AN INVESTIGATION OF VOICE ONSET TIME AND THE FACTORS THAT AFFECT IT IN L1 AND L2 MANDARIN

by

Chiu-ching Tseng
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Linguistics

Committee:

______________________________________ Director

____

____

______________________________________ Department Chairperson

______________________________________ Program Director

______________________________________ Dean, College of Humanities and Social Sciences

Date: _________________________________ Spring Semester 2021
George Mason University
Fairfax, VA

An Investigation of Voice Onset Time and the Factors that Affect It in L1 and L2 Mandarin

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Chiu-ching Tseng
Master of Arts
Florida International University, 2015
Master of Science
Florida International University, 2014
Bachelor of Arts
City University of New York - Queens College, 2012
Associate in Science
City University of New York - LaGuardia Community College, 2000

Director: Steven Weinberger Associate Professor
Department of English, Linguistics Program

Spring Semester 2021
George Mason University
Fairfax, VA

DEDICATION

This dissertation is dedicated to my wife Rina, my mother, Chia-jung, father Wen-tai, my older brother and sister, Chiu-bin and Chiu-hua, and my forever-canine buddy, Woody.

ACKNOWLEDGEMENTS

I must take this opportunity to thank many people who have, in many ways, helped me with this dissertation along the way; without them, I could have never accomplished my academic goals this far.

I am deeply indebted to my dissertation chair and advisor, Dr. Steven Weinberger, who introduced me to the field of Foreign Accents and Phonology. He has advised, urged on, and guided me every step of the way for my career. I can never thank him enough for his time and effort in helping me grow as a scholar. He is a great role model for me as a scholar, mentor, and university professor. This dissertation would not have been possible without Dr. Weinberger. Thank you so much, Professor!

I must thank Dr. Harim Kwon, who showed me the right path to the field of Phonetics and speech perception. Her expertise, knowledge, and professionalism have been the most excellent resource for my achievement. Particularly, without her support and guidance on acoustic analysis and statistical reasoning, this dissertation would not have completed.

I am extremely grateful to Dr. Douglas Wulf for taking me as his teaching assistant for three years. His guidance, advice, and kindness since day one of my graduate training have been priceless, and learning under his supervision has been my most valuable experience. He is the greatest semanticist whom I have ever met. He is also an extraordinary applied linguist whom I want to grow to be. His writing expertise, insightful comments, and valuable feedback throughout the development of this dissertation have been my greatest learning pleasure.

I especially must thank Arai-sensei from Sophia University, Tokyo, and professors and colleagues from Fuguang University, Providence University, and Tunghai University in Taiwan, for using their acoustic labs and conducting my research there,

Most importantly, I owe a debt of gratitude to all 164 participants who kindly and willingly participated in this study. They underpinned this dissertation. Without them, this study could never be possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

AN INVESTIGATION OF VOICE ONSET TIME AND THE FACTORS THAT AFFECT IT IN L1 AND L2 MANDARIN

Chiu-ching Tseng, Ph.D.

George Mason University, 2021

Dissertation Director: Dr. Steven Weinberger

This dissertation investigates Voice Onset Time (VOT), which serves as an essential property for differentiating plosive consonants in L1 and L2 Mandarin Chinese. It surveys VOT variations and demonstrates that they are affected by several phonetic and phonological properties, e.g., lexical tone, place of articulation (POA), speech rate, phrase-position, pitch register, and gender. While researchers disagree on the relationship between lexical tone and VOT, this study investigates this relationship explicitly. Moreover, while it has been suggested that VOT varies because of different lexical tones, the question has not been comprehensively explored as to which of the tonal properties are responsible for this effect on VOT. This dissertation also tests whether non-native Mandarin speakers exhibit similar effects.

Four experiments were conducted—two for native Taiwan Mandarin speakers and a parallel set of two for L2 Mandarin speakers. Two experiments elicited stop-initial

words produced with one open-unrounded vowel, three articulation places, four lexical tones, three different speech rates, and three utterance positions. The other two experiments elicited the same set of stimuli and conditions in one tone with three pitch-levels and at a natural speech rate.

A series of linear mixed-effects regression models were employed to model the effects of the properties mentioned earlier on VOT duration. We wanted to know whether these factors affect VOT in native and in non-native speech.

Testing 164 participants (68 Taiwanese, 34 Spanish, 40 Japanese, and 22 English speakers of Mandarin), the results reveal that when other factors were kept constant, tone indeed influenced VOT, and the higher the onset tone pitch, the shorter the VOT. POA and the speech rate were also found to be highly significant factors. The results reveal that all non-native groups showed the same effects regardless of their L1. This finding suggests that the tone effect on VOT in Mandarin is a universal tendency due to the physiology of the vocal tract rather than due to language-specific phonology.

However, the Spanish and Japanese groups showed extended VOT values, which were not from their native VOTs, but their Mandarin VOTs were still significantly shorter than the native Mandarin speakers. Thus, the significant VOT differences between groups indicate some degree of L1 influence, which suggests that L2 VOT delay is probably mediated by language-specific phonological grammar.

This dissertation provides empirical evidence that an acoustic property, such as VOT, is not an isolated phenomenon but is involved with other complex phonological categories such as lexical tone. It also discusses how the effects operate within phonetic and phonological theories. Additionally, it compares Mandarin learners' L1 and L2 VOT directly. This cross-linguistic survey offers insight for L2 performance variations regarding phonetics, which may provide Mandarin instructors with multi-dimensional comparisons and confirmation of the interlanguage process as relevant to Second Language Acquisition. The observed phenomena may offer L2 classrooms insights into Mandarin accent variations for L2 English, Japanese, and Spanish learners.

CHAPTER 1. INTRODUCTION

1.1 Overview

Voice Onset Time (VOT) of prevocalic stops has been one of the essential topics in phonetics and phonology and has been extensively studied in many languages (e.g., Lisker & Abramson, 1964; Abramson & Lisker, 1972; Shimizu, 1990; Silva, 1992; Cho & Ladefoged, 1999; Lai et al., 2004; Pearce 2005; Liu et al., 2008; Narayan & Bowden, 2013). The consensus is that VOT is affected by place of articulation (POA) (e.g., Lisker & Abramson, 1964), consonantal voicing (e.g., Cho & Ladefoged, 1999), lexical tone (e.g., Liu et al., 2008), fundamental frequency (F0) (e.g., Lai et al., 2004), vowel context (e.g., Rochet and Fei, 1991; Chen et al., 2009), vowel duration (e.g., Port & Rotunno, 1979), gender (e.g., Whiteside & Blumstein, 1997; Oh, 2011; Li, 2013), speech rate (e.g., Kessinger & Blumstein, 1998), phrase-position (e.g., Lisker & Abramson, 1967), lung volume (e.g., Hoit et al., 1993) and language background (e.g., Flege, 1991). However, there are disagreements among scholars on the precise detail of each of these effects. Some of the issues found in the literature might be due to experimental design rather than linguistic factors.

For instance, in Lisker and Abramson's (1964) study, they found that the further back of the place of articulation, the longer the VOT value. However, they also observed that the aspirated stops in Cantonese (a Sinitic tonal language with two levels of register tones and contour tones) do not seem to follow this tendency. Lisker and Abramson did not consider a tone effect in their study; therefore, the exception found in their study might have been due to the tone effect. There are clearly many effects potentially in play with VOT variations. It is then important to attempt to control or otherwise keep track of the various effects when paying attention to any others.

The present study investigates lexical tone, POA, speech rate, phrase-position, pitch register, and gender effects on the prevocalic stop VOTs in L1 and L2 Mandarin and observes the tone effect within four of these six factors. It also explores which phonetic/phonological aspect of a lexical tone is responsible for the tone effect.

Cho and Ladefoged (1999) propose that VOT variations caused by POA can be due to physiological, aerodynamic, and temporal factors. This dissertation also endeavors to understand the tone effect within these accounts. While VOT variation due to POA can be explained by physiology and aerodynamics, tone effect on VOT requires considering vocal folds' stiffness relevant to raising and lowering pitch in producing the lexical tone. This dissertation proposes that vocal fold tensity is the leading cause of the tone effect (McCrea & Morris; 2005; Narayan & Bowden, 2013).

1.2 Background

The growing interest in speaking Mandarin has ushered the language to the forefront of global linguistics, becoming one of the leading foreign languages taught at many colleges and universities. However, Mandarin is a tonal language that can present great difficulty for non-tonal speakers, such as native English, Japanese, and Spanish speakers (Ding et al., 2010).

