length conditions will not be significantly different for the non-Mandarin speaker group (see Figure 7).

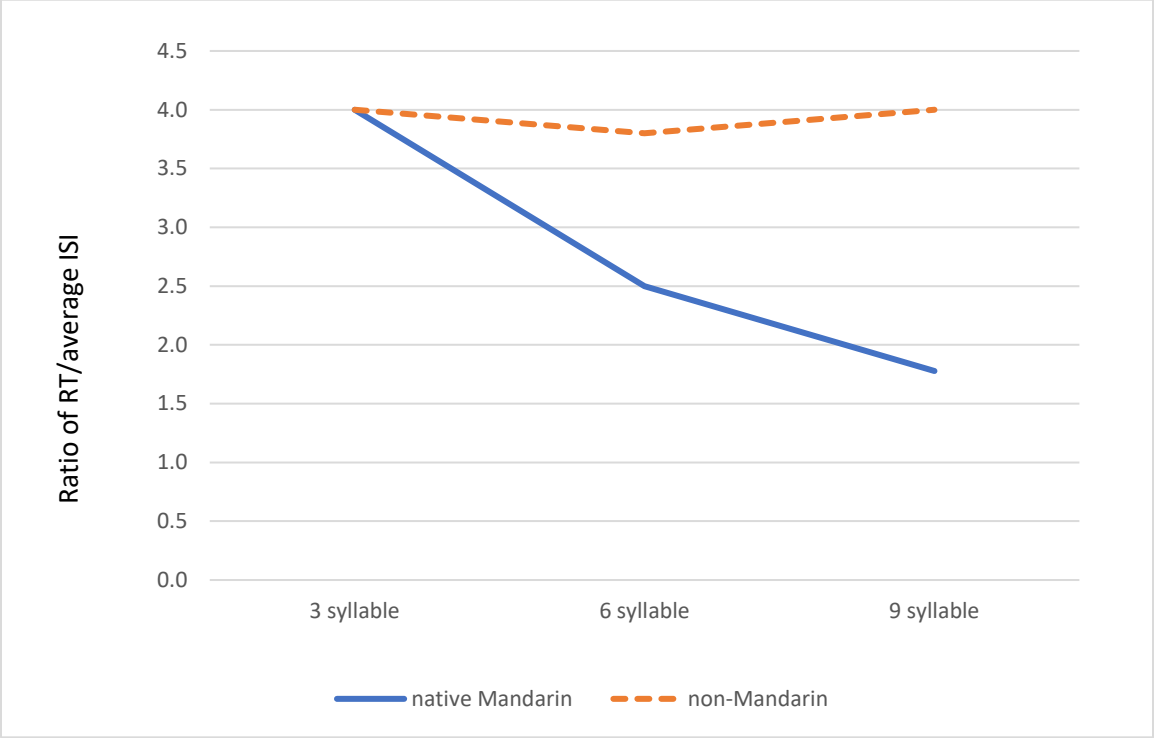

Figure 7 Hypothesized plot of the interaction between Group and Sequence Length on the Ratio of RT/average ISI

Complexity Effect Assumed

Reaction Time (RT)

RQ22: Is there a significant interaction between group and sequence length conditions for the RT?

Hypothesis22: There will be a significant interaction between group and sequence length conditions for RTs (see Figure 8).

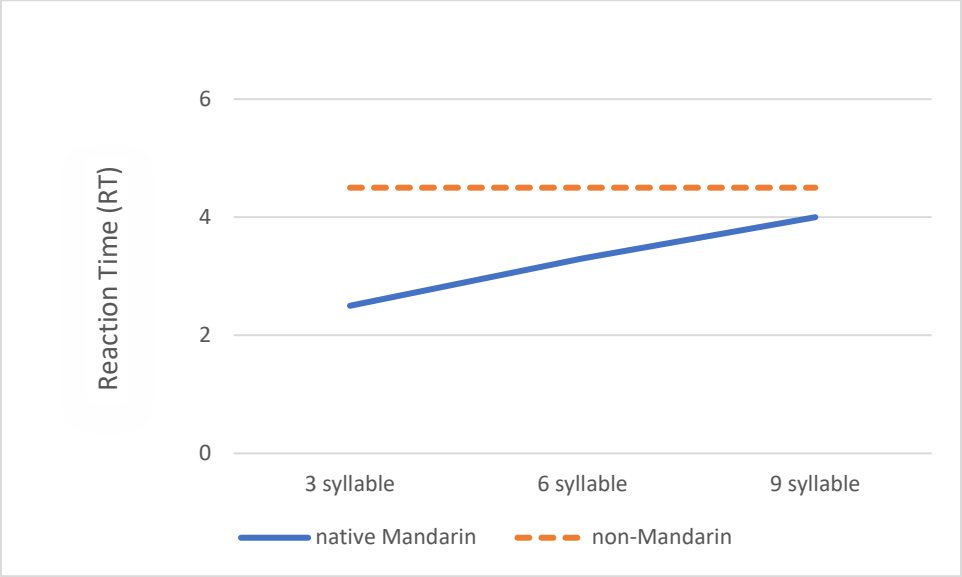

Figure 8 Hypothesized plot of the interaction between Group and Sequence Length on the RT

RQ23: Is there a significant difference on the RT between native Mandarin group and non-Mandarin group at each sequence length condition?

Hypothesis23: There will be a significantly shorter RT for the native Mandarin group than for the non-Mandarin group at all sequence length conditions (see Figure 8).

RQ24: Is there a significant difference in RTs among sequence length conditions for each speaker group?

Hypothesis24: There will be significantly longer RTs in the long sequence length condition than the short sequence length condition for the non-Mandarin group. However, RT will not be

significantly different among sequence length conditions for the native Mandarin group (see Figure 8).

Average ISI

RQ25: Is there a significant interaction between group and sequence length conditions on the average ISI?

Hypothesis25: There will be a significant interaction between group and sequence length conditions on the average ISI (see Figure 9).

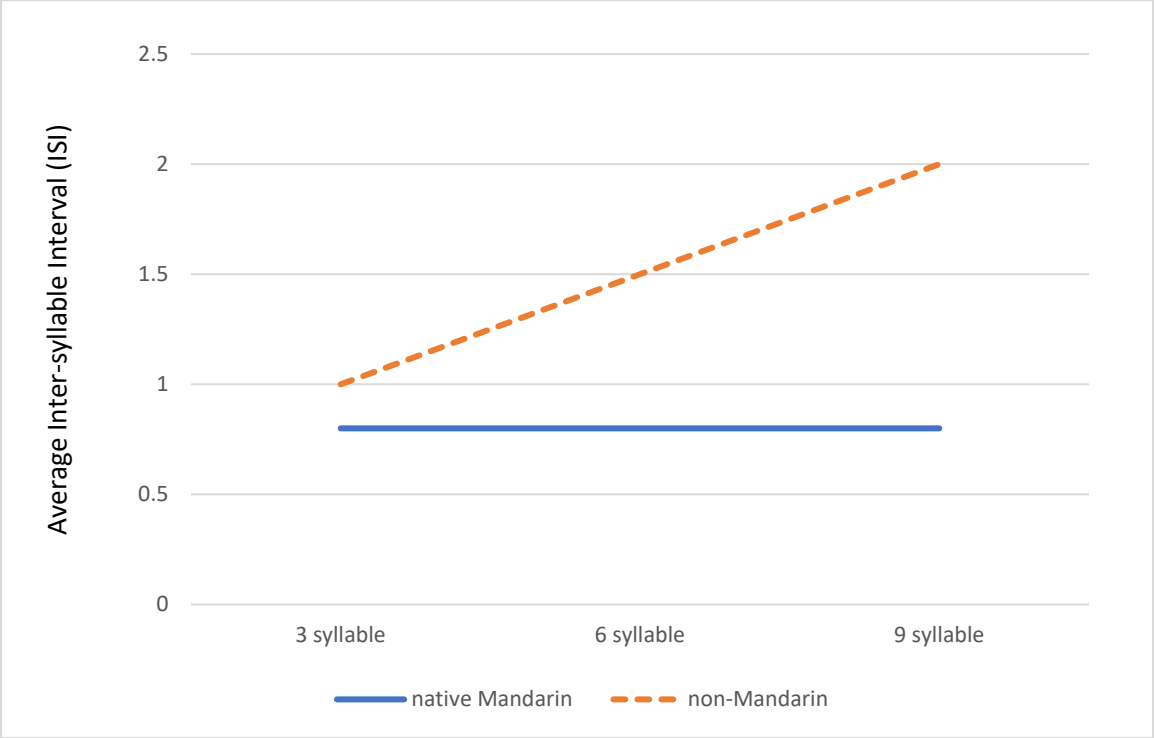

Figure 9 Hypothesized plot of the interaction between Group and Sequence Length on the average ISI

RQ26: Is there a significant difference on the average ISI between the native Mandarin and non-Mandarin groups at each sequence length condition?

Hypothesis26: There will be a significantly shorter average ISI for the native Mandarin group than for the non-Mandarin group at all sequence length conditions (see Figure 9).

RQ27: Is there a significant difference on the average ISI among sequence length conditions for each speaker group?

Hypothesis27: There will be a significantly longer average ISI in the long sequence length condition than the short sequence length condition for the non-Mandarin group. However, the average ISI will not be significantly different among sequence length conditions in the native Mandarin group (see Figure 9).

Ratio of RT/Average ISI

RQ28: Is there a significant interaction between group and sequence length conditions on the ratio of RT/average ISI.

Hypothesis28: There will be a significant interaction between group and sequence length conditions on the ratio of RT/average ISI (see Figure 10).

RQ29: Is there a significant difference on the ratio of RT/average ISI between native and non-Mandarin speakers at each sequence length?

Hypothesis29: There will be a significantly higher ratio of RT/average ISI for the native Mandarin group than for the non-Mandarin group at all sequence length conditions. The difference in the ratio of RT/average ISI between two groups will be larger in the long sequence condition than in the short sequence condition (see Figure 10).

RQ30: Is there a significant difference on the ratio of RT/average ISI among sequence length conditions for each speaker group?

Hypothesis30: There will be a significantly higher ratio of RT/average ISI in the short sequence length condition than in the long sequence length condition for the non-Mandarin speaker

group. However, the ratio of RT/average ISI will be significantly lower in the short sequence length condition than in the long sequence length condition for the native Mandarin speaker group (see Figure 10).

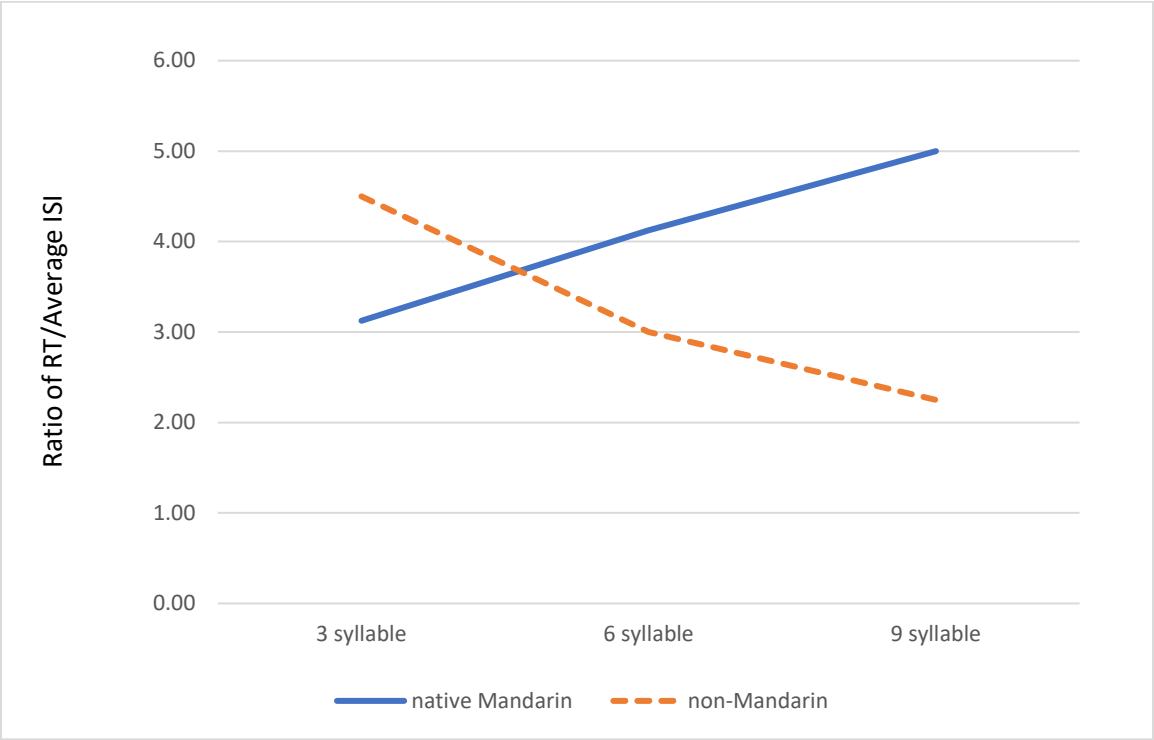

Figure 10 Hypothesized plot of the interaction between Group and Sequence Length on the Ratio of RT/average ISI

2.0 METHODS

2.1 SUBJECTS

Twenty-four Mandarin speakers and twenty-four native English (non-Mandarin) speakers were recruited. Participants were between 19 and 30 years of age (Mean age, native Mandarin: 23.75 (SD 2.82), non-Mandarin: 22.25 (SD 3.45)). This narrow age range was selected because studies have demonstrated a linear increase in the RT across the entire age range when a choice RT paradigm was used (Der and Deary, 2006). Although gender differences were not observed in the choice RT (Der and Deary, 2006), the variability may increase when the f_0 trajectory of each trial is compared to the gender-combined template trajectory during data analysis. To reduce this kind of variability during data analysis, only male speakers were enrolled in this study. All participants were naïve with regards to the purpose of the study. All participants consented to participate according to the approved consent procedure by the Institutional Review Board of the University of Pittsburgh. All participant screening procedures and inclusion and exclusion criteria are described in section 2. 4.

2.1.1 Calculation of Sample Size

Spencer and Rogers (2005) and Reilly and Spencer (2013) examined sequence length effects on choice RT and inter-syllable intervals (ISIs) while increasing the number of syllables from one to five. They also explored practice effects while participants produced each speech task twice in

each block over five practice blocks (total 10 trials for each speech task). The partial eta square for Spencer and Roger's RT study for the sequence length effect for the five syllable length conditions was .88. The partial eta square for Reilly and Spencer's ISI study for the sequence length effect for four syllable length (2-5 syllable length) conditions was .365. The partial eta square for Reilly and Spencer's ISI measure for the five practice blocks was .192. To be conservative, the lowest eta square of .192 was used for power analysis. Because hierarchical generalized linear model was used for the current study, the sample size was calculated using G-power for the repeated measure ANOVA. To calculate an adequate sample size, an effect size based on Cohen's *f* of .488, which corresponds to the eta square .192, was used. A total of 6 participants appeared to be an adequate sample to estimate a practice effect on the ISI measurement. However, this number was considered to be too small for the hierarchical generalized linear model. Therefore, Cohen's *d* effect sizes for the comparison of only the 3 and 5 syllable length conditions and of only the first and the last practice blocks were obtained using the Reilly and Spencer's study data. The correlations for both of these comparisons were .915 and .905 for sequence length effect and practice effect respectively. To obtain a large enough sample size for this current study, the determination was made to use .5 as a correlation coefficient value (r). Cohen's *d* effect size for the sequence length effect was .43 and the practice effect was .257. To be conservative, the lower effect size of .257 was used for a power analysis. To calculate an adequate sample size using G-power, the effect size for Cohen's *f* of .129, which corresponds to Cohen's *d* of .257, was used. A total of 48 participants was estimated to observe practice effects due to repeated trials on ISI measurements. Thus, according to the power analysis, 24 native Mandarin speakers and 24 non-Mandarin speakers were required for this study to have a power of 0.80 at $\alpha = 0.05$ (Cohen, 1988) (see Figure 11).

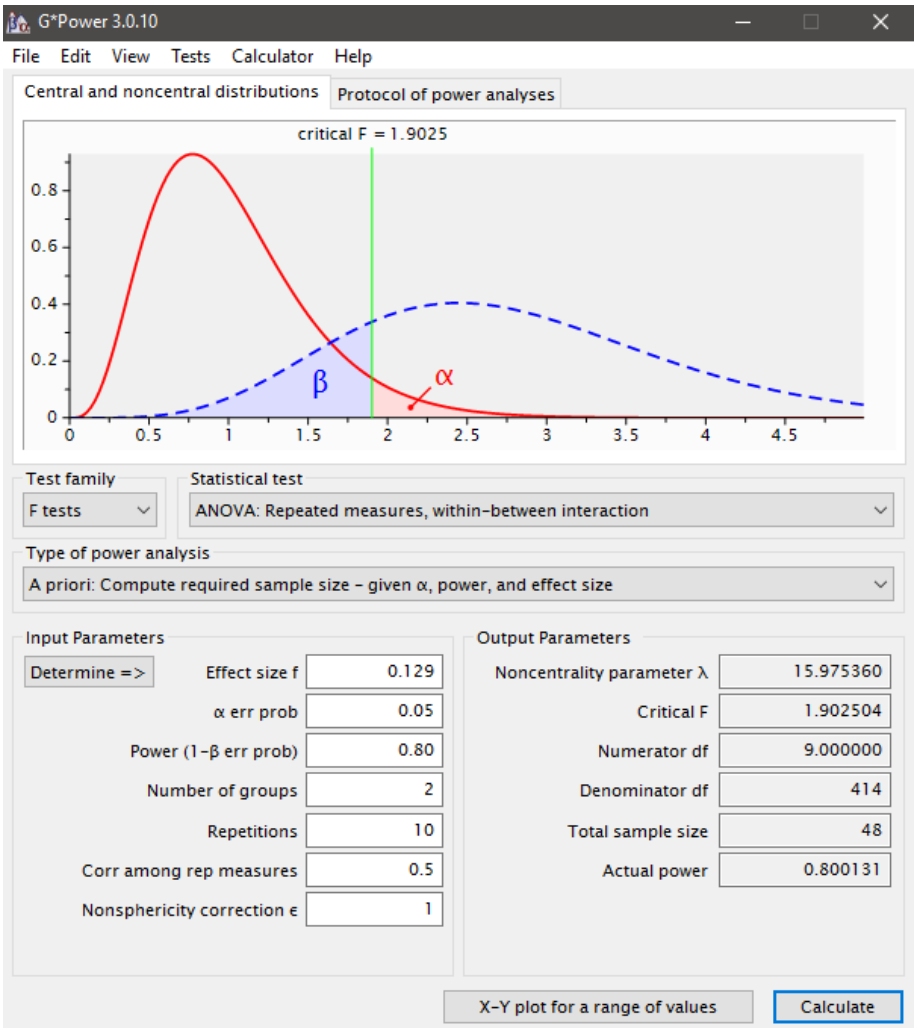

Figure 11 Snapshot of the calculation estimating sample size using G*Power Win 3.0.10

2.2 REMUNERATION

All participants received $40 for their participation and completion of the study. Participants received $8 if they were screened out or discontinued their participation for any reason. The entire experimental procedures took between two and four hours to complete.

2.3 STIMULI

2.3.1 Target Stimuli

As addressed in Section 1.2, Mandarin speakers use the relativeness information in the acoustic signals to identify Mandarin tones. Because the relative values of the acoustic signals can be used to infer existence of GMPs for speech, this study used Mandarin tones as speech targets and studied the relative changes in the fundamental frequency contours to evaluate the possible existence of GMPs for Mandarin tone productions. This study used three, six, and nine syllable length utterances. Utterances of four-syllable lengths have been found to be the longest utterance length in which individual tone contour is not influenced by intonation contour according to Chao (1965) and Tseng (1981). Although there is a risk of interaction between intonation and tone in utterances longer than four syllables, this study used three, six, and nine syllable length utterances. There are four identified tones in Mandarin: high-level (Tone 1), mid-rising (Tone 2), low falling-rising (Tone 3), and high-falling (Tone 4) (Xu, 1997). Figure 12 shows the f_0 trajectories for each Mandarin tone, and Figure 13 demonstrates f_0 range when two different gender groups of Mandarin speakers produce five-syllable Mandarin tone sentences.

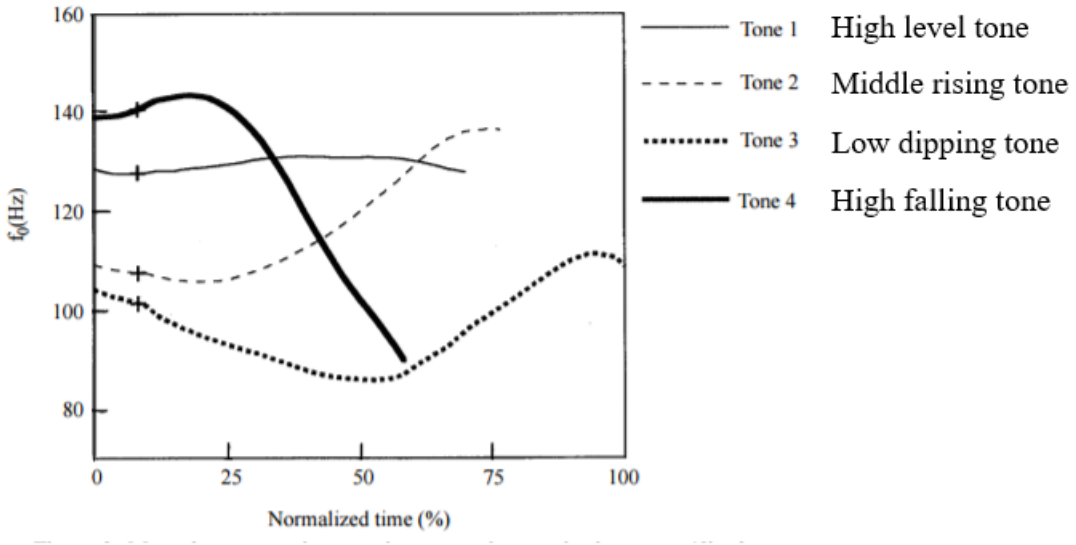

Figure 12 Mean f_0 contours (averaged over eight male native speakers of Mandarin and tokens; n=48) of four Mandarin tones in the mono-syllable /ma/ produced in isolation. The time is normalized, with all tones plotted with their average duration proportional to the average duration of Tone 3. (figure borrowed and modified from Xu (1997) with permission)

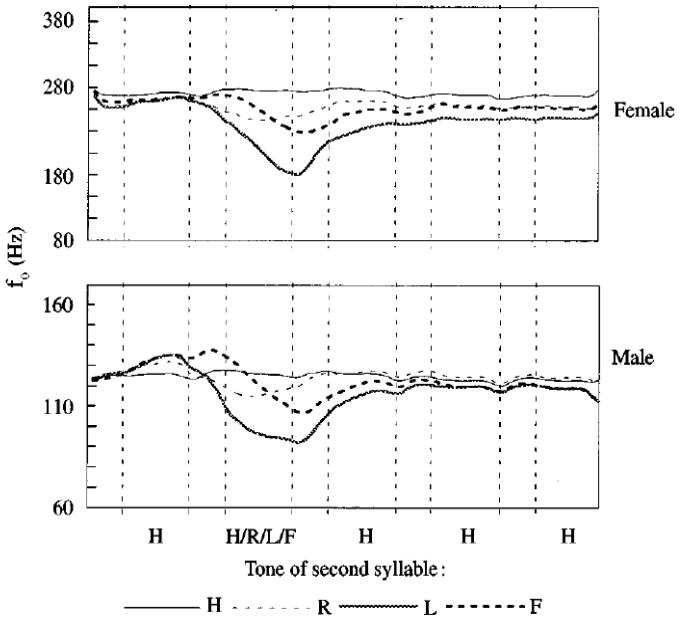

Figure 13 f_0 curves for 5 syllable sentences averaged across four female and four male Mandarin speakers separately (figure borrowed from Xu (1999) with permission)

There were two experimental phases: the mono-syllable tone practice phase and the tone sequence production phase. Four different Mandarin tones were presented at mono-syllabic levels during the mono-syllable practice phase. Mandarin tones were put into sequences and were used during the tone sequence production phase. These four tones can create 1,048,576 ($=4^9$) combinations of nine-syllable tone sequences, when they share segmental contexts (the same consonants and vowels). Only one tone sequence (Tone 4-2-1-3-4-1-4-2-1) was chosen to form target stimuli and this tone sequence was combined with the consonant /b/ and the vowel /a/ to form three different syllable lengths: The target tones consisted of the three-syllable length sequences, Tone 4-2-1 (bàbábā);, the six-syllable length, Tone 4-2-1-3-4-1 (bàbábābǎbàbā);, and the nine-syllable length, Tone 4-2-1-3-4-1-4-2-1 (bàbábābǎbàbābàbábā). Three other pairs of three, six, and nine-syllable length tone sequences were used as fillers: Tone 1-3-2 (bābǎbá), Tone 1-3-2-4-1-2 (bābǎbábàbābá), Tone 1-3-2-4-1-2-1-3-2 (bābǎbábàbābábābǎbá), Tone 3-2-1 (bǎbábā),

Tone 3-2-1-2-1-4 (bǎbábābábābà), Tone 3-2-1-2-1-4-3-2-1 (bǎbábābábābàbǎbábā), Tone 2-4-1 (bábàbā), Tone 2-4-1-3-1-4 (bábàbābǎbābà), and Tone 2-4-1-3-1-4-2-4-1 (bábàbābǎbābàbábàbā). Altogether twelve tone sequences were used in this study.

The selection criteria for tones to form a tone sequence were quasi-random; four criteria were considered during this selection process. First, the third tone is often perceived as "creaky" (Chao, 1965; Davison, 1991) and results in discontinuity in the tone contour when the f_0 trajectory is examined in a spectrogram using the Praat acoustic analysis program. In addition, the f_0 levels of neighboring tones influence the f_0 level of one tone. Therefore, in order to avoid a creaky voice or too low f_0 contours for the third tone, the target tone sequence was chosen to include a third tone between the two high f_0 levels (i.e., Tone 1-3-4, bābǎbà) to create the target stimuli. Second, any three tones which appeared sequentially did not reappear in the same order in the other tone sequences, if it was not repeated for the sequence length manipulation (i.e., 3 → 6 → 9 syllables). Third, the first three tones reappeared in the 7th, 8th, and 9th tone positions to create nine-syllable sequences. As a result, the following basic tone sequences, Tone 4-2-1-3-4-1 (bàbábābǎbàbā), Tone 1-3-2-4-1-2 (bābǎbábàbābá), Tone 3-2-1-2-1-4 (bǎbábābábābà), Tone 2-4-1-3-1-4 (bábàbābǎbābà) were selected and manipulated to create different tone sequence lengths. Last, the tone sandhi was avoided. The tone sandhi refers to the phonological phenomenon, where one tone changes to another in the influence of neighboring tones when two tones are produced in a sequence. For example, when two third tones are produced in a sequence (i. e., Tone 3-Tone 3), the first Tone 3 is pronounced as Tone 2 due to the tone sandhi in Mandarin.

Strand (1987) reported decreased reaction times (RT) and inter-word intervals (IWI) when five-syllable English sentences were produced, as compared to when a single-syllable word was repeated five times. This result could be due to articulatory constraints to prepare the same

articulatory movement repeatedly. However, the motor programs for five-syllable sentences might also have been prepared as a whole motor response reducing the IWIs. To avoid any semantic and syntactic processing influences on tone production, the target stimuli for this study were presented only in pinyin, in which diacritics to denote tones were marked. One tone syllable usually represents more than one related Chinese character and associated meanings. Thus, if the native Mandarin speakers are not provided with Chinese characters, it was predicted to be difficult for them to retrieve related meanings automatically. Thus, only the pinyin with tone diacritics was presented without Chinese characters in order to eliminate possible semantic and syntactic contextual (higher level processing) influences on the phonological, phonetic, and articulatory processes (lower level processing) involved in tone production. Table 1 demonstrates four different pinyins that were used for the current study. Presenting only the pinyin without Chinese characters was also expected to prevent any influence of intonation on the individual tone contour, whose possibility was addressed above.

Table 1 Examples of target tones to be used

Tone	1	2	3	4
Pinyin	bā	bá	bǎ	bà

2.3.2 Recording of Target Stimuli for Mono-Syllable Practice Phase

One male Ph.D. student at University of Pittsburgh, native speaker of Mandarin Chinese (age 27), recorded target audio stimuli for the mono-syllable tones. He was born and grew up in Shandong, China, where Mandarin is spoken as the regional language. He moved to Beijing at the age of 16 and stayed there until he came to U. S. for his doctoral study. He identified himself as having zero

accent (standard) in speaking Mandarin. His English proficiency level met the level of proficiency outlined in the next section 2. 4 as criteria for research participants. He produced four different Mandarin tones with /ba/ combination ten times composing forty productions. These productions were recorded acoustically. Three other native Mandarin speakers were recruited to verify that those forty productions were produced correctly. All three judges agreed that all forty productions were correct and without production errors.

To select one sound file that represents each tone, first, the total duration for each production was measured and thirty three time points, at equal intervals were extracted from f_0 trace for each trial. The f_0 values were obtained for each time point and interpolation was performed using the neighboring f_0 values if there was no f_0 value corresponding to each time point. After extracting 33 data points from each tone production, the mean f_0 trajectory was obtained from ten trials. Then, one representative production was selected that provided the highest correlation with the mean f_0 trajectory for each tone type. Again, three listening judges verified that those selected sounds were representative sounds for each tone. These representative recordings were used to provide participants with auditory models and the template (or target) f_0 trajectory for each tone as visual feedback during the mono-syllable practice phase. The details of instrumentation used for data collection are provided below.

2.4 SCREENING PROCEDURE

The study was advertised (as approved by the IRB) through personal contact and on-line and off-line boards. After waivers for written consent for recruitment purposes were received from the IRB at the University of Pittsburgh, potential participants were pre-screened for eligibility during

phone calls, when they first contacted the author. The potential participants provided their verbal consent to collect their personal information during this call, which might have included personally identifiable information (PII) or protected health information (PHI). The sensitive information to be collected during this study included information such as name, contact information, any self-reported history of developmental or acquired language and learning disorders, any neurologic and psychological illness, and any history of substance abuse. Participant's voice was also recorded during the screening and experimental procedures.

During pre-screening, the examiner collected demographic data including a potential participant's age, gender, primary and secondary languages, general vision/hearing ability. Participants were also pre-screened for any history of neurologic and psychological (e.g., depression) problems, and current status of any substance abuse (alcohol or drug), as determined by self-report (see demographic questionnaires in Appendix A). Evidence has shown that these factors can influence speech and voice characteristics (Cohn et al., 2009; Dietrich & Abbott, 2012; Magri, Ferry, & Abela, 2007; Martin, 1988; Sapir & Aronson, 1985; Theodoros & Murdoch, 1994). Thus, to avoid any confounding effects of these conditions on the experimental task performances, these factors were pre-screened. Additionally, because this study involved non-Mandarin speakers' production of novel speech, participants were pre-screened for their history of communication or learning disorders. All participants were male speakers between the ages 19 and 30. Participants self-reported either English or Mandarin as their first language (L1). The non-Mandarin speaker group reported that they had no experience learning Mandarin. Speakers in both language groups self-reported that their L1 proficiency demonstrated following language profiles in the 0-5 scale questionnaires on the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld, & Kaushanskaya, 2007): understanding $\geq$ 4.4 (=4.86-0.46), speaking $\geq$ 4.05 (=4.65 -

0.60), reading ≥ 2.73 (4.25 - 1.52), and writing ≥ 2.6 (=4.08 - 1.48) (see adapted LEAP-Q Questionnaire in Apprendix A). All instructions to all participants were provided in English during the screening and experimental procedures. Thus, Mandarin speakers also were required to have the ability to use English proficiently as a secondary language. This was determined by scores of 7 or higher on the academic module of International English Language Testing System (IELTS); or a score of 80 or higher on the Internet Based Test (iBT), a score of 213 or higher on the Computer Based Test (CBT), or a score of 550 or higher on the Paper Based Test (PBT) of English as a Foreign Language (TOEFL) test. These thresholds were consistent with the admission criteria for universities in the United States.

In order to avoid a dialectal influence on Mandarin productions, questions were asked to include those native Mandarin speakers whose dialect or regional accents were minimal. The criteria of 3.36 or less on a scale of 0 to 5 (0: none or never, 5: heavy or always) for self-perceived degree of dialect/accent in speaking Mandarin or 3.61 or less when the degree of dialect/accent in speaking Mandarin was identified by others were used. These criteria were also borrowed from LEAP-Q questionnaire (Marian et al., 2007). While the same criteria were applied to the native English speakers, only one out of twenty four native English speakers in this study reported that he was exposed to another language (French) at home before learning English. He still identified himself as a native English speaker with no French accent in speaking English.