Mandarin, or Putonghua, comprises many varieties that differ in minimal ways from each other. One of these Mandarin varieties is Taiwan Mandarin, or Guoyu, the official language of Taiwan. For this dissertation, we will be considering Taiwan Mandarin (Guoyu) and the official Mandarin language of Mainland China (Putonghua) essentially the same and will henceforth label them as simply Mandarin.

Languages feature different sets of stop categories. In theory, negative VOT (voice-lead) at about 100 milliseconds (msec), zero VOT, and positive VOT (voice-lag) at around 75 msec are three cross-language phonetic categories summarized by Lisker and Abramson (1964) for voiced and voiceless stops. In reality, these numbers shift significantly according to many factors. For instance, one of them is the POA effect, where voiceless stops' VOT values can be around a positive VOT of 58-80 msec from [p^h] to [k^h] in English (Lisker & Abramson, 1964).

"Lexical tone ... is the systematic modulation of the pitch ... [carried by the syllable of a word]" (Sun, 1998). It generally refers to a distinctive pitch level carried in a syllable. Massaro et al. (1985) reveal that tone can be found in the domain of pitch, whose primary acoustic correlate is the fundamental frequency (F0) (p. 272). For instance, there are four basic tones in Mandarin. So, the syllable /ma/ has four different representations depending on which tone it carries. The four tones are tone one or high-level tone, tone two or mid-rising tone, tone three or falling-rising tone, and tone four or high-falling tone (henceforth T1, T2, T3, & T4, respectively). Figure 1 is a schematic tone contour and an example of /ma/ in four tones and their corresponding Chinese characters and English translations.

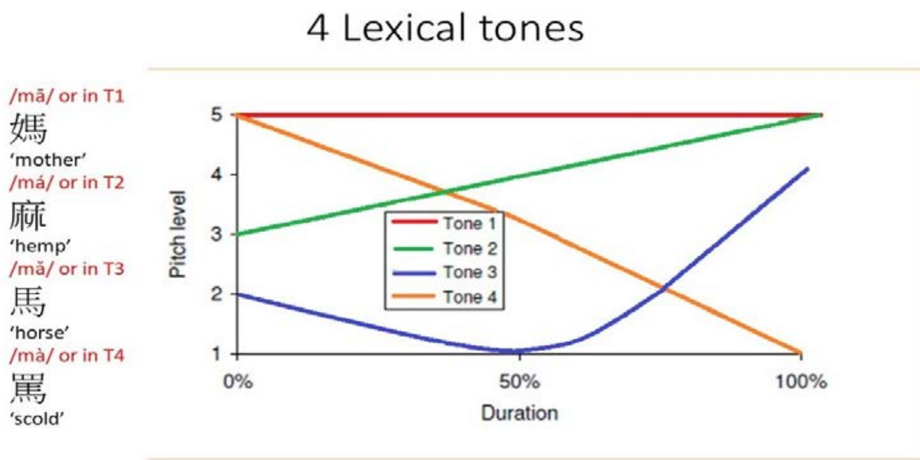

Figure 1. Schematic Lexical Tone Contour of Mandarin

Studies have reported that Mandarin lexical tone affects VOT values (Liu et al., 2008; Chen et al., 2009; Tseng, 2018). It should be noted that, however, there are controversies among studies of Mandarin on the issue. For instance, Liu et al. (2008), Chen et al. (2009), and Peng et al. (2009) reported that lexical tone affects the voiceless stop aspiration in L1 Mandarin production while Chen et al. (1998), Ran (2005), and Tse (2005) showed no significant differences in their analyses. This dissertation aims to explore the tone effect and investigates which of the tonal properties are responsible for this effect. Subsequently, it extends the investigation into L2 Mandarin.

1.3 Current Study & Rationale

This dissertation features four experiments (Experiment 1a & 2a for the native Taiwan Mandarin speakers and Experiment 1b & 2b for the L2 speakers) and one baseline VOT test to investigate the relationships between VOT-lexical tone and VOT-POA as a function of speech rate, pitch register, phrase-position, and gender in L1 and L2 Mandarin. It surveys VOT variations and demonstrates that they are affected by several phonetic and phonological properties. While there are inconsistent results reported in the literature (e.g., Liu et al., 2008; Chen et al., 2009; Peng et al., 2009; Chen et al., 1998; Ran, 2005; Tse, 2005; Tseng 2018), this study assumes a tone effect and looks at the effect in the conditions mentioned above. Moreover, it tests the findings from the native speakers with non-native Mandarin speakers from three language groups.

In this study, the central question addresses the acoustic properties in a tone that affects VOT. It examines two of the main properties of the lexical tone: F0 onset frequency and post-stop vowel duration. The study discusses the effect from a universal perspective by discussing several theories that have been put forth to explain vocal fold oscillation. It proposes that the effect is due to physiology, aerodynamics, but the main cause of the tone effect is the vocal cord tensity (Cho & Ladefoged, 1999; Hombert, Ohala, & Ewan, 1979; Ven dam Berg, 1958; McCrea & Morris, 2005; Narayan & Bowden, 2013). It proposes that the effects of lexical tone on VOT in Mandarin are not due to language-specific phonology. Instead, it is due to the physiology of the vocal tract.

Furthermore, this dissertation aims to gather empirical data from L2 production to broaden the discussion. To address relevant questions, it examines L1 Mandarin speakers' production in one post-stop vowel, three places of articulation, four lexical tones, three-pitch conditions, three speech rates, three utterance phrase-positions, and three different language backgrounds of L2 learners of Mandarin, who possess distinct voiceless stop series in their L1.

The three research questions are:

1. What are the roles of the listed factors below in VOT variation in Mandarin?
2. What is the specific component of lexical tone responsible for the tone effect on VOT?
3. Do we find the same effects in L2 learners of Mandarin?
    a. Does L2's native language matter? Do we find L1 VOTs for their L2 production?

The first focus examines how VOT varies when surrounded by the following factors: lexical tone, POA, speech rate, phrase-position, pitch register, post-stop vowel duration, and gender in L1 Mandarin. It expects to find a tone effect on VOT in different above-mentioned phonetic and phonological conditions and provides empirical evidence to understand the VOT variations in a tonal language.

The second focus drills down into the tone components, i.e., what specific lexical tone component is responsible for the tone effect on VOT. This follow-up question has not been well explored and documented in the literature. Therefore, the current study's second goal is to explore the tone effect and look at its acoustic properties of pitch-level and vowel duration first.

Fundamental frequency (F0) is the most crucial acoustic feature of tone. Vowel duration is also found to be different between tones. This dissertation focuses on these two critical tone components, i.e., post-stop F0 onset and post-stop vowel duration in semi-citation tone production (Yang, 2015). Given that T1 and T4 generally associate with high onset pitch and T2 and T3 usually exhibit low onset pitch, this study hypothesizes a negative correlation that as a speaker's pitch increases, the length of VOT in aspirated stops decreases (Lai et al., 2004).

As for vowel duration, Tseng (2018) reports an overall weak correlation (r = 0.184) between VOT and the following vowel duration. However, when analyzing tones separately, he reports a positive correlation for VOT and the post-stop vowel duration in T4 (r = 0.355), but a negative correlation in T3 (r = -0.164); the correlations in T1 (r = 0.031) and T2 (r = 0.066) were both positive but weak. This finding is somewhat unclear; therefore, this study aims to provide further empirical evidence to this incongruence. Given that studies (Kessinger & Blumstein, 1997; Kalveram & Jancke, 1989) have reported a correlation between VOT and vowel duration, we assume a vowel duration effect on VOT, and the longer vowel duration would lead to a longer VOT value.

Furthermore, the disagreement may be a methodological issue because most studies reviewed focused on one or two variables and treated other factors as random variables. Thus, this dissertation aims to broaden the discussion by adding a few more of the previously ignored random variables as the independent variables. Moreover, the shadowing task, which was employed in one of the previous studies (i.e., Tseng, 2018), but not all, has been found to have an effect of spontaneous phonetic imitation on VOT production (Goldinger, 1998; Babel, 2012; Kwon, 2015). Thus, the discrepancies reviewed here could potentially explain the inconsistent results found in previous studies. This study addresses some of the methodological gaps by keeping the same stimulus conditions, using modified carrier phrases, and removing the shadowing task.