After passing the pre-screening criteria, participants underwent the screening and experimental data collection procedures. The screening data were collected in quiet rooms in Forbes Tower or in Scaife Hall at the University of Pittsburgh. Participants were required to pass a corrected or uncorrected binocular vision screening with the Snellen chart (≤ 40) and were required to detect 500Hz, 1000Hz, 2000Hz and 4000Hz pure tones in one ear at 25 dB when tested

using a portable audiometer. Participants needed to demonstrate normal discrimination ability in the pure tone discrimination task which is explained in the next paragraph. Participants demonstrated normal oral movements when assessed using the Oral Speech Mechanism Screening Examination-Revised (OSMSE-R) (St Louis & Ruscello, 1981), as determined by the experimenter. Participants demonstrated normal language comprehension on the listening version of Computerized Revised Token Test (CRTT-L) (McNeil et al., 2015) in their native language. An overall score of ≥14.17, the lower end of the 95% confidence interval, was used as the criteria for normal performance in the CRTT-L English version (McNeil et al., 2015). The CRTT-L Mandarin Chinese version has not undergone extensive standardization, so there are no set criteria for the normal group for this test. However, an average overall score of ≥13.54 (SD .59) was obtained in a study of normal Taiwanese speakers (S.-H. K. Chen, McNeil, Hill, & Pratt, 2013). Thus, an overall score of 12.36 (=mean - 2 × SD=13.54 - 2 × .59) was used as the criterion score for normal Mandarin speakers in this study. Participants also were required to demonstrate adequate vocal function to achieve the experimental tasks as evidenced by their performance on the Cepstral Spectral Index of Dysphonia for Rainbow passege ($CSID_R$) with a cutoff score of 24.3 (Awan, Roy, Zhang, & Cohen, 2016). Additionally, the author, a 1st class Speech-Language Pathologist (SLP) certified by the Korean Health Personnel Licensing Examination board, perceptually judged the respiratory, phonatory, articulatory, and resonance abilities while participants performed sustained phonation and diadochokinetic tasks during OSMSE-R. The vocal quality was also perceptually judged while participants produced vocal sweeps (i. e., “continuous changes from the lowest to the highest pitch that one could comfortably produce” (Pfordresher & Brown, 2009)) or when participants retold three stories of the Story Retell Procedure (SRP) task (Form A) (McNeil et al., 2007) (see Appendix B for screening form). When any pathologic symptoms were detected,

the examiner recommended participants to visit a speech-language pathologist and these participants discontinued the study participation. The summary of the screening test results appear in section 3.1

Tone Discrimination Task

A frequency difference limen (DL) is defined as the minimal difference in frequency that a listener discriminates at a 75% accuracy level (Moore, 2012). The DL values at 125Hz, 1000Hz, and 2000Hz levels were assessed in this current study for the following reasons. A male model speaker produced the target tone sequence of nine syllable lenths such as /bàbábābăbàbābàbábā/. The mean fundamental frequency (f_0) and the mean first formant (F1) and second formant (F2) frequencies of ten trials of these productions were 138Hz, 1038Hz, and 1998Hz respectively. However, literature typically reported mean DLs at 125Hz, 1000Hz, and 2000Hz levels. Thus, 125Hz, 1000Hz, and 2000Hz levels were chosen for testing instead of 138Hz, 1038Hz, and 1998Hz. Thus, these frequency values were considered representative for the current study and appropriate for screening the participants' frequency discrimination ability as determined by their frequency DLs. Spiegel and Watson (1984) reported that the frequency discrimination threshold ($\Delta f/f$) for a single tone was between 0.001 and 0.0045 for musicians. When nonmusciains were investigated, half of these individuals demonstrated the same range of thresholds as musicians, but the other half produced five times higher thresholds than musicians. Thresholds of ≤ 2.81 for 125Hz, ≤ 22.5 for 1000Hz, ≤ 45 for 2000Hz were used as screening criteria. Because these thresholds were strigent for untrained listeners in particular at 125Hz level, up to two practice trials were provided at each frequency level before obtaining the actual results.

Frequency DLs were measured using custom-built software. The tones were presented between 68-70dB SPL level at the ear, in sound field using a set of speakers. If a participant

complained about the sounds being too loud, the intensity was adjusted. All pure tones were presented for 200ms in duration with 20ms rise/fall times and with a 600ms inter-stimulus interval. Frequency discrimination was measured in a two alternative-forced choice paradigm. On each trial, participants determined which of the two stimuli contained the tone with a higher frequency. A two-down, 1-up adaptive bracketing procedure was used, which tracks the 70.7 % correct point on the psychometric function (Levitt, 1970). That is, two consecutive correct responses resulted in a decrease in the frequency difference, and a single incorrect response resulted in an increase. Frequency changes were 5 Hz for the first two reversals, and 2 Hz for the final six reversals. The average frequency difference at the last 6 reversals was taken as the DL for that run.

2.5 INSTRUMENTS

All tasks were performed in a quiet room either in Forbes Tower or in Saife Hall at the University of Pittsburgh. A quiet room in Saife Hall was used only once because there was a black-out in Forbes Tower. Since both rooms had similar attributes, the participant's performance did not appear to be affected based on this change. A SAMSUNG laptop computer featuring an Intel ® Core™ i7 Processor, Windows 10 Home OS, 15.5" LED Display (1920 X 1080 resolution) was used. The participant sat erect facing the laptop display, and the distance between the participant and the laptop display was between 42 and 45 cm. The "Stimulate" program, designed specifically for this study, was used to present acoustic and visual stimuli, to collect participants' vocal responses, and to provide visual feedback about those responses during the experiment. Hearing screening was accomplished with a portable audiometer and testing was conducted in a quiet, distraction free room or in a soundbooth. Acoustic stimuli were presented through two Logitech

multimedia speakers Z200 (total watt RMS: 5W, frequency response: 80Hz to 20kHz), positioned bilaterally at the level of the computer screen, between 45 and 48 cm from the participant. The acoustic target were presented between 65 and 70 dB SPL after calibration, measured by a BAFX 3370 digital sound level meter (measuring range: 30-130dB(A), frequency response: 31.5Hz-8KHz, sampling rate: 2 times/sec). The calibration of intensity was performed with the sound level meter placed at the level of each participant's ear while speech noise was presented from the laptop computer and while the intensity of the speaker was adjusted to meet the 65-70 dB criterion at the sound level meter.

For acoustic recording, a calibrated fixed distance microphone (Dayton Audio EMM-6 Electret Measurement Microphone) with an analog-to-digital converter (MXL Mic Mate $_{TM}$) was used. The microphone was positioned perpendicularly to the participant's air flow at 3 to 5 cm away from the subject's lips. A built-in recorder in the laptop computer (Realtek High Definition Audio Driver version 6.0.1.8105) was used to process the participant's production on-line and to provide visual feedback based on this information during the mono-syllable practice phase. Each participant's mean pitch and intensity levels were recorded while placing the sound level meter between the microphone and the participant's mouth before starting experimental procedures and while participants spoke /abababababababababa/. The mean pitch level was used to adjust participant's visual feedback screen to his own mean pitch level. The sound level meter was placed in front of the computer screen to ensure participant's productions ranged between 65 and 85 dBA levels during data collection. The intensity level was monitored on-line using the sound level meter during the experiment. Because of this process, the intensity level was not revisited to exclude any data post-experimentally based on the intensity levels. Praat 6023 (window 64bit edition) was used to extract the fundamental frequency (f_0) values over time from the audio files for additional

acoustic analysis. While the acoustic target stimuli were recorded from the native Mandarin model speaker, the target stimuli were presented using power point slides. For this recording, a Tascam DR-40 recorder (sampling rate: 48kHz, amplitude resolution: 24bit) and a head-mounted directional microphone (AKG C 420) were used. E-Prime software version 2. 0. 10. 356 (Schneider, Eschman, & Zuccolotto, 2002) was used for presentation of the acoustic stimuli and recording of the key-pressing responses while the author was trained to make correctness judgment on the visual and auditory information of Mandarin tones and while native Mandarin listeners judged the correctness of speech sounds. A custom-built python program 3.6.1 was used for the pure tone discrimination test. Custom-built Praat script (explained in section 2. 7. 2), python (version WinPython-32bit-2.7.10.3) programs, and Matlab (version 2016a) programs were used for data analysis. The details about how each of these programs contributed are described in the next section.

2.6 EXPERIMENTAL PROCEDURE

After participants met the inclusion/exclusion criteria during screening, the experimental data collection was initiated. Each participant engaged in a mono-syllable practice phase and a tone sequence production phase composed of five production blocks.

2.6.1 Mono-Syllable Practice Phase

The practice phase replicated the acquisition condition of Almelaifi's (2013) study, with some modifications. Almelaifi (2013) examined instruction effects while English speakers practiced monosyllabic Mandarin tones. However, this study did not explore instruction effects, rather only general instructions were provided to the participants (see Appendix C for more details about general instructions used). In addition, Almelaifi (2013) randomly presented target stimuli and each tone was practiced 50 times. The degree of learning for each tone varied, depending on the tone property or instruction condition. Therefore, Almelaifi (2013) advised to use a blocked design in future studies to ensure that participants learned to produce all tones accurately. Therefore, the current study used a blocked design to ensure that participants learned to produce each tone accurately during the practice phase so that the learned tones could be used for the sequence production phase. Also, the number of trials included in the practice phase was determined by the rate of acquisition of each individual. During this mono-syllabic tone practice phase, the participants sat in front of the computer monitor displaying the target pinyin stimuli ("bā," "bá," "bǎ," and "bà") along with acoustic models. These acoustic models were pre-recorded target stimuli produced by the model native Mandarin speaker. During this phase, each block started with general instructions and each tone was presented at the single syllable level. The participants listened to an acoustic model for each tone and repeated the tones one at a time. Each tone syllable was presented 10 times in each block. To challenge the participants when they learned each tone, these model speech sounds were provided only 70% of the time.

The author, a non-Mandarin speaker, judged the accuracy of each participant's tone production through on-line auditory and visual examination of the f_0 contours. In order to make reliable auditory and visual judgments, the author was trained to examine the correctness of each

production by observing three native Mandarin speakers judged speech samples. These speech samples were gathered after making a model speaker and eleven non-Mandarin speakers produced four different Mandarin tones and twelve different tone sequences five times. From these speech samples, 100 mono-syllable productions and 100 Mandarin tone sequence productions were randomly selected for the E-prime-constructed training program. The author then made judgments of correct or incorrect productions. This training was done until the author demonstrated a 90% and 80% correspondence level at mono-syllable and tone sequence listening tasks respectively with the auditory perceptual judgment results of three native Mandarin listeners.

During the mono-syllable practice phase, when the accuracy of the production reached 80% in two consecutive blocks, the experimental participant moved to the next block that presented a different tone. If the participant did not reach the 80% criteria, another block of 10 trials was presented for the same tone. There was a one-minute break after each block, unless participants ask for more time. This practice was repeated up to 10 blocks for one type of tone. If the participant still did not reach 80% accuracy for one tone in two consecutive blocks after practicing one hundred times, the participant was to be excluded from the study. However, no participant in this study was excluded for this reason. Feedback (knowledge of performance) about participants' performance was provided pseudo-randomly 60% of the time upon the completion of the trial. As feedback, the actual f_0 trajectory produced by a participant as well as the target f_0 trajectory were presented on the screen. During this phase, the participants' voluntary rehearsal behavior was allowed between trials. Sometimes, participants requested to listen to the target sounds which they had learned in the previous practice blocks because they no longer remembered them as they began to practice of another tone. In this case, the examiner permitted the participants to re-listen to those target sounds. Additionally, the examiner and participants discussed the

characteristics of each tone before starting the mono-syllable practice phase or during the breaks if needed. If a one-minute break was not enough, extra time was given to complete the conversation. This process helped participants remember each tone and related pinyin in the next tone sequence production phase. The mono-syllable practice phase took between 1 and 1.5 hours.

2.6.2 Visual Feedback

The y-axis of the visual feedback window presented the f_0 level of each participant's production and the x-axis informed of the duration. The window was adjusted at the average f_0 level produced by each participant which was obtained after the participants produced /abababababababababa/ observing the Praat program display. This speech task was chosen because it was similar to the target response of this current study except the fact that tones were not specified. The y-axis was rearranged to have an equal interval around the average f_0 level for each individual so that each participant would not be required to adjust their pitch level to that of the model speaker. Then, the absolute f_0 values of each tone production was logarithmically transformed to semitone values so that the exponentially increasing frequency values had equal intervals on the display for the same perceptual tone changes. After this adjustment, the target f_0 trajectory and each participant's production was aligned in terms of x-axis (time) to have the same initial points (or left alignment). Then, the semitone converted f_0 and duration of each production was presented on the screen (see Figure 14). The y-axis of the window covered +/- 12 semitones and the x-axis covered 1.5 sec after initation of each speech production. Any production out of this range was not captured.

Figure 14 Example display of the visual feedback of the target f_0 trajectory for /bā/ (top) and participant's f_0 production (bottom)

2.6.3 Practice Before Tone Sequence Production Phase

Participants had two blocks of six practice trials (two trials for each syllable length) before beginning the tone sequence production phase. During these two practice blocks, participants

became familiar with this production condition in which they were asked to verbally produce a visually presented (pinyin) Mandarin tone sequence. The reaction time of their production was measured between the visual presentation of the target stimulus and the voice onset. All participants were given 25000 ms to produce each tone sequence, and they indicated their readiness to move on to the next trial if they completed each production within that time by pressing the spacebar. These long durations were chosen to ensure that novice speakers had enough time to produce each tone sequence. If any participant could not complete producing the target stimuli within 25000 ms duration, that trial was considered as errored speech and no adjustment was provided to this aspect. A vocal intensity in the 65 to 85 dB SPL range was encouraged. Visual feedback of the vocal intensity was provided using the sound level meter placed in front of the screen. Verbal instruction from the experimenter was provided from time to time if the participants exceeded this range. The tone sequences used in this practice phase did not overlap with the tone sequences used for the actual experimental phase.

In the simple RT paradigm, the internal structure of the motor response (INT process) is prepared before the “Go” signal. Thus, the SEQ process (sequencing the elements) is reflected in the simple RT. The simple RT informed the number of elements in the sequence to be prepared during RT, but the simple RT may not include the programming process because this paradigm cannot prevent prior programming of the response before the “Go” signal. Thus, to examine the naturally existing motor program unit and motor programming process, the choice RT paradigm was utilized. The practice before tone sequence production phase took about 10 minutes.

2.6.4 Tone Sequence Production Phase

There were five blocks in the experimental tone sequence production phase. Two groups of speakers repeated the same target utterances (three target stimuli and nine fillers) ten times across five blocks, yielding 120 utterances for each speaker. Each experimental Mandarin tone sequence was presented two times pseudo-randomly in each block, yielding 24 trials in each block. The same feedback regarding speech rate and intensity was provided during this production phase as in the practice phase. Knowledge of performance (KP) was provided verbally by the examiner only when the participants constantly made substitution errors between two tones, such as Tone 2 and Tone 4. The total amount of this feedback differed based on each participant's performance. The inter-trial interval was 2000ms. There was a two-minute break after each block, unless participants asked for more time. The tone sequence production phase took between 30 and 40 minutes, and the entire experimental procedure including screening took between two and four hours.

2.7 DATA ANALYSIS

2.7.1 Error analysis

After ensuring that participants produced each speech sound within 65-85dB SPL during the experiment, no further analyses were performed to exclude responses based on intensity levels. The data were not excluded based on extreme RT values because variability in RT was expected and was of interest in this study. The trials with phonological speech sound errors, such as

substitution, addition or deletion of the phonemes or tones were marked on-line by the examiner and confirmed post-experimentally using the recorded speech samples. Once the errors were confirmed, the trials with errors were removed from subsequent data analyses of time measurements and were treated as missing values during statistical analyses for other dependent variables.

2.7.2 Interpolation

The custom-built Praat script using a Pitch listing function was used to extract the f_0 values at each time point for each production. These values were obtained every 5 ms (sampling rate 200/sec) across the utterance and were saved in an Excel spreadsheet in the comma delimited (*.csv) format. The custom-built python program named "1st_program_modified.py" plotted f_0 trajectory for each production visually and the spectrogram and waveform were examined in the Praat program if needed to identify gaps within a tone trajectory that require interpolation. It was observed that these gaps occurred in all four tones when there were discontinuous f_0 signals. These gaps were filled using linear interpolation with the help of this python program whenever the author identified the gap as a within-syllable gap. However, this process was not completely automatic because the examiner visually examined the need for interpolation of each gap. Also, during this process any environmental noise including echo from the microphone was identified and removed from the signal. Any noise higher than normal f_0 signals conveying voicing characteristics were considered as due to lip gestures for consonant production or as noises filling the gaps for the third tones. These signals were considered as meaningful and were included in the analyses. Whenever within-syllable gaps were identified, the linear interpolation was performed using the following formula: when (x_0, y_0) and (x_1, y_1) were known, the (x, y) was obtained from the following formula:

$$y = y_0 + (y_1-y_0)*(x-x_0/x_1-x_0)$$

2.7.3 Obtaining Dependent Variables

After gaps were filled, two different analyses were performed.

2.7.3.1 Time Measurements

RT, average ISI, and the ratio of the two intervals were obtained. The time measurements were obtained using a custom-built python program (1st_program_modified.py) while the interpolation process was performed. In this case, the duration of time points that did not have corresponding f_0 values was identified using the python program and recorded in corresponding columns in the excel spread sheet under the headings of RT, ISI1, ISI2, ISI3, ISI4, ISI5, ISI6, ISI7, and ISI8. This python program generated an excel file and named it as "results.csv." From these data, the ratio of RT/average ISI was calculated. The time measurements were obtained following the criteria below:

1) RT: the interval between the time the target stimuli was presented on the screen and the time a speech response was initiated.
2) ISI: the interval between the offset of the prior verb and the onset of the stop burst for the next voiced consonant (see Figure 15). The acoustic landmarks were identified by this python program (1st_program_modified.py) and the author visually examined and confirmed those points whenever necessary.
3) Average ISI: All ISIs in the tone sequence were summed and divided by the number of ISIs.
4) Ratio of RT/average ISI: the ratio of the RT over the averaged ISI was calculated.

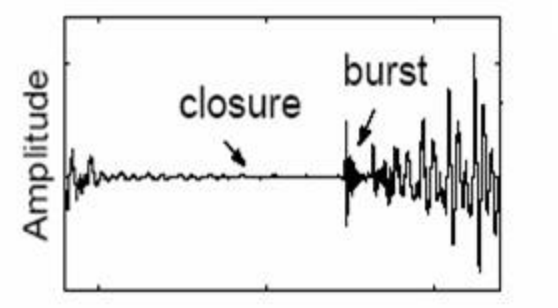

Figure 15 Example of a stop burst in the waveform

2.7.3.2 GMP and Parameter Error Measurements

The f_0 trajectory for each tone sequence was utilized to determine GMP errors and parameter factors using the following steps:

Obtaining a template f_0 trajectory for each target tone sequence

The f_0 trajectories for each tone were plotted next to each other after excluding pause intervals among neighboring tones.

While existing data points appeared at every 5 ms in the f_0 tracking file for each production, it was determined to restrict the number of data points for easy comparison among different trials and speakers. Thus, 33, 100, 200, and 300 data points were obtained for one-syllable, three-syllable, six-syllable, and nine-syllable length tone seuqences, respectively, using a custom-built matlab program named "program2.m." The f_0 value at every $1/33^{th}$, $1/100^{th}$, $1/200^{th}$, and $1/300^{th}$ points of the whole trajectory was obtained and recorded as coordinate values for each syllable length condition; i.e., (x, y) = (*time, f_0 value*). The x coordinate value was obtained after measuring

the whole sequence duration divided by 33, 100, 200, and 300 for each sequence length. The corresponding *y* coordinate value was obtained using the two neighboring f_0 values, and the interpolation rule was applied again in this process. The template f_0 trajectory for each tone sequence was the averaged f_0 trajectory derived from all qualifying trials that were produced by all native Mandarin participants during the tone sequence production phase. A second custom-built python program ("ex.py") was used to obtain mean f_0 values of all trials along the data points and standard deviations (SDs) around the means to create mean template f_0 trajectories. These template f_0 trajectories were obtained for each corresponding length condition. This template f_0 trajectory was generated for data analysis purposes and was compared to the production of each speaker's trial. This way of obtaining a template trajectory followed Smith et al. (1995), who obtained a kinematic trajectory for each rate condition after averaging fifteen trials of lip displacement at each time point after normalization. Smith et al. then averaged template trajectories for three different conditions to obtain one composite trajectory across all rate conditions. For the current study, the mean trajectory was obtained before normalization process occurred because the normalization process for this study required a template with which to compare (see Figure 16).

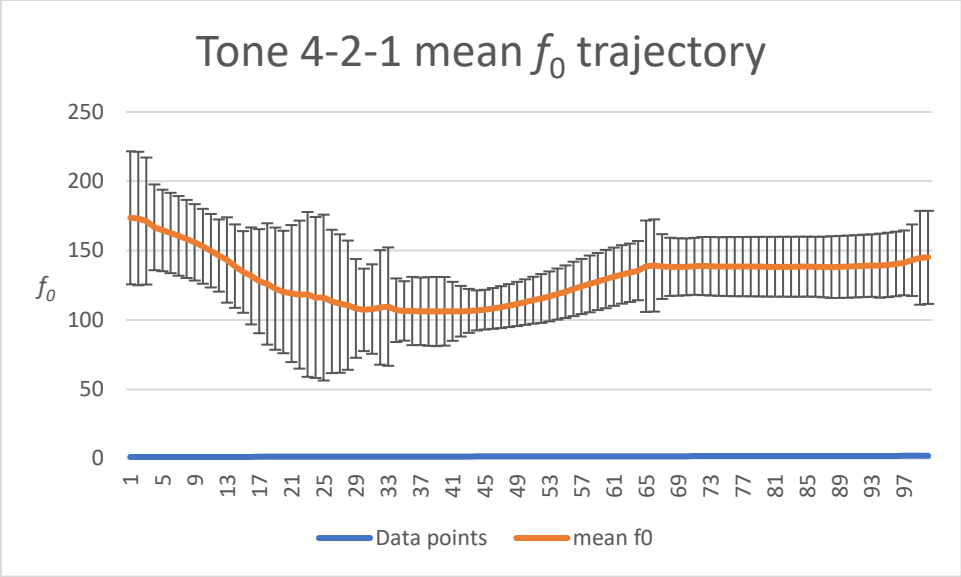

Figure 16 Example of a mean f_0 trajectory obtained for three-syllable tone sequence (Cf. black bar: standard deviation at each data point)

Between-subject GMP errors for each tone sequence

All the measurements explained from this section were produced by a custom-built matlab program named "program3.m" and "program3_ver4.m."

1) After obtaining an f_0 template for each tone sequence length, the f_0 trajectory of each trial for each participant was compared to the f_0 template trajectory.
2) The degree to which the two trajectories matched was determined by the method used by Wulf et al. (1993). The f_0 trajectory of each trial in the tone sequence production phase was scaled proportionally to fit the template trajectory by expanding or compressing overall production patterns along the time axis. First, the initial coordinates of the two (template and participant's) trajectories were synchronized by moving the initial coordinate (e.g., (time, f_0)) of the participant's f_0 trajectory to meet the initial coordinate of the template f_0 trajectory. All subsequent coordinates that

composed the participant's f_0 trajectory were recalculated and moved while maintaining its overall contour. Next, a single scaling factor for time, which optimizes the match between the two trajectories, was determined. After time scaling was completed, the scaling was performed once again over the variance of the magnitude dimension of the f_0 changes until the participant's time-scaled f_0 trajectory for each trial optimally fit the f_0 template trajectory. The optimal fit between two trajectories was determined based on the scaling factor that resulted in the minimal sum of Euclidean distance values. These distance values were obtained by comparing two corresponding data points from two trajectories. The scaling for magnitude of f_0 change was performed in the same way as the time scaling while maintaining the time scaling factor constant. The proportion used to expand or compress the participant's f_0 trajectory of each trial changed from 0.1 to 2.0 in increments of 0.1. Accordingly, the participant's f_0 trajectory for the tone sequence production phase was sometimes compressed to as small as one tenth of its original size and sometimes it was expanded to twice its original size.

3) This process determined a single time and a single f_0 magnitude scaling factor that optimally matched the f_0 trajectory of each trial from the tone sequence production phase to each tone sequence f_0 template trajectory. This scaling factor for the time dimension was called as "time parameter" and the scaling factor for the magnitude of f_0 change dimension was called as "f_0 magnitude parameter". The Euclidean distances between two corresponding data points of the two f_0 trajectories were obtained and summed, and this sum of Euclidean distances was considered as the between-subject

GMP error. This value determined the degree to which each participant's f_0 trajectory deviated from the template trajectory informing the degree of inaccuracy of the GMP.

Within-subject GMP errors for each tone

The f_0 trajectories for the same tones that were produced in different positions in the sequence were segregated from the f_0 trajectories of the whole tone sequences. The above procedure of obtaining GMP errors (1-3) was repeated for these segregated f_0 trajectories for individual tone syllables.

GMP errors per syllable

After GMP errors were obtained, the error values were plotted across different sequence length conditions. Then, the GMP error values were divided by the number of syllables. The GMP errors per syllable were compared statistically among different sequence length conditions or groups.

Within-utterance parameter variability

The parameter variability was obtained using the following procedure. First, the first-syllable f_0 trajectory was segmented from the whole f_0 trajectory while maintaining the time coordinate information. Second, the template for the first-syllable f_0 trajectory was obtained after averaging the first-syllable f_0 trajectory of the thirty target sequence productions from all native Mandarin speakers. The time and f_0 magnitude scaling factors (or parameter factors) for the first syllable tone were obtained after comparing the first-syllable f_0 trajectory of each production to the template f_0 trajectory of the first syllable tone. The same scaling process described above in the section *Between-subject GMP errors for each tone sequence* was utilized. Those parameter factors for the first syllable tone were subtracted from the parameter factors obtained from the whole tone

sequence that were obtained while calculating between-subject GMP errors for each tone sequence. This value was entered into statistical analysis to evaluate the parameter variability.

Distance values per syllable

GMP errors were obtained using Euclidean distance values between two f_0 trajectories. However, two additional distance values were obtained between the template f_0 trajectory and the participant's f_0 trajectory from the tone sequence productions. The distance values were obtained after coding f_0 trajectories using Slope and Parsons' code measurements. These measurements were similar to those used by Ramadoss (2012). However, the results were interpreted differently from Ramadoss' study. In this investigation, it was assumed that the f_0 Slope at a given interval reflected the proportional change in the acoustic signal. This measurement was expected to reflect the pre-set movement structure, which was proposed by Schema theory. In contrast, Parsons' code was used to capture the direction and magnitude of change above a set threshold at two consecutive time points. A concurrent judgment is required to assign Parsons' code values. Thus, this measurement was used to examine f_0 value changes in a concurrent manner.

The goal of this analysis was to understand whether Parsons' code measurement reflected participants' performance better than Slope measurement when participants changed their production mode from retrieving a whole movement sequence to controlling a speech movement in a concurrent manner. It was examined whether the Hamming distance for Slope increased as the sequence length became longer and whether Hamming distance for Parsons' code measurements is similar between the 9 syllable condition than the 3 syllable condition for the native Mandarin speakers, while the difference remains similar among different sequence length conditions in the non-Mandarin speakers.

Hamming Distance for Slope Measurement

The first Hamming distance value was obtained as follows. After the two trajectories (template vs. participant's) were normalized, the Slope values were calculated between the two consecutive time points that composed an f_0 trajectory. If the f_0 magnitude change was divided by the time interval at second level, such as 0.01385 sec, as in table 2, the Slope values became exaggerated due to a very small time interval denominator. As a result, the Slope values between participant's and target f_0 trajectories rarely matched as in table 2, and the Slope value reached a ceiling effect receiving "1" for almost all data points. For example, if there were 33 data points on average for a syllable production, the Hamming distance of Slope values were near 33 on almost all files. Therefore, in order to avoide this ceiling effect, the time interval at 10 msec level (e.g., 1.385) was used as a denominator for the Slope formula (see table 3). After this adjustment was made, these Slope values were obtained for each f_0 trajectory at the integer level after rounding up the values from the first digit beyond the decimal. The degree of difference in the Slope values between participant's trial trajectory and the template trajectory was expressed using the Hamming distance. When the two Slope values (template vs. participant's) matched exactly at one time point, the value "0" was assigned as a Hamming distance. When two Slope values did not match exactly at one time point, the value "1" was assigned as a Hamming distance.

Table 2 Hamming distance for Slope when magnitude of f_0 change was divided by the time interval at second level

Data Point num	Time	Interpolated f_0	Slope Subject	Slope Target	Hamming Slope	Hamming Slope Sum	Parsons Subject	Parsons Template	Hammig Parsons	Hamming Parsons Sum
1	1.52	243.00	0	0	0	99	0	0	0	54
2	1.53	241.94	-64	-60	1	0	-1	-1	0	0
3	1.55	240.39	-93	-211	1	0	-1	-1	0	0
4	1.56	239.19	-73	-523	1	0	-1	-1	0	0
5	1.58	236.20	-180	-262	1	0	-1	-1	0	0
6	1.59	231.21	-301	-211	1	0	-1	-1	0	0
7	1.60	226.49	-284	-243	1	0	-1	-1	0	0
8	1.62	221.22	-318	-254	1	0	-1	-1	0	0
9	1.63	213.56	-461	-284	1	0	-1	-1	0	0
10	1.64	202.38	-673	-331	1	0	-1	-1	0	0
11	1.66	187.46	-898	-370	1	0	-1	-1	0	0
12	1.67	173.19	-858	-400	1	0	-1	-1	0	0
13	1.69	159.65	-815	-372	1	0	-1	-1	0	0
14	1.70	148.46	-673	-512	1	0	-1	-1	0	0
15	1.71	139.26	-553	-489	1	0	-1	-1	0	0
16	1.73	131.99	-437	-331	1	0	-1	-1	0	0
17	1.74	124.44	-454	-439	1	0	-1	-1	0	0
18	1.76	118.71	-345	-230	1	0	-1	-1	0	0
19	1.77	113.62	-306	-393	1	0	-1	-1	0	0
20	1.78	108.78	-292	-278	1	0	-1	-1	0	0
21	1.80	104.60	-251	-133	1	0	-1	-1	0	0
22	1.81	102.29	-139	-86	1	0	-1	-1	0	0
23	1.82	99.19	-187	16	1	0	-1	1	1	0
24	1.84	101.11	116	-259	1	0	1	-1	1	0
25	1.85	96.69	-266	-6	1	0	-1	-1	0	0

Table 3 Hamming distance for Slope when magnitude of f_0 change was divided by the time interval at 10 msec level

Data Point Num	Time	Interpolated f_0	Slope Subject	Slope Target	Hamming Slope	Hamming Slope Sum	Parsons Subject	Parsons Template	Hammig Parsons	Hamming Parsons Sum
1	1.52	243.00	0.0	0.0	0	64	0	0	0	30
2	1.53	241.94	-0.5	-0.5	0	0	0	0	0	0
3	1.55	240.39	-0.8	-1.8	1	0	0	0	0	0
4	1.56	239.19	-0.6	-4.5	1	0	0	-1	1	0
5	1.58	236.20	-1.5	-2.3	1	0	0	-1	1	0
6	1.59	231.21	-2.5	-1.8	0	0	-1	0	1	0
7	1.60	226.49	-2.4	-2.1	0	0	-1	-1	0	0
8	1.62	221.22	-2.6	-2.2	1	0	-1	-1	0	0
9	1.63	213.56	-3.8	-2.5	1	0	-1	-1	0	0
10	1.64	202.38	-5.6	-2.9	1	0	-1	-1	0	0
11	1.66	187.46	-7.5	-3.2	1	0	-1	-1	0	0
12	1.67	173.19	-7.1	-3.5	1	0	-1	-1	0	0
13	1.69	159.65	-6.8	-3.2	1	0	-1	-1	0	0
14	1.70	148.46	-5.6	-4.4	1	0	-1	-1	0	0
15	1.71	139.26	-4.6	-4.2	1	0	-1	-1	0	0
16	1.73	131.99	-3.6	-2.9	1	0	-1	-1	0	0
17	1.74	124.44	-3.8	-3.8	0	0	-1	-1	0	0
18	1.76	118.71	-2.9	-2.0	1	0	-1	0	1	0
19	1.77	113.62	-2.5	-3.4	0	0	-1	-1	0	0
20	1.78	108.78	-2.4	-2.4	0	0	-1	-1	0	0

Hamming Distance for Parsons' Code Measurement

The second Hamming distance value was obtained as follows. After the two trajectories (template vs. participant's) were normalized, the f_0 values of two consecutive time points were compared. The above obtained Slope values were used to make judgments about the magnitude of change. If a change greater than 2 Hz was observed in the f_0 value in the upward direction (> +2 Hz), the value of "+1" was assigned. If a change greater than 2 Hz was observed in the f_0 value in the downward direction (-2 Hz>), the value of "-1" was assigned. If a change equal to or less than 2Hz ($\pm$2Hz $\geq$ $\geq$ -2Hz) was made in the f_0 values between two consecutive time points, the value of "0" was assigned. The 2 Hz criterion was chosen because 2 Hz was the median value used in the just noticeable difference (JND) testing method by Tervaniemi et al. (2005). They examined ERP responses to JND tasks and used a 1 Hz decrease after correct discrimination between two consecutive tones and a 3 Hz increase following a false answer. The Hamming distance between Parsons' code measurements of two trajectories (template vs. participant's) was calculated using the same method (Hamming distance, HD) that was used for Slope measurement.

When the differences in the measurement values between two trajectories were obtained using Hamming distance, the Hamming distance values were divided by the number of syllables. Again, the Hamming distance difference per syllable was obtained between the hamming distance values of Slope measurement and Parsons' code measurement. These Hamming distance difference values per syllable were obtained for each trial of each participant and were used for comparison between sequence length conditions and groups.

Table 4 Hamming distance for Parsons' codes after comparing participant' and template f_0 trajectories

Data Point Num	Time	Interpolated f_0	Slope Subject	Slope Target	Hamming Slope	Hamming Slope Sum	Parsons Subject	Parsons Template	Hammig Parsons	Hamming Parsons Sum
1	1.52	243.00	0.0	0.0	0	64	0	0	0	30
2	1.53	241.94	-0.5	-0.5	0	0	0	0	0	0
3	1.55	240.39	-0.8	-1.8	1	0	0	0	0	0
4	1.56	239.19	-0.6	-4.5	1	0	0	-1	1	0
5	1.58	236.20	-1.5	-2.3	1	0	0	-1	1	0
6	1.59	231.21	-2.5	-1.8	0	0	-1	0	1	0
7	1.60	226.49	-2.4	-2.1	0	0	-1	-1	0	0
8	1.62	221.22	-2.6	-2.2	1	0	-1	-1	0	0
9	1.63	213.56	-3.8	-2.5	1	0	-1	-1	0	0
10	1.64	202.38	-5.6	-2.9	1	0	-1	-1	0	0
11	1.66	187.46	-7.5	-3.2	1	0	-1	-1	0	0
12	1.67	173.19	-7.1	-3.5	1	0	-1	-1	0	0
13	1.69	159.65	-6.8	-3.2	1	0	-1	-1	0	0
14	1.70	148.46	-5.6	-4.4	1	0	-1	-1	0	0
15	1.71	139.26	-4.6	-4.2	1	0	-1	-1	0	0
16	1.73	131.99	-3.6	-2.9	1	0	-1	-1	0	0
17	1.74	124.44	-3.8	-3.8	0	0	-1	-1	0	0
18	1.76	118.71	-2.9	-2.0	1	0	-1	0	1	0
19	1.77	113.62	-2.5	-3.4	0	0	-1	-1	0	0
20	1.78	108.78	-2.4	-2.4	0	0	-1	-1	0	0

2.8 ACOUSTIC ANALYSIS

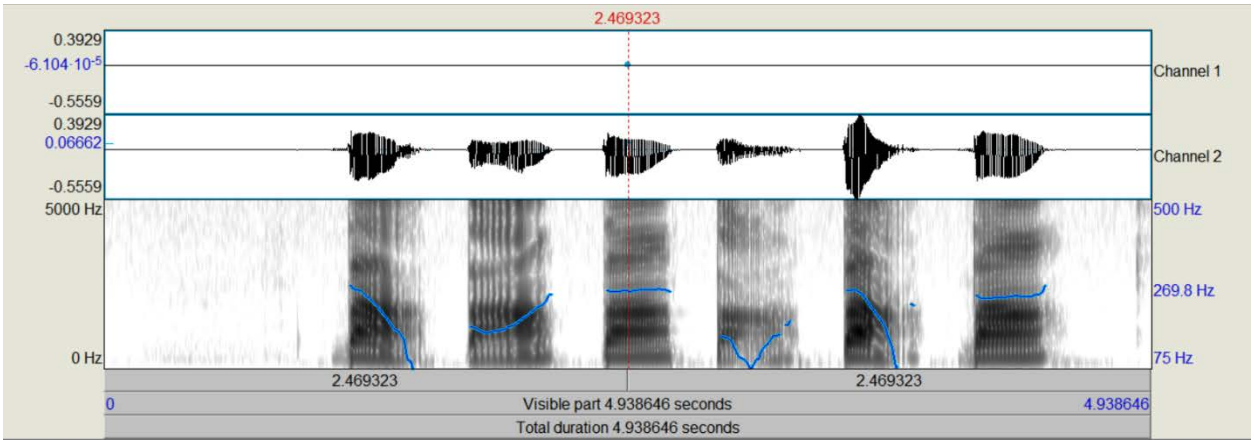

Figure 17 A waveform and spectrogram of Praat program for Tone 4-2-1-3-4-1 sequence (/bàbábābǎbàbā/)

The f_0 contour was obtained for further analysis to answer the research questions. The f_0 value for each lexical tone was obtained from the vowel segment of the syllable, because this part was known to bear the tone (Xu et al., 2004). Praat software displayed the f_0 contour (the blue line in Figure 17), and the f_0 value at each time point was collected at every 5 ms interval using the Pitch listing function. From these data, the time and f_0 value coordinates were obtained (e.g., (x, y) = (*time*, f_0 *value*)).

In addition, each syllable in the target tone sequences produced by participants was isolated for the analysis for the within-subject GMP error for each tone (the third research question in section 1.3.3.3). Most of the procedures used to identify boundaries for each tone syllable was performed semi-automatically by the python program, however, both spectrograms and waveforms in the Praat software were examined to identify noise in the signal and meaningful boundaries of each tone syllable whenever needed. The syllable onset begins with the aspiration period of the bilabial stop consonant /b/, which corresponds to the voice onset time (VOT). The

closure duration for the bilabial stop consonant /b/ was not included in the total duration for the /ba/ syllable because the onset of the closure was not reliably identifiable in the acoustic signals. The syllable offset was determined when the pitch extracting program did not detect an f_0 value or when the end of vowel /a/ was located during visual and auditory examination in both spectrogram and waveforms. The end of the vowel was indicated by the point where the complex waveform ended and where the amplitude of the speech signal attenuated abruptly on a wideband spectrogram and when no meaningful voicing was detected auditorily.

2.9 PERCEPTUAL ANALYSIS

The accuracy of tone production was judged perceptually by the author and by a native Mandarin Chinese speaker during both the mono-syllable practice phase and the tone sequence production phase. The author judged the accuracy of the tone production based on visual and auditory examination of the f_0 contours during the mono-syllable practice phase. These judgments were used to determine whether the study participant acquired the ability to produce each tone accurately or required additional practice. The author was trained to make this visual and auditory judgment before the data collection began. A 90% correspondence level for mono-syllable and 80% correspondence level for tone sequences with those judged by three native Mandarin listeners were achieved in the auditory perceptual judgment tasks.

The procedure to train the first author for correctness judgement was as following: Eleven randomly selected males, produced each target speech sound (4 mono-syllable, 12 tone sequences) five times, generating eighty utterances for each speaker. The sound files were presented, via E-prime, to three native Mandarin speakers who met all screening criteria. Each judge also

orthographically transcribed their responses. Errored speech productions were determined when two out of three judges agreed. This correctness judgment result was used for the visual and auditory training of the author before the data collection process began.

After data collection was completed, five percent of the tone sequence of non-Mandarin participants' productions were presented to untrained native Mandarin listeners for reliability testing. They listened to instructions about the task and judged for accuracy. Each production was categorized as "correct" or "incorrect" based on whether the production matched the intended production. This perceptual judgment was used to calculate inter-judger reliability.

2.9.1 Reliability

Three listeners made the off-line judgment after the experiment for the target speech samples from the tone sequence production condition. Two sound files (5% of the target speech samples) were randomly selected from each participant from the tone sequence production condition. The total number of speech stimuli evaluated by these native Mandarin listeners were 96 files (48*2=96). The author's intra-rater reliability was 97% and inter-rater reliabilities between the author and each of these three native Mandarin listeners were 92%, 99%, and 93% respectively.

2.10 STATISTICAL ANALYSIS

A Hierarchical generalized linear model (HGLM) was used to evaluate the main and interaction effects of sequence length condition and group. The HGLM analysis was performed after treating the data structure as a means model in which observations (or trials) were nested within

individuals. This model was chosen instead of an individual growth model, in which 'trial' is treated as a growth factor, so that the model can be parsimonious following the law of Occam's razor. Each observation (or 'trial') was considered as a first level unit and each individual was treated as a second level unit because having at least 24-30 levels for the second level variable was recommended by Maas and Hox (2005). The 'sequence length condition' (SLC) was entered as a level-1 predictor and the 'group' was treated as a level-2 predictor as the "sector" notion in the fourth model of Singer's (1998) article. The dependent variables were GMP errors per syllable, parameter variability, Hamming distance difference between two measurements per syllable, RT, average ISI, and the Ratio of RT/average ISI. Separate analyses were performed for each dependent variable. The p-value was set at 0.05 for all comparisons for each model testing. Bonferoni adjustment was made on alpha level for post-hoc analyses. The equations for the HGLM analyses were as following:

Level 1 equation: $Y_{ij}=\beta_{0j}+\beta_{1j}(SLC_{ij})+ r_{ij}$ (i: trial, j: each individual) (1)

Level 2 equations:

$\beta_{0j}= \gamma_{00}+ \gamma_{01}(Group_j)+u_{0j}$ (2)

$\beta_{1j}= \gamma_{10}+ \gamma_{11}(Group_j)+u_{1j}$

$r_{ij} \sim N(0, \sigma^2)$ and $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$

Combined equation:

$Y_{ij}=[\gamma_{00}+\gamma_{01}(Group_j)+\gamma_{10}(SLC_{ij})+\gamma_{11}(Group_j)(SLC_{ij})]_{fixed\ \ effect}+[u_{0j}+u_{1j}(SLC_{ij})+r_{ij}]_{random\ \ effect}$ (3)

Example model building when dependent variable is "GMP error":

γ_{00}: the mean GMP error for native Mandarin speaker group

γ_{01}: the mean GMP error difference between two speaker groups

γ_{10}: the mean relationship between SLC and GMP error in native Mandarin speaker group

γ_{11}: the mean difference in SLC and GMP error Slopes between two speaker groups

u_{0j}: the unique effect of j^{th} individual on mean GMP error holding group constant (variation in intercepts between individuals) (τ_{00})

u_{1j}: the unique effect of j^{th} individual on the SLC and GMP error Slope holding group constant (variation in Slopes) (τ_{11})

r_{ij}: variation across trials within-individual (σ^2)

τ_{10} or τ_{01}: correlation between intercept and Slope

First, the intra-class correlation (ICC) was obtained in the unconditional model, in which no predictor was included for the analysis. From the results of this unconditional model, the following formula was used to check ICC value; ICC= $\tau_{00}/(\tau_{00} + \sigma^2)$. If the ICC was larger than zero, it meant that there was variability due to the grouping effect and the use of HGLM was justified.

Second, the assumptions for normality and homoscedasiticity were investigated and any necessary adjustment was made using SAS 9.4 codes. Because all dependent variables were continuous variables and the residuals for dependent variables were non-normally distributed, PROC GLIMMIX commands with a gamma distribution were used for HGLM analyses. As noted by brackets and small descriptors in the equation (3), Group, SLC, Group*SLC were entered as fixed effects in the model statement and intercept and SLC were entered as random effects in the random statement. Example SAS codes used for analysis appear in Appendix E.

Third, the appropriate covariance structure was investigated. The information criteria of AIC and BIC were used to determine the best fitting covariance structure.

Last, the significance of each effect caused by each predictor was examined in the tables labeled “Type III Tests of Fixed Effects” and “Solutions for Fixed Effects” in the SAS output.

3.0 RESULTS

3.1 SCREENING RESULTS

Thirty-nine non-Mandarin speakers and twenty-five native Mandarin speakers were recruited. After these participants consented to participate, thirteen non-Mandarin speakers and one Mandarin speaker were screened out or discontinued their participation. Two non-Mandarin speakers quit their participation after they passed the screening tests. Six non-Mandarin participants and one native Mandarin participant were screened out in the tone discrimination test. Four non-Mandarin participants did not pass the CRTT-L screening criteria. One non-Mandarin participant did not meet the CSID criteria. One non-Mandarin speaker's data were collected in a different protocol using different model speaker's sound files for the mono-syllable practice phase, so this participant's data were not included in the final analysis. One non-Mandarin speaker completed all parts of the study, but his data were excluded from the data analysis because this participant produced errors on more than 50% of the target responses when substitution, addition or deletion of the phonemes or tones were judged. After excluding all these participants, twenty-four non-Mandarin speakers and twenty-our native Mandarin speakers' data were included in the final data analysis.

The screening results of these participants are summarized in tables 5 and 6. Information about the degree of musical training experience in both groups was collected for descriptive purposes and the results are provided in Appendix D. Because normality and homogeneity of variance assumptions were not satisfied after a two-step normalization process in many screening measures, a Mann-Whitney U non-parametric analysis was performed to compare the screening

results between native Mandarin and non-Mandarin speaker groups. As a result, the total years of education were significantly higher in the native Mandarin group than the non-Mandarin group (*p=.000*). AMR performances were also significantly faster in the native Mandarin group than the non-Mandarin group (/pΛ/, *p=.002*, /tΛ/, *p=.021*, /kΛ/, *p=.001*) while the SMR performances were not different (*p=.439*). Although the experimenter intended to perceptually judge respiratory, phonatory, articulatory and resonatory functions while participants produced connected speech in English during the story retell procedure (SRP), the contents of the speech samples were also analyzed. The native Mandarin speakers produced significantly fewer information units (IU) than non-Mandarin speakers (total number of IU, *p=.006*, % IU, *p=.006*). However, this finding was expected considering that the SRP task was performed in their second language (English) for the native Mandarin speakers. The statistical group comparison results are summarized in Table 7.

Table 5 Screening results for native Mandarin speaker group

Mandarin Group	Age (Yrs.)	Education (Yrs.)	DL (125Hz)	DL (1000Hz)	DL (2000Hz)	Average Sustained Phonation Duration	AMR /pʌ/ (num/sec)	AMR /tʌ/ (num/sec)	AMR /kʌ/ (num/sec)	SMR /pʌtʌkʌ/ (num/sec)	SMR /pʌtʌkʌ/ (syllable/sec)	CRTT-L Mean Score	CSID_R	Total Infor-mation Units (IU)	%IU (%)
M1	29	20	1.33	6.33	8.00	17.92	6.92	7.16	5.59	2.18	6.54	14.54	-5.14	178.00	39.73
M2	19	14	2.33	6.17	9	24.91	6.8	7.33	6.39	2.51	7.53	14.29	1.67	203	45.31
M3	19	13	1	1.33	2	26.14	6.91	6.9	6.69	2.7	8.1	14.67	6.35	231	51.56
M4	24	18	2.33	5	24	11.36	6.85	7.82	7.13	2.44	7.32	14.68	2.81	211	47.10
M5	22	16	1.33	1	21.67	12.48	6.5	6.66	5.93	2.35	7.05	14.32	13.65	146	32.59
M6	27	21	1	1.33	2.67	15.99	6.88	7.1	6.22	2.38	7.14	13.87	20.81	126	28.13
M7	22	16	1.67	7.67	18	20.85	7.56	7.75	7.82	3.06	9.18	14.26	0.61	176	39.29
M8	27	21	1.33	3.67	2	17.13	7.2	7.39	6.46	2.31	6.93	13.83	12.96	147	32.81
M9	23	17	1	9.33	18.67	11.65	6.98	6.95	6.1	2.24	6.72	14.89	4.77	127	28.35
M10	23	17	1.67	11.67	11	15.89	6.6	6.21	6.38	2.25	6.75	14.64	-9.75	219	48.88
M11	25	19	1	6.67	9.33	16.61	7.33	7.3	7.37	2.4	7.2	14.26	17.48	222	49.55
M12	29	23	1.33	4.67	8.67	25.14	7.07	8.15	8.52	2.92	8.76	14.84	6.33	204	45.50
M13	23	20	1.67	8.33	2	20.29	7.95	6.53	6.25	2.25	6.75	14.75	19.04	142	31.70
M14	22	16	1.33	5	8	20.66	7.27	7.84	7.38	2.46	7.38	14.47	5.86	210	46.88
M15	24	16	1.67	2.67	10.67	10.67	8.64	7.25	5.87	2.21	6.63	13.79	16.95	145	32.37
M16	25	19	1	2.33	2.33	8.7	6.43	7.64	6.72	2.3	6.9	14.65	20.13	200	44.64
M17	22	16	1	1	4.33	18.07	7.46	7.62	6.69	2.53	7.59	14.17	15.84	217	48.44
M18	22	17	1	1.33	5.67	21.93	7.24	7.33	6.97	2.46	7.38	14.67	18.65	146	32.59
M19	21	15	2.33	13.67	8.67	13.31	6.65	7.19	6.69	2.29	6.87	14.67	-10.56	155	34.60
M20	22	15	2	12	21.33	20	7.44	7.87	6.78	2.29	6.87	14.27	3.33	189	42.19
M21	23	17	1	3.33	6.33	20.47	6.76	8.83	7.36	2.51	7.53	14.47	-4.69	161	35.94
M22	29	22	2	1.67	7.33	12.08	7.68	7.25	7.05	2.57	7.71	14.61	23.78	234	52.23
M23	25	19	1.33	5.33	3.33	17.54	6.83	7.64	6.79	2.46	7.38	14.12	7.57	148	33.04
M24	23	18	2	2.67	18.67	31.1	6.89	8.06	6.04	2.5	7.5	14.59	5.08	202	45.09
Mean	23.75	17.71	1.49	5.17	9.74	17.95	7.12	7.41	6.72	2.44	7.32	14.43	8.06	180.79	40.35
SD	2.82	2.56	.47	3.71	6.94	5.54	.50	.57	.67	.21	.64	.31	9.89	34.84	7.77

Table 6 Screening results for non-Mandarin speaker group

non-Mandarin Group	Age (Yrs.)	Education (Yrs.)	DL (125Hz)	DL (1000Hz)	DL (2000Hz)	Average Sustained Phonation Duration	AMR /pʌ/ (num/sec)	AMR /tʌ/ (num/sec)	AMR /kʌ/ (num/sec)	SMR /pʌtʌkʌ/ (num/sec)	SMR /pʌtʌkʌ/ (syllable/sec)	CRTT-L Mean Score	CSID_R	Total Infor-mation Units (IU)	%IU (%)
E1	26	16	1.33	6.33	8.33	15.82	6.2	6.8	6.2	1.56	4.68	14.85	15.29	240	53.57
E2	29	18	1	2	7	21.42	6.64	7.73	6.29	2.2	6.6	14.68	6.79	211	47.10
E3	23	17	1	2.67	16.33	15.89	5.87	7.33	6.03	2.56	7.68	14.35	9.24	237	52.90
E4	26	14	2	14	22.33	26.69	7.45	8.26	7.64	2.51	7.53	14.50	19.27	178	39.73
E5	28	19	1	2.33	3.33	14.4	6.79	8.75	7.13	2.9	8.7	14.81	19.27	263	58.71
E6	19	14	2.67	9	11	24.56	7.1	6.69	5.65	1.78	5.34	14.67	17.4	169	37.72
E7	19	14	1	1.33	3.67	19.25	6.43	6.71	5.93	2.27	6.81	14.37	13.11	215	47.99
E8	19	14	2.67	7	6.33	9.1	6.24	6.78	5.77	2.59	7.77	14.64	11.28	248	55.36
E9	19	14	2.33	2.67	7	19.07	6.11	6.03	5.57	2.48	7.44	14.50	-2.68	198	44.20
E10	19	16	2	3	6	30.37	6.09	7.1	5.42	2.74	8.22	14.40	-0.26	246	54.91
E11	21	15.5	2	2	8	14.14	7.1	7.49	6.11	2.11	6.33	14.72	20.15	277	61.83
E12	20	14	1.33	2.67	7.33	14.24	6.67	6.89	5.84	2.39	7.17	14.94	24.3	178	39.73
E13	19	14	1	2.67	18.33	12.87	7.2	7.02	5.94	2.18	6.54	14.53	13.72	122	27.23
E14	23	16	1	2	2	38.17	6.28	7.27	6.22	2.59	7.77	14.70	0.16	203	45.31
E15	21	15	2.67	3.67	3.33	14.68	6.67	6.1	5.29	2.34	7.02	14.69	19.27	282	62.95
E16	23	16	2	19.67	13	23.4	7.28	7.41	6.81	2.64	7.92	14.61	-9.27	235	52.46
E17	25	13	1.33	17.33	16	11.27	5.89	5	4.85	1.79	5.37	14.63	16.65	131	29.24
E18	19	13	1	4.67	1.33	13.08	7.56	8.24	7.32	2.8	8.4	14.67	16.59	209	46.65
E19	20	13	1.33	1	8	13.13	6.78	6.79	5.98	1.9	5.7	14.58	16.35	176	39.29
E20	30	16	1	1.67	1.67	14.23	7.02	6.83	5.67	1.92	5.76	14.43	10.59	242	54.02
E21	21	15	2.33	2	11	15.8	6.13	6.47	5.96	2.51	7.53	14.74	17.56	213	47.54
E22	20	15	1.83	7.83	7.33	18.33	6.56	7.16	5.62	2.4	7.2	14.37	17.52	185	41.29
E23	24	16	1.33	4.33	11.33	19.79	6.64	6.35	6.08	2.13	6.39	14.24	13.08	241	53.79
E24	21	15	1	16	17.67	32.04667	6.16	6.13	5.56	2.4	7.2	14.68	10.13	222	49.55
Group Mean	22.25	15.10	1.59	5.74	9.07	18.82	6.62	6.97	6.04	2.32	6.96	14.60	12.31	213.38	47.63
SD	3.45	1.53	0.62	5.51	5.74	7.18	0.49	0.80	0.64	0.35	1.04	0.17	8.19	41.21	9.20

Table 7 Mann-Whitney U group comparisons on screening measures

Variable Name	Passing Criteria	native Mandarin		non-Mandarin		p-value
		Mean	SD	Mean	SD	
Age	19-30	23.75	2.82	22.33	3.45	.055
Education		17.71	2.56	15.06	1.55	.000*
DL (125Hz)	≤2.81Hz	1.49	.47	1.59	0.62	.733
DL (1000Hz)	≤22.50Hz	5.17	3.71	5.13	5.12	.934
DL (2000Hz)	≤45.00Hz	9.74	6.94	8.68	5.44	.757
Average Sustained Phonation Duration	9.45-33.13	17.95	5.54	18.44	6.67	.975
AMR /pʌ/ (num/sec)	4.3-7.22	7.12	.50	6.61	0.51	.002*
AMR /tʌ/ (num/sec)	4.4-8.8	7.41	.57	6.97	0.80	.021*
AMR /kʌ/ (num/sec)	4.3-9.1	6.72	.67	6.04	0.64	.001*
SMR /pʌtʌkʌ/ (num/sec)		2.44	.21	2.31	0.35	.439
SMR /pʌtʌkʌ/ (syllable/sec)	4.72-8.68	7.32	.64	6.94	1.04	.439
CRTT-L Mean Score	≥14.17 for non-Mandarin ≥12.36 for native Mandarin	14.43	.31	14.60	0.17	.058
CSID-R	≤24.3	8.06	9.89	12.56	8.21	.132
Total Information Unit (IU)		180.79	34.84	212.54	41.23	.006*
% IU		40.35	7.77	47.63	9.20	.006*

3.2 DESCRIPTIVE AND STATISTICAL RESULTS

The research questions below are organized by the dependent variable examined. For each dependent variable, the interaction effect was examined between two independent variables: the native Mandarin and non-Mandarin speaker groups and sequence length conditions (3, 6 and 9 syllables) or between group and Whole_1st or GMP_2_8 conditions. The Whole_1st and GMP_2_8 variables will be explained later. If a significant interaction was detected, a post-hoc analysis of simple main effects was explored to determine the locus of the difference. A simple main effect question asks if each independent variable is significantly different across one level of the other independent variable (e.g., if the sequence length effect is significantly different for native Mandarin speakers). If there was no significant interaction, main effects for each independent variable, averaged across the levels of the other independent variable, were computed. All the statistical analysis results are summarized in Appendix G.

3.2.1 Parameter Variability

In the analysis of parameter variability, the parameter used for scaling was measured from the whole sequence and compared to another parameter, measured from the first syllable of each sequence. In this comparison, a new variable named "Whole_1st" was used to indicate the contextual condition from which the parameters were measured; whether from the whole sequence or from the 1st syllable. For simplicity of modeling, first, the time and amplitude parameter of the 1st syllable were subtracted from the parameters of the whole sequence, and the interaction between group and sequence length condition was tested. Next, because this subtracted value determines only the difference in the parameter values between the 1st syllable and the whole sequence, the

differences in the absolute parameter values across different Groups and Whole_1st conditions were also tested. The interactions between the Whole_1st and Group in the time and f_0 magnitude parameters were examined from the nine-syllable sequence length condition.

3.2.1.1 Time Parameter Variability When Parameter Difference Was Examined

The time parameter, obtained from the 1st syllable, was subtracted from the time parameter obtained from the whole sequence. This variable then underwent square-root transformation and was considered to represent the time parameter variability for statistical analysis (Time_param_sub_sqrt). The intra-class correlation (ICC) for the Time_param_sub_sqrt in the unconditional model was 39.13% (intercept: 0.018, residual: 0.028). This result justified the need for using the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. The homoscedasticity assumption was satisfied (*p=.7525* for interaction term Group*Sequence Length Condition).

As in Figure 18, the interaction of Group and Sequence Length Condition on Time_param_sub_sqrt was not significant (*F(2,60)=.76, p=.4725*). Thus, the pattern of change in this measurement was similar between the two groups. The main effect of Group on Time_param_sub_sqrt was significant (*F(1,36)=10.89, p=.0022*). Counter to preditictions, native Mandarin speakers were significantly higher on Time_param_sub_sqrt by .1742 than the non-Mandarin speakers. The main effect of Sequence Length Condition on Time_param_sub_sqrt was not significant (*F(2,60)=1.88, p=.1615*). The differences in the time parameters between whole sequence and 1st syllable were maintained similarly across the Sequence Length Conditions. In Figure 18, the black bar represents the standard deviation for the native Mandarin speakers at each condition, and the red bar represents the standard deviation for the non-Mandarin speakers at each

condition. In the subsequent figures, the same colored bars will represent the standard deviations of the measured dependent variables of the respective group.

Because the main effect of Sequence Length Condition was not significant, only the nine-syllable sequence condition was examined to investigate the interaction effect between Whole_1st time parameter conditions and the group in section 3.2.1.2.

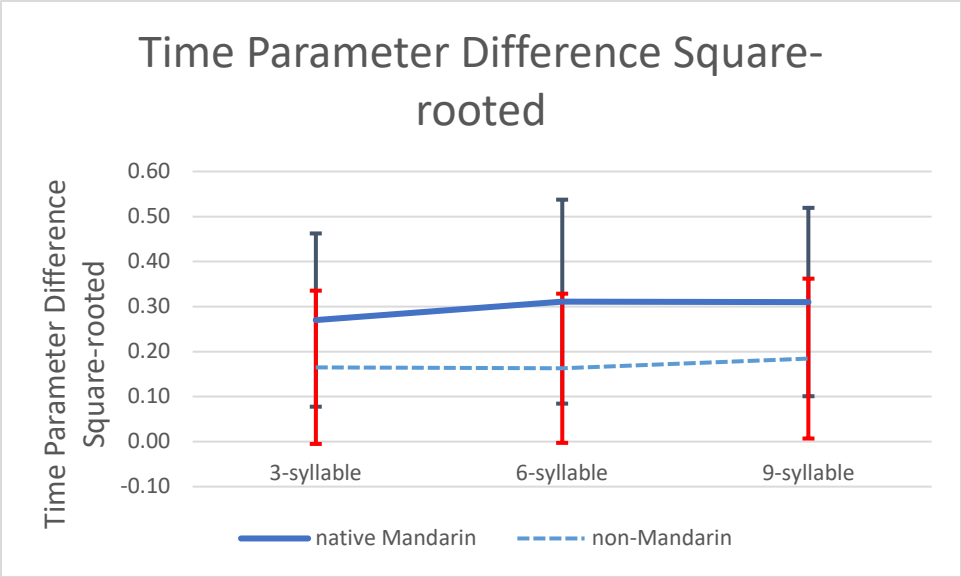

Figure 18 Square-root transformed time parameter difference between 1st syllable and whole squence

3.2.1.2 Time Parameter Variability in Nine-Syllable Sequence Length Condition

The time parameter measured in the first syllable was larger than the one found in the whole sequence for the non-Mandarin speaker group. Similarly, the time parameter during the first syllable production was similar to or larger than the one for the whole sequence production in the native Mandarin speaker group. These patterns were observed in all three sequence length conditions. Because a time parameter below 1 means that the productions were slower than the speech rate used for the template production, a time parameter for the whole sequence that is

smaller than the time parameter for the 1st syllable means that the non-Mandarin speaker group was slower in producing longer tone sequences. These data are summarized in Table 8 and displayed in Figure 19. The statistical significance of this pattern was examined at the nine-syllable sequence level in the following paragraphs.

Table 8 Time parameters in the Whole sequence vs. 1st syllable productions across Sequence Length Conditions and Groups (Cf., Parameter value ranges between 0.1-2.0 where '1' represents no need for scaling between participant's f_0 trajectory and the template f_0)

		3-syllable		6-syllable		9-syllable	
		native Mandarin	non-Mandarin	native Mandarin	non-Mandarin	native Mandarin	non-Mandarin
Whole Sequence	Mean	1.05	0.68	1.06	0.64	1.05	0.67
	SD	0.26	0.15	0.27	0.15	0.25	0.14
1st Syllable	Mean	1.08	0.85	1.06	0.84	1.10	0.89
	SD	0.39	0.43	0.37	0.45	0.39	0.47

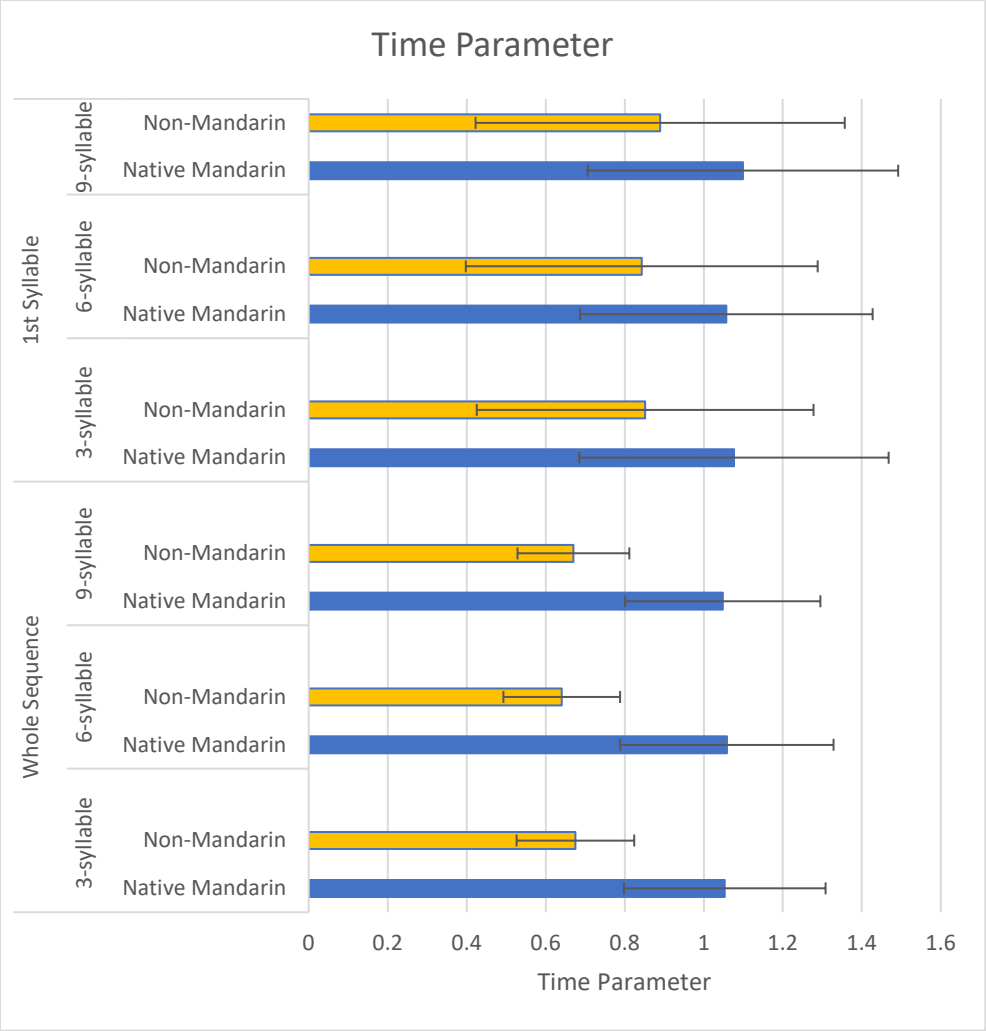

Figure 19 Time parameters in the 1st syllable and whole sequence productions

To reduce the complexity of the model, time parameter variability was examined in the nine-syllable sequence only by comparing time parameters (Time_param) obtained from the whole sequence and that from the 1st syllable. These data were log-transformed for statistical analysis (Time_param_log). The intra-class correlation (ICC) for the Time_param_log in the unconditional model was 68.86% (intercept: 0.106, residual: 0.048). This result justified the need for using the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. The homoscedasticity assumption

was tested in the full model in which covariates were specified. The Whole_1st, Group, Whole_1st*Group were entered in the model statement. The intercept and Whole_1st were entered in the random statement. The homoscedasticity assumption was satisfied (*p=.0846* for interaction term Whole_1st*Group).