The third question deals with L2 speech production. It postulates that the tone effect on VOT is mainly due to vocal fold tensity and hypothesizes that not only does tone effect exhibit in L1 Mandarin, it also exhibits in L2 Mandarin, thus universal. More specifically, do we find the same behaviors and effects in L1 Mandarin in L2 Mandarin? There seems to be a paucity in the literature that provides insights into L2 Mandarin production. In order to enhance our understanding of the phenomenon, it is necessary to examine both the production in L1 and L2 Mandarin. We expect to find the same tone effect in L2 production with other things being equal; i.e., L2 learners must be able to produce the tones correctly. Concerning the correctness of L2 tone production, this dissertation employs four native Mandarin speakers to evaluate L2 speech production (see section 5.4).

Moreover, we expect to find similar effects from different L2 backgrounds, e.g., Spanish, Japanese, and English learners of Mandarin. This dissertation examines data collected from these groups of L2 participants. The motivation for choosing these languages is that their native VOT values could be categorized by Cho and Ladefoged’s (1999, p.223) taxonomy, e.g., no aspirated, weak aspirated, and aspirated VOT values vs. strongly aspirated VOT values. This setting for the Spanish and Japanese speakers can potentially limit the possibility of their L1 transfer since their native languages do not feature strong aspiration. To provide a more comprehensive comparison of L1 and L2 VOT values, the same L2 participants also recorded their native speech for the present study (see section 5.12).

1.4 Organization of the Dissertation

This dissertation proceeds as follows. Chapter 2 covers a literature review of focused acoustic properties and factors and their effects. It summarizes noteworthy findings in previous studies and discusses methodological shortcomings.

Chapter 3 and Chapter 4 focus on Experiment 1a and Experiment 2a of the native Taiwan Mandarin, including significant findings, statistical analyses, and a preliminary discussion of the results. Chapters 5 and 6 are dedicated to the L2 production and the group comparison of Experiment 1b and Experiment 2b; the evaluation for the L2 speakers' tone productions and their L1 VOTs are reported. Chapter 7 concludes this dissertation with a discussion and a conclusion. Future directions and implications are provided in Chapter 7 as well.

CHAPTER 2. LITERATURE REVIEW

2.1 Introduction

Researchers of Mandarin speech production have found that the lexical tone affects the prevocalic voiceless stop aspiration in L1 Mandarin (e.g., Liu et al., 2008; Chen et al., 2009; Peng et al., 2009; Tseng, 2018); that is, with other things being equal, VOT is significantly shorter preceding Tone 1 and Tone 4 than Tone 2 and Tone 3 (henceforth T1, T4, T2, & T3, respectively). It is unclear what in the lexical tone affects VOT duration and how the tone effect behaves in different POAs, speech rates, phrase-positions, and pitch registers. It is also unclear whether L2 Mandarin speakers also exhibit a similar tone effect. This study aims to contribute to the discussion by examining tones' effects on VOT and their relationships in L1 and L2 Mandarin productions.

The central questions address the acoustic properties in a tone that affects VOT. Furthermore, this dissertation discusses the effect from the perspectives of Physiological, Aerodynamic, and Myoelastic theories (Cho & Ladefoged, 1999; Hombert, Ohala, & Ewan, 1979; Ven dam Berg, 1958). To explore relevant questions, we examine VOT variations in L1 Mandarin speakers' production to further broaden our understanding of the effects from the native perspective. Subsequently, it tests the findings with L2 learners of Mandarin from three language backgrounds. Chapter 2 reviews the focused laryngeal properties, available sources to VOT variations, tone types, and the target L2 language backgrounds.

2.2 Review of VOT and Lexical Tone

2.2.1 Voice Onset Time

VOT is the interval between the release of a stop consonant and the onset of voicing for the following segment (e.g., Cho & Ladefoged, 1999). This short interval serves as a significant perceptual cue to distinguish laryngeal contrast and aspiration in Mandarin (Liu et al., 2008). The delay of the vocal fold oscillation is also referred to as "voice lag". Cross-linguistically, the amount of lag may be a *short-lag*, i.e., less than 30 milliseconds (msec), or a *long lag*, i.e., greater than 30 msec (Yavas, 2011). The vocal cord configuration for the relative stop may be 1) fully voiced, 2) partly voiced, 3) voiceless unaspirated, 4) weak or voiceless aspirated, or 5) strongly aspirated (Ashby & Maidment, 2005). In Mandarin, stops are phonetically voiceless across the board (Iwata & Hirose, 1976) and can be considered as being in the category of "strongly aspirated" (Cho & Ladefoged, 1999). Figure 2 shows the possible VOTs proposed, and this dissertation focuses on the fifth category and measures the VOT values in milliseconds.

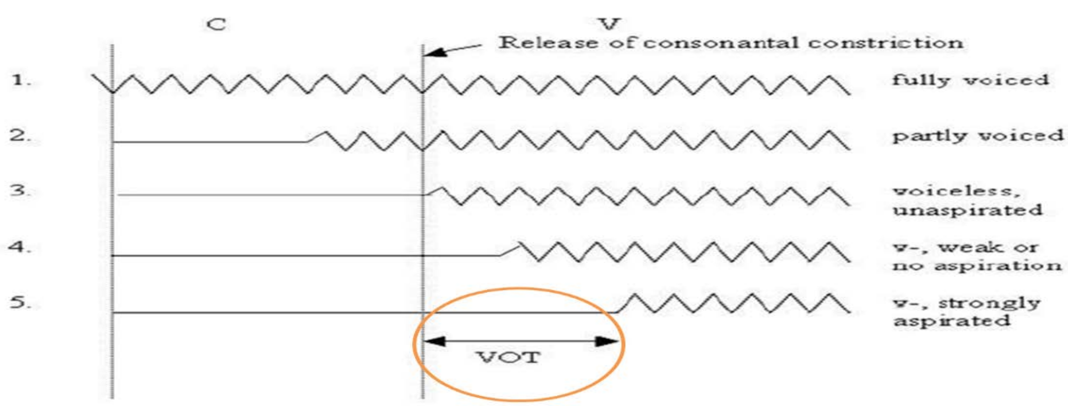

Figure 2. Schematic possible VOTs adapted from Ashby & Maidment (2005, p. 95)

2.2.2 Lexical Tone in Mandarin

Lexical tone is the systematic use of the phonemic F0 distinction in a word's syllable for a lexical meaning (Sun, 1998). It generally refers to a unique pitch movement in a syllable, and the physical attribute of pitch is the fundamental frequency (Yang, 2015). There are four basic tones plus a neutral tone in Mandarin; the neutral tone is ignored in this study because it is phonologically weak and environmentally predictable. For instance, the syllable /ma/ has four different representations depending on which tone it carries. As shown in Figure 1 (reinserted in Figure 3 below for convenience), the four tones are T1 or high-level tone, T2 or mid-rising tone, T3 or falling-rising-tone, and T4 or high-falling tone. They are conventionally denoted in a five-tone notation system: 55, 35, 214, and 51 for T1, T2, T3, and T4, respectively (Sun, 1998; Chao 1930).

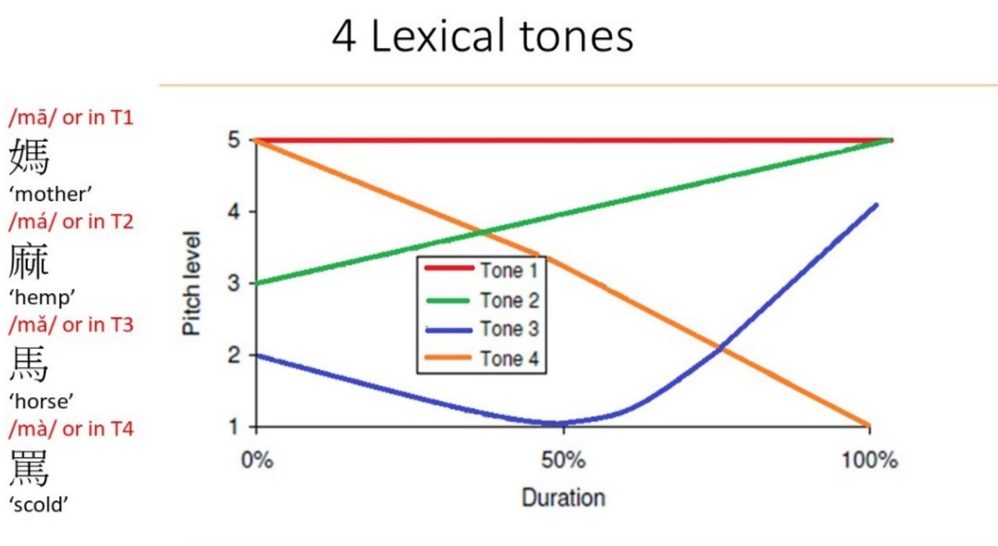

Figure 3. Schematic Lexical Tone Contour of Mandarin

Massaro et al. (1985) reveal that tone can be found in the domain of pitch, whose primary acoustic correlate is the F0 (p. 272). This study analyzes the tone contours from the first detectable/measurable F0 to the end of the post-stop vowel's second formant (see section 3.5).