The interaction of the Whole_1st and the Group on Time_param_log was not significant (*F(2,14)=3.38, p=.0875*). The main effect of Whole_1st on Time_param_log was not significant (*F(1,14)=.09, p=.7674*). Thus, the observed difference in time parameter was not significantly different between the first syllable and whole sequence condition, at least at the nine-syllable length condition. However, the main effect of Group on Time_param_log was significant (*F(2,14)=4.79, p=.0462*). When averaged across Whole_1st conditions, the Time_parm_log obtained from the native Mandarin speakers was significantly higher than that obtained from the non-Mandarin speakers in the nine-syllable sequence length condition (see Figure 20).

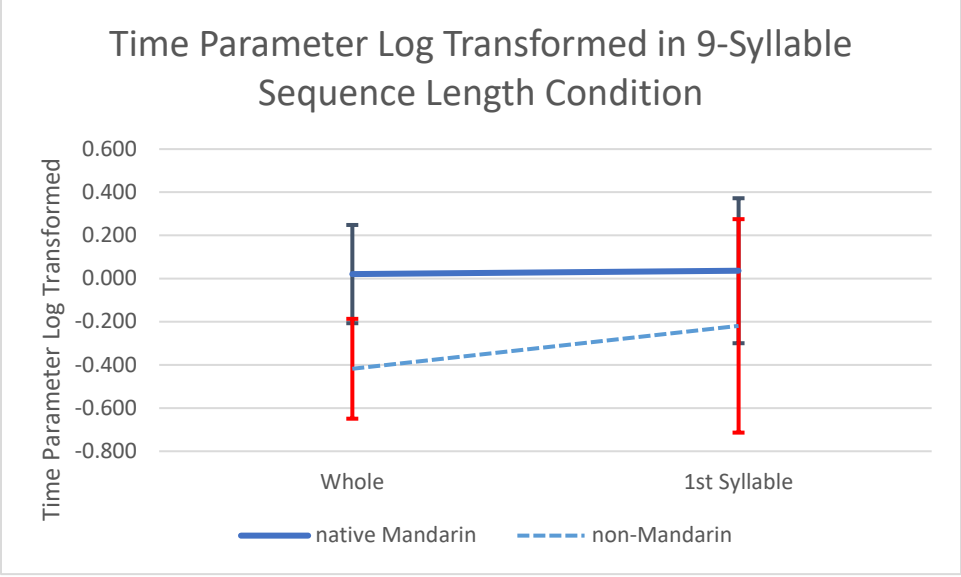

Figure 20 Log-transformed time parameter change across Group and Whole_1st conditions

In summary, when the interaction between Group and Whole_1st conditions was examined for Time_param_log after restricting the analysis to nine-syllable sequence length condition, neither the interaction between the Whole_1st and Group, nor the main effect of Whole_1st conditions was significant. This fact that neither the interaction nor the main effect was significant suggested that the time parameter related to speech rate did not change significantly within the nine-syllable utterances. However, the main effect of Group was significant. After examing the original time parameter values in Table 8, it was concluded that native Mandarin speakers spoke faster than non-Mandarin speakers across all syllable length conditions.

3.2.1.3 f_0 magnitude Parameter Variability When Parameter Difference Was Examined

The amplitude parameter, which was obtained from the 1st syllable, was subtracted from the f_0 magnitude parameter obtained from the whole sequence (f_0_param_sub). The f_0_param_sub was square-root transformed (f_0_param_sub_sqrt). The intra-class correlation (ICC) for the f_0_param_sub_sqrt in the unconditional model was 29.90% (intercept: 0.018, residual: 0.043). This result justified the need for using the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. The homoscedasticity assumption was satisfied (*p=.4713* for interaction term Group* Sequence Length Condition).

The interaction of Group and Sequence Length Condition on f_0_param_sub_sqrt was not significant (*F(2,79)=2.47, p=.0909*). The main effect of Group on f_0_param_sub_sqrt was not significant (*F(1,46)=.15, p=.6969*). The main effect of Sequence Length Condition on f_0_param_sub_sqrt was significant (*F(2,79)=3.14, p=.0489*) (see Figure 21).

Post-hoc tests of marginal comparisons among sequence length conditions after the Bonferoni correction (α=.0167=.05/3) revelaed that none of the marginal comparisons were

significant. When averaged across groups, the f_0_param_sub_sqrt in Sequence Length Condition 1 (three-syllable sequence length condition) was not significantly different from that of Sequence Length Condition 2 (six-syllable sequence length condition) (t(79)=1.57, *p=.1209*). The f_0_param_sub_sqrt in Sequence Length Condition 1 was not significantly different from that in Sequence Length Condition 3 (nine-syllable sequence length condition) (t(79)=2.13, *p=.0359*). Finally, the f_0_param_sub_sqrt in Sequence Length Condition 2 was not significantly different from that in Sequence Length Condition 3 (t(79)=.67, *p=.5017*).

Similarly to the time parameter, a significant interaction between Sequence Length Condition and group was not observed in the f_0_param_sub_sqrt, and the pattern of change in the f_0 magnitude parameter was parallel between two groups. Although the f_0_param_sub_sqrt was significantly different among Sequence Length Conditions, the marginal comparisons during post-hoc tests did not locate the differences in the f_0_param_sub_sqrt among Sequence Length Conditions. Next, the f_0 magnitude parameter variability was compared between the 1st syllable and the whole sequence in the nine-syllable sequence only.

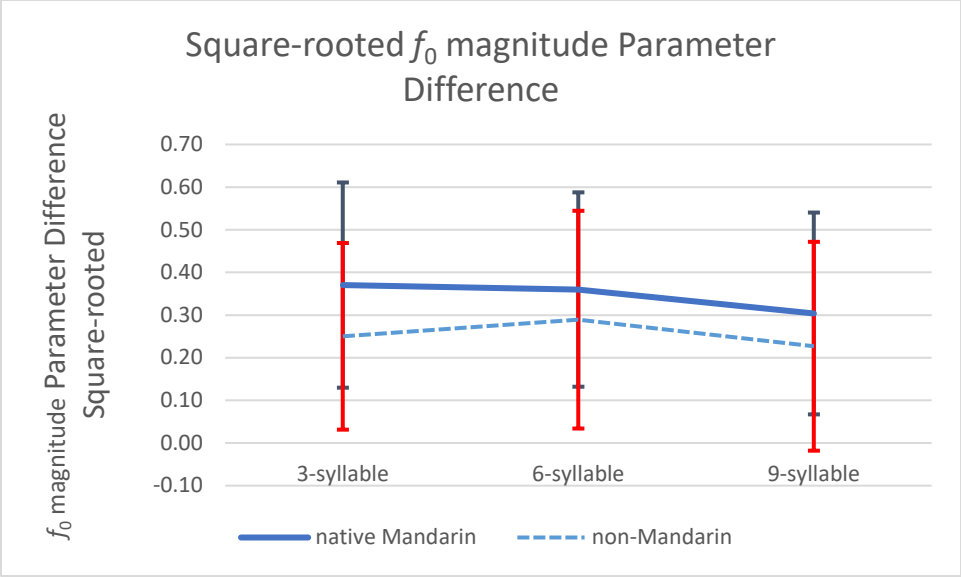

Figure 21 Square-root transformed f_0 magnitude parameter difference between 1st syllable and whole sequence

3.2.1.4 f_0 magnitude Parameter Variability in Nine-Syllable Sequence Length Condition

Because the above variable, f_0_param_sub_sqrt, examined only the difference in the f_0 magnitude parameters between 1st syllable and whole sequence, in order to examine the differences in the absolute f_0 magnitude parameter values across different Groups and Whole_1st conditions, additional analysis was performed for the actual f_0 magnitudeparameter variable at the nine-syllable sequence condition.

Similar to the time parameters, the f_0 magnitude parameters greater/smaller than a value of "1" means that the participant's f_0 trajectory required more scaling in terms of f_0 magnitude when trying to match their f_0 trajectory to the template trajectory. The f_0 magnitude parameters, measured from the whole sequence, were slightly greater than the ones measured from the 1st syllable productions in the three- and six-syllable sequence length conditions in native Mandarin speakers. Thus, it appeared that native Mandarin speakers increased their f_0 magnitude variability

toward the end of the utterance by lowering their pitch levels. The f_0 magnitude parameters were slightly smaller when measured from the whole sequence than the ones measured from the 1st syllable productions in the three- and nine-syllable conditions in the non-Mandarin speakers. Thus, it appeared that non-Mandarin speakers increased their f_0 magnitude variability toward the end of the utterance, in particular, for the three- and nine-syllable length utterances by raising their pitch levels. The f_0 magnitude parameter data are summarized in Table 9 and displayed in Figure 22. The statistical significance of this pattern was examined only at the nine-syllable sequence level.

Table 9 f_0 magnitude parameter in the Whole sequence vs. 1st syllable productions across Sequence Length Conditions and Groups (Cf., Parameter value ranges between 0.1-2.0 where '1' represents no need for scaling between participant's f_0 trajectory and the template f_0 trajectory)

		3-syllable		6-syllable		9-syllable	
		native Mandarin	non-Mandarin	native Mandarin	non-Mandarin	native Mandarin	non-Mandarin
Whole Sequence	Mean	1.07	0.64	1.04	0.69	0.99	0.62
	SD	0.42	0.39	0.38	0.43	0.37	0.36
1st Syllable	Mean	0.98	0.73	0.96	0.68	0.99	0.73
	SD	0.44	0.50	0.44	0.50	0.43	0.53

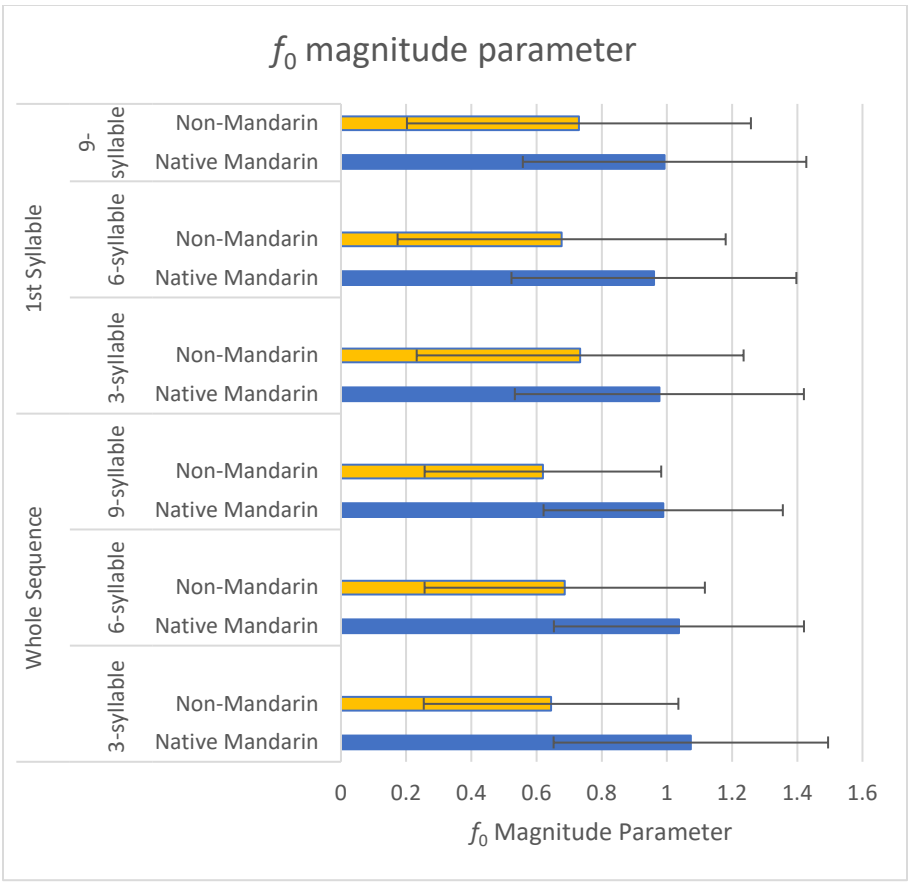

Figure 22 f_0 magnitude parameter in the 1st syllable and whole sequence productions

To reduce the complexity of the model, the f_0 magnitude parameter variability was examined in the nine-syllable sequence only by comparing the f_0 magnitude parameters (f_0_param) obtained from the whole sequence and that from the 1st syllable. The intra-class correlation (ICC) for the f_0_param in the unconditional model was 53.54% (intercept: 0.1156, residual: 0.1003). This result justified the need for using the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. The homoscedasticity assumption was tested in the full model in which covariates were specified. The Whole_1st, Group, Whole_1st*Group were entered in the model statement. The intercept and

Whole_1st were entered in the random statement. The homoscedasticity assumption was satisfied (*p=.1128* for interaction term Whole_1st*Group).

The interaction of Whole_1st and Group for the f_0_param was not significant (*F(1,46)=1.54, p=.2203*). The main effect of Whole_1st on the f_0_param was not significant (*F(1,46)=.83, p=.3669*). However, the main effect of Group on the f_0_param was significant (*F(1,46)=10.39, p=.0023*). When averaged across Whole_1st conditions, the f_0_parm obtained from native Mandarin speakers was significantly higher by .3397 than that obtained from the non-Mandarin speakers in the nine-syllable sequence length condition (see Figure 23).

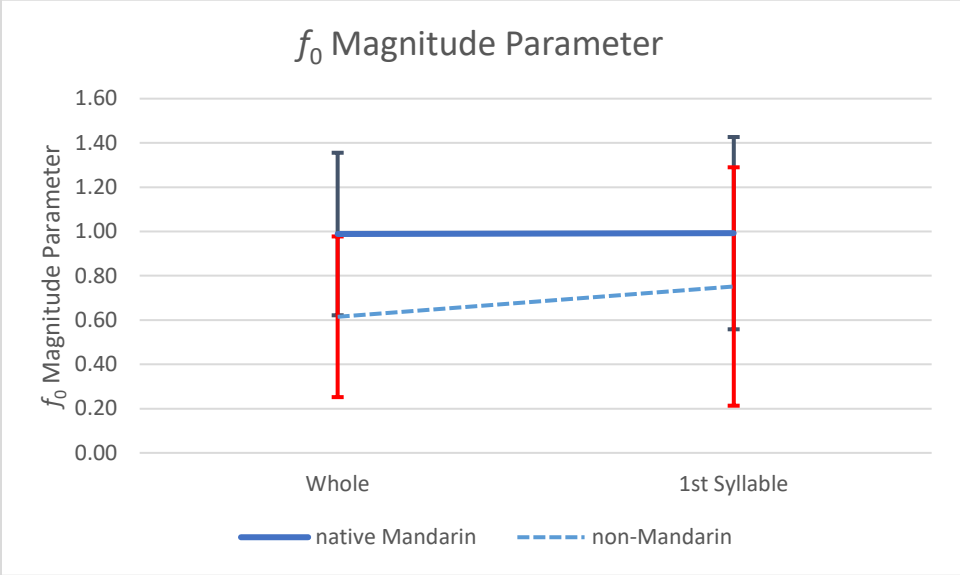

Figure 23 f_0 magnitude parameter change across Group and Whole_1st conditions

Similar to the time parameters, the result that neither the interaction between Whole_1st and Groups nor the main effect of Whole_1st conditions for the f_0 magnitude parameters were significant suggested that there was no significant f_0 parameter change during the execution of the prepared speech motor responses regardless of the learning experience.

3.2.2 GMP Errors per Syllable

3.2.2.1 Sum of Euclidean Distances per Syllable

The mean GMP errors, as measured by the Sum of the Euclidean distances per syllable, increased and then slightly decreased as the sequence length became longer in both native Mandarin and non-Mandarin speaker groups. Table 10 and Figure 24 summarize and display these data. The significance of this observed change is statistically examined in the next paragraph.

Table 10 Sum of Euclidean distances per syllable across Group and Sequence Length conditions

Sum of Euclidean distances		3-syllable	6-syllable	9-syllable
native Mandarin	Mean	320.76	492.09	452.43
	SD	255.96	270.37	260.30
non-Mandarin	Mean	705.54	879.80	854.25
	SD	379.22	418.23	394.05

The GMP errors per syllable (GMP_syl) were indexed by the sum of Euclidean distances per syllable. The intra-class correlation (ICC) for the GMP_syl in the unconditional model was 63.78% (intercept: 99044, residual: 56252). This result justified the need for using the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution. The homoscedasticity assumption was satisfied ($p=.1163$) for the interaction term Group*Sequenece Length Condition.

A significant Group by Sequence Length Condition interaction ($F(2,92)=9.31$, $p=.0002$), main effect of Group ($F(1,46)=48.10$, $p<.0001$), and main effect of Sequence Length Condition ($F(2,92)=93.62$, $p<.0001$) were observed. The native Mandarin speaker group produced

significantly lower GMP_syl by -.6543 than that of the non-Mandarin speaker group when averaged across sequence length conditions (see Figure 24).

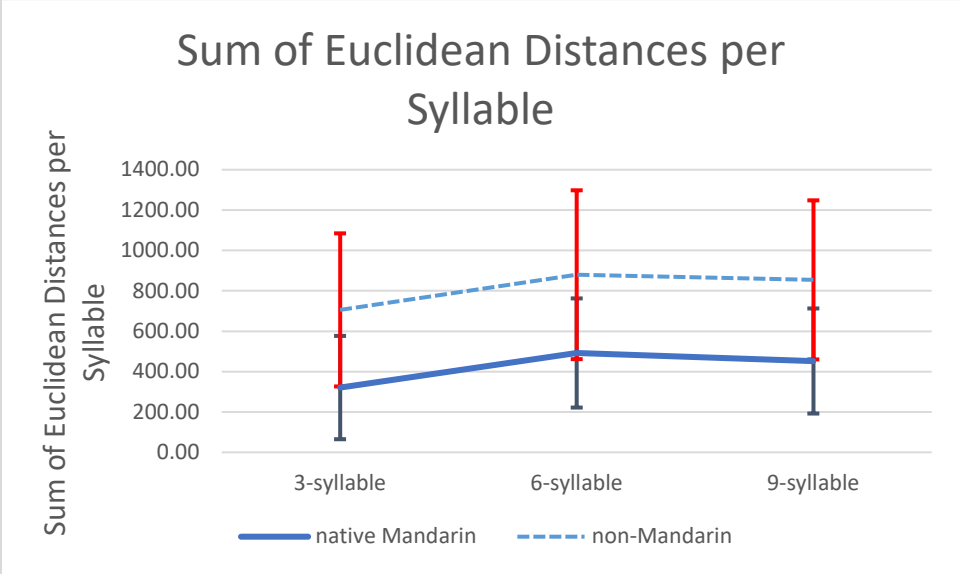

Figure 24 GMP errors per syllable across Group and Sequence Length conditions

A Bonferoni adjustment was made on the alpha level for post-hoc testing. An alpha level of .0083 (=.05/6)) was used for simple main effect comparisons. When the simple main effect of Sequence Length Condition was investigated in each group, the GMP_syl in three-syllable condition (Sequence Length Condition 1) was significantly lower by -0.4674 than that in six-syllable condition (Sequence Length Condition 2) in the native Mandarin group (t(92)=-12.28, *p<.0001*). The GMP_syl in Sequence Length Condition 1 was significantly lower by -0.3835 than that in nine-syllable condition (Sequence Length Condition 3) in the native Mandarin group (t(92)=-10.08, *p<.0001*). The GMP_syl in Sequence Length Condition 2 was not significantly different from that in Sequence Length Condition 3 in the native Mandarin group (t(92)=2.20, *p=.0300*).

On the other hand, the GMP_syl in Sequence Length Condition 1 was significantly lower by -.2438 than that in Sequence Length Condition 2 in the non-Mandarin group (t(92)=-6.04, *p<.0001*). The GMP_syl in Sequence Length Condition 1 was significantly lower by -.1987 than that in Sequence Length Condition 3 in the non-Mandarin group (t(92)=-4.98, *p<.0001*). The GMP_syl in Sequence Length Condition 2 was not significantly different from that in Sequence Length Condition 3 in the non-Mandarin group (t(92)=1.09, *p=.2780*).

A Bonferoni adjustment was made on the alpha level for post-hoc testing, so that an alpha level of .0167 (=.05/3) was used for simple main effect comparisons for group. When the simple main effect of group was investigated at each Sequence Length Condition, the GMP_syl in the native Mandarin group was significantly lower by -.8390 than that in the non-Mandarin group at the three-syllable length condition (t(92)=-7.90, *p<.0001*). The GMP_syl in the native Mandarin group was significantly lower by -.6154 than that in the non-Mandarin group at six-syllable length condition (t(92)=-5.79, *p<.0001*). The GMP_syl in the native Mandarin group was significantly lower by -.6543 than that in the non-Mandarin group at nine-syllable length condition (t(92)=-6.14, *p<.0001*).

3.2.3 GMP Errors for Tone 2 between Two Syllable Positions

Two template trajectories were derived from native Mandarin speakers to form a Tone 2 template trajectory. First, thirty productions were obtained from the second syllable position from all sequence length conditions in the native Mandarin speaker's productions to form a Tone 2 template trajectory. Then, ten productions were obtained from the second syllable position from the nine-syllable length conditions to form a Tone 2 template trajectory. The GMP errors increased when the 30 production template for Tone 2 was used as compared to the 10 production template in both

groups and in both syllable positions. The pattern of differences were similar between templates from the thirty and ten trial productions. Thus, it was decided to use the template which was obtained from the ten trials for the actual statistical analyses.

The sum of Euclidean distances for Tone 2 (ED2) was obtained from the 2nd and 8th syllable positions of the target nine-syllable sequence (Tone 4-2-1-3-4-1-4-2-1). In this analysis, a new independent variable called "GMP_2_8" was used to indicate the two syllable positions in which ED2 was observed. Because Tone 2 was produced in the 8th syllable position only in the nine-syllable sequence, this analysis was restricted to Sequence Length Condition 3 only.

The intra-class correlation (ICC) for ED2 in the unconditional model was 20.57% (intercept: 6438.49, residual: 24868). This result justified the need for using the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. The homoscedasticity assumption was tested in the full model in which covariates were specified. The Group, GMP_2_8, Group*GMP_2_8 were entered in the model statement. The intercept and GMP_2_8 were entered in the random statement. The homoscedasticity assumption was satisfied ($p=.3880$ for interaction term Group*GMP_2_8).

The interaction of Group and GMP_2_8 on ED2 was not significant ($F(1,46)=.54$, $p=.4643$). The main effect of Group was significant ($F(1,46)=50.66$, $p<.0001$). Native Mandarin speakers had significantly smaller ED2 by -.5336 than that of non-Mandarin speakers when averaged across GMP_2_8 syllable positions. Although the ED2 decreased in the 8th syllable position as compared to the 2nd syllable position in the non-Mandarin participants, and the ED2 increased in the 8th syllable position as compared to the 2nd syllable position in native Mandarin speakers (see Table 11), the main effect of GMP_2_8 on the ED2 was not significant ($F(1,46)=.02$, $p=.8806$) (see Table 11, Figure 25).

Table 11 Sum of Euclidean distances between 2nd syllable or 8th syllable positions of the nine-syllable sequency productions when ten 2nd Tone productions were used to obtain 2nd Tone template

	native Mandarin		non-Mandarin	
	Mean	SD	Mean	SD
2nd Syllable Position	166.39	119.46	294.04	199.23
8th Syllable Position	173.48	134.37	284.90	204.83

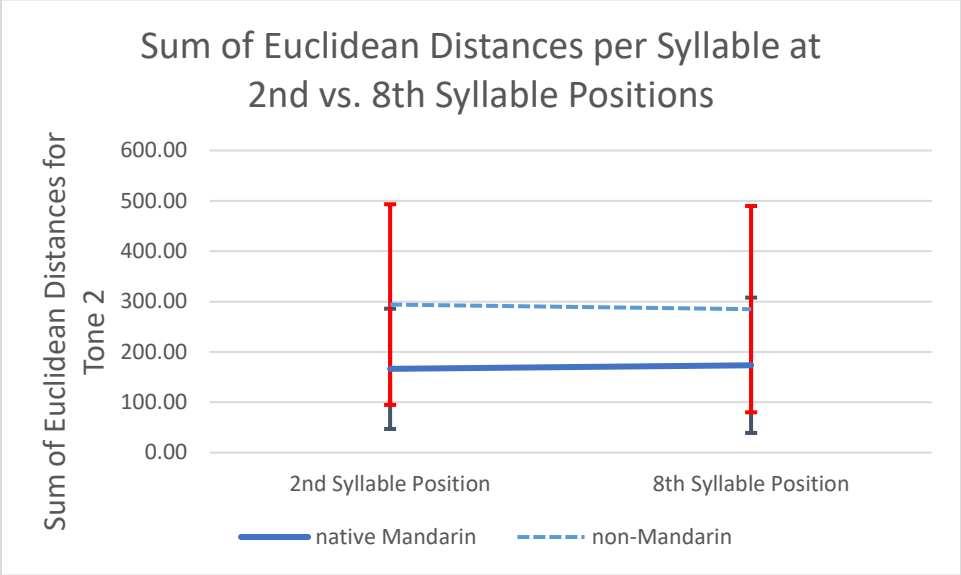

Figure 25 GMP errors for Tone 2 across Group and 2nd and 8th syllable positions

In summary, the variability of the motor program was higher in non-Mandarin speakers than in native Mandarin speakers as evidenced by the finding that the GMP error for Tone 2 was significantly higher in the non-Mandarin speaker group than the native Mandarin speaker group. The magnitude of the GMP error for Tone 2 was similar between the 2nd and 8th syllable positions in the nine-syllable sequence. While the non-Mandarin group had a significantly higher GMP error for Tone 2 than native Mandarin group, the variability within the group for the same sound was maintained relatively constant within the utterance in both speaker groups.

3.2.4 Hamming Distance Difference per Syllable between Slope Measurement and Parsons' Code Measurement

3.2.4.1 Hamming Distance Difference per Syllable between Slope and Parson's Code Measurements

The differences between hamming distance per syllable between Slope and Parsons' code measurements (Diff_Hamm_Slope_Par) were examined after subtracting hamming distance per syllable for Parsons' code measurement from hamming distance per syllable for Slope measurement. The intra-class correlation (ICC) for the Diff_Hamm_Slope_Par in the unconditional model was 23.88% (intercept: .8314, residual: 2.6507). This result justified the need for using the hierarchical linear model. Because the normality assumption was met, the proc glimmix command was used with gaussian distribution for model testing. The homoscedasticity assumption was satisfied ($p=.1503$ for the interaction term Group*Sequence Length Condition).

The interaction of Group and Sequence Length Condition on Diff_Hamm_Slope_Par was significant ($F(2, 92)=10.94, p<.0001$). The main effect of Group on Diff_Hamm_Slope_Par was significant when averaged across Sequence Length Conditions ($F(1,46)=6.51, p=.0141$). In other words, the native Mandarin speaker group produced significantly higher (by .1484) Diff_Hamm_Slope_Par values than the non-Mandarin speaker group when averaged across Sequence Length Conditions. The main effect of Sequence Length Condition on Diff_Hamm_Slope_Par was significant when averaged across groups ($F(2, 92)=20.33, p<.0001$) (see Figure 26). In table 12, the hamming distance difference between Slope and Parsons' code measurements increased as the sequence length became longer in both speaker groups. To learn where the differences originated from, the simple main effects were examined at the post-hoc analyses.

Table 12 Comparison of Hamming distance difference between Slope and Parsons' code measurements

		Hamming Distance Difference Between Slope and Parsons' Code Measurements		
		3-syllable	6-syllable	9-syllable
native Mandarin	Mean	13.19	13.24	13.45
	SD	2.27	1.52	1.36
non-Mandarin	Mean	11.75	12.79	13.24
	SD	2.10	1.58	1.60

A Bonferoni adjustment was made on the alpha level for the post-hoc test, so an alpha level of .0083 (=.05/6) was used for simple main effect comparisons. When a simple main effect for Sequence Length Condition was investigated for each group, the Diff_Hamm_Slope_Par in Sequence Length Condition 1 was not significantly different from that in Sequence Length Condition 2 for the native Mandarin group (t(92)=-.27, *p=.7861*). The Diff_Hamm_Slope_Par in Sequence Length Condition 1 was not significantly different from that in Sequence Length Condition 3 in the native Mandarin group (t(92)=-1.34, *p=.1820*). The Diff_Hamm_Slope_Par in Sequence Length Condition 2 was not significantly different from that in Sequence Length Condition 3 in the native Mandarin group (t(92)=-1.07, *p=.2866*).

The Diff_Hamm_Slope_Par in Sequence Length Condition 1 was significantly lower by -1.0148 than that in Sequence Length Condition 2 in the non-Mandarin group (t(92)=-4.99, *p<.0001*). The Diff_Hamm_Slope_Par in Sequence Length Condition 1 was significantly lower by -1.5294 than that in Sequence Length Condition 3 in the non-Mandarin group (t(92)=-7.48, *p<.0001*). The Diff_Hamm_Slope_Par in Sequence Length Condition 2 was not significantly different from that of Sequence Length Condition 3 in the non-Mandarin group (t(92)=-2.49, *p=.0146*).

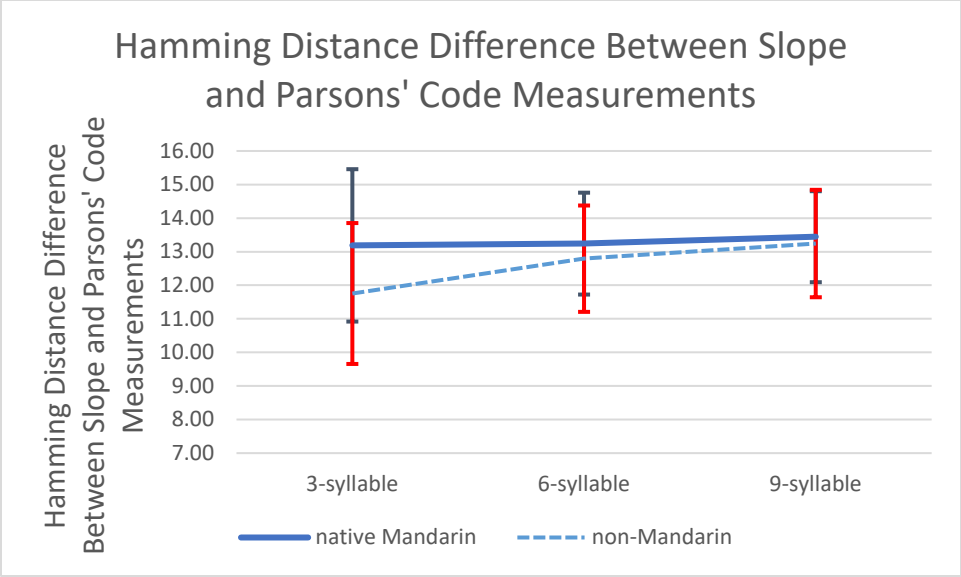

Figure 26 Hamming distance difference per syllable between Slope and Parsons' code measurements

When the simple main effect of the Group was investigated in each Sequence Length Condition, an alpha level of .0167 (=.05/3) was used. The Diff_Hamm_Slope_Par for the native Mandarin group was significantly higher by 1.4143 than that in the non-Mandarin group in Sequence Length Condition 1 (t(92)=4.58, $p<.0001$). The Diff_Hamm_Slope_Par for the native Mandarin group was not significantly different from the non-Mandarin group in Sequence Length Condition 2 (t(92)=1.46, $p=.1479$). The Diff_Hamm_Slope_Par in group1 was not significantly different from the non-Mandarin group in Sequence Length Condition 3 (t(92)=.1484, $p=.6344$).

In summary, the Hamming distance differences between Slope and Parsons' code measurement was significantly higher in the native Mandarin speakers than non-Mandarin speakers in the three-syllable condition. The magnitude of this hamming distance difference between Slope and Parsons' code measurement was maintained across sequence length conditions for the native Mandarin speakers while it increased and then remained at a similar level as the

sequence length became longer for the non-Mandarin speakers. In order to explain why this happened, Slope and Parsons' code hamming distances were examined separately.

3.2.4.2 Hamming Distance per Syllable for Slope Measurement

The sum of the hamming distances, comparing Slope measurements between the participant's f_0 trajectories and template f_0 trajectories, was divided by the number of syllables in the utterance. In this measurement, the more deviated the participant's f_0 trajectory was from the template f_0 trajectory, the higher hamming distance values were generated.

The sum of Slope hamming distances per syllable increased slightly as the sequence length became longer in both native Mandarin and non-Mandarin speaker groups. The degree of change was small, in particular, between six-syllable and nine-syllable sequence length conditions. The Slope hamming distances per syllable were larger in the non-Mandarin speakers than native Mandarin speakers. Table 13 summarizes these data and Figure 27 displays the group by syllable length contrasts.

To learn the statistical significance of the observed change, the statistical analyses were performed on the Hamming distances for Slope per syllable (Hamm_Slope_syl). The intra-class correlation (ICC) for the Hamm_Slope_syl in the unconditional model was 25.18% (intercept: 1.3212, residual: 3.9265), justifying use of the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. Because the variance component of Sequence Length Condition in the random statement was estimated to be zero, the Sequence Length Condition was not included in the random statement. The homoscedasticity assumption was satisfied (*p=.5429* for interaction term Group* Sequence Length Condition). The interaction of Group and Sequence Length Condition on Hamm_Slope_syl was not significant (*F(2, 92)=.45, p=.6412*). The main effects for Group (*F(1,*

46)=7.29, p=.0097) and Sequence Length Condition on Hamm_Slope_syl were significant (*F(2, 92)=141.55, p<.0001*) (see Figure 27).

Table 13 Comparison of Hamming distance per syllable for Slope and Parsons' code measurements

		3-syllable		6-syllable		9-syllable	
		Slope Hamming	Parsons' code Hamming	Slope Hamming	Parsons' code Hamming	Slope Hamming	Parsons' code Hamming
native Mandarin	Mean	22.14	8.95	24.11	10.87	24.22	10.77
	SD	2.62	2.17	2.06	1.91	1.60	1.69
non-Mandarin	Mean	22.78	11.02	25.03	12.23	25.20	11.96
	SD	2.27	1.81	1.59	2.27	1.54	2.27

A Bonferoni adjustment was made for post-hoc testing, with an alpha level of .0167 (=.05/3) for marginal effect comparisons. When the marginal effect of Sequence Length Condition was investigated after averaging across groups, the Hamm_Slope_syl in Sequence Length Condition 1 was significantly lower by -0.09115 than Sequence Length Condition 2 (t(92)=-13.95, *p<.0001*). The Hamm_Slope_syl in Sequence Length Condition 1 was significantly lower by -0.09855 than Sequence Length Condition 3 (t(92)=-15.06, *p<.0001*). The Hamm_Slope_syl in Sequence Length Condition 2 was not significantly different from Sequence Length Condition 3 (t(92)=-1.12, *p=.2643*).

The marginal effect of group was also examined after averaging across Sequence Length Conditions. The Hamm_Slope_syl for the native Mandarin group was significantly lower by -.03894, than that of the non-Mandarin group (t(46)=-2.70, *p=.0097*).

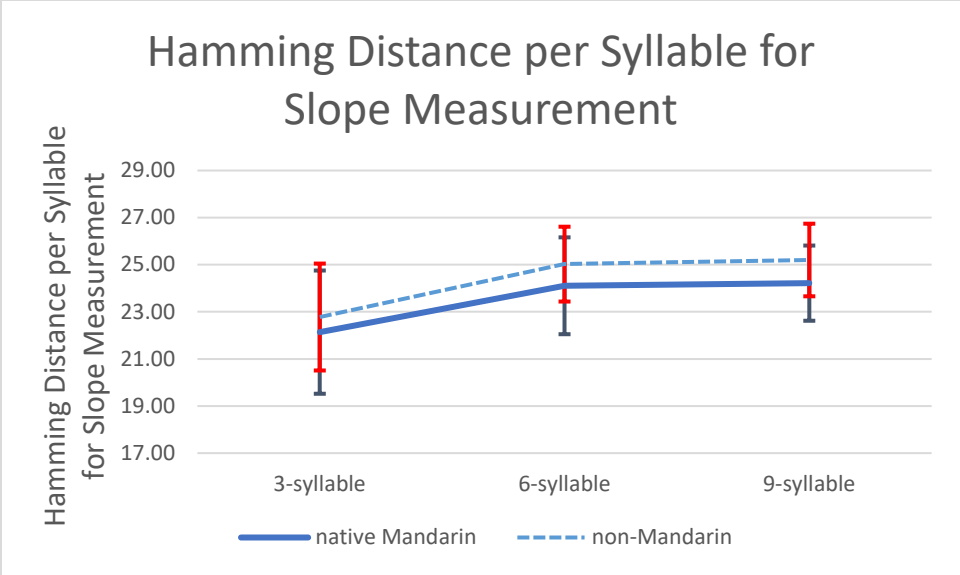

Figure 27 Hamming distance per syllable for Slope measurement

In summary, while the Hamming distance per syllable for Slope measurement changed in a parallel manner between the two speaker groups, the hamming distance itself was larger in non-Mandarin speakers than that of native Mandarin speakers. The hamming distance per syllable for the Slope measurement significantly increased as the sequence length changed from the three- to six-syllable length. However, there was no significant increase in the Slope hamming distance when the sequence length increased from six to nine-syllables. This pattern across sequence length conditions was evidenced when the results were averaged across two groups.

3.2.4.3 Hamming Distance per Syllable for Parson's Code Measurement

The sum of hamming distances was obtained after comparing Parsons' code measurements of the participant's f_0 trajectories to those of template f_0 trajectories. This value was divided by the number of syllables. When absolute value changes were examined in Table 13, the sum of the Parsons' code hamming distances per syllable increased as the sequence length changed from the three- to the six-syllable condition and slightly decreased as the sequence length changed from the

six- to the nine-syllable condition for both speaker groups (Table 13, Figure 28). The non-Mandarin speaker group demonstrated higher hamming distances per syllable for the Parsons' code measurement than those of the native Mandarin speaker group.

In general, the sum of hamming distances per syllable was higher for Slope than that for Parsons' code in both speaker groups.

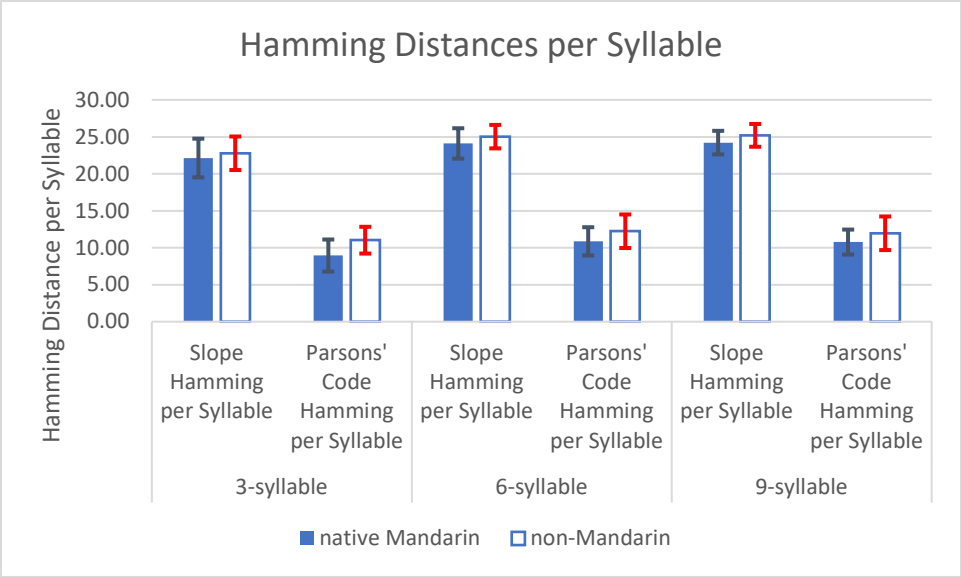

Figure 28 Comparison of Hamming distance per syllable for Slope and Parsons' code

To learn the significance of this observed change, a statistical analysis was performed. This Hamming distances for Parsons' code per syllable was calculated for analyses (Hamm_par_syl). The intra-class correlation (ICC) for the Hamm_par_syl in the unconditional model was 32.23% (intercept: 1.6866, residual: 3.5462), justifying the use of the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. The homoscedasticity assumption was not satisfied in the initial model with gaussian distribution ($p<.0001$ for interaction term Group*Sequence Length Condition). However, after running generalized linear mixed model (GLMM) with gamma link

for this non-negative, continuous data, the model fit with the data well showing a normal distribution. Also, the boxplots indicated that the variability was fairly consistent across different conditions. Therefore, it was concluded that the new model with gamma distribution was well-specified and any inferences from this model result were trustworthy. Only the intercept was included in the random statement.

The interaction of Group and Sequence Length Condition on Hamm_par_syl was significant (*F(2, 1264)=14.48, p<.0001*). The main effect of Group on Hamm_par_syl was significant (*F(1, 46)=24.48, p<.0001*). The Hamm_par_syl for the native Mandarin speaker group was significantly lower by -.05183 than that for the non-Mandarin speaker group when averaged across Sequence Length conditions. The main effect of Sequence Length Condition on the Hamm_Par_Syl was significant when averaged across groups (*F(2, 1264)=120.35, p<.0001*) (see Figure 29).

A Bonferoni adjustment was made with an alpha level of .0083 (=.05/6) for simple main effect comparisons. The simple main effect for the Sequence Length Condition was investigated for each group. The Hamm_par_syl for Sequenc Length Condition 1 was significantly lower by -.1994 than for Sequence Length Condition 2 for the native Mandarin group *(t(1264)=-13.44, p<.0001)*. The Hamm_par_syl for Sequence Length Condition 1 was significantly lower by -.1916 than for Sequence Length Condition 3 for the native Mandarin group *(t(1264)=-12.93, p<.0001)*. The Hamm_par_syl for Sequence Length Condition 2 was not significantly different from Sequence Length Condition 3 for the native Mandarin group *(t(1264)=.52, p=.6003)*.

The Hamm_par_syl for Sequence Length Condition 1 was significantly lower by -.1063 than for Sequence Length Condition 2 for the non-Mandarin group *(t(1264)=-6.68, p<.0001)*. The Hamm_par_syl for Sequence Length Condition 1 was significantly lower by -.0837 than for

Sequence Length Condition 3 for the non-Mandarin group *(t(1264)=-5.21, p<.0001)*. The Hamm_par_syl in Sequence Length Condition 2 was not significantly different from Sequence Length Condition 3 for the non-Mandarin group *(t(1264)=1.38, p=.1665)*.

When the simple main effect of the Group was investigated for each Sequence Length Condition, the Hamm_par_syl for the native Mandarin group was significantly lower by -.2149 than for the non-Mandarin group for Sequence Length Condition 1 *(t(1264)=-6.65, p<.0001)*. The Hamm_par_syl for the native Mandarin group was significantly lower by -.1219 than for the non-Mandarin group for Sequence Length Condition 2 *(t(1264)=-3.75, p=.0002)*. Likewise, the Hamm_par_syl for the native Mandarin group was significantly lower by -.1070 than for the non-Mandrin group for Sequence Length Condition 3 *(t(1264)=-3.29, p=.0010)*.

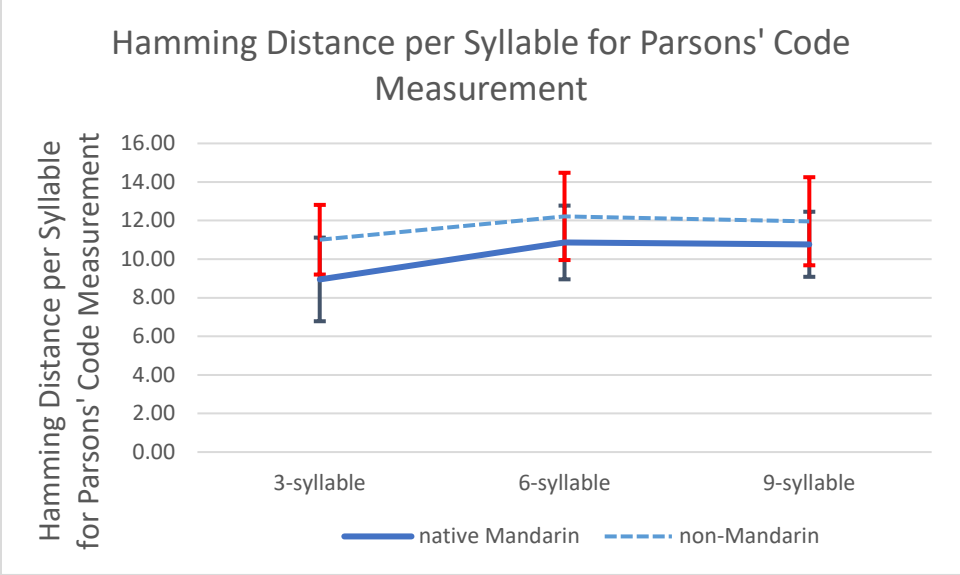

Figure 29 Hamming distance per syllable for Parsons' code measurement

In summary, the hamming distance per syllable for the Parsons' code measurement (Hamm_par_syl) was significantly larger in non-Mandarin speakers than native Mandarin speakers, as it was for Slope measurement. The Hamm_par_syl significantly increased between

three-syllable and six-syllable sequence length conditions, and remained about the same level between six-syllable and nine-syllable sequence length conditions for both speaker groups.

3.2.4.4 Supplementary Analyses

The Slope hamming distance increased more in non-Mandarin speakers (9.79%) than in the native Mandarin speakers (8.89%) between three- and six-syllable length condition. Also, the Parsons' code hamming distance increased less in the non-Mandarin speakers (10.97%) than the native Mandarin speakers (21.41%) between three- and six-syllable length conditions (see Figure 30).

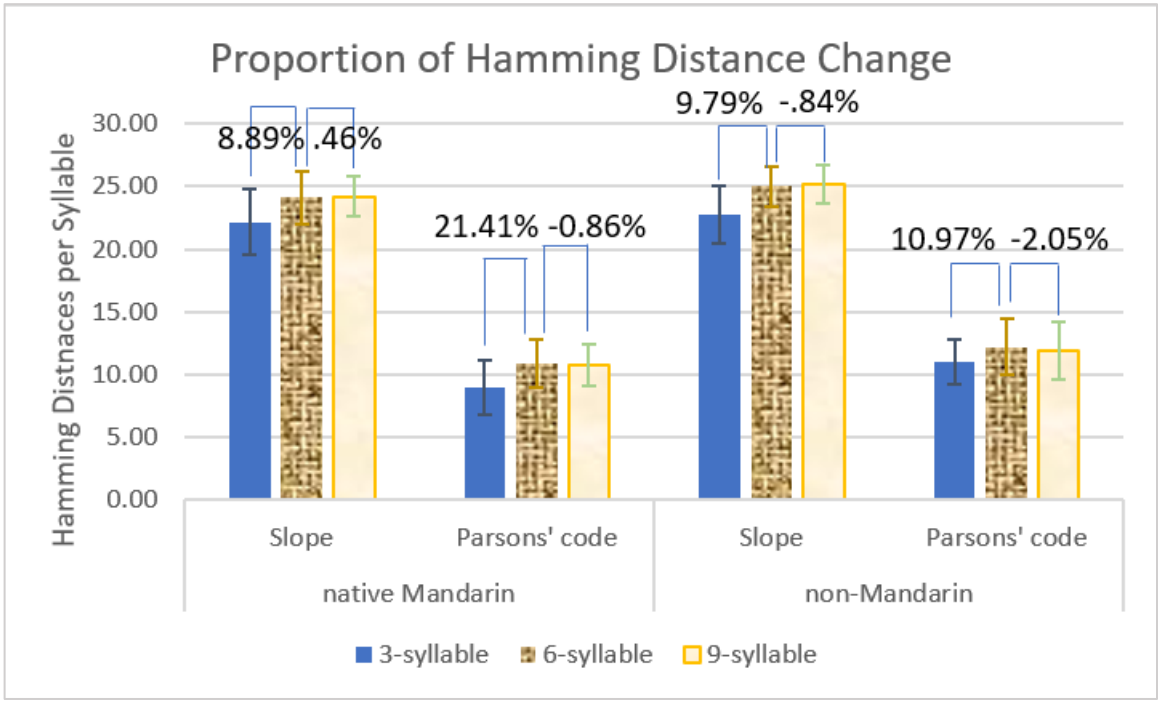

Figure 30 Proportional change in the Hamming distnance per syllable

3.2.5 Reaction Time (RT), Average ISI, Ratio of RT/Average ISI

Prior to this analysis, all the trials with errored productions were excluded. The mean and standard deviation (SD) for RT (Figure 31), average ISI (Figure 32), and the Ratio of RT over average ISI (Figure 33) across different Sequence Length Conditions and Groups are summarized in Table 14. When raw values in Table 14 were examined, the RT and ISI were higher in the non-Mandarin speaker group than the native Mandarin speaker group in all sequence length conditions. The variability as indexed by SD around mean RT or ISI also appeared to be greater in the non-Mandarin speaker group than that in the native Mandarin speaker group.

Table 14 Mean and SD of RT, average ISI and Ratio of RT over average ISI (Unit: seconds)

RT	3-syllable		6-syllable		9-syllable	
	Mean	SD	Mean	SD	Mean	SD
native Mandarin	1.18	0.26	1.21	0.35	1.24	0.28
non-Mandarin	1.81	0.93	1.94	1.57	1.77	0.85

Average ISI	3-syllable		6-syllable		9-syllable	
	Mean	SD	Mean	SD	Mean	SD
native Mandarin	0.48	0.27	0.48	0.21	0.49	0.21
non-Mandarin	1.09	0.56	1.13	0.48	1.10	0.43

Ratio of RT over average ISI	3-syllable		6-syllable		9-syllable	
	Mean	SD	Mean	SD	Mean	SD
native Mandarin	3.57	3.07	3.19	1.91	3.13	1.81
non-Mandarin	1.92	1.23	1.92	2.41	1.75	0.83

Both RT and ISI increased as the sequence length became longer and the Ratio of RT over average ISI decreased slightly as the sequence length increased in the native Mandarin speaker

group. In contrast, both the RT and ISI increased from the three- to six-syllable length condition and dropped from the six- to nine-syllable length conditions in the non-Mandarin speaker group. The Ratio of RT over average ISI was stable between the three- and six-syllable length conditions and dropped between the six- to nine-syllable length conditions in the non-Mandarin speaker group. The Ratio of RT over average ISI in the native Mandarin speaker group was overall higher than that in the non-Mandarin speaker group across all sequence length conditions. The statistical significance of these differences is explored in section 3.2.5.1, 3.2.5.2, and 3.2.5.3.

3.2.5.1 Reaction Time (RT)

The square-root transformed RT (RT_sqrt) was used for all RT analyses. The intra-class correlation (ICC) in the unconditional model was 56.18% (intercept: 0.033, residual: 0.026), justifying the use of the hierarchical linear model. The normality assumption was not met. Thus, the proc glimmix command with gamma distribution was used for model testing. Because the variance component of Sequence Length Condition in the random statement was estimated to be zero, Sequence Length Condition was not included in the random statement. The homoscedasticity assumption was satisfied ($p=.6640$ for interaction term Group*Sequence Length Condition).

The interaction of Group and Sequence Length Condition was not significant for RT_sqrt ($F(2, 1264)=1.22, p=.2959$). The main effect of Group on RT_sqrt was significant ($F(1, 46)=27.87, p<.0001$) with the native Mandarin speaker group responding with significantly shorter reaction times by -0.1542 for the RT_sqrt than the non-Mandarin speaker group when averaged across Sequence Length Conditions. The main effect of Sequence Length Condition on RT_sqrt was not significant when averaged across groups ($F(2, 1264)=1.12, p=.3259$) (see Figure 31).

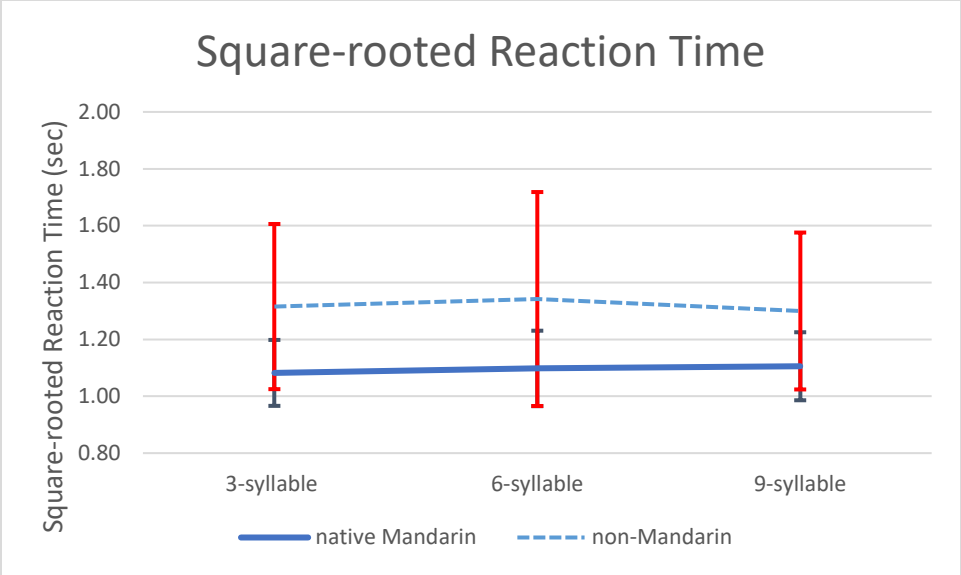

Figure 31 RT change across Sequence Length Conditions and Groups

3.2.5.2 Average ISI

The intra-class correlation (ICC) for the average ISI (avg_ISI) in the unconditional model was 83.30% (intercept: 0.164, residual: 0.197). This result justified the need for using the hierarchical linear model. Because the normality assumption was not met, the proc glimmix command was used with gamma distribution for model testing. The autoregressive covariance structure was used. The homoscedasticity assumption was satisfied (*p=.0647* for the interaction term of Group* Sequence Length Condition).

The interaction of Group and Sequence Length Condition for the avg_ISI was not significant (*F(2,92)=1.53, p=.2225*). The main effect of Group was significant (*F(1,46)=46.38, p<.0001*) with the native Mandarin speaker group producing significantly shorter avg_ISI by -0.8325 than the non-Mandarin speaker group when averaged across Sequence Length Conditions

(t(46)=-6.81, *p<.0001)*. The main effect of Sequence Length Condition on the avg_ISI was also significant (*F(2, 92)=3.64, p=.0301*) (see Figure 32).

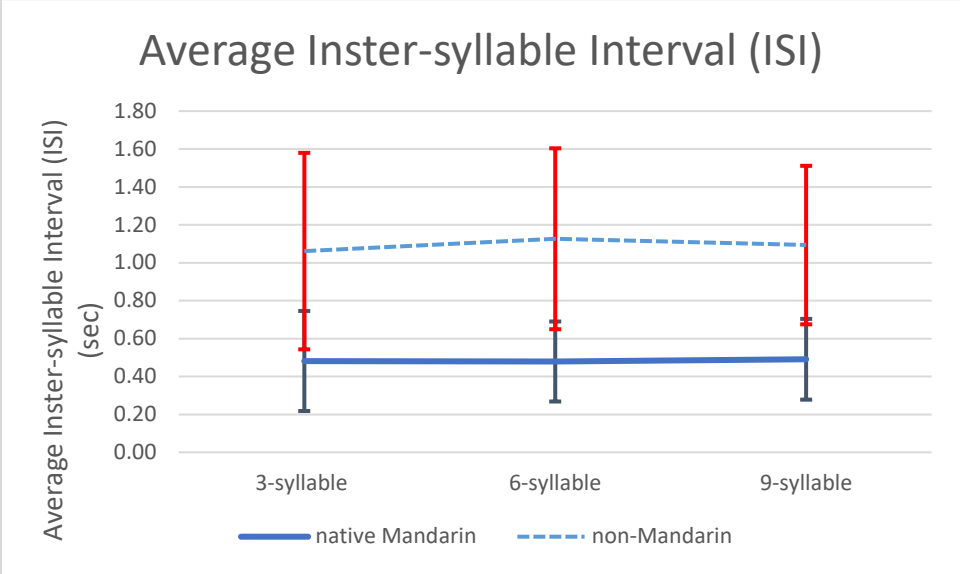

Figure 32 Average ISI change across Sequence Length Conditions and Groups

The marginal effect of Sequence Length Conditions was investigated with an alpha level of .0167 after averaging across groups. The avg_ISI in Sequence Length Condition 1 was not significantly different from that in Sequence Length Condition 2 (t(92)=-2.68, *p=.0196*). The avg_ISI for Sequence Length Condition 1 was not significantly different from that in Sequence Length Condition 3 (t(92)=-2.83, *p=.0193*). The avg_ISI for Sequence Length Condition 2 was not significantly different from that in Sequence Length Condition 3 (t(92)=-.31, *p=.7569*).

3.2.5.3 Ratio of RT/Average ISI

A square-root transformation was performed on the variable named Ratio of RT over average inter-syllable interval (Ratio_RT_AvgISI_sqrt). The intra-class correlation (ICC) for the Ratio_RT_AvgISI_sqrt in the unconditional model was 71.21% (intercept: 0.169, residual: 0.068). This result justified the use of the hierarchical linear model. Because the normality assumption

was not met, the proc glimmix with gamma distribution was used for model testing. The homoscedasticity assumption was satisfied (*p=.2541* for interaction term Group*Sequence Length Condition).

As a result, an interaction of Group and Sequence Length Condition for the Ratio_RT_AvgISI_sqrt was not significant (*F(2,92)=0.38, p=.6854*). The main effect of Group was significant (*F(1,46)=18.41, p<.0001*) suggesting that the Ratio_RT_AvgISI_sqrt for the native Mandarin speaker group was significantly higher by .2522 than that for the non-Mandarin speaker group when averaged across Sequence Length Conditions (t*(46)=4.29, p<.0001*). The main effect of Sequence Length Condition on the Ratio_RT_AvgISI_sqrt was also significant (*F(2,92)=3.95, p=.0227)* (see Figure 33).

The marginal effect of Sequence Length Conditions was investigated with an adjusted alpha level of .0167 after averaging across groups. The Ratio_RT_AvgISI_sqrt for Sequence Length Condition 1 was not signficantly different from that for Sequence Length Condition 2 (t*(92)=2.08, p=.0407*). The Ratio_RT_AvgISI_sqrt for Sequence Length Condition 1 was significantly higher by 0.03168 than that for Sequence Length Condition 3 (t*(92)=2.67, p=.0090*). The Ratio_RT_AvgISI_sqrt for Sequence Length Condition 2 was not significantly different from that for Sequence Length Condition 3 (t*(92)=.60, p=.5529*). Thus, the significant effect of Sequence Length Condition came from the difference in Ratio_RT_AvgISI_sqrt between Sequence Length Condition 1 and Sequence Length Condition 3.

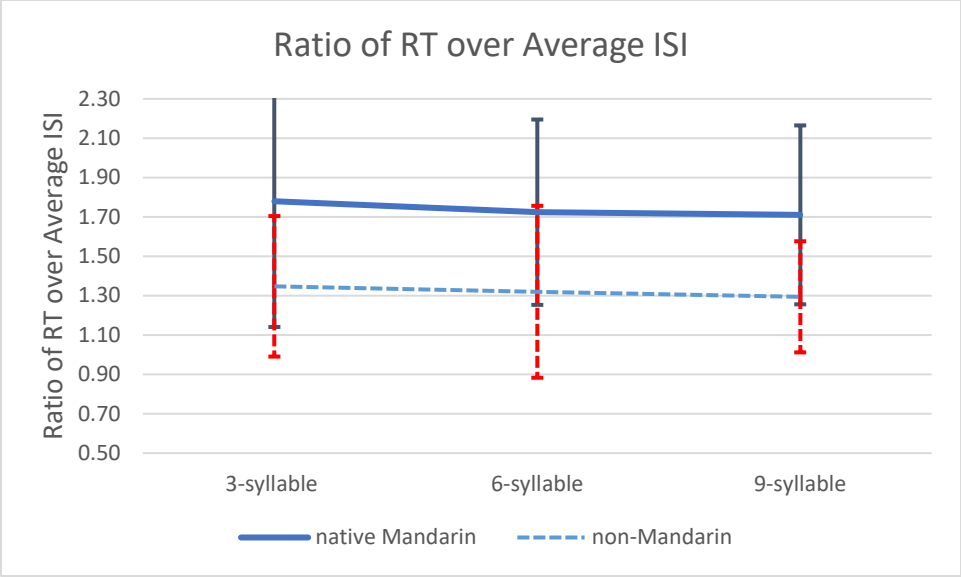

Figure 33 Ratio of RT over average ISI across Sequence Length Conditions and Groups

4.0 DISCUSSION

4.1.1 Paramter Variability

It was hypothesized that there would be significantly different parameter values between the first syllable and whole syllable sequences, which would support the notion of parameter variability if more than one parameter is utilized in long utterances of multiple syllable lengths. It was predicted that native Mandarin speakers would have lower parameter variability than non-Mandarin speakers and the parameter variability would increase as the sequence length increased for native Mandarin speakers, while it would not differ across sequence length conditions for non-Mandarin speakers. That is, native Mandarin speakers would have lower parameter variability than the non-Mandarin speakers because native Mandarin speakers have more experience with long utterances and would prepare and execute whole utterances with less frequent parameterization. On the other hand, it was expected that non-Mandarin speakers would utilize a concurrent method of motor control, parameterizing more frequently while they produce whole utterances.

The subtracted parameter values (or parameter variability) indicated differences between the whole sequence and the 1st syllable in the parameters. The time parameter variability, which underwent square-root transformation, was maintained similarly across different sequence length conditions for both groups. Although the f_0 magnitude parameter variability was significantly different among sequence length conditions, the post-hoc tests failed to locate the differences among sequence length conditions. Therefore, the parameter variability was examined further only at the nine-syllable sequence length condition.

Counter to the hypothesis, when the parameter variability was examined in the nine-syllable sentence length condition, no interaction between the Whole_1st condition and Group was observed. Significant parameter differences were observed only between speaker groups. That is, the native Mandarin speakers produced higher time and f_0 magnitude parameters than the non-Mandarin speakers, indicating that the native Mandarin speakers spoke faster than the non-Mandarin speakers. Also, native Mandarin speakers demonstrated smaller variabilities of f_0 change or lower pitch than the non-Mandarin speakers. While speakers from both groups produced speech in a concurrent manner, it appeared that the parameters measured in this study were pre-determined before speech production began and those parameter values were used for that whole utterance. This result was contrary to what Fujimura (1987) assumed about a constant scaling factor because he pointed out the need to account for the nonlinear-lengthened intervals in the later portion of an utterance.

Few studies have experimentally investigated whether speakers parameterized their speech motor programs in an on-line manner. This question is particularly important because the result could inform current models of speech production. For example, according to van der Merwe (1997, 2009); van der Merwe and Steyn (2018), the spatial and temporal specifications of the core motor plan (aka parameterization process) have to meet the phonetic requirements of the task and take place prior to specifying muscle-specific motor programs. Van der Merve labeled this “speech motor planning”. If this is accurate, the parameters (or contexts) in which the motor programs are embedded are pre-determined and remain the same while more than one motor program is activated. However, Guenther (2006) has proposed that motor programs for complex actions such as phonemes or frequently used syllables are stored in a speech sound map, which exists in “the left ventral premotor cortex and/or posterial Broca’s area” (Guenther et al., 2006, p. 4). Once a

motor program is retrieved from this area, the information is sent to the motor cortex to activate specific muscle commands which are then integrated with other feedback information that is forwarded from sensory areas. The movement specifications for muscles, such as velocity or force, are made through a 16-dimensional vector in this motor cortex. Thus, based on Guenther's model, it was speculated that a rough format of a stored motor program is retrieved first and movement specification (or parameters) is determined later. Similarly, the Internal model also proposed that movements are controlled in a concurrent manner by retrieving a rough format of motor programs and that movement parameters are specified in a rapid manner before they are executed (Wolpert et al., 1995; Wolpert et al., 1998). If this later view is accurate, it is possible to determine movement specifications, such as speech rate, pitch, and intensity, in a concurrent manner. Then, parameter change can be observed during on-going speech.

The results from the current study are more consistent with van der Merwe (1997, 2009) according to whom parameters for speech are prepared in advance and do not change during the course of speech movement sequences. Therefore, it appears that once the parameter aspects are prepared, they remain unchanged over time while executing the motor responses regardless of the learning experience. However, because the later view does not prevent the possibility of unchanging movement specifications (or parameters) during the production of speech sequences, the finding of this current study does not completely refute the later view of Guenther's (2006). Additionally, because of the particular experimental environment in this study, the participants might have utilized a more constrained speech production strategy, such as maintaining a consistent speech rate across productions. As this study used relatively novel speech production tasks that carry no linguistic content for either speaker group, future studies need to re-examine parameter variability in a more natural speech production environment. In addition, further

examination of this aspect is required using longer utterances. The current study may motivate future studies to a better understanding about the relationship between GMP and movement parameters.

4.1.2 GMP Errors per Syllable

As hypothesized, a significant interaction between Group and Sequence Length Condition was observed for the GMP errors per syllable. The native Mandarin speaker group demonstrated lower GMP errors per syllable than the non-Mandarin speaker group in all sequence length conditions. It was also hypothesized that these group differences in GMP errors per syllable would be smaller in the nine-syllable condition than in the three- and six-syllable length conditions because this condition was where the native Mandarin speaker group was least different from the non-Mandarin speakers. However, the GMP error did not increase as the sequence length increased from six- to nine-syllable sequence length conditions for either speaker group. This result might have been due to the repetition of the first three syllables in the 7^{th}, 8^{th}, and 9^{th} syllable positions. Because all four sequences used for the nine-syllable sequence condition (one target sequence and three filler sequences) followed this pattern of repetition, the participants may have figured out that they needed to repeat the first three-syllables at the last three syllable positions for the longest tone sequence condition. This could have facilitated motor programming for the last three syllables in the nine-syllable sequence length condition.

Nonetheless, the GMP errors per syllable was still higher in the six- and nine-syllable sequence length conditions than in the three-syllable sequence length condition suggesting that there was an increase in GMP errors due to an increase in the sequence length. In other words, there was a cost for producing longer utterances when indexed by GMP errors per syllable after

the three-syllable sequence length production. This finding suggested that speakers utilized GMPs while producing the first three-syllables and they turned to a concurrent production manner for longer utterances as evidenced by increased GMP errors per syllable. This cost explanation is consistent with the proposed reasons for failing to find consistent proportional relationships across different contexts, which might be due to increased variability in the trajectory as a result of the concatenation effort of the motor programs. Verwey (1996) also reported that the errors in the key-pressing tasks increased as the sequence length increased. This is because an earlier part of the sequence appeared to be prepared as a whole unit, but the later part of the sequence seemed to be prepared in a concurrent manner. While this current study was consistent with what Verwey (1996) observed, finding increased GMP errors per syllable at the six- and nine-syllable levels was meaningful because it was observed in speech tasks.