2.3 Theoretical Accounts of Vocal Fold Vibration and VOT

To understand the physical nature of VOT, many theories have been proposed to explain how the vocal folds vibrate and how post-stop voicing is delayed (Van den Berg, 1958; Jaeger, 1978; Titze, 1980; Browman & Goldstein, 1984; Cho & Ladefoged, 1999; Jansen, 2004; Ladefoged & Johnson, 2010; Reets & Jongman 2014). Physiologically, the vocal fold vibration rate is related to the vocal fold ligaments' length and the tissue tension linked to the vocal tract. Aerodynamically, the vocal folds oscillate when they are adducted, not tightly, and there is sufficient air force pushing through them. However, this simulation can be passively delayed by a consonantal closure. That is, a stop closure at a syllable-initial position breaks the equilibrium of the cavity air pressures, and the subglottal pressure becomes insufficient to initiate the vibration "on time." This is VOT.

VOT has been found to correlate to the place of oral occlusion, stop closure duration, the folds' tenseness, the airstream's constriction area and volume, the balance of supraglottal/subglottal air pressure, and lung volume (e.g., Ohala, 1983; Hoit & Solomon, 1993; Stevens 2000; Boersma, 1998; Jansen, 2004). Moreover, speech rate, utterance's phrase-position, gender, sociolinguistic factors, and language background have also been suggested to influence VOT (e.g., Flege, 1991, 1995; Kessinger &

Blumstein, 1997, 1998; Chen et al., 2009; Stölten, K., & Engstrand, 2002; Ma et al., 2018). The following section reviews some of these VOT influencers.

2.4 Sources of VOT Variations

2.4.1 Introduction

VOT is highly sensitive to various factors. The review here is cross-linguistic because of the shortage of studies in Mandarin. The factors reported may include the place of articulation (henceforth POA; e.g., Lisker & Abramson, 1964; Cho & Ladefoged, 1999), lexical tone (e.g., Liu et al., 2008; Chen et al., 2009; Peng et al., 2009; Tseng, 2018), vowel context (e.g., Chen et al., 2009; Rochet and Fei,1991), vowel duration (e.g., Port & Rotunno, 1979; Kessinger & Blumstein, 1997), phrase-position (e.g., Lisker & Abramson, 1967), the onset of the fundamental frequency (henceforth F0 onset; e.g., McCrea & Morris, 2005; Lai et al., 2009; Narayan & Bowden, 2013), gender (e.g., Ma et al., 2018; Li, 2013; Swartz, 1992; Whiteside & Irving, 1997; Oh, 2011), speech rate (e.g., Kessinger & Blumstein, 1997, 1998; Allen & Miller, 1999), language background (e.g., Flege, 1991), and lung volume (e.g., Hoit et al., 1993). This is not an exhaustive list, and although these factors have been proposed, they are not all without any disagreements amongst studies.

2.4.2 Place of Articulation Effect

Place-dependent VOT variations can generally be explained by physiological and aerodynamic accounts. That is that VOT may depend on the places of constriction contacts and occlusion area plus the intraoral air pressure (see Cho & Ladefoged, 1999 and Eshghi et al., 2016, for more details).

A study by Lisker and Abramson (1964) reports that POA affects VOT. They reveal that the further back of the POA, the longer the VOT. However, they also report exceptions, such as the aspirated stops in Cantonese and Eastern Armenian that do not abide by this tendency. Here, Cantonese is a tonal language, and it might be the case that they did not consider the possibility of the tone effect.

Similarly, Cho and Ladefoged (1999) compare 18 languages' VOT data in the UCLA Phonetics Lab database. Most languages follow the POA effect tendency, but some ejective languages, Hupa, Montana Salish, and Navajo, do not. They show that the aspirated alveolar stop's VOT was longer than that in the velar stop in Hupa and Navajo, and Montana Salish's bilabial stop had longer VOT than the alveolar and velar counterparts. On the other hand, some data from tonal language speakers show that velar stops generally associate with longer VOT values than /p/ and /t/, but /p/ sometimes associates with longer VOTs than /t/ (Rochet & Fei, 1991; Chen et al., 2007; Liu et al., 2008; Lai et al., 2009). These observations lead us to wonder whether the POA effect applies in Mandarin. If it is a common trend, Mandarin should present the same effect where the aspirated VOTs in a /k/ > /t/ > /p/ pattern. However, for instance, Chen et al.'s data show a /k/ > /p/ ≈ /t/ pattern. There may be some other factors (e.g., lexical tone) that need to be accounted for when examining VOT variations. This dissertation adds empirical data from Mandarin and aims to look at the POA effect in each lexical tone.

2.4.3 Lexical Tone Effect

Pitch value, tone category, and tone contour are generally used to describe lexical tones, and F0 is the most important acoustic property of tones (Yang, 2015). While pitch height is generally used to explore level tones, duration is used to compare contour tones (Chen et al., 2009; Yang, 2015). Two of these main characteristics for distinguishing lexical tones have been suggested to influence VOT value in Mandarin (Lai et al., 2004; Chen et al., 2009).

Many studies have looked at the relationship between VOT and tones across different tonal languages. While some researchers have claimed that there is no significant influence of tone on VOT duration (e.g., Chen et al., 1998; Tse, 2005; Ran, 2005), others have reported a significant tone effect (e.g., Liu et al., 2008; Pearce, 2005; Chen et al., 2009; Tse, 2005; Tseng, 2018).

Liu et al. (2008) study the effect of tonal changes on VOT between native Mandarin speakers and superior esophageal (SE) Mandarin speakers who have undergone a laryngectomy. They show that both groups of subjects had VOT patterns of T3 > T2 > T1 > T4, and VOTs in T3 and T2 pairs were significantly longer than those in T1 and T4 were (normal speakers also had significantly longer mean VOTs than the SE speakers did). Liu et al. propose that both groups revealed the tone effect on VOT; i.e., even speakers with an artificial speech resource showed a tone effect on VOT. Their results seem to lend more support to the accounts of physiology and aerodynamics than the vocal cord tensity. The present study assumes the tone effect and only tests non-pathological participants.

Modeling Liu et al.'s (2008) study, Chen et al. (2009) examine the tone effect in Mandarin and Sixian Hakka (a variety of Chinese that has six contrasted lexical tones; Chen et al., 2009, p. 553). They look at two voiceless stop series with three vowels in four contrasting matching tones, although the matched tones they used do not have the matched tone contours and onset pitches[1]. They report a significant tone effect in both groups, and their results reveal that Mandarin VOTs were in the T2 > T3 > T1 > T4 pattern, where only VOTs in T4 were significantly shorter than those in T3, but the VOT differences between T4 and T2 were not significant; this was not the same findings as in Liu et al. (2008).

Following Liu et al.'s (2008) and Chen et al.'s (2009) studies, Tseng (2018) investigates the tone effect on VOT for word-initial aspirated stops (i.e., [p^ha], [t^ha], and [k^ha]) in L1 and L2 Mandarin production. He looks at the production data from eight native Mandarin speakers (4 males & 4 females) and finds results in keeping with Liu et al.'s, where T3 and T2 exhibit significantly longer VOT values than T1 and T4. Tseng's results also show that the same results were observed in the production of his L2 subjects (15 native English speakers). He shows that VOT values were significantly affected by tones in both groups, and the results revealed that his L2 learners used non-English VOT for L2 production (i.e., 58~80 msec; c.f., Cho & Ladefoged (1999) vs. 88~93 msec).