Furthermore, overall GMP errors per syllable were larger in the non-Mandarin speaker group than that of the native Mandarin speaker group, suggesting that GMPs for lexical tones were developed, strengthened, and made more available for production as a product of their production experiences in the native Mandarin speaker group. Although both speaker groups initially increased and then slightly decreased in the GMP errors per syllable as the sequence length increased, these results support the hypothesized differential production patterns. That is, the changes in GMP errors were relatively larger in the native Mandarin speaker group than that of the non-Mandarin speaker group. There was a 9% increase in GMP errors per syllable between three- and six-syllable conditions and a 8% increase between three- and nine-syllable conditions in the native Mandarin speaker group. Conversely, there was a 4% increase in GMP errors per syllable between three- and six-syllable conditions and a 3% increase between three- and nine-syllable conditions in the non-Mandarin speaker group.

It is speculated that this trend occurred because the native Mandarin speaker group used a stored GMP in the three-syllable sequence and changed their production mode to programming and executing motor responses in a more concurrent manner as the sequence reached the six-syllable length. The degree of change in GMP errors per syllable was relatively smaller in the non-Mandarin speaker group between three- and six-syllable sequences than in the native Mandarin speaker group, possibly because these non-Mandarin speakers may have prepared the subsequent motor response relying on a concurrent manner before completing the three-syllable level utterances athroughout different sequence length conditions.

4.1.3 GMP Errors for Tone 2 between Two Syllable Positions

It was hypothesized that the GMP errors for Tone 2 would not be significantly different between the 2nd and 8th syllable positions in the nine-syllable sequence for the native Mandarin speakers, but would be significantly different for the non-Mandarin speakers because of their fluctuating performances. It was also expected that GMP errors for Tone 2 would be higher in non-Mandarin speakers than the native Mandarin speakers.

As hypothesized, GMP errors for Tone 2 were significantly higher in the non-Mandarin speaker group than the native Mandarin speaker group. Thus, variability of motor control, as measured by GMP errors for Tone 2, was significantly higher in non-Mandarin speakers than the native Mandarin speakers. However, unexpectedly, the magnitude of GMP error for Tone 2 was not significantly different between the 2nd and 8th syllable positions for either speaker group. This may be due to the fact that the same GMP was utilized for Tone 2 in the two different syllable positions within an utterance, which were considered transfer conditions.

Both findings support the existence of the GMP. First, the fact that the GMP accuracy was lower in the non-Mandarin speaker group than that of native Mandarin speakers suggests that the GMP error, as measured in this study, was a valid index of the GMP accuracy as evidenced by fewer errors in the well-practiced native Mandarin speakers. Second, once the non-Mandarian speakers acquired a certain level of experience and increased accuracy at a syllable level, through the mono-syllable practice phase, they were able to utilize relatively stable motor programs for the practiced tone in the connected speech condition. In other words, even the less-experienced speakers appeared to develop accessible motor programs for practiced speech gestures. These non-Mandarin speakers might have utilized a rough format of motor programs to produce Tone 2, which was borrowed from their own motor program storage. However, these motor programs still might be imprecise as compared to those of native Mandarin speakers.

Hao (2018) also reported imperfect tone productions in non-Mandarin speakers although they could "imitate the general shape of Mandarin tones" (Hao, 2018, p. 40). Borrowing from Bent's (2005) interpretation, it was speculated that this could be due to the influence of their first language (L1) on the tone productions. For example, English has only intonations rather than lexical tones. Thus, the English speakers were unable to make accurate changes in pitch at a syllable level, resulting in inaccurate or imprecise tone heights. Hao also suggested that this inaccurate production might be due to imprecise representation in non-Mandarin speakers. Therefore, the non-Mandarin speakers in the current study might also have utilized or modified the existing motor programs in their motor program storage which would generate the closest speech sounds to what they heard during the mono-syllable practice phase.

Furthermore, this result might be due to the constant level of GMP errors at a syllable level within a speaker, regardless of whether he is producing the same sound or not. This aspect was

investigated additionally by comparing GMP errors per syllable among the 1st syllable, 2nd syllable, and 8th syllable positions. Among these syllable positions, the degree of GMP error for the Tone produced in the 1st syllable position was different from those errors for Tones used for the other two syllable positions. The result revealed that there was a significant syllable position effect, which meant that the GMP errors per syllable were not maintained at the same level across different syllable positions. This finding is inconsistent with the above stated possibility of having the same level of GMP errors per syllable no matter which speech sound is being produced. The level of GMP errors per syllable appeared to be tone-specific, and this result supported the interpretation that the same level of GMP errors at two different syllable positions was a result of learning Tone 2.

4.1.4 Difference in the Hamming Distance per Syllable between Slope and Parsons' Code Measurements

The difference in the Hamming distance per syllable between Slope and Parsons' code measurements (Diff_Hamm_Slope_Par) was compared across groups and sequence length conditions (SLC). The Diff_Hamm_Slope_Par was obtained after subtracting hamming distance per syllable for Parsons' code measurement from hamming distance per syllable for Slope measurement. It was hypothesized that the interaction of Group and SLC would exist in the Diff_Hamm_Slope_Par such that the native Mandarin speaker group would demonstrate significantly larger Diff_Hamm_Slope_Par than the non-Mandarin speaker group at all sequence length conditions (see Figure 4). That is, the non-Mandarin speakers were expected to produce all syllable sequences in the same concurrent manner and demonstrate similar Diff_Hamm_Slope_Par across all sequence length conditions. Conversely, it was expected that the native Mandarin

speakers would produce Mandarin tone sequences in a concurrent manner only when the sequence length became long enough to exceed their stored motor programs for these gestures. Thus, it was expected that the Slope hamming distance would be smaller for the native Mandarin speakers for short utterances, and it would increase as the sequence length becomes longer. It was also expected that the Parsons' code measurement would capture the changes in the f_0 trajectory with more sensitivity when sequences were produced in a concurrent manner. Thus, the hamming distance per syllable for Parsons' code would not change much by the speech production manners (advance vs. concurrent programming) for both groups. Thus, the Diff_Hamm_Slope_Par was predicted to have a diverging pattern between the two speaker groups as the sequence length increased because the native Mandarin speaker group would demonstrate significantly greater Diff_Hamm_Slope_Par as the sequence length becomes longer, and the value would not change significantly among the syllable length conditions for the non-Mandarin speaker group.

As hypothesized, it was observed that the interaction between Group and Sequence Length Condition and the main effects of Group and Sequence Length Condition were significant on Diff_Hamm_Slope_Par. The Diff_Hamm_Slope_Par was higher in the native Mandarin speaker group than that of the non-Mandarin speaker group. However, the Diff_Hamm_Slope_Par did not change across Sequence Length Conditions as hypothesized. The hypothesized larger Diff_Hamm_Slope_Par with increased sequence length in the native Mandarin speakers did not occur. In constrast, the Diff_Hamm_Slope_Par value increased between three- and six-syllable conditions and stayed about the same level between six- and nine-syllable conditions in the non-Mandarin speaker group. Thus, unexpectedly, the difference between the two groups on this measure decreased as the sequence length became longer (see Figure 26). To explain this finding,

the hamming distance per syllable for Slope and Parsons' code measurements were examined separately.

It was expected that the Slope hamming distance per syllable would increase as the sequence length became longer in the native Mandarin speaker group, and would not change significantly as the sequence length increased for the non-Mandarin speaker group. However, it was observed that both groups changed in a parallel manner. The Slope hamming distance per syllable increased as the sequence length changed from three-syllable to six-syllable length conditions and remained constant from the six- to nine-syllable length conditions. Considering that the hamming distance was compared at a syllable level, any relative increase in this measurement within a speaker reflected a cost due to changes in the sequence length. It was concluded that both speaker groups had a cost due to an increase in the sequence length as measured by the Slope hamming distance. If the Slope hamming distance per syllable reflected any digression from the proportional change of the template trajectory, this divergence was significantly larger in the six- or nine-syllable length conditions than the three-syllable length condition, and there was no difference in Slope hamming distance per syllable between six- and nine-syllable length conditions. Thus, as speculated, the outcome f_0 trajectory deviated more from the proportional change of the template trajectory at the six-syllable length condition because speakers no longer had a GMP available for this length.

Interestingly, but counter to predictions, this pattern was parallel in both speaker groups. Consistent with the results of the GMP errors for Tone 2 in the previous section, it was speculated that non-Mandarin speakers would also utilize some forms of stored motor programs in order to produce three-syllable level tone sequences. However, because the Slope hamming distance per syllable was larger in the non-Mandarin speakers than in the native Mandarin speakers, the

outcome f_0 trajectory of the non-Mandarin speakers appeared to be proportionally less similar to the template trajectory than that of the native Mandarin speakers. It can be concluded that non-Mandarin speakers utilized GMPs that generate less similar (or less accurate) trajectories from the template trajectory than native Mandarin speakers.

Two possibilities can account for the lack of change in the Slope hamming distance per syllable between six- and nine-syllable conditions. First, there was no additional increase in the Slope hamming distance per syllable after a certain sequence length (six-syllable length in this case). Second, the degree of divergence from the template did not linearly increase between six- and nine-syllable length conditions because production of the last three syllables in the nine-syllable sequence production condition appears to have been facilitated by the repetition of the first three syllables. If the latter interpretation is accurate, this result would mean that there was an unanticipated artifact in the experimental design in the current study. A future study including different tones for all syllable positions in the nine-syllable sequence condition should be able to address this speculation.

Similar to the Slope measurement, the hamming distance per syllable for Parsons' code measurement (Hamm_par_syl) was significantly larger in non-Mandarin speakers than that in native Mandarin speakers. Because Hamm_par_syl was expected to track the changes in the f_0 trajectory in a more concurrent manner, this measurement was expected to be a more sensitive measure of changes in the f_0 trajectory. Also, this measurement was expected to be less influenced by the proportional mismatch between the f_0 outcome trajectory and the template f_0 trajectory. It was hypothesized that there would be a constant hamming distance per syllable for Parsons' code measurement across different sequence length conditions for both speaker groups. Contrary to this prediction, the Hamm_par_syl significantly increased between the three-syllable and six-syllable

sequence length conditions and remained constant between the six-syllable and nine-syllable sequence length conditions for both speaker groups. This pattern was similar to Slope measurement results across sequence length conditions. There appears to have been an additional cost in motor programming when the sequence length changed from the three- to the six-syllable length conditions that was not realized between the six- and nine-syllable length conditions.

Not surprisingly, the Hamm_par_syl was higher for the non-Mandarin speakers than for the native Mandarin speakers. This is interpreted to mean that the non-Mandarin speakers' trajectories varied more from the template trajectory than the native Mandarin speakers. This result suggested that the non-Mandarin speakers produced less-refined sequences and utilized less accurate motor programs than the native Mandarin speakers.

In summary, the Slope hamming distance per syllable increased in the native Mandarin speakers as the sequence length became longer. This result supports the hypothesized concurrent production of longer utterances in native Mandarin speakers. However, this trend was also true for the non-Mandarin speakers, so it was speculated that non-Mandarin speakers also utilize some forms of stored motor programs for short utterances. The Parsons' code hamming distance per syllable also increased as the sequence length became longer in both speaker groups. This result suggests that there was a cost in increasing sequence lengths from three-syllable to six-syllable lengths when measured by Parsons' code hamming distance. The hamming distance values were lower in the native Mandarin speaker group than in the non-Mandarin speaker group, supporting the retrieval of more accurate GMPs in native Mandarin speakers than non-Mandarin speakers.

A different manner of production between the two speaker groups was more evident in the results of the Diff_Hamm_Slope_Par than in the Parsons' code hamming distance measurement. It was initially hypothesized that the native Mandarin speaker group would demonstrate a

significantly greater Diff_Hamm_Slope_Par value as the sequence length became longer. No significant differences in the Diff_Hamm_Slope_Par across sequence length conditions were predicted in the non-Mandarin speaker group. However, the results were not consistent with these predictions. Both the Slope and Parsons' code measurements increased as the sequence length became longer, and the Diff_Hamm_Slope_Par was not significantly different across the sequence length conditions in the native Mandarin speaker group. The Diff_Hamm_Slope_Par was significantly smaller in non-Mandarin speakers than native Mandarin speakers at the three-syllable length condition, but the non-Mandarin speakers' Diff_Hamm_Slope_Par increased and approached a level similar to the Diff_Hamm_Slope_Par value of the native Mandarin speakers' in the nine-syllable length condition.

In order to explain the underlying reasons for the observed findings that both Slope and Parsons' code measures were lower in the native Mandarin speakers than non-Mandarin speakers, supplementary analyses/comparisons were computed (see Figure 30). The degree of difference between the two measurements (Slope vs. Parsons' code) was larger in the native Mandarin speaker group as compared to the non-Mandarin speaker group at three-syllable length condition. The Slope hamming distance saw a greater change in non-Mandarin speakers (9.88%) than in the native Mandarin speakers (8.89%) between three- and six-syllable length condition. Also, the Parsons' code hamming distance demonstrated a smaller change in the non-Mandarin speakers (10.98%) than the native Mandarin speakers (21.45%) between three- and six-syllable length conditions. Thus, the hamming distance difference between Slope and Parsons' code measurement of the non-Mandarin speakers increased substantively and reached almost the same level as that of the native Mandarin speakers' at the six-syllable length condition.

Earlier, it was mentioned that the Slope measurement might increase when there is no proportional change because there is no GMP available at the six-syllable length condition. If speakers relied more heavily on a concurrent production manner, the Slope measurement would be expected to deviate more from the template Slopes than if they didn't use a concurrent production mode. On the other hand, it was predicted that the Parsons' code measurement would involve a concurrent judgment to assign proper values and this measurement would fit better with the f_0 trajectory if the utterance was controlled concurrently for longer utterances. This was the case for both speaker groups. The increase in the Slope values was relatively higher in the non-Mandarin speakers than native Mandarin speakers, and the increase in the Parsons' code was relatively smaller in the non-Mandarin speakers than native Mandarin speakers between three- and six-syllable length conditions. It is speculated that this was because non-Mandarin speakers did not utilize GMPs as much and controlled their speech production more concurrently than the native Mandarin speakers at the six-syllable length condition. This result is consistent with the interpretation that speakers changed their production manner from advanced programming to concurrent programming in longer tone sequences. This evidence of change in production mode is consistent with the key-pressing studies by Verwey and colleagues (1995, 1996; 1996). Also, this result supports the interpretation that speech production manner depends on speech production experience. This group difference also suggests that the information in the GMP may be stored as proportional information.

The Slope and Parsons' code measurements were adapted from Ramadoss' (2012). The Hamming distance measurements were used as tertiary measurements to compare Slope and Parsons' code results directly because Slope and Parsons' code measurements have different scales. Using the Hamming distance measurements was novel for this literature. Furthermore,

while Ramadoss (2012) used Slope and Parsons' code measurements to compare perception simulation models using data from different measurements, this study utilized the same measurements to assess speech motor control. Therefore, a direct comparison between results from Hamming distances for Slope and Parsons' code and results from previous studies was not possible. Given that the differences in the results validly capture the differences in the the production manner for the two speaker groups as concluded, the Slope and Parsons' code measurements served as valid and sensitive metrics with which to capture those differences and changes in the production manner.

4.1.5 Reaction Time (RT), Average ISI, Ratio of RT/Average ISI

Counter to the prediction that RT would increase with an increase in sequence length, the RT did not change significantly as the sequence length became longer for either speaker group. Thus, the complexity effect, which refers to the increased RT as the sequence length becomes longer, was not evidenced in these results. Previous literature was not consistent in predicting the complexity effect on RT. Wright, et al. (2009) expected that the number of syllables would influence the SEQ process and the subsequent simple RT but not the choice RT. Based on Klapp, et al.'s findings (2003), it was expected that the choice RT would increase with an increase in the number of syllables, because the number of syllables determines the complexity of motor responses. Although the absolute RT values slightly increased as the sequence length increased in native Mandarin speakers, the difference in RT was not significant in this current study. This result supported consistent with Wright, et al.'s findings that the number of syllables does not determine the complexity of the speech motor response.

It was initially hypothesized that if the complexity effect was not found, there would be a significant interaction between Group and Sequence Length Conditions on time measurements. A shorter RT and shorter average ISI was expected for the native speakers than the non-Mandarin speakers in all Sequence Length Conditions. A significant decrease in the RT, increase in the average ISI, and decrease in the Ratio of RT over average ISI were expected as the sequence length became longer in the native Mandarin group. However, no such change was predicted among Sequence Length Conditions in the non-Mandarin group.

Interestingly, the interactions between Group and Sequence Length Conditions were not observed for any of the time measurements (RT, average ISI, and Ratio of RT over average ISI). That is interpreted to mean that the overall pattern of change on these time measurements was generally parallel between the two groups across the sequence length conditions. However, as hypothesized, the RT and average ISI were shorter for native Mandarin speakers than for the non-Mandarin speaker group. Despite this shorter RT and ISI in the native Mandarin speaker group, the Ratio of RT over average ISI was higher in the native Mandarin speaker group than the non-Mandarin speaker group when averaged across sequence length conditions. Thus, it was presumed that native Mandarin speakers spent relatively more time to prepare motor responses in advance of execution. This finding is consistent with the evidence provided by the native Mandarin speakers of utilizing GMPs more efficiently than by the non-Mandarin speakers. Furthermore, when averaged across groups, the Ratio of RT over average ISI significantly decreased between the three-syllable and nine-syllable sequence length conditions. This result supports the interpretation that both speakers relied more on a concurrent manner of production than advance programming as the sequence length increased.

The results of the RT, average ISI, and Ratio of RT over average ISI analyses were consistent with previous research findings (Levelt & Wheeldon, 1994; Reilly & Spencer, 2013; Spencer & Rogers, 2005; Verwey, 1995, 1996). Levelt and Wheeldon (1994) mentioned that RT was shorter before well-practiced syllable sequences than less-practiced speech sequences. Verwey (1995) observed a shorter ISI after practice as concurrent programming became more automatic. Verwey (1996) also reported a shorter RT and Inter-Response Interval (IRI) and a larger Ratio of RT over IRI in a key-pressing task after practice. He addressed the loading of larger and fewer motor units to produce the same length of key-pressing tasks after practice. Verwey (1999) also proposed that motor units appeared to be loaded faster after practice and concurrent loading of motor responses seemed to occur during the IRI. Reilly and Spencer (2013) and Spencer and Rogers (2005) also found shorter RT and ISI in practiced speakers. All of these previous study findings are consistent with the findings of the current study observed from the three time measurements.

To conclude, participants from both speaker groups demonstrated similar patterns of timing control in speech production as measured by RT, average ISI, and Ratio of RT over average ISI. Despite shorter RT and ISI in the native Mandarin speaker group than the non-Mandarin group, it appeared that the relative degree of advance planning was stronger in the native Mandarin speaker group than non-Mandarin speaker group as evidenced by higher Ratio of RT over average ISI. Nonetheless, the Ratio of RT over average ISI decreased significantly as the sequence length became longer, in particular, between three- and nine-syllable lengths. Thus, this result supports the idea of a transition in the manner of control from advance programming to concurrent programming as the sequence length increased for both speaker groups.

4.1.6 General Discussion

This study compared the mechanisms by which speech movements are controlled in more-practiced native Mandarin speakers and less-practiced non-Mandarin (English) speakers. Several acoustic measures including reaction time (RT), inter-syllable intervals (ISIs), and a few f_0 trajectory measurements that were obtained while participants produced three-syllable, six-syllable, and nine-syllable Mandarin tone sequences were compared. It was hypothesized that native Mandarin speakers would prepare the sequences in advance for short utterances, but they would change their production mode to a concurrent preparation manner as the sequence length increased, as has been reported in key-pressing task studies (Verwey, 1995, 1996; Verwey & Dronkert, 1996). It was also predicted that less-practiced, non-Mandarin speakers would utilize a concurrent production manner regardless of the length of the utterance because they would not have had a chance to develop motor programs for utterances longer than a syllable.

While attempting to answer the above questions, the type of information stored in motor programs was also investigated. Schmidt (1975) proposed in Schema theory that information about the proportional relationships among the components of a movement trajectory is stored in a Generalized Motor Program (GMP). This is the core and invariant information which generalizes across different speech production contexts as long as the same speech task is performed. Maas, et al. (2008) proposed that the GMP for speech may correspond to motor programs for phonemes, words, and phrases. However, the notion of GMP was challenged when some studies failed to observe consistent proportional relationships among acoustic, kinematic and electro-myographic (EMG) speech signals (e.g., Max & Caruso, 1997).

In the current study, it was speculated that the reason why previous researchers did not observe invariant proportional relationships among the acoustic, kinematic, and EMG signals was

because the movement outcome trajectory is produced after concatenating different numbers of motor programs which have various unit sizes. It was also expected that these motor programs would be prepared and executed differently for each individual. Thus, less variable and larger sized motor programs (GMPs) were predicted in the native Mandarin speakers who have more experience with the target language than the non-Mandarin speakers. To explore these questions, speech movement outcomes were measured acoustically using the information that tracked the changes of the fundamental frequency (f_0), which is known to inform the Mandarin lexical tones over time.

Two reasons why previous researchers failed to find the consistent proportional relationships among acoustic, kinematic, and EMG signals were proposed. First, variability in the movement trajectory is an expected phenomenon if the speakers utilize various sized GMPs and concatenate these GMPs differently. Second, if all participants perform the same movement tasks but parameterize them differently, then variability between individuals in the movement trajectories would be expected. In other words, if each speaker reparameterizes differently during the execution of motor programs, at different times, it would increase the variability in the speech movement outcomes. The first possibility was supported better than the second because parameter variability was not observed within the nine-syllable length utterances. Although future studies will be required to explore this aspect in more varied contexts, for now, it appears that the variability in the trajectory was due to the concatenation of several motor programs rather than due to reparameterization.

A few previous studies that examined motor learning either did not take into account the effect of parameterization in the speech motor outcomes (e. g., Adams & Page, 2000; Max & Caruso, 1997), or their experimental methods were not sensitive enough while attempting to tease

apart the effect of parameter specification from the global movement outcome (e. g., Lai et al., 2000). Thus, there was a risk of reporting mixed effects of GMPs and parameters while addressing these two aspects. The method borrowed from Wulf, Schmidt, and Duebel (1993) facilitated the separation of GMP's influence from the parameter change in the current study. This represents a novel approach for speech production research as this method has been used primarily in kinematic studies involving limb movements (J.-H. Park & C. H. Shea, 2003; Wulf & Schmidt, 1997). Those speech production studies that did include normalization processes (Almelaifi, 2013; Hao, 2018; Smith et al., 1995) did not normalize the speech production data based on proportional scaling as proposed by Schema theory. The current study attempted to separate the effects of GMP errors from parameter errors in the acoustic information. Future studies using similar methods will provide additional evidence about the relationships between GMP and parameters.

The data from this study provide support for the existence of GMPs for Mandarin lexical tones in native Mandarin speakers. These native Mandarin participants appeared to retrieve and execute movements in advance of initiating speech production at the three-syllable length condition. The findings from this study suggested that the production mode switched from a pre-programmed one to a concurrent production mode for longer utterances. The following evidence supported the interpretation that these native Mandarin speakers retrieved these gestures from stored motor programs, at least for three-syllable length condition: 1) the relatively higher increase in GMP errors per syllable, 2) relatively higher increase in the hamming distance per syllable for the Parsons' code measurement by the native Mandarin speakers than the non-Mandarin speakers, 3) along with their overall lower GMP errors per syllable, and 4) higher Ratio of RT over average ISI across all sequence length conditions as compared to non-Mandarin speakers. In particular, their overall lower GMP errors per syllable and higher Ratio of RT over average ISI consistently

across all sequence length conditions as compared to non-Mandarin speakers suggested consistent use of GMPs or advance programming to some extent for longer utterances in the native Mandarin speaker group. These results are consistent with the hypothesis that stored motor programs are retrieved and used but not at all length sequences, and they are determined by the experience of the speakers (Levelt & Wheeldon, 1994; Reilly & Spencer, 2013; Spencer & Rogers, 2005; Verwey, 1995, 1996).

There was little change in GMP errors per syllable between the six- and the nine-syllable length conditions, potentially due to the nature of the task in which the first three syllables repeated in the 7^{th}, 8^{th}, and 9^{th} syllable positions in the nine-syllable sequence condition. This repeating environment might have facilitated the development of motor programs for the last three syllables in the nine-syllable sequence length condition. This cost of the increasing sequence length in GMP errors will be discussed later.

Few studies have directly explored the existence of a GMP for lexical tones. Overall, the findings of the current study supported the existence of a GMP for lexical tones, in particular, when native Mandarin speakers produced Mandarin tone sequences in the three-syllable length condition. The fact that the GMP errors per syllable were not significantly different between two different syllable positions within an utterance (i. e., 2^{nd} and 8^{th} syllable positions in nine-syllable sequence condition), even in non-Mandarin speakers, may challenge this conclusion. However, the following findings supported this conclusion. First, the GMP errors per syllable were smaller in the native Mandarin speakers than the non-Mandarin speakers. Second, the hamming distance per syllable for Slope measurement was smaller in the native Mandarin speakers than the non-Mandarin speakers. Third, the increase in the Slope measurement between the three- and six-syllable length conditions was relatively smaller in the native Mandarin speakers than the non-

Mandarin speakers. These findings support the interpretation that native Mandarin speakers utilized well-developed GMPs to produce well-practiced lexical tone sequences, whereas the non-Mandarin speakers were less efficient at preparing these motor programs before utterance initiation.

There is the possibility that the motor programs used in these two different syllable positions (i. e., 2nd and 8th syllable positions in nine-syllable sequence condition) by the non-Mandarin speakers were also some type of stored motor programs (aka GMPs). Although these motor programs were not as accurately or efficiently executed as the ones used by native Mandarin speakers, as indicated by higher GMP errors per syllable in this group, these speakers might also have assembled/retrieved consistent stored motor programs for lexical tones. As addressed earlier, these imperfect motor programs might have been borrowed from their motor program storage to generate speech sounds that resemble target sounds as closely as possible. Also, these motor programs might have yielded a certain level of accuracy in the non-Mandarin speakers since all errored speech sounds were excluded from the analysis.

The evidence for a transition from advanced programming to concurrent programming production modes was derived from the presence of increased GMP errors per syllable and Slope and Parsons' code hamming distances per syllable, and from the decreased Ratio of RT over average ISI between three- and nine-syllable length conditions in both speaker groups. This finding is consistent with the interpretation of a transition in the speech production manner from advance programming to concurrent programming.

Increasing variability in the trajectories between three- and six-syllable length conditions reflect a cost of concurrent concatenation of several motor programs in longer utterances. In speech synthesis studies, concatenation cost was used to "reflect a level of perceived discontinuity

between two consecutive units" (Legát & Matoušek, 2010). However, in this current study, the term is used to refer to the increased variability in the speech movement outcomes due to the effort to concatenate more than one speech motor program. This cost is expected to occur when speakers prepare speech movements concurrently, preparing the upcoming motor responses while executing the prior motor responses. The fact that there was a relatively smaller increase in the Slope measurement and a relatively larger increase in the Parsons' code measurement between three- and six-syllable length conditions in the native Mandarin speakers as compared to non-Mandarin speakers, supports this cost interpretation in the native Mandarin speaker group. This refers to the cost due to concurrent movement control rather than due to changes in the information stored in the motor programs. In other words, relatively smaller changes in the Slope hamming distances per syllable in the native Mandarin speaker group implies a lesser degree of deviation in their f_0 trajectories from the proportionally scaled template trajectories than the degree of deviation in their f_0 trajectories produced by the non-Mandarin speakers. This result also may explain why previous studies did not find consistent relative timing or force relationships in the outcomes of connected speech motor tasks. This failure could be due to the effort of concatenating motor programs while the motor programs themselves maintained information about proportional relationships among the components of the movement structure. This is information known to be stored in GMPs (Schmidt, 1975, 2003).

In addition to the fact that there appeared to be a cost for concatenating motor programs, speakers' loss of proportional information with an increase of sequence length was examined. This loss seems unlikely because both speaker groups maintained similar levels of GMP errors in the two syllable positions within an utterance in nine-syllable length utterances. Thus, as long as the

stored motor programs were accessed, it seems that information about proportional relationships was maintained throughout the utterance.

The current study provides evidence of different unit sizes of speech motor programs among different speakers. The increase in GMP errors between three- and six-syllable length conditions was smaller in the non-Mandarin speaker group than that of native Mandarin speakers. Because the current study did not examine the GMP errors per syllable at each syllable level within an utterance, but rather obtained GMP errors at the whole sequence level and divided it by the number of syllables, this method was not sensitive to changes in the GMP errors that occured within the target sequence length. Therefore, the smaller increase in the GMP errors per syllable between three- and six-syllable length conditions also implied that the non-Mandarin speakers changed their production mode from advance programming to concurrent programming during the production of the three-syllable length utterances with little change observed between three- and six-syllable length conditions. Thus, it is speculated that changes in the GMP errors per syllable due to changes in the production mode were relatively smaller in the non-Mandarin speaker group as compared to native Mandarin speakers between three- and six-syllable length conditions. Also, as mentioned above, the higher Ratio of RT over average ISI in the native Mandarin speakers as compared to non-Mandarin speakers suggested that the native Mandarin speaker group relied on advanced programming more than the non-Mandarin speaker group in the longer utterances. Overall, GMPs of different unit sizes were expected based on learning experience.

The pattern of changes was similar between native Mandarin speakers and non-Mandarin speakers, particularly for the time measurements. Initially, it was predicted that RT, average ISI and Ratio of RT over average ISI would differ significantly across sequence length conditions for the native Mandarin speakers while they would not for non-Mandarin speakers. Unexpectedly, RT

in both speaker groups did not differ significantly across sequence length conditions. The Ratio of RT over average ISI decreased significantly between the three-syllable and the nine-syllable length conditions for both speaker groups. This suggests that both speaker groups relied on a concurrent production manner more than advance programming as the sequence length increased. Overall, the findings suggested that speech production behaviors were similar across all speakers to some extent, regardless of their learning experience when producing Mandarin tone sequences. This unexpected similarity in speech production behavior suggests that advanced programming could occur in non-Mandarin speakers with comparatively less-practice for short utterances to some extent. Similarly, well-practiced native Mandarin speakers may begin concurrent programming sooner than expected after speech initiation.

Despite similarities in performance, there were both quantitative and qualitative differences between the two speaker groups in their utilization of motor programs and how they switched between two production modes. It appeared that speakers with more tonal language production experience produced motor programs in larger units and more in advance as Verwey (1996) observed in the key-pressing task. This interpretation is supported by the fact that the native Mandarin speakers produced shorter RTs and shorter average ISIs, but higher Ratios of RT over the average ISI than non-Mandarin speakers in all sequence length conditions. The native Mandarin speakers spent relatively longer time to prepare motor responses in advance of movement initiation and spent less time to prepare the next motor responses during the execution of prior motor responses.

Additionally, the effect of learning experience was also supported when GMP errors per syllable were examined. In general, the GMP errors per syllablewere smaller in the native Mandarin speakers than non-Mandarin speakers. Also, the relative increase in GMP errors per

syllable was larger in the native Mandarin group than in the non-Mandarin group as the sequence length changed from three-syllable to six-syllable lengths. This was presumably because the native Mandarin speakers retrieved and executed stored motor programs in advance of movement initiation in the three-syllable condition and then changed their production mode to a concurrent production manner as the sequence length increased to the six-syllable condition. The degree of change in GMP errors per syllable was smaller in the non-Mandarin speaker group possibly because they relied more on a concurrent production manner throughout different sequence length conditions. It is therefore concluded that the degree of advance programming is influenced by the degree of motor control experience for that specific language.

Because this study included only three different sequence length conditions, based on previous key-pressing tasks (Verwey, 1995, 1996; Verwey & Dronkert, 1996), and the nine-syllable length condition failed to add complexity due to length because of an artifact in the experimental design, it is not possible to determine where exactly the transition of production mode occurred. As addressed earlier, because GMP errors per syllable at the three-syllable unit level were obtained instead of at a single syllable level, it is not possible to explain where a breakdown may have occurred within a three-syllable unit. Despite this limitation, this study provides evidence of a transition between two different production modes in the two speaker groups. The transition occurred between the three- and six-syllable length conditions for both speaker groups and it is speculated that the concurrent production mode was utilized by the non-Mandarin speakers earlier than by native Mandarin speakers during three-syllable sequence productions. These findings support a continuum view that any speaker may be able to utilize two different production modes depending on their need, as proposed by Miller (2001), rather than exclusively use direct and indirect production routes at a specific situation.

4.1.7 Limitation and Future Studies

The findings of this study may have limited generalization to other speech production tasks and conditions for the following reasons. First, target sounds were presented in Pinyin without related Chinese characters. Pinyin expresses sounds using the Roman alphabet system combined with diacritics to denote tones (e.g., bā, bá, bǎ, bà). This pinyin target sound presentation was employed to minimize semantic and syntactic activation while participants were preparing phonological and phonetic aspects of Mandarin tones. However, this condition may not have fully prevented the activation of representative linguistic meaning of that specific pinyin in native Mandarin speakers. It is also possible that motor control patterns for natural speech tasks and unnatural nonsense speech tasks may be different. Additional study using Chinese characters with their attached meanings for the native speakers will be required to sort out exactly how much meaning was derived and whether these linguistically meaningful productions carry accessible motor programs with them. Thus, the generalization of the current study's findings to natural speech production awaits additional experimentation. This experimentation might include one speech production session using a nonsense speech task, and the next session might include Chinese characters to activate syntactic and semantic processing. In this case, the experimental condition should include many filler stimuli and enough time between sessions to prevent learning effects. Alternatively, participants could be asked to produce nonsense pinyin sequences in one session and then, produce the same sound sequencess in a priming paradigm, again to examine the degree of influence of the semantic and syntactic activations. A facilitating or distracting Chinese character, which can be interpreted only in one way by the context or by the most frequently activated semantics, might be presented as a prime prior to production of the nonsense pinyin sequences. Reaction times would be expected to slow-down or speed-up based on the prime condition compared to the control (non-

prime) condition. The performance might be compared between the primed and non-primed conditions or between the first and second speech production sessions. In this way, the effect of linguistic context (semantic and syntactic activations) can be examined while producing nonsense speech stimuli for utterances of various lengths from which the existence of advanced programming versus concurrent programming can be inferred.

Second, in this study, the target speech sounds were presented in written form. This involved a reading process: the decoding of graphemes and the encoding of phonemes. However, motor programming for reading can be different from that for natural speech (Liberman, 1992). Thus, it is necessary to further examine the influence of reading on the results of this study. A picture could be used instead of graphemes to induce natural speech as in the Lee, et al. study (2015). A story retelling task could also be used to induce more natural speech production while the context is constrained.

Last, the current study argued that participants produced speech movements in a concurrent manner in the longer sequences. This could be due to an efficiency in producing motor responses in a concurrent manner. However, because the size of a programmable unit may also rely on a memory limitation, how much motor command information a motor control system may hold at a time, warrants further attention. The relationship between the size of the motor program (and hence the need for concurrent productions) and the size of the motor buffer is worthy of additional research. Bohland, et al. (2010) proposed a potential physical area where a motor program buffer may exist in their GODIVA (Gradient Order DIVA) model. According to their model, the basal ganglia mediates between the Inferior Frontal Sulcus (IFS) where phonological representations live, and the speech sound map, where phonetic representations live. Also, they proposed that motor plan cells in the motor cortex play the role of motor program buffers and receive inputs from

the speech sound maps. Therefore, we also need to extend this study to explore the relationship between the processing unit and individual differences in motor program buffer size (or short-term or working memory). The speech production may break down when a programming unit is larger than the buffer size despite the processing itself being intact. Knowing the buffer size as well as the processing unit is important because it provides information about which aspect is impaired and what the remaining function or strategies will be within an individual with speech production disorders. Previous studies had prior hypothesis about the size of processing units such as syllable, phoneme, descriptive features, or gestures (Browman & Goldstein, 1986; Browman & Goldstein, 1988; Browman & Goldstein, 1989; Guenther, 1995; Levelt et al., 1999; van der Merwe, 1997; Ziegler, 2013; Ziegler & Ackermann, 2013; Ziegler et al., 2011). This aspect has been controversial. We need to extend this question about processing units further while considering the relationship between the processing unit size and the motor program buffer size. These aspects could be studied by manipulating stimuli lengths or by comparing reaction times or processing times at different reaction time paradigms.

5.0 CONCLUSION

This study was initiated with the intent to explain previous studies' (e. g., Löfqvist, 1991; Max & Caruso, 1997) inability to find evidence for Generalized Motor Programs (GMPs), the notion which Schmidt (1975) proposed in the Schema theory. It was believed that information about the relative timing and relative force of movements and the order of motor events is stored in these GMPs. The above researchers did not observe this presumably constant relative timing and force information in the acoustic, electro-myographic (EMG), or kinematic data across different speech rate conditions. However, instead of discarding the notion of the GMP, this study attempted to explain the potential reasons for finding more variable outcome trajectories than proportionally scaled trajectories by noting possible influences of various unit sizes of motor programs and differences in production modes (advance programming vs. concurrent programming) among different individuals. Because these two aspects may be affected by language production experience, this study included well-practiced native Mandarin speakers and less-practiced non-Mandarin speakers, in order to determine the differences in motor program size and production manner in the acoustic measures of Mandarin tones.

A consistent proportional relationship was expected to be more readily observed if it is examined in a single motor program, which may vary in size depending on each performer's speech target and experience. Verwey and colleagues (1995, 1996; 1996) observed changes in production modes from preparing a motor response in advance to preparing it in a concurrent manner as the sequence length became longer in key-pressing tasks. Thus, this current study hypothesized to find proportional relationships in the trajectories of speech movements by the experienced speakers, which would support the existence of GMPs. More experienced speakers

were expected to show more flexible production modes as well. Thus, they would control speech related movements by flexibly switching between advance programming and concurrent programming based on the contextual needs (sequence length conditions). In contrast, it was predicted that non-Mandarin speakers would have little flexibility in the speech production manner and rely more on concurrent movement control to produce Mandarin tone sequences.

Twenty-four native Mandarin and twenty-four non-Mandarin male speakers (19-30 years of age) produced three-syllable, six-syllable, and nine-syllable Mandarin tone sequences. The following dependent variables were examined: parameter variability, which is based on changes in time and f_0 magnitude; GMP errors (variability around the f_0 trajectory measured by sum of Euclidean distance values); Hamming distance between Slope and Parsons' code measurements; reaction time (RT); inter-syllable intervals (ISIs); and Ratio of RT over average ISI. The results supported the hypotheses of this current study.

First, this study examined the effect of reparameterization during execution of motor programs on the movement outcome that may compromise the proportional relationship information. No significant parameter change was observed in either group in the current study. Thus, it was concluded that the variability in the outcome trajectory may not be due to reparameterization, but due to the concatenation of several motor programs.

Second, the results supported the existence of GMP for lexical tones. Native Mandarin speakers produced lower GMP errors per syllable as compared to non-Mandarin speakers. Hamming distance per syllable for slope measurement was smaller in the native Mandarin speakers than the non-Mandarin speakers. Also, there was a smaller increase in the slope measurement (less deviation from template trajectory) between three- and six-syllable length conditions in the native Mandarin speakers than the non-Mandarin speakers. This evidence corroborated that native

Mandarin speakers retrieved and executed stored motor programs (aka GMPs) to produce lexical tones at least for the three-syllable length utterances.

Third, the findings also suggested that the production mode switched from advance programming to concurrent programming as the sequence length increased. This possibility was supported when GMP errors per syllable increased as the sequence became longer, Slope and Parsons' code hamming distances per syllable increased, and the Ratio of RT over average ISI between three- and nine-syllable length conditions decreased in both speaker groups. Thus, it was presumed that the production mode had changed based on utterance length.

Fourth, the timing of this switch occurred later in more-experienced speakers. Non-Mandarin speakers relied on concurrent movement control earlier than the native Mandarin speakers because the increase in the GMP error per syllable between three- and six-syllable length conditions was relatively smaller in this speaker group as compared to the native Mandarin speakers. The increase in the slope hamming distance values was relatively higher in the non-Mandarin speakers than native Mandarin speakers between three- and six-syllable length conditions, and the increase in the parsons' code was relatively smaller in the non-Mandarin speakers than native Mandarin speakers between three- and six-syllable length conditions. It was speculated that this was because non-Mandarin speakers utilized GMPs less and controlled their speech production more concurrently than native Mandarin speakers in the three- and six-syllable length conditions. Thus, the mode changes in speech production manner depended on speech production experience. These group differences in hamming measurements also suggested that the information stored in the GMP might be proportional information. Overall, the different unit size of GMP and different changes in the production mode were supported by the language production experience.

Last, there appeared to be a cost for concatenating motor programs because the GMP error per syllable still increased between three- and six- or between three- and nine-syllable length conditions. This concatenation cost potentially explained why previous studies failed to find consistent proportional relationships. That is, the variability possibly increased in the f_0 trajectory due to concatenation efforts of the motor programs.

However, this increased variability in the outcome f_0 trajectory did not seem to be due to the information change in the motor programs. This was because both speaker groups maintained similar levels of GMP errors in the two syllable positions within an utterance in nine-syllable length utterances. The fact that the variability of the motor programs for the same sound was relatively constant within an utterance in both speaker groups suggested that even the non-Mandarin group might have utilized a rough format of constant motor programs borrowed from their motor program storage. Even non-Mandarin speakers could use the stable GMPs for each speech sound once they acquired the ability to produce each tone relatively accurately at a syllable level. However, the GMP might still have been less precise (aka more variable motor control) in this group than the native-Mandarin group as evidenced by larger GMP errors per syllable in the non-Mandarin group. As long as the GMPs were accessed, the information about proportional relationships appeared to be maintained throughout the utterance.

There are many aspects to improve in this study and further examinations are required in more varying contexts. Despite this fact, this study was the first attempt to support the existence of GMP for lexical tones and provide an alterantive explanation for why previous studies resulted in variable movement outcomes that did not support the existence of GMPs. It might have been, in reality, due to concatenation of more than one varying sized GMP.

APPENDIX A

PRE-SCREENING QUESTIONNARIE

Subject # _____ Date: _____

Examiner Initials: _____

Discontinue the questionnaires if not meeting the criteria below.

Table 15 Pre-screening questionnaire

	Questionnaires	Inclusion Criteria
1	Age (yrs.)? _____	19-30
2	Sex? Male Female	Male
3	Years of formal education?	(Descriptive)
4	Please list all the languages you know in order of acquisition (your native language first): Language A: Language B: Language C: Language D: Language E:	Mandarin (Go to Q 6, 7) or English (Go to Q 5, 6)
5	If you are a non-Mandarin speaker, did you have any experience of learning Mandarin? Yes No	No

Table 15 Pre-screening questionnaire (Continued)

<table>
<tr><td>6</td><td>On a scale from zero to five (0: none, 5: proficient) please select your level of proficiency in speaking, understanding, reading, and writing in the first and second languages you listed above:<table><tr><td></td><td>Language A:</td><td>Language B:</td></tr><tr><td>Speaking:</td><td>0 1 2 3 4 5</td><td>0 1 2 3 4 5</td></tr><tr><td>Understand spoken language:</td><td>0 1 2 3 4 5</td><td>0 1 2 3 4 5</td></tr><tr><td>Reading:</td><td>0 1 2 3 4 5</td><td>0 1 2 3 4 5</td></tr><tr><td>Writing:</td><td>0 1 2 3 4 5</td><td>0 1 2 3 4 5</td></tr></table></td><td>First language proficiency:
Understanding ≥ 4.4,
Speaking ≥ 4.05,
Reading ≥ 2.73,
Writing ≥ 2.6</td></tr>
<tr><td>7</td><td>If you are a Mandarin speaker, what is your English proficiency level as measured by International English Language Testing System (IELTS) or Test of English as a Foreign Language (TOEFL) score?
Pass / Fail</td><td>IELTS (academic module): ≥ 7;
TOEFL:
iBT ≥ 80;
CBT ≥ 213;
PBT ≥ 550.</td></tr>
<tr><td>8</td><td>If you are a Mandarin speaker, what province are you originally from?</td><td>If the province is where Mandarin is spoken, go to Q 12.
If the regional language is not Mandarin, go to Q 9.</td></tr>
<tr><td>9</td><td>Did you learn the regional language before you learned Mandarin/English?</td><td>If yes, go to Q 10&11.
If no, go to Q 12.</td></tr>
<tr><td>10</td><td>On a scale from zero to five (0: none, 5: heavy or pervasive), rate your perception of how much of a regional language accent you have in speaking Mandarin/English.</td><td>≤3.36</td></tr>
</table>

Table 15 Pre-screening questionnaire (Continued)

11	On a scale from zero to five (0: never, 5: always) please rate how frequently others identify you as a non-Mandarin/English speaker based on your regional language accent in Mandarin/English.	≤3.61
12	How many years of formal education in Mandarin/ English do you have?	(Descriptive)
13	Do you have any problems with vision after correction? Yes No	No
14	Do you have any problems with hearing? Yes No	No
15	Have you ever had any kind of language or learning disability? Yes No	No
16	Do you have any history of neurologic or psychological problems (e.g. closed head injury, dementia, degenerative nervous system illness, schizophrenia, manic-depressive disorder, depression, or schizoaffective disorder, etc.)? Yes No	No
17	Do you currently have a problem with alcohol or drug abuse? Yes No	No

Cf. The question 4, 6, 10 and 11 are borrowed and modified from the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al., 2007).

APPENDIX B

SCREENING FORM

1. **Vision Screen (Snellen Card):**

Subject must be able read, copy, or match criterion line (≥ 20/40)

Yes No

2. **Audiogram:** (≤ 25 dB in one ear)

	500	1000	2000	4000
Right				
Left				

3. **Pure Tone Discrimination Task**:

	Participant's DL	Adapted from Spiegel and Watson (1984)
125 Hz		≤ 2.81Hz
1000 Hz		≤ 22.50Hz
2000 Hz		≤ 45.00Hz

4. **Oral Speech Mechanism Screening Examination-Revised (OSMSE-R)** (St Louis & Ruscello, 1981)

Pass / Fail

5. C-RTT-L

English version: Overall score: _____ (≥ 14.17) (McNeil et al., 2015)

Mandarin version: Overall score: _____ (≥ 12.36) (S.-H. K. Chen et al., 2013)

6. Cepstral Spectral Index of Dysphonia for Rainbow passege ($CSID_R$)

$CSID_R$ Score: _____ (≤ 24.3) (Awan et al., 2016)

7. General Speech Production Ability:

	Sustained phonation	Diadochokinetic	Vocal sweeps	Story Retell (Form A)
Respiratory	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal
Phonation	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal
Articulation	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal
Resonance	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal	Normal/Abnormal

8. Musical Experiences (Descriptive Purposes)

1) Have many years have you received a musical training?
   If the answer is not 1, proceed to the next questions.

   ① None, ② 0-5 years, ③ 6-10 years, ④ more than 11 years

2) Do you perceive yourself as a professional musician?
3) Are you a singer or an instrumental player? 1. Singer 2. Instrument Player 3. Both
4) What type of instrument do you play?

_____

Table 16 DDK Criteria

	Mean	SD	Mean ± 2SD	
Maximum Sustained Phonation	21.29 for age 18-39 years (n=78)	5.92	9.45-33.13	(Zraick, Smith-Olinde, & Shotts, 2012)
Comfortable Sustaind Phonation	5.76 for age 18-39 years (n=78)	0.73	4.3-7.22	
/pʌ/	6.6	1.1	4.4-8.8	(Westbury & Dembowski, 1993)
/tʌ/	6.7	1.2	4.3-9.1	
/kʌ/	6.1	1.0	4.1-8.1	
/pʌtʌkʌ/(syllable per second for male)	6.7	0.99	4.72-8.68	(Topbas, 2010)

APPENDIX C

GENERAL INSTRUCTIONS

General Instructions for Mono-Syllable Practice Phase:

You will practice on producing four Mandarin tones: bā, bá, bǎ, and bà. Listen carefully and repeat after the model speaker. You will practice on one type of tone at a time. Sometimes a visual feedback will appear on the screen that tracks your pitch change while you imitate the model speaker. This will inform you of how close your production was to the target sound. You will move onto another type of tone when you reach an 80% accuracy level with one tone in two consecutive practice blocks. If you don't reach this level of accuracy even after practicing on one tone about 100 times, you will discontinue participating in this experiment.

General Instructions for Practice Phase before Tone Sequence Production Phase:

Now you will produce Mandarin tone sequences. For example, the tone sequence of Tone 2-3-4-4-3-1 will appear on the screen as following:

bábǎbàbàbǎbā

Respond as soon as you see the target sequence on the screen. However, please do not hurry because you will have enough time to produce each tone sequence. You may press the x button to indicate that you completed your production. Please remember to press the button only when you completed your production. You will also be asked to produce each sequence at the 65-85dB level. Please monitor the sound level meter in front of you. The examiner will keep track of

your intensity level and will provide you with feedback when you exceed this range. There will be two blocks in this practice phase. A two-minute break will be given between these two blocks unless you ask for more time.

Do you have any questions? If not, let's try a few trials for practice.

General Instructions for Tone Sequence Production Phase:

The idea is the same. Now you will produce Mandarin tone sequences. For example, the tone sequence of Tone 2-3-4-4-3-1 will appear on the screen as following:

bábǎbàbàbǎbā

Respond as soon as you see the target sequence on the screen. However, please do not hurry because you will have enough time to produce each tone sequence. You may press the x button to indicate that you completed your production. Please remember to press the button only when you completed your production. You will also be asked to produce each sequence at the 65-85dB level. Please monitor the sound level meter in front of you. The examiner will keep track of your intensity level and will provide you with feedback when you exceed this range. There will be five blocks in this experimental phase. Each block will contain 24 trials. A two-minute break will be given between blocks unless you ask for more time.

Do you have any questions? If not, let's start!

Sustained phonation task: "Take a deep breath, then produce a sustained open vowel "aaaah~~~" at a comfortable pitch and loudness for as long as possible."

DDK task: e.g., "Take a deep breath, then produce "/p^ p^ p^ p^ p^ p^ ~~~/" as fast as you can at your comfortable pitch and loudness level for 5 seconds."

CSID-R task: "Take a deep breath, then read these sentences on the screen at your comfortable pitch and loudness level. Remember this is to test your vocal function, so please try to make a clear speech sound as much as you can."

CRTT task: In this task, you will listen to commands and respond to those commands. Please respond in the order the objects were presented. For example, when you hear a command like, “touch the blue circle and the red square,” please touch the blue circle first and the red square later. Also, in this task not only the accuracy but also the response time matters. Thus, please try to respond as fast as you can as well as as accurate as possible. Furthermore, each trial is independent, so remember that you do not have to memorize the prior trial to respond to the current trial. When you are done responding, please place your mouse pointer in the circle of the bottom screen, this will help you move to the next trial quickly. Otherwise, you will have to wait long for the next trial. Do you have any questions? If not, let’s start!

Tone discrimination task: In this task, you will hear two different tones in a row. You have to indicate which tone is higher. For example, you will hear “pi (high) -pi (low)” (with a vocal modeling by the examiner), what tone do you think is higher? Yes. Likewise, you will have to

indicate which tone is higher by clicking on the corresponding number (1 or 2) on the screen. In this task, the accurate response is important and the response time does not matter. Thus, try not to rush and spend all time you need before responding. Also, please feel free to hum after you hear the tones, this helps you have the better idea about what you are hearing. Are you ready? Let's start!

APPENDIX D

MUSICAL TRAINING EXPERIENCES

Table 17 Screening results-Musical training experiences

non-Mandarin Group	Years of musical training (unit: years) 1: none 2: 0-5 3: 6-10 4: 11 or more	Professional ? 1: Yes 2: No	1: Singer 2: Instrument Player 3: Both 4: Neither	Type of Instrument	Mandarin Group	Years of musical training (unit: years) 1: none 2: 0-5 3: 6-10 4: 11 or more	Professional ? 1: Yes 2: No	1: Singer 2: Instrument Player 3: Both 4: Neither	Type of Instrument
E1	3	2	3	Violin	M1	2	2	2	Violin, Chinese flute (bamboo)
E2	4	1	2	Clarinet, Saxophone	M2	1	2	2	Guitar
E3	4	2	3	Piano	M3	2	2	2	Piano, guitar
E4	2	2	2	Recorder	M4	2	2	2	Pipe
E5	4	2	2	Jazz Trumpet and Ukulele	M5	2	2	2	Er-hu
E6	1	2	2	Harmonica	M6	2	2	2	Accordion
E7	2	2	2	Trumpet	M7	2	2	2	Flute
E8	2	2	2	Saxophone	M8	2	2	2	Traditional bamboo flute
E9	2	2	2	Trombone, Tuba	M9	1			
E10	3	2	2	Trumpet	M10	1	2		
E11	2	2	2	Piano	M11	1	2		
E12	2	2	2	Guitar	M12	2	2	3	Drum
E13	2	2	2	Saxophone, Piano	M13	2	2	2	Piano
E14	2	2	2	Recorder, Guitar	M14	2	2	2	Guitar
E15	2	2	2	Guitar	M15	2	2	2	Guitar
E16	2	2	2	Guitar	M16	1	2	4	
E17	2	2	4	Guitar	M17	2	2	2	Violin
E18	2	3	2	Drum, trumpet	M18	1	2	4	
E19	2	2	3	Drum	M19	2	2	3	Guitar, Piano, Singing lessons (ten times)
E20	2	2	2	Piano, baritone	M20	2	2	2	Piano
E21	4	2	3	Guitar	M21	3	2	4	Piano
E22	3	2	2	Trumpet	M22	1	2	4	Guitar
E23	3	2	2	Trumpet, guitar	M23	2	2	2	Flute
E24	1	2	4		M24	1	2	4	

APPENDIX E

EXAMPLE SAS CODE

Example SAS code for GMP error per syllable variable:

```
/*Unconditional model*/
proc mixed noclprint covtest;
class subjectid group;
model GMP_syl = /solution;
random intercept /subject=subjectid(group);
run;

/*proc mixed*/
proc mixed data=dissertation1 noclprint covtest;
class subjectid group slc;
model GMP_syl=group|slc/solution ddfm=bw;
random intercept slc/subject=subjectid(group);
run;

/*proc glimmix*/
proc glimmix data=dissertation1 plots=(studentpanel boxplot(fixed student));
class subjectid group slc;
model GMP_syl= group|slc/solution dist=gamma;
random intercept slc/subject=subjectid(group);
run;
```

APPENDIX F

ADDITIONAL REACTION TIME INTER-SYLLABLE INTERVAL RESULTS

Table 18 Mean and SD of RT and ISI across Groups and Sequence Length Conditions (Unit: seconds)

			RT	ISI1	ISI2	ISI3	ISI4	ISI5	ISI6	ISI7	ISI8
3 syllable condition	native Mandarin	Mean	1.18	0.54	0.43						
		SD	0.26	0.36	0.22						
	non-Mandarin	Mean	1.81	1.05	1.05						
		SD	0.93	0.83	0.39						
			RT	ISI1	ISI2	ISI3	ISI4	ISI5	ISI6	ISI7	ISI8
6 syllable condition	native Mandarin	Mean	1.21	0.51	0.44	0.47	0.46	0.53			
		SD	0.35	0.25	0.21	0.24	0.19	0.24			
	non-Mandarin	Mean	1.94	1.03	1.15	0.86	1.34	1.14			
		SD	1.57	0.56	0.51	0.44	0.75	1.06			
			RT	ISI1	ISI2	ISI3	ISI4	ISI5	ISI6	ISI7	ISI8
9 syllable condition	native Mandarin	Mean	1.24	0.51	0.44	0.47	0.47	0.55	0.50	0.55	0.45
		SD	0.28	0.24	0.22	0.20	0.20	0.25	0.23	0.26	0.23
	non-Mandarin	Mean	1.77	1.05	1.11	0.85	1.38	1.14	1.06	1.11	1.09
		SD	0.85	0.68	0.49	0.42	0.93	0.58	0.67	0.60	0.42

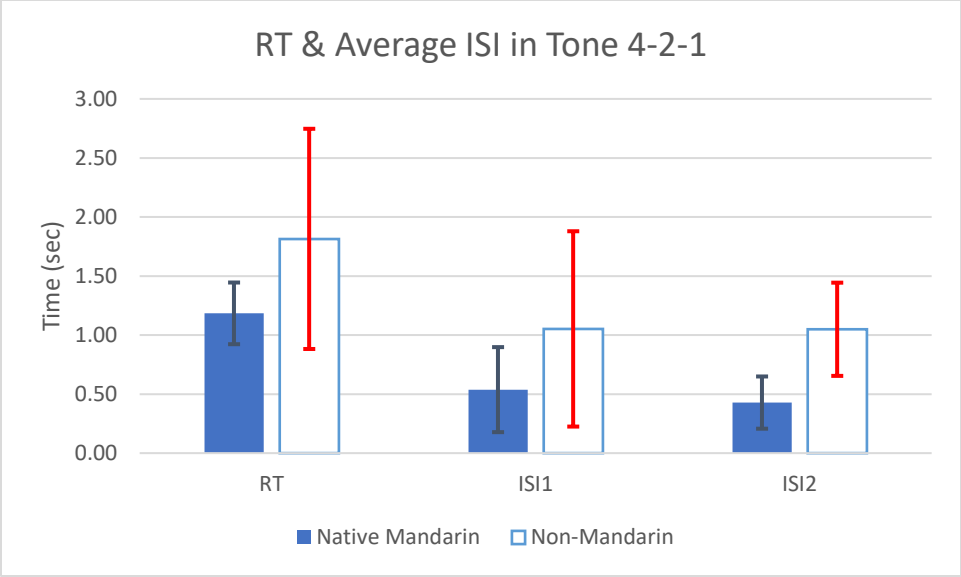

Figure 34 Group comparison of Mean and SD of RT and ISIs for three-syllable sequence length condition

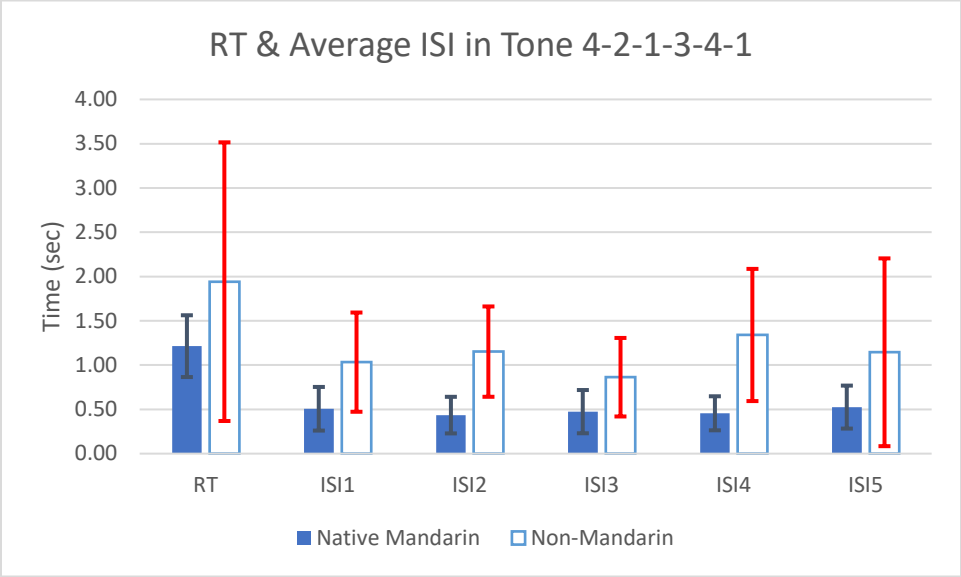

Figure 35 Group comparison of Mean and SD of RT and ISIs for six-syllable sequence length condition

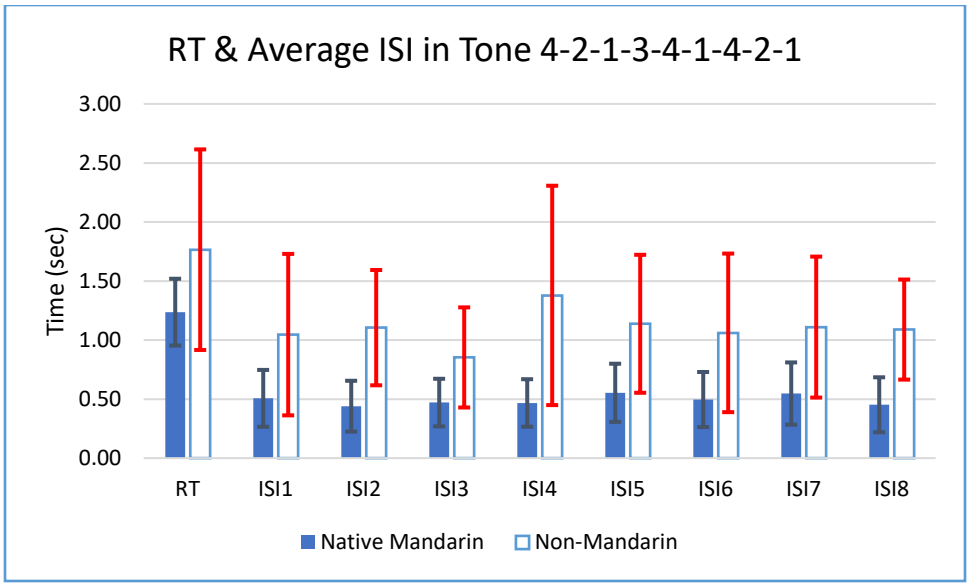

Figure 36 Group comparison of Mean and SD of RT and ISIs for nine-syllable sequence length condition

APPENDIX G

SUMMARY OF STATISTICAL RESULTS

Table 19 Summary of statistical results

	Effect	*F*	*p-value*	Post-hoc
Time Parameter Difference between 1st Syllable and the Whole Sequence (Square-root Transformed)	Interaction	*.76*	*.4725*	
	Group Main Effect	*10.89*	*.0022**	native Mandarin > non-Mandarin*
	Sequence Length Condition (SLC) Main Effect	*1.88*	*.1615*	

Table 19 Summary of statistical results (Continued)

Time Parameters in Nine-Syllable Sequence Length Condition (Log Transformed)	Interaction	*3.38*	*.0875*	
	Group Main Effect	*4.79*	*.0462**	native Mandarin > non-Mandarin*
	Whole_1st Main Effect	*.09*	*.7674*	
f_0 Magnitude Parameter Difference between 1st Syllable and the Whole Sequence (Square-root Transformed)	Interaction	*2.47*	*.0909*	
	Group Main Effect	*.15*	*.6969*	
	Sequence Length Condition (SLC) Main Effect	*3.14*	*.0489**	Marginal Comparison: SLC1=SLC2 SLC1=SLC3 SLC2=SLC3

Table 19 Summary of statistical results (Continued)

f_0 Magnitude Parameters in Nine-Syllable Sequence Length Condition	Interaction	*1.54*	*.2203*	
	Group Main Effect	*10.39*	*.0023**	native Mandarin > non-Mandarin*
	Whole_1st Main Effect	*.83*	*.3669*	

Table 19 Summary of statistical results (Continued)

GMP Errors per Syllable	Interaction	*9.31*	*.0002**	
	Group Main Effect	*48.10*	*<.0001**	Simple Main Effects: SLC1: native Mandarin < non-Mandarin* SLC2: native Mandarin < non-Mandarin* SLC3: native Mandarin < non-Mandarin*
	Sequence Length Condition Main Effect	*93.62*	*<.0001**	Simple Main Effects: native Mandarin: SLC1 < SLC2* SLC1 < SLC3* SLC2 = SLC3 non-Mandarin: SLC1 < SLC2* SLC1 < SLC3* SLC2 = SLC3

Table 19 Summary of statistical results (Continued)

GMP Errors per Syllable for Tone 2 in Nine-Syllable Sequence Length Condition	Interaction	*.54*	*.4643*	
	Group Main Effect	*50.66*	*<.0001**	native Mandarin < non-Mandarin*
	2^{nd}_8^{th} Syllable Position Main Effect	*.02*	*.8806*	

Table 19 Summary of statistical results (Continued)

Hamming Distance Difference per Syllable between Slope and Parsons' code Measurements	Interaction	*10.94*	*<.0001**	
	Group Main Effect	*6.51*	*.0141**	Main Effect: native Mandarin > non-Mandarin* Simple Main Effects: SLC1: native Mandarin > non-Mandarin* SLC2: native Mandarin = non-Mandarin SLC3: native Mandarin = non-Mandarin
	Sequence Length Condition Main Effect	*20.33*	*<.0001**	Simple Main Effects: native Mandarin: SLC1 = SLC2 SLC1 = SLC3 SLC2 = SLC3 non-Mandarin: SLC1 < SLC2* SLC1 < SLC3* SLC2 = SLC3

Table 19 Summary of statistical results (Continued)

Hamming Distance per Syllable for Slope	Interaction	*.45*	*.6412*	
	Group Main Effect	*7.29*	*.0097**	native Mandarin < non-Mandarin*
	Sequence Length Condition Main Effect	*141.55*	*<.0001**	SLC1 < SLC2* SLC1 < SLC3* SLC2 = SLC3

Table 19 Summary of statistical results (Continued)

Hamming Distance per Syllable for Parsons' Code	Interaction	*14.48*	*<.0001**	
	Group Main Effect	*24.48*	*<.0001**	Main Effect: native Mandarin < non-Mandarin* Simple Main Effects: SLC1: native Mandarin < non-Mandarin* SLC2: native Mandarin < non-Mandarin* SLC3: native Mandarin < non-Mandarin*
	Sequence Length Condition Main Effect	*120.35*	*<.0001**	Simple Main Effects: native Mandarin: SLC1 < SLC2* SLC1 < SLC3* SLC2 = SLC3 non-Mandarin: SLC1 < SLC2* SLC1 < SLC3* SLC2 = SLC3

Table 19 Summary of statistical results (Continued)

Reaction Time (RT) (Square-root Transformed)	Interaction	*1.22*	*.2959*	
	Group Main Effect	*27.87*	*<.0001**	native Mandarin < non-Mandarin*
	Sequence Length Condition Main Effect	*1.12*	*.3259*	
Inter-Syllable Interval (ISI)	Interaction	*1.53*	*.2225*	
	Group Main Effect	*46.38*	*<.0001**	native Mandarin < non-Mandarin*
	Sequence Length Condition Main Effect	*3.64*	*.0301**	Marginal Effects: SLC1 = SLC2 SLC1 = SLC3 SLC2 = SLC3
Ratio of RT over average ISI	Interaction	*.38*	*.6854*	
	Group Main Effect	*18.41*	*<.0001**	native Mandarin > non-Mandarin*
	Sequence Length Condition Main Effect	*4.29*	*<.0001**	Marginal Effects: SLC1 = SLC2 SLC1 > SLC3* SLC2 = SLC3

BIBLIOGRAPHY

Abbs, J. H. (1986). *Invariance and variability in speech processes: A distinction between linguistic intent and its neuromotor implementation.* (Vol. xxiii, 604). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology, 51*, 705-723.