Peng, Chen, and Lee (2009) also look at 15 males and 15 females of Taiwan Mandarin speakers and six Hakka speakers in Taiwan for a tone effect on aspirated and

---

[1] They used match tones because Mandarin has four lexical tones while Hakka has seven. The four tones from Mandarin are& 55, 35, 214, & 51, and 6 tones from Hakka are& 24, 31, 55, <u>32</u>, 11, & <u>55</u>. They use the underlined two for matching to Mandarin 55 and 35.

unaspirated stops in real words and non-words. Peng et al. report that VOTs in non-words were more prolonged than those in real words. Their analysis shows that only in non-word stimuli, tone affects VOT, and it does not influence stop's VOT in real words (p.351). They reveal that while their Mandarin results showed no tone effect in real words but non-words, their Hakka's results exhibited significant VOT differences in both conditions. This raises the perplexing question of why only in the non-word conditions VOT varied.

Moreover, they report that VOTs in Hakka were longer in the low-low tone than those in the mid-falling tone, which in turn were shorter than those in the low-rising tone. This result was contrary to the general findings in the literature, where the low-onset tone usually exhibits longer VOT than the mid-onset tone does. To complicate the picture further, they bring results from two more tones: Hakka's low-onset-entering and high-onset-entering level tones. Here, we would expect the VOT value in the high-entering tone to have shorter VOT values than the low-entering tone due to the high F0 onset effect (e.g., McCrea & Morris, 2005; Lai et al., 2009; Narayan & Bowden, 2013); however, they report non-significant VOT differences in these two tones from Hakka. It is unclear what might have caused their non-significant finding.

Tse (2005) scrutinizes a possible tone effect on VOT in Cantonese. He examines six subjects (5 males & 1 female) with two sets of the voiceless stops and affricates in two to four-syllable words with two vowels in four tones (i.e., the 55, 21, 33, & 25 Cantonese tones). He shows that VOT values in tones were 78.58, 77.17, 65.78, and 62.70 milliseconds for 21-tone, 25-tone, and 33-tone 55-tone, respectively. Individually, Tse reveals that the tone effect was only observed in three male participants; it was not in the other two males and the female subjects. Overall, Tse reports that only VOTs in level tones were significantly shorter than those in contour tones in Cantonese, and VOT differences were non-significant between contour tones.

On the other hand, examining 27 Cantonese speakers, Lam (2010) reports that low-onset register tones had longer VOT than high or mid onset register tones, but Lam's low-low tone had shorter VOT than the high-rising tone did. In addition to the Hakka data, Cantonese, too, does not help to clarify the investigation.

Two studies look at the comparisons of level tones. Pearce (2005) looks at Kera, a Chadic language with three distinct register tones, e.g., high 55, mid 33, and low 11 lexical tones, and shows that the higher the tone, the longer the VOT. Geissier (2018) looks at 19 native Central Tibetan speakers, a standard Tibetan branch dialect that uses two distinctive high and low register tones. He reports longer VOTs in high-tone aspirated stops than low-tone aspirated stops. Register-tone languages seem to suggest a different pattern of tone effect on VOT than contour-tone languages. If we assume the tone effect is exclusively due to vocal fold tensity, we would expect some consistency across languages; however, both findings here differ from the results found in Mandarin mentioned above.

The general patterns in Mandarin seem to propose a tone effect in which the high onset pitch leads to short VOT. This observation can be explained by the Myoelastic-aerodynamic theory, which postulates that the airstream delivered by the lungs and the trachea activate the vocal folds, and therefore the fundamental frequency of the vibration depends on the effective mass and stiffness of the vocal folds (Van dam Berg, 1958, p.230). However, unlike the POA effect, the outcomes reported here do not provide enough confirmation to support a robust inference for the lexical tone effect across different tonal languages.

The discrepancy could have been due to the methodology. For example, Tseng (2018) uses the stimulus carrier phrase, ***pa**, Wǒ xiàn zài shuō* **pa (*pa,** I now say, **pa***) that has been questioned for the lack of control on pitch variation effect of the carrier phrase. Furthermore, the stimulus token is not in an actual embedded condition due to the phrasal final pitch falling. In order to provide additional supportive evidence to the questions, this dissertation uses a modified version of the carrier phrases to supply a better control of pitch variation and vowel duration (see Chapter 3 & 4 for details).

2.4.4 Vowel Context Effect

The general report concerning the relationship between the vowel context and VOT is that the higher the tongue height of the vowel, the longer the VOT value[2] (e.g., Port, 1979; Rochet & Fei, 1991; Chao et al., 2006; Chen et al., 2009; Li, 2013) although Lisker and Abramson (1967) report that the vowel seems not to influence the preceding stop VOT. Reexamining Lisker and Abramson's study, Port and Rotunno (1979) report that the high vowel does seem to cause a longer VOT value.

---

[2] Jansen (2004) proposes that the tenseness of the vocal tract walls determines the amount of passive expansion of the vocal tract behind the constriction, thereby changing the intraoral pressure, and the higher pressure leads to a longer delay of voicing.

In English, Chang et al. (1999) show that VOT of voiceless stops is typically long before high vowels and short before low ones. Nearey and Rochet (1994) confirm this with data from French and English. Nearey and Rochet's data show that the trends for VOT values with vowels were tense > lax, front > back, and high > low, but a few exceptions, where VOTs with /a/, which is a low vowel, was longer than with /ɪ/, which is a high vowel.

In Mandarin, two studies show the opposite results. Rochet and Fei (1991) report that the VOT values were the longest with /i/, the second with /u/, and the third with /a/. In contrast, looking at the vowel-VOT relationship in two genders, Li's (2013) reports that VOTs in /u/ were shorter than those in /a/ were in both genders; VOTs with /i/ were longer than those in /u/, but no significant difference between VOTs with /i/ and /a/.

The majority of the studies show that tense vowels do lead to longer VOT values. However, investigating this tendency is not the focus of the present study. The present study controls the vowel context effect by using only /a/, the open-unrounded vowel.

2.4.5 Vowel duration and Speech Rate Effects

The general observation concerning VOT and speech rate is that the faster the speech rate, the shorter the VOT value; however, while long-lag VOT shortens as a function of the fast speech rate, short-lag and lead VOT values remain relatively stable regardless of the speed (e.g., Miller et al. 1986; Pind, 1995; Kessinger & Blumstein, 1998).

In addition to the theories of physiology and aerodynamics, Cho and Ladefoged (1999) suggest a temporal account for VOT alterations, which proposes that vowel duration can affect VOT, and speech rate determines the differences of vowel duration. The temporal account suggests that the change is proportional, i.e., VOT and vowel duration would change for the same or a similar percentage. However, evidence from Port and Rotunno (1979) shows that it is not the case; VOT changed around 13-14 % while vowel duration changed roughly about 50% according to the speech rate.

One thing to keep in mind is that speech rate and vowel duration are co-variants to VOT. Vowel duration can be calculated for the speed, but vowel duration does not determine the speech rate. Kessinger and Blumstein (1997) test a speech rate effect on VOT in Thai, French, and English. Kessinger and Blumstein (1998) further propose a speech rate categorization in three speeds where syllable durations are 100-299 msec, 300-499 msec, and 500-799 msec for *fast*, *medium*, and *slow* speeds, respectively. Although they suggest that speech rate affects VOT and vowel duration values, their results show that the changes were not systematic. For instance, testing VOT in /pi/ and /pæ/at different speeds, they report that VOT values increased systematically in the /pi/ instances as a function of different speeds, but not with /pæ/ instances; VOT increased at the *fast* and *medium* speeds but decreased at the *slow* speed. In other words, VOT in /pæ/ at the *slow* speed had shorter values than those at *fast* and *medium* speeds. Nonetheless, their finding suggests the presence of an additional factor to VOT variations.

While scholars have looked at the relationship between speech rate and VOT in different languages, no study was found for Mandarin, which is another reason to consider VOT from the temporal account in this study. Since languages may have intrinsic speech rates, the suggested speech rates by Kessinger and Blumstein (1998) can only be a guide. Based on the suggestion, this study predicts that the faster the speech rate, the shorter the VOT value in Mandarin. It tests VOT changes in three speeds, using Kessinger and Blumstein's (1998) method.

2.4.6 F0 Onset Effect

The lexical tone effect mentioned above is largely related to F0 onset. Studies have addressed the correlation between F0 onset and VOT; that is, generally, a high F0 onset causes a shorter VOT value (McCrea & Morris, 2005; Lai et al., 2009; Narayan & Bowden, 2013)

McCrea and Morris (2005) examine the effects of the F0 level on VOT in six adult English male speakers. They ask the participants to produce the stimulus sentences in monotone in three F0 levels. They report no significant difference between these pitch levels in aspirated stops (M= 62.8, 60.7, and 56.9 milliseconds for *low*-, *mid*-, and *high*-F0s, respectively; although they did report that in unaspirated stops, high F0 displayed significantly shorter VOTs than *low*- and *mid*-F0s).

Replicating the study of McCrea and Morris (2005), Narayan and Bowden (2013) compare ten female English speakers and ten female Korean speakers producing the same stimuli as those of McCrea and Morris (2005). They ask the participants to produce plosives with /a/ in one phonation continually: 1) in their regular speaking pitch, 2) raise the pitch to their highest pitch, and then 3) drop to their lowest pitch. They report that in both languages, shorter VOTs were associated with higher F0. However, they report that this association only occurred in the *high*-pitch condition in both languages; in *low*- and *mid*-pitch conditions, the association showed in Korean, not in English. The difference in pitch frequency between *high*-, *mid*-, and *low*- was not reported.

In order to raise a pitch (or F0), one must elevate the larynx, which in turn decreases the cavity volume behind the constriction (see Eshghi et al., 2016 and McCrea & Morris, 2005). This simulation of laryngeal movement increases the supra-glottal air pressure, which should further delay voicing, as the instance shown with velar versus bilabial and alveolar plosives. From the physiological and aerodynamic perspective, the elevation of the larynx to raise pitch or produce voiceless stops decreases the intraoral volume, thereby reducing the velocity of the airflow, which in turn inhibits vocal vibration (Cho & Ladefoged, 1999).

As reviewed earlier, Myoelastic-aerodynamic theory postulates four factors to explain the vibration of vocal folds (Van dem Berg, 1958; Reetz & Jongman, 2014, p. 77). It illustrates that a) the Bernoulli Effect allows the vocal folds to be sucked together and b) the folds vibrate faster when they are elongated than they are shortened, tensed than relaxed, and thin than thick. In other words, long vocal folds oscillate at a slower rate than short ones. Therefore, the vocal folds should be short when producing high F0 frequency. Thus, the delay of voicing should be shorter with short vocal folds than long ones. The present study replicates the pitch rate application using Narayan and Bowden’s (2013) approach for this investigation.

In sum, the pitch has been suggested to alter the VOT value. This dissertation investigates the tone effect on VOT and aims to gain further insight into how tone affects VOT. The Myoelastic-aerodynamic account provides a sounding explanation for a simulation in producing Mandarin T1 and T4, which both start with a higher pitch. Therefore, we expect them to exhibit shorter VOT values than T2 and T3. Chapter 4 explores this topic.

2.4.7 Phrase-position Effect

This effect has not been explored much. In other words, many scholars have treated the phrase-position effect on VOT as a random variable. Lisker and Abramson (1967) look at some effects on VOT's utterance environment in English stops in running speech. They report a significant increase in voicing-lag in syllables at the sentence-*final* over those at the sentence-*medial*. They suggest that the extended VOT values may be due to sentence-*final* stress.

On the other hand, testing L1 and L2 Mandarin, Tseng (2018) compares VOT values at the phrasal *initial* and phrasal *final* and reports no significant VOT difference in the two conditions. Even though Lisker and Abramson suggest a phrase-position effect on VOT duration, our understanding of this effect is unclear. This dissertation assumes the effect and hypothesizes that the phrase-position affects VOT, and VOT values will be longer at the utterance-*final* position than those at the utterance-*initial* and utterance-*medial* positions.

2.4.8 Gender Effect

In addition to the possible factors mentioned above, another subsequent, less apparent effect on VOT is gender. Intuitively, gender can cause VOT differences because of the oral cavity and laryngeal sizes. However, if VOT would be affected by gender due to the cavity size, we would expect to find long VOTs in females, analogous to the long VOT caused by the small cavity size of a velar stop. Secondly, F0 can also be another possibility. Female generally has a higher pitch than male; therefore, females should exhibit shorter VOT than males. Studies have reported incongruent results of VOT value with respect to gender.

Swartz (1992) tests eight-male and eight-female American English speakers reading sentences. Focused only on different voiceless and voiced alveolar plosives, Swartz reports that females produced significantly longer VOTs than males. However, Morris, McCrea, and Herring (2008) report no significant differences between genders for both voiced and voiceless plosives from 40 male and 40 female native English speakers (p.311). Still, in English speech, Whiteside & Irving (1997) report that their female participants produced longer VOT than their male peers, but Eshghi, Alemi, & Zajac (2016) report no significant differences between genders.

If VOT varies as a function of the speaker's gender, it can be thought of as attributed to the physiological account. However, Oh (2011) shows that while Whiteside and Irving (1997) report females produced longer VOTs than the males, Oh's 19 native male Korean speakers produced longer VOT values than her 19 female native Korean participants did. She argues that there is no inherence reason why VOTs are longer in English females but shorter in Korean females.

Furthermore, Peng, Chen, and Lee (2014) report that their females had significantly longer VOT than males did in aspirated stops, but the inverse pattern in unaspirated stops in both Mandarin and Hakka, and in Mandarin, this was only true in the alveolar and bilabial stops, not in the velar stops. Li (2013) reports similar patterns to that of Peng, Chen, and Lee's, yet after factoring out the speech rate, the gender effect was only significant for the voiced stops; males spoke faster than the females in their study. Moreover, Ma et al. (2018) report significant VOT differences due to gender in aspirated stops, but not those in unaspirated stops. They report that females displayed longer VOTs for aspirated stops but shorter VOTs in unaspirated counterparts than the males. Although other factors such as speech rate and pitch register were attended to in these studies, it is unclear why females who generally have higher pitch showed significantly longer VOTs than the males. This study includes gender as another mixed-effect to understand further the effect in different tones, POAs, speech rates, pitch registers, and vowel durations.

2.4.9 Language Background Effect

Finally, VOT has been reported to be different across languages and proficiencies. Second language studies have revealed that L2 learners often use their L1 system for L2 production (Ioup & Weinberger, 1987; Flege, 1991, 1995; Major, 2001, 2005; Harada, 2004; Escudero & Boersma, 2004; Gass, 2013).

Flege (1991) shows that early L2 Spanish learners of English were able to produce longer and more English-like VOTs than the late learners do. Shimizu (2011) reports that his Korean, Thai, and Mandarin Chinese speakers tend to use their native stop VOTs to produce English stops, but the ESL Korean learners' production was more native-like than the Thai and Chinese counterparts.

Harada (2004) compares Japanese and English VOT productions by Japanese monolinguals, Japanese bilinguals, English bilinguals, and English monolinguals and shows that L1 influence existed in the bilingual groups. His Japanese bilinguals produced English VOTs shorter than English monolinguals, and the Japanese bilinguals' Japanese VOTs were shorter than the Japanese monolinguals. English Bilinguals produced similar English VOTs to the English monolinguals but longer Japanese VOTs than Japanese bilinguals and Japanese monolinguals.

In Mandarin, Tseng (2018) reports that his L2 English learners of Mandarin produced similar Mandarin VOT values to native Mandarin speakers, which seem to be a counterexample to the general findings. He reports that in a closer examination, the length of learning Mandarin for his L2 group was 5.93 years, which might be the reason why they were able to produce comparable VOTs to that of the natives (slightly shorter, but statistically non-significant). Tseng notes that the English VOTs of the L2 learners were not elicited from the same participants; therefore, the comparison of English aspirated VOT values indirectly could be mistaken. The current study aims to bring additional evidence to the questions from different languages that do not feature aspiration in their native VOTs, such as Spanish and Japanese, and elicit the native VOTs from the same group of L2 subjects.

2.5 Summary

This dissertation investigates the effects of lexical tone and other factors on VOT of Mandarin aspirated stops. As mentioned earlier, four experiments (two for the native Mandarin speakers and a parallel set of two for the L2 speakers) are designed to look at how VOT behaves with different lexical tones, POAs, speech rates, pitch registers, phrase-positions, vowel durations, genders, and native language backgrounds. Table 1 summarizes the factors that each experiment attends to and the expected patterns of VOT distribution for each factor. A baseline VOT test is also conducted to elicit the native VOTs from the same L2 participants. Their length of learning Mandarin is also reported. Starting with the first experiment, the next chapter investigates the tone effect on VOT in native Mandarin speakers in Taiwan.