Abry, C., Orliaguet, J.-P., & Sock, R. (1990). Patterns of speech phasing. Their robustness in the production of a timed linguistic task: single vs. double (abutted) consonants in French. *Cahiers de psychologie cognitive, 10*(3), 269-288.

Adams, S., & Page, A. (2000). Effects of selected practice and feedback variables on speech motor learning. *Journal of Medical Speech-Language Pathology, 8*(4), 215-220.

Almelaifi, R. (2013). *The role of "focus of attention" on the learning of non-native speech sounds: English speakers learning of mandarin Chinese tones.* University of Pittsburgh,

Awan, S. N., Roy, N., Zhang, D., & Cohen, S. M. (2016). Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: development of clinical cutoff scores. *Journal of Voice, 30*(2), 130-144.

Bailly, G. (1998). Cortical dynamics and biomechanics. *Les Cahiers de l'ICP. Bulletin de la communication parlée*(4), 35-44.

Ballard, K. J., Barlow, J. A., & Robin, D. A. (2001). The underlying nature of apraxia of speech: A critical evaluation of Varley and Whiteside's dual route speech encoding hypothesis. *Aphasiology, 15*(1), 50-58.

Ballard, K. J., Maas, E., & Robin, D. A. (2007). Treating control of voicing in apraxia of speech with variable practice. *Aphasiology, 21*(12), 1195-1217.

Ballard, K. J., Savage, S., Leyton, C. E., Vogel, A. P., Hornberger, M., & Hodges, J. R. (2014). Logopenic and nonfluent variants of primary progressive aphasia are differentiated by acoustic measures of speech production. *PloS one, 9*(2), e89864.

Barlow, S., & Bradford, P. (1992). Measurement and implications of orofacial muscle performance in speech disorders. *Journal of Human Muscle Performance, 1*(4), 1-31.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook, 3*, 255-309.

Bent, T. (2005). Perception and production of non-native prosodic categories. *Unpublished Ph. D. thesis, Department of Linguistics, Northwestern University, Evanston, IL*.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America, 66*(4), 1001-1017.

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. *Journal of Cognitive Neuroscience, 22*(7), 1504-1529.

Borden, G. J. (1979). An interpretation of research on feedback interruption in speech. *Brain and language, 7*(3), 307-319.

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

Browman, C. P., & Goldstein, L. M. (1988). Some notes on syllable structure articulatory phonology. *Haskins Laboratories Status Report on Speech Research, SR-93/94*, 85-102.

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

Browman, C. P., & Goldstein, L. M. (1992). Articulatory phonology: an overview. *Phonetica, 49*(3-4), 155-180.

Cai, S., Ghosh, S. S., Guenther, F., & Perkell, J. S. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *Journal of Acoustic Society of America, 128*(4), 2033-2048.

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *The Journal of Neuroscience, 31*(45), 16483-16490.

Caramazza, A. (1991). Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. In *Issues in reading, writing and speaking* (pp. 15-44): Springer.

Carter, M. C., & Shapiro, D. C. (1984). Control of sequential movements: evidence for generalized motor programs. *Journal of Neurophysiology, , 52*(5), 787-796.

Chamberlin, C. J., & Magill, R. A. (1992). A note on schema and exemplar approaches to motor skill representation in memory. *Journal of motor behavior, 24*(2), 221-224.

Chan, Y. Y. F. (1974). A perceptual study of tones in Cantonese. *Hong Kong: Centre of Asian Studies, University of Hong Kong*.

Chao, Y. R. (1965). *A grammar of spoken Chinese*: Univ of California Press.

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

Chen, S.-H. K., McNeil, M. R., Hill, K., & Pratt, S. R. (2013). Translating and validating a Mandarin Chinese version of the computerized revised token test. *Speech, Language and Hearing, 16*(1), 37-45.

Chen, S. H., Liu, H., & Xu, Y. (2007). Voice F_0 responses to pitch-shifted voice feedback on voice F_0 contours in syllables. *Journal of Acoustic Society of America, 111*, 357-366.

Cholin, J., Levelt, W. J., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99*(2), 205-235.

Cohen, J. (1988). *Statistical power analysis: A computer program*: Routledge.

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., . . . De la Torre, F. (2009). *Detecting depression from facial actions and vocal prosody.* Paper presented at the Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and speech, 45*(3), 207-228.

Crompton, A. (1982). Syllables and segments in speech production. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 109-162). Berlin: Mouton.

Croot, K. (2001). Integrating the investigation of apraxic, aphasic and articulatory disorders in speech production: A move towards sound theory. *Aphasiology, 15*(1), 58-62.

Crystal, D. (1979). Prosodic development. In P. Fletcher & M. Garman (Eds.), *Language Acquisition*: Cambridge University Press.

Cummins, F. (1998). Limit cycle dynamics in prosody. *The Journal of the Acoustical Society of America, 103*(5), 2852-2852.

Damian, M. F. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 416-431.

Das, P., & McCollum, G. (1988). Invariant structure in locomotion. *Neuroscience, 25*(3), 1023-1034.

Davison, D. S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics, 78*, 50-57.

De Jong, K. J. (2001). Effects of syllable affiliation and consonant voicing on temporal adjustment in a repetitive speech-production task. *Journal of speech, language, and hearing research, 44*(4), 826-840.

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

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in cognitive sciences, 4*(11), 423-431.

Diedrichsen, J., & Kornysheva, K. (2015). Motor skill learning between selection and execution. *Trends in cognitive sciences, 19*(4), 227-233.

Dietrich, M., & Abbott, K. V. (2012). Vocal function in introverts and extraverts during a psychological stress reactivity protocol. *Journal of Speech, Language, and Hearing Research, 55*(3), 973-987.

Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages.* (Doctor of Philosophy), Massachusetts Institute of Technology,

Ellis, A. W., & Young, A. W. (2013). *Human cognitive neuropsychology: A textbook with readings*: Psychology Press.

Feng, Y. (2008). *Dissociating the role of auditory and somatosensory feedback in speech production: Sensorimotor adaptation to formant shifts and articulatory perturbations*: ProQuest.

Fennell, A. M., & Weismer, G. (1984). When (if ever) does speaking rate become an essential variable in speech production. *The Journal of the Acoustical Society of America, 76*(S1), S16-S16.

Fujimura, O. (1987). A linear model of speech timing. *R. Channon & L. Shockey*, 109-123.

Gandour, J. (1979). Perceptual dimensions of Cantonese tones: a multidimensional scaling reanalysis of Fok's tone confusion data. *Southeast Asian Linguistic Studies, 4*, 415-429.

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

Garcia-Colera, A., & Semjen, A. (1987). The organization of rapid movement sequences as a function of sequence length. *Acta Psychologica, 66*(3), 237-250.

Garcia-Colera, A., & Semjen, A. (1988). Distributed planning of movement sequences. *Journal of motor behavior, 20*(3), 341-367.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *The journal of the Acoustical society of America, 63*(1), 223-230.

Gentner, D. R. (1987). Timing of skilled motor performance: tests of the proportional duration model. *Psychological Review, 94*, 255-276.

Ghez, C., Gordon, J., & Ghilardi, M. F. (1995). Impairments of reaching movements in patients without proprioception. II. Effects of visual information on accuracy. *Journal of Neurophysiology, 73*(1), 361-372.

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage, 55*(3), 1324-1338.

Gracco, V. L. (1988). Timing factors in the coordination of speech movements. *The Journal of Neuroscience, 8*(12), 4628-4639.

Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65*, 156-166.

Gracco, V. L., & Abbs, J. H. (1988). Central patterning of speech movements. *Experimental Brain Research, 71*, 515-526.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102*(3), 594-621.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39*, 350-365.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and language, 96*(3), 280-301.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: Cambridge University Press.

Guthrie, B. L., Porter, J. D., & Sparks, D. L. (1983). Corollary discharge provides accurate eye position information to the oculomotor system. *Science, 221*(4616), 1193-1195.

Hankamer, J. (1989). Morphological parsing and the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process.* Cambridge, MA: MIT Press.

Hao, Y.-C. (2018). Contextual effect in second language perception and production of Mandarin tones. *Speech Communication*.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology.* (pp. 201-260). New York: Academic Press.

Heuer, H. (1988). Testing the invariance of relative timing: Comment on Gentner (1987).

Heuer, H., & Schmidt, R. A. (1988). Transfer of learning among motor patterns with different relative timing. *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 241.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135-145.

Higgins, J. R., & Spaeth, R. K. (1972). Relationship between consistency of movement and environmental condition. *Quest, 17*(1), 61-69.

Hu, X., Murray, W. M., & Perreault, E. (2010). *Biomechanical constraints on the control of endpoint stiffness.* . Paper presented at the 34th Annual Meeting of American Society of Biomechanics, Providence, RI.

Hughes, O. M., & Abbs, J. H. (1976). Labial-mandibular coordination in the production of speech: Implications for the operation of motor equivalence. *Phonetica, 33*(3), 199-221.

Hula, S. N. A., Robin, D. A., Maas, E., Ballard, K. J., & Schmidt, R. A. (2008). Effects of feedback frequency and timing on acquisition, retention and transfer of speech skills in acquired apraxia of speech. *Journal of Speech, Language and Hearing Research, 51*, 1088-1113.

Hula, W. D., & McNeil, M. R. (2008). Models of attention and dual-task performance as explanatory constructs in aphasia. In *Seminars in speech and language* (Vol. 29, pp. 169-187): Thieme Medical Publishers.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*, 67-103.

Ito, T., Kimura, T., & Gomi, H. (2005). The motor cortex is involved in reflexive compensatory adjustment of speech articulation. *NeuroReport, 16*(16), 1791-1794.

Jacquemot, C., Dupoux, E., & Bachoud-Lévi, A.-C. (2007). Breaking the mirror: Asymmetrical disconnection between the phonological input and output codes. *Cognitive Neuropsychology, 24*(1), 3-22.

Jun, S.-A. (2005). Prosody in sentence processing: Korean vs. English. *UCLA Working Papers in Phonetics, 104*, 26-45.

Kawahara, H. (1993). Transformed auditory feedback: Effects of fundamental frequency perturbation. *Journal of the Acoustical Society of America, 94*, 1883-1884.

Keele, S. W. (1968). Movement control in skilled motor performance. *Psychological Bulletin, 70*(6 (1)), 387-403.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective on speech production- data and theory. *Journal of Phonetics, 14*, 29-59.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11*, 201-258.

Kent, R. D. (2015). Nonspeech Oral Movements and Oral Motor Disorders: A Narrative Review. *American journal of speech-language pathology, 24*(4), 763-789.

Kim, H. S., Shaiman, S., & McNeil, M. R. (n.d.). Does Learned Motor Control for Lexical Stress Generalize Across Different Consonantal Contexts?

Klapp, S. T. (1995). Motor response programming during simple and choice reaction time: The role of practice. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 1015-1022.

Klapp, S. T. (2003). Reaction time analysis of two types of motor preparation for speech articulation: action as a sequence of chunks. *Journal of motor behavior, 35*(2), 135-150.

Klapp, S. T., Abbott, J., Coffman, K., Greim, D., Snider, R., & Young, F. (1979). Simple and choice reaction time method in the study of motor programming. *Journal of motor behavior, 11*(2), 91-101.

Knock, T., Ballard, K. J., Robin, D. A., & Schmidt, R. A. (2000). Influence of order of stimulus presentation on speech motor learning: A principled approach to treatment for apraxia of speech. *Aphasiology, 14*, 653-668.

Kozhevnikov, V. A., & Chistovich, L. A. (1965). Speech: Articulation and perception.

Kugler, P., Kelso, J. S., & Turvey, M. (1982). On the control and coordination of naturally developing systems. *The development of movement control and coordination, 5*, 78.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. *The development of movement control and coordination, 5*, 78.

Lai, Q., & Shea, C. H. (1998). Generalized motor program (GMP) learning: effects of reduced frequency of knowledge of results and practice variability. *Journal of motor behavior, 30*(1), 51-59.

Lai, Q., Shea, C. H., Wulf, G., & Wright, D. L. (2000). Optimizing generalized motor program and parameter learning. *Research Quarterly for Exercise and Sport, 71*(1), 10-24.

Lee, J., Yoshida, M., & Thompson, C. K. (2015). Grammatical planning units during real-time sentence production in speakers with agrammatic aphasia and healthy speakers. *Journal of Speech, Language, and Hearing Research, 58*(4), 1182-1194.

Legát, M., & Matoušek, J. (2010). *Collection and analysis of data for evaluation of concatenation cost functions.* Paper presented at the International Conference on Text, Speech and Dialogue.

Lehiste, I., & Lass, N. J. (1976). Suprasegmental features of speech. *Contemporary issues in experimental phonetics*, 225-239.

Leinen, P., Green, M. F., Esat, T., Wagner, C., Tautz, F. S., & Temirov, R. (2015). Virtual reality visual feedback for hand-controlled scanning probe microscopy manipulation of single molecules. *Beilstein journal of nanotechnology, 6*(1), 2148-2153.

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

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

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

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

Levitt, H. (1970). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America, 49*(2B), 467-477.

Liberman, A. M. (1992). The relation of speech to reading and writing. In *Advances in psychology* (Vol. 94, pp. 167-178): Elsevier.

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

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243*(4890), 489-494.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439): Springer.

Lisker, L., & Abramson, A. S. (1970). *The voicing dimension: some experiments in comparative phonetics.* Paper presented at the The 6th International Congress of Phonetic Sciences (1967), Prague.

Liu, F. (1924). Szu Sheng Shih Yen Lu. *Ch'un Yi, Shanghai*.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*(2-4), 70-87. doi:10.1159/000090090

Löfqvist, A. (1991). Proportional timing in speech motor control. *Journal of phonetics, 19*(3-4), 343-350.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 1*(3), 86.

Maas, E. (2006). *the nature and time course of motor programming in apraxia of speech.* (Ph. D.), University of California, San Diego, San Diego. (b6635519)

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17*, 277-298.

Maas, E., Robin, D. A., Hula, S. N. A., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American journal of speech-language pathology, 17*(3), 277-298.

Maas, E., Robin, D. A., Wright, D. L., & Ballard, K. J. (2008). Motor porgramming in apraxia of speech. *Brain and Language, 106,* 107-118.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological REview, 77*(3), 182-196.

Magri, C. J., Ferry, P., & Abela, S. (2007). A review of the aetiology and management of vocal behaviour in dementia. *Malta Medical Journal, 19*(3), 30-35.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940-967.

Martin, F. G. (1988). Drugs and vocal function. *Journal of Voice, 2*(4), 338-344.

Maslovat, D., Klapp, S. T., Jagacinski, R. J., & Franks, I. M. (2014). Control of response timing occurs during the simple reaction time interval but on-line for choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance, 40*(5), 2005.

Max, L., & Caruso, A. J. (1997). Acoustic Measures of Temporal Intervals Across Speaking RatesVariability of Syllable-and Phrase-Level Relative Timing. *Journal of Speech, Language, and Hearing Research, 40*(5), 1097-1100.

McCann, R. S., & Johnston, J. L. (1992). Locus of single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 471-484.

McNeil, M. R., Pratt, S. R., & Fossett, T. R. (2004). The differential diagnosis of apraxia of speech. *Speech motor control in normal and disordered speech*, 389-413.

McNeil, M. R., Pratt, S. R., Szuminsky, N., Sung, J. E., Fossett, T. R., Fassbinder, W., & Lim, K. Y. (2015). Reliability and Validity of the Computerized Revised Token Test: Comparison of Reading and Listening Versions in Persons With and Without Aphasia. *Journal of Speech, Language, and Hearing Research, 58*(2), 311-324.

McNeil, M. R., Sung, J. E., Yang, D., Pratt, S. R., Fossett, T. R. D., Doyle, P., J., & Pavelko, S. (2007). Comparing connected language elicitation procedures in persons with aphasia: Concurrent valiation of the story retell procedure. *Aphasiology, 21*(6), 775-790.

Meyer, A. S. (1990). The time course of phonological encoding in language production:The encoding of successive syllables of a word. *Journal of Memory and Language, 29,* 525-545.

Miall, R., Weir, D., Wolpert, D. M., & Stein, J. (1993). Is the cerebellum a smith predictor? *Journal of motor behavior, 25*(3), 203-216.

Miller, N. (2001). Dual or duel route? *Aphasiology, 15*(1), 62-68.

Moore, B. C. (2012). *An introduction to the psychology of hearing*: Brill.

Moser, D., Fridriksson, J., Bonilha, L., Healy, E. W., Baylis, G., Baker, J. M., & Rorden, C. (2009). Neural recruitment for the production of native and novel speech sounds. *Neuroimage, 46*(2), 549-557.

Müller, E., & MacLeod, G. (1982). Perioral biomechanics and its relation to labial motor control. *The Journal of the Acoustical Society of America, 71*(S1), S33-S33.

Müller, E. M., Abbs, J. H., Kennedy, J. G., & Larson, C. R. (1977). *Significance of biomechanical variables in lip movements for speech.* Paper presented at the American Speech and Hearing Association National Convention, Chicago.

Munhall, K. G. (1985). An examination of intra-articulator relative timing. *The Journal of the Acoustical Society of America, 78*(5), 1548-1553.

Nespor, M., & Vogel, I. B. (1986). *Prosodic phonology*. Dordrecht: Foris.

Ning, L. H., Loucks, T. M., & Shih, C. (2015). The effects of language learning and vocal training on sensorimotor control of lexical tone. *Journal of Phonetics*.

Ning, L. H., Shih, C., & Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. *Speech Communication, 63*, 55-69.

Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behavior*. Retrieved from

Park, J.-H., & Shea, C. H. (2003). Effect of practice on effector independence. *Journal of motor behavior, 35*(1), 33-40.

Park, J. H., & Shea, C. H. (2002). Effector independence. *Journal of motor behavior, 34*, 253-270.

Park, J. H., & Shea, C. H. (2003). The effects of practice on effector transfer. *Journal of Motor Behavior, 35*, 33-40.

Park, J. H., & Shea, C. H. (2005). Sequence learning: Response structure and effector transfer. *The Quarterly Journal of Experimental Psychology, Section A, 58*(3), 387-419.

Parkinson, A. L., Korzyukov, O., Larson, C. R., Litvak, V., & Robin, D. A. (2013). Modulation of effective connectivity during vocalization with perturbed auditory feedback. *Neuropsychologia, 51*, 1471-1480.

Parsons, D. (1975). *The directory of tunes and musical themes*: Cambridge, Eng.: S. Brown.

Pashler, H. (1994). Overlapping mental operations in serial performance with preview. *The quarterly journal of experimental psychology. A, Human experimental psychology, 47*, 161–191.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., . . . Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28*(3), 233-272.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, perception, & psychophysics, 71*(6), 1385-1398.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(04), 329-347.

Pierrehumbert, J. B. (1980). *The phonetics and phonology of English intonation.* (Ph. D. thesis), MIT, Available from Published by Garland Press, New York (1990)

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese Tone Structure*. Cambridge, M. A.: MIT Press.

Pike, K. L. (1948). *Tone Languages: A Technique for Determining the Number and Type of Pitsch Contrasts in a Language, with Studies in Tonemic Substitution and Fusion*. Ann Arbor, M. I.: University of Michigan Press.

Prablanc, C., & Martin, O. (1992). Automatic control during hand reaching at undetected two-dimensional target displacements. *Journal of Neurophysiology, 67*(2), 455-469.

Pulvermüller, F. (1996). Hebb's concept of cell assemblies an the psychophysiology of word processing. *Psychophysiology, 33*(4), 317-333.

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *Journal of Acoustic Society of America, 120*, 966-977.

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *Journal of Acoustic Society of America, 119*, 2288-2297.

Ramadoss, D. (2012). *The phonology and phonetics of tone perception*: THE JOHNS HOPKINS UNIVERSITY.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460-499.

Raymer, A. M., & Thompson, C. K. (1991). Effects of verbal plus gestural treatment in a patient with aphasia and severe apraxia of speech *Clinical Aphasiology*, 285-298.

Reilly, K. J., & Spencer, K. A. (2013). Speech serial control in healthy speakers and speakers with hypokinetic or ataxic dysarthria: Effects of sequence length and practice. *Front Hum Neurosci*.

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and cognitive processes, 14*(2), 173-200.

Rogers, M., & Spencer, K. (2001). Spoken word production without assembly: Is it possible? *Aphasiology, 15*(1), 68-74.

Rogers, M. A., & Storkel, H. L. (1998). Reprogramming phonologically similar utterances: The role of phonetic features in pre-motor encoding. *Journal of Speech, Language, and Hearing Research, 41*(2), 258-274.

Rogers, M. A., & Storkel, H. L. (1999). Planning speech one syllable at a time: The reduced buffer capacity hypothesis in apraxia of speech. *Aphasiology, 13*, 793-805.

Roth, K. (1988). Investigations on the basis of generalized motor program hypothesis. In O. G. Meijer & K. Roth (Eds.), *Complex movement behavior: The motor-action controversy* (pp. 261-288). Amsterdam: North Holland.

Russell, D. G. (1976). Spatial location cues and movement production. In G. E. Stelmach (Ed.), *Motor control: Issues and trends.* New York: Academic Press.

Sainburg, R. L., Poizner, H., & Ghez, C. (1993). Loss of proprioception produces deficits in interjoint coordination. *Journal of neurophysiology, 70*(5), 2136-2147.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological bulletin, 95*(3), 355.

Sapir, S., & Aronson, A. E. (1985). Aphonia after closed head injury: aetiologic considerations. *International Journal of Language & Communication Disorders, 20*(3), 289-296.

Schmidt, R. A. (1975). A Schema Theory of Discrete Motor Skill Learning. *The American Psychological Association, 82*(4), 225-260.

Schmidt, R. A. (1982). The schema concept. *Human motor behavior: An introduction*, 219-235.

Schmidt, R. A. (2003). Motor schema theory after 27 years: Reflections and implications for a new theory. *Research Quarterly for Exercise and Sport, 74*(4), 366-375.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological science, 3*(4), 207-217.

Schmidt, R. A., & Lee, T. D. (1988). *Motor Control and Learning*. Champaign, IL: Human Kinetics.

Schmidt, R. A., & Lee, T. D. (1999). *Motor Control and Learning, A Behavioral Emphasis*. Champain, IL: Human Kinetics.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis* (4th ed. ed.). Champaign, IL: Human Kinetics.

Schmidt, R. A., & Russell, D. G. (1972). Movement velocity and movement time as determiners of the degree of preprogramming in simple movements. *Journal of Experimental Psychology, 96*, 315-320.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime (Version 2.0). *Computer software and manual]. Pittsburgh, PA: Psychology Software Tools Inc*.

Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of cognitive neuroscience*.

Selkirk, E. (1978). On prosodic structure and its relation to syntactic structure. *Linguistic Inquiry, 11*, 563-605.

Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience, 14*(5), 3208-3224.

Shaffer, L. H. (1980). 26 Analysing Piano Performance: A Study of Concert Pianists. *Advances in Psychology, 1*, 443-455.

Shaffer, L. H. (1984). Timing in solo and duet piano performances. *The Quarterly Journal of Experimental Psychology, 36*(4), 577-595.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw *Journal of Acoustic Society of America, 86*, 78-88.

Shaiman, S., & Gracco, V. L. (2002). Task-specific sensorimotor interactions in speech production. *Exp Brain Res, 146*, 411-418.

Shapiro, D. (1977). A preliminary attempt to determine the duration of a motor program. *Psychology of motor behavior and sport, 1*, 17-24.

Sharkey, S. G., & Folkins, J. W. (1985). Variability of lip and jaw movements in children and adults: implications for the development of speech motor control. *Journal of Speech and Hearing Research, 28*, 8-15.

Shea, C. H., Lai, Q., Wright, D. L., Immink, M., & Black, C. (2001). Consistent and variable practice conditions: Effects on relative and absolute timing. *Journal of motor behavior, 33*, 139-152.

Shea, C. H., & Wulf, G. (2005). Schema theory: A critical appraisal and reevaluation. *Journal of motor behavior, 37*(2), 85-102.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of Acoustic Society of America, 125*(2), 1103-1113.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of educational and behavioral statistics, 23*(4), 323-355.

Smith, A., Goffman, L., Zelaznik, H., Ying, G., & McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research, 104*(3), 493-501. doi:10.1007/BF00231983

Spencer, K. A., & Rogers, M. A. (2005). Speech motor programming in hypokinetic and ataxic dysarthria. *Brain and language, 94*(3), 347-366.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America, 76*(6), 1690-1695.

St Louis, K. O., & Ruscello, D. M. (1981). The Oral Speech Mechanism Screening Examination (OSMSE).

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor program representation* (pp. 117-152). New York: Academic.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66).

Strand, E. A. (1987). *Acoustic and response time measures in utterance production: A comparison of apraxic and normal speakers.* (Doctor of Philosophy), University of Wisconsin-Madison, Wisconsin-Madison.

Studdert-Kennedy, M. (1980). Speech perception. *Language and Speech, 23*(1), 45-66.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental brain research, 161*(1), 1-10.

Terzuolo, C. A., & Viviani, P. (1979). The central representation of learned motor patterns. *Posture and movement*, 113-121.

Theodoros, D. G., & Murdoch, B. E. (1994). Laryngeal dysfunction in dysarthric speakers following severe closed-head injury. *Brain injury, 8*(8), 667-684.

Topbas, O. (2010). *Effects of diadochokinetic rate on vocal fundamental frequency and intensity in normally speaking young adults*: West Virginia University.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*, 1429-1443.

Tseng, C.-y. (1981). *An acoustic phonetic study on tones in Mandarin Chinese*: Brown University.

Tuller, B., Harris, K. S., & Kelso, J. A. S. (1981). Phase relationships among articulator muscles as a function of speaking rate and syllable stress. *The Journal of the Acoustical Society of America, 69*(S1), S55-S55.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance, 8*(3), 460.

Turvey, M. (1977). Preliminaries to a theory of action with reference to vision. *Perceiving, acting and knowing*, 211-265.

van der Merwe, A. (1997). A theoretical framework for the characterization of pathological speech sensorimotor control. In M. R. McNeil (Ed.), *Clinical management of sensorimotor speech disorders*. New York: Thieme Medical Publishers.

van der Merwe, A. (2009). A theoretical framework for the characterization of pathological speech sensorimotor control. In M. R. McNeil (Ed.), *Clinical Management of Sensorimotor Speech Disorders.* . New York: Thieme Medical Publishers.

van der Merwe, A., & Steyn, M. (2018). Model-Driven Treatment of Childhood Apraxia of Speech: Positive Effects of the Speech Motor Learning Approach. *American journal of speech-language pathology, 27*(1), 37-51.

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

Van Wijk, C., & Kempen, G. (1985). From sentence structure to intonation contour; An algorithm for computing intonation. *Sprachsynthese; zur Synthese von natürlich gesprochener Sprache aus Texten und Konzepten, 0*, 158-180.

Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive psychology, 19*, 403-440.

Varley, R. A., & Whiteside, S. P. (2001a). Exploring the enigma. *Aphasiology, 15*(1), 78-84.

Varley, R. A., & Whiteside, S. P. (2001b). What is the underlying impairment in acquired apraxia of speech. *Aphasiology, 15*(1), 39-49.

Varley, R. A., Whiteside, S. P., Hammill, C., & Cooper, K. (2006). Phases in speech encoding and foreign accent syndrome. *Journal of Neurolinguistics, 19*(5), 356-369.

Varley, R. A., Whiteside, S. P., & Luff, H. (1999a). Apraxia of speech as a disruption of word-level schemata: Some durational evidence. *Journal of Medical Speech-Language Pathology, 7*(2), 127-132.

Varley, R. A., Whiteside, S. P., & Luff, H. (1999b). Dual-route speech encoding in normal and apraxic speakers: Some durational evidence. *Journal of Medical Speech and Language Pathology, 7*, 127-132.

Varley, R. A., Whiteside, S. P., Windsor, F., & Fisher, H. (2006). Moving up from the segment: A comment on Aichert and Ziegler's syllable frequency and syllable structure in apraxia of speech. *Brain and language, 96*, 235-239.

Verwey, W. B. (1995). A forthcoming key press can be selected while earlier ones are executed. *Journal of motor behavior, 27*(3), 275-284.

Verwey, W. B. (1996). Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 544-562.

Verwey, W. B. (1999). Evidence for a multistage model of practice in a sequential movement task. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1693-1708.

Verwey, W. B., & Dronkert, Y. (1996). Practicing a structured continuous key-pressing task: Motor chunking or rhythm consolidation? *Journal of motor behavior, 28*(1), 71-79.

Verwey, W. B., Groen, E. C., & Wright, D. L. (2015). The stuff that motor chunks are made of: Spatial instead of motor representations? *Experimental brain research*, 1-14.

Verwey, W. B., Shea, C. H., & Wright, D. L. (2015). A cognitive framework for explaining serial processing and sequence execution strategies. *Psychonomic bulletin & review, 22*(1), 54-77.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of Acoustical Society of America, 122*(4), 2306-2319.

Vu, V. H., Isableu, B., & Berret, B. (2016). On the nature of motor planning variables during arm pointing movement: Compositeness and speed dependence. *Neuroscience, 328*, 127-146.

Walker, G. M., & Hickok, G. (2015). Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM). *Psychonomic bulletin & review*, 1-14.

Wambaugh, J. L., Martinez, A. L., McNeil, M. R., & Rogers, M. A. (1999). Sound production treatment for apraxia of speech: Overgeneralization and maintenance effects. *Aphasiology, 13*(9-11), 821-837.

Weismer, G., & Fennell, A. M. (1985). Constancy of (acoustic) relative timing measures in phrase-level utterances. *The Journal of the Acoustical Society of America, 78*(1), 49-57.

Westbury, J., & Dembowski, J. (1993). Articulatory kinematics of normal diadochokinetic performance. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 27*, 13-36.

Whalen, D. H. (1990). Coarticulation is largely planned. *Haskins Laboratories Status Report on Speech Research, SR-101/102*, 149-176.

Whiteside, S. P., & Varley, R. A. (1998). A reconceptualisation of apraxia of speech: A synthesis of evidence. *Cortex, 34*(2), 221-231. doi:http://dx.doi.org/10.1016/S0010-9452(08)70749-4

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *nature neuroscience, 3*, 1212-1217.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science-AAAS-Weekly Paper Edition, 269*(5232), 1880-1882.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in cognitive sciences, 2*(9), 338-347.

Wright, D. L., Robin, D. A., Rhee, J., Vaculin, A., Jacks, A., Guenther, F. H., & Fox, P. (2009). Using the self-select paradigm to delineate the nature of speech motor programming. *Journal of Speech, Language, and Hearing Research, 52*, 755-765.

Wulf, G., & Schmidt, R. A. (1989). The learning of generalized motor programs: Reducing the relative frequency of knowledge of results enhances memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 748-757.

Wulf, G., & Schmidt, R. A. (1997). Variability of practice and implicit motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 987.

Wulf, G., Schmidt, R. A., & Deubel, H. (1993). Reduced feedback frequency enhances generalized motor program learning but not parameterization learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1134-1150.

Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review, 9*(2), 185-211.

Xu, Y. (1997). Contextual tonal variations in mandarin. *Journal of Phonetics, 25*, 61-83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f_0 contours. *Journal of Phonetics, 27*, 55-105.

Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *Journal of Acoustic Society of America, 116*(2), 1168-1178.

Young, D. E., & Schmidt, R. A. (1990). Units of motor behavior: Modifications with practice and feedback.

Young, D. E., & Schmidt, R. A. (1991). *Motor programs as units of movement control.* Paper presented at the Making them move.

Ziegler, W. (2013). The rhythmic organization of speech gestures and the sense of it. *Language, Cognition, and Neuroscience*, 1-3.

Ziegler, W., & Ackermann, H. (2013). Neuromotor speech impairment: It's All in the Talking. *Folia Phoniatrica et Logopaedica, 65*, 55-67.

Ziegler, W., Ackermann, H., & Kappes, J. (2011). From Phonology to Articulation: A Neurophonetic View. In *The handbook of psycholinguistic and cognitive processes: Perspectives in communication disorders* (pp. 329-346).

Zraick, R. I., Smith-Olinde, L., & Shotts, L. L. (2012). Adult normative data for the KayPENTAX phonatory aerodynamic system model 6600. *Journal of Voice, 26*(2), 164-176.