Table 1. Expected general VOT trend suggested in the literature. Experiments 1a & 2a are for the native speakers, and Experiments 1b & 2b are replications of 1a & 2a for the L2 speakers.

Attended to in:	Factors	General expected patterns of VOT values in each factor's condition
Experiment 1a	Lexical tone:	T3 ≈ T2 > T1 ≈ T4
	POA:	/k/ > /t/ > /p/
	Speech rate:	*slow* > *natural* > *fast*
	Phrase-position:	*final* > *medial* & *initial* ≈ *final*?
	Gender:	*female* > *male*? or *male* > *female*?
Experiment 2a	Fundamental frequency:	low F0 onset > high F0 onset
	Vowel duration:	*long* > *medium* > *short*
Experiment 1b & 2b	Non-native Language Background:	Mandarin Chinese > English > Japanese > Spanish

>: VOT to be greater than, ≈: VOT to be approximately equal to

CHAPTER 3. TONE EFFECT IN NATIVE SPEAKERS – EXPERIMENT 1A

3.1 Introduction

This chapter investigates the relationship between lexical tones and VOTs by examining the production of various phonetic properties in native Mandarin. The central inquiry addresses whether there is a relationship between lexical tone and VOT in L1 Mandarin when the following factors are individually attended to: the place of articulation (POA), speech rate (speed), phrase-position, and gender.

This chapter starts with a report of the participants' background information, followed by subsections of the stimuli and data measurement. It then summarizes the statistical results and provides a preliminary discussion of the VOT influencers in the L1 Mandarin data.

3.2 Participants

The study subjects were 68 Taiwanese participants who self-identified as native speakers of Mandarin[3]. This Taiwan group had 36 females and 32 males. Their ages ranged from 18 to 68 years old, with an average age of 29.09 (SD = 13.19). They reported having been born and raised in Taiwan. Sixty out of 68 participants reported that they also regularly speak Taiwanese Minnan, a variety of Hokkien languages spoken

---

[3] Among the 68 native participants, 52 of them were born and raised in Taichung, 6 in Taipei, 3 in Yilan, 1 in Jilong, 1 in Taiyuan, 1 in Miauli, 1 in Zhanghua, 1 in Nantou, 1 in Koushong, and 1 in Pintong; generally speaking, 12 were from the north of Taiwan, two from the south of Taiwan, and 54 were from the central Taiwan.

natively by many Taiwanese. None of the participants self-reported any known history of a speech disorder or a hearing impairment.

The author recruited most participants at the National Library of Public Information (NLPI) in Taichung city by handing out recruitment flyers at the entrance. Fifty-four participants recorded their production in this library. Fourteen additional native speakers recorded their speech at three different universities - Fuguang University in Yilan city, Providence University in Shalu city, and Tunghai University in Taichung city.

3.3 Speech Samples

The study's stimuli were monosyllabic Mandarin words, which included the aspirated voiceless plosives /p, t, k/ followed by the low-central-unrounded-vowel /a/. Each stimulus appeared twice in two carrier phrases, once with a sentential ending classifier and once without the classifier to create open/closed-ending sentence conditions. Each sentence was produced three times at different speech rates.

The change of phrase-position conditions was done by adding the classifier 七次 (qīcì or [tɕʰītsʰì]); "*seven times*" after the second stimulus. The adding of the 七次, ([tɕʰītsʰì]) created a token-embedded condition for the second stimulus, and without it, the sentence created a token-ended condition. All Chinese characters in the carrier phrases were in T1 between the phrase-*initial* and the phrase-*medial* stimuli. Notice that 七次, [tɕʰītsʰì] enclosed the stimulus with a tone-one character as well. Table 2 lists the stimuli and the carrier phrases.

Table 2. Monosyllabic stimuli & the carrier phrase plus the classifier

Tone	POA		
T1	[pʰā]	[tʰā]	[kʰā]
T2	[pʰá]	[tʰá]	[kʰá]
T3	[pʰǎ]	[tʰǎ]	[kʰǎ]
T4	[pʰà]	[tʰà]	[kʰà]
Carrier phrase:	____先生說 (xiān shēng shūo) ____七次(qīcì); "*Mr.*___ *says seven times*"		

The entire sentence was presented with a combination of p*inyin*[4] and Chinese characters. For instance, **pā** 先生說 (xiān shēng shūo) **pā** 七次 (qīcì) had the target stimulus in pi*nyin*, but the carrier words were in Chinese characters. There were two reasons for this. The first reason was that the monosyllabic word might not have a written character, which does not have a semantic meaning across all three articulatory gestures and four tones. For instance, neither /kī/ nor /kì/ has a linguistic meaning or a matched character in Mandarin. However, /kī/ and /kì/ are possible pronunciations for Mandarin speakers, which are accidental gaps, and /kì/ can be a surname for a foreigner when introduced (e.g., Key 先生 as *Mr. Key*). Secondly, the entire sentence was an attempt to create a meaningful interpretation (e.g.," *Mr.* *Key* *says* *Key* *seven times*") to provide more contextual support instead of producing [pʰā], or [tʰā], or [kʰā] monotonically (e.g., "*Pa, I now say Pa"*).

4 In Taiwan, the *pinyin* system is not officially taught in school, but the translation of street names and other foreigner-friendly signs are written in *pinyin.*

The elicited data were attended to the following conditions:

1. Place of Articulation (POA): the aspirated voiceless stops of Mandarin, /p, t, k/, were used.
2. Post-stop vowel: only the low-central-unrounded vowel /a/ was selected. Vowel duration control was done by adding 七次 (qīcì or [tɕ^{h}īts^{h}ì]); "*seven times*," which allowed for measurements of *long* and *short* vowel durations from the phrase-m*edial* token.
3. Four lexical tones: high-level (or T1), mid-rising (or T2), falling-rising (or T3), and high-falling (or T4).
4. Three speech rates: *slow*, *natural*, and *fast*. This study adapts Kessinger and Blumstein's (1998) method to ask the participants to speak as *natural*/*slow*/*fast* as possible without forsaking accuracy. See the next section for detail.
5. Three types of phrase-positions: utterance-i*nitial* (i.e., the initial stimulus), utterance-*medial* (i.e., the medial stimulus due to the adding of 七次 (qīcì or [tɕ^{h}īts^{h}ì]); "sev*en times*," as a classifier), and utterance-*final* (i.e., the final stimulus without the classifier). The participants read the stimulus sentence once with the classifier, which generates two stimulus tokens of utterance-*initial* and utterance-*medial*, and once without the classifier, which generates two stimulus tokens of utterance-*initial* and utterance-*final*. There are twice as many *initial* tokens as *medial* and *final* tokens for analysis.

3.4 Data Elicitation

Experiment 1a instructed the participants to read 24 stimuli (i.e., /p t k/ in four tones, and each stimulus appeared twice) in their carrier phrases at three different speech rates to collect the speech data. The subjects' task was simply to read the sentences on the computer screen, using PsychoPy (version 1.83.04; Peirce 2009), which displayed the 24 stimuli and the carrier phrase in a random order within each carrier phrase type. The participants recorded the carrier phrase without the classifier first and then recorded the one with the classifier. Figure 4 shows two instances of the screenshot for what the participants saw without and with the classifier.

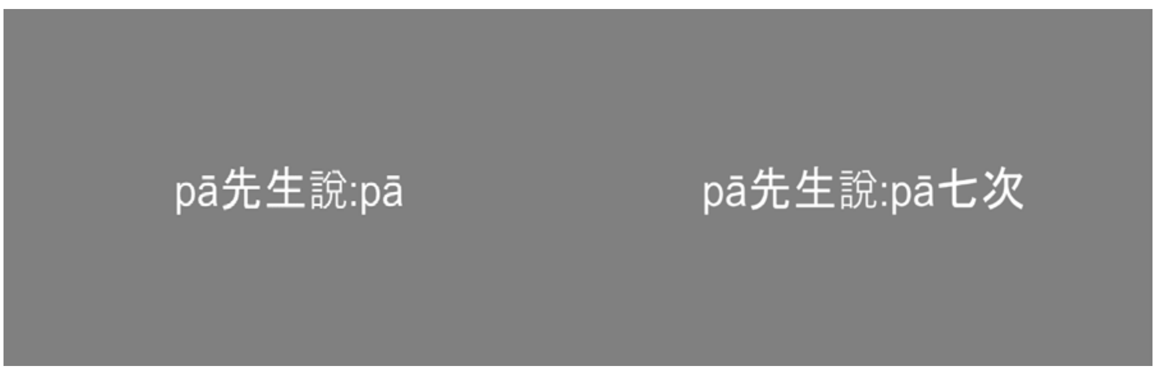

Figure 4. Screenshot examples of the visual presentation of stimulus sentence without (on the left) and with (on the right) the classifier (qīcì). The second stimulus without the classifier creates the utterance-*final* instance, and the utterance-*medial* instance is on the right.

The main experiment asked participants to read the sentences at different speeds to investigate the possible effect of speech rate on VOT productions. The main task, therefore, contained three blocks. The first block required participants to read the sentences at their most comfortable speech rate (*natural*). The second block was to speak at a noticeably slower speed than his or her regular speech rate (*slow*), and the third block was to speak as *fast* as s/he could. A short self-paced break was allowed in between blocks. The entire procedure went as follows.

1. Initial greeting and a minute of casual chatting.
2. Reading the consent form and the recording instructions in Mandarin (Appendix A).
3. Repeating the instructions orally in Mandarin by the experimenter to confirm the participants' understanding of the task.
4. Starting the training phase. This phase asked the participants to practice the task format with the training stimuli to familiarize themselves with the task; the instructions were the same as the main task. The training stimuli consist of /s, l, n/ as word onsets in four tones, rather than /p, t, k/ to avoid possible effects of training phase stimuli on the main experiment.
5. Beginning the main experiment phases.

Data collection was done using an SSD laptop (Lenovo 110S) with a USB-connected Zoom Handy H2 recorder. All recordings were made using Audacity (version 2.1.3 of Audacity®, Audacity Team), stored as WAV files (44.100 kHz, 16 bit, mono). During the tasks, the participants were seated in front of the computer and the microphone. The experimenter was seated on the left of the computer, controlling the keyboard. If the subject's utterance was perceptually unclear to the experimenter, the participant was asked to repeat the sentence before moving on to the subsequent trial (there were several requests for repeating, but it was not kept for a record).

The speech samples were counterbalanced to avoid any order effects; stimulus presentation was also randomized for each participant. Participants could ask questions or stop the experiment at any time, although none of the participants did so. The participants filled out the questionnaire after Experiment 2a (see Chapter 4). Both Experiment 1a and Experiment 2a together lasted about 20 minutes for the native Mandarin speakers. All participants were paid $250 NTW dollars (equivalent to seven US dollars).

Regardless of the recording sites, the recording equipment and the procedures were the same. Overall, 9820 stimulus tokens were collected from 68 native Mandarin speakers for further examination and evaluation.

3.5 Measurements

VOT values were obtained following previous studies (Liu et al., 2008; Lai et al., 2009; Chen et al., 2009; Lam, 2010; Tseng, 2018). Traditionally, VOT, the delay of voicing, is defined as the duration from the release of the stop to the start of vocal fold oscillation for the vowel. When measuring VOT values, each stop consonant has three possible realizations: positive VOT, negative VOT, and zero VOT. This dissertation focuses on the positive VOT of the aspirated plosives /p, t, k/ of Mandarin.

Each stimulus response was digitally analyzed using Praat (Boersma & Weenink, 2018). The VOT and the following vowel duration values for the stops were labeled and measured. The vowel duration, which is a covariant to the speech rate, was exclusively used to test a correlation using Pearson’s correlation coefficient in Experiment 1a, discussed in section 7.2.4. For the statistical analysis for VOT's temporal effect, a categorical variable, the speech rate, was used.

Figure 5 below shows the waveform and spectrogram of an example and the labeling following the practice of previous studies (Lou, 2018; Kwon, 2015; Ding et al., 2010; Chen et al., 2009; Liu et al., 2008; Chen et al., 1998; Silva, 1992). Four acoustic measurements for this study are defined below:

1. Voice onset time (VOT): the duration in millisecond (ms) from A to B in Figure 5: the stop release burst of the articulator to the beginning of the visible voice bar in the spectrogram and/or the first vertical spike in the waveform
2. Fundamental frequency (F0) onset before the vowel: the F0 at point B plus/minus 5 ms where frequency in Hertz (Hz) can be detected/measured
3. Post-stop vowel duration (vLength) from B to C: the start of aperiodic energy to the end of vowel's periodic energy
4. Vowel-end and F0 offset: the ending of the vowel's periodic energy or F0 offset at point C.

Following these criteria, each stimulus was manually edited to label for A, B, and C points one by one on the TextGrid file. In this dissertation, the VOT value is determined from point A to B, and the vowel duration is from B to C in Figure 5. Acoustic measurement proceeds in this manner for all productions of the native speakers of Mandarin.

Several Praat scripts were used to extract the aforementioned acoustic measurements automatically. These values were entered into a spreadsheet alongside phonological variables and demographic information. Finally, 9722 data points[5] were imported into R (R Development Core Team, 2014) for the statistical analysis.

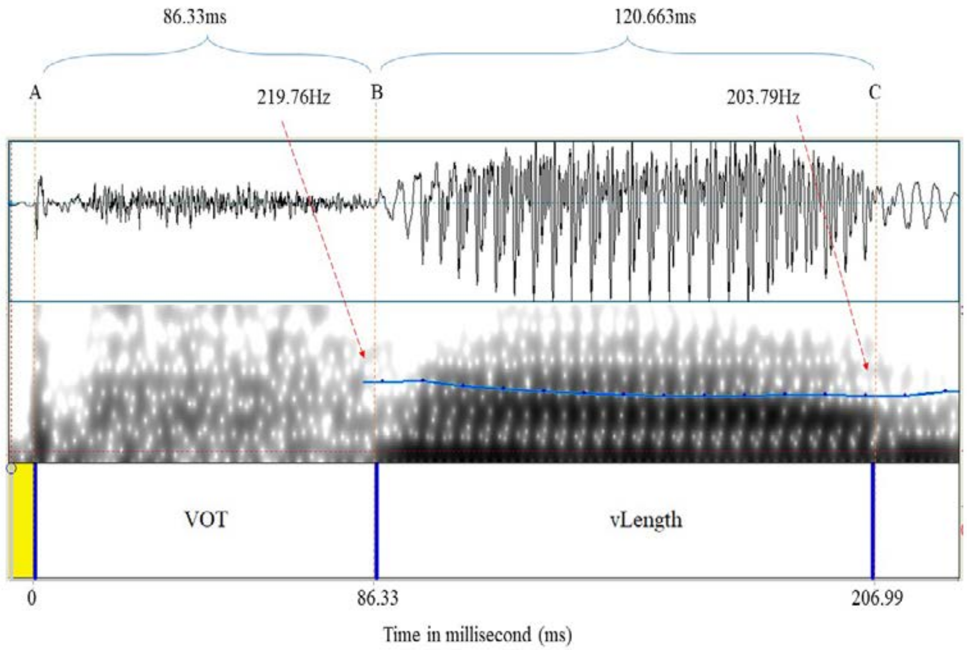

Figure 5. Waveform and spectrogram of an example of [pʰā] in T1. A is the stop release point; B is the start of voicing (or the aperiodic energy in the waveform); C is the end of F2 of the vowel (or the end of the periodic energy). The blue line depicts the overall pitch contour.

3.6 Descriptive Statistics Overview

This section provides descriptive and inferential statistics of Experiment 1a. The summaries of each data point of the measurements are provided.

VOT was measured in msec, and the categorical variables of tone, POA, speech rate, phrase-position, and gender were dummy-coded accordingly. The final set of the data's relevant acoustic properties were measured from each stimulus, and the results

[5] 98 (or about 1%) stimuli were excluded due to mispronunciation, undetectable boundaries in the waveform/spectrogram, and/or overlapped segments

collapsed across variables were graphed in the figures below. The data reported here were then analyzed for their significance using a series of mixed-effect models discussed in sections 3.7 and 3.8.

3.6.1 Overall VOT in Lexical Tones

Figure 6 shows the overall VOT values by tone and by the POA. As can be seen from the figure, T3 had the longest average VOT value at 92.32 msec (SD = 26.42), T2 the second at 87.32 msec (SD = 24.51), T4 the third at 80.32 msec (SD = 22.45), and T1 the shortest at 79.22 msec (SD = 22.15). Descriptively, VOTs in T3 and T2 were similar, and so were in T4 and T1.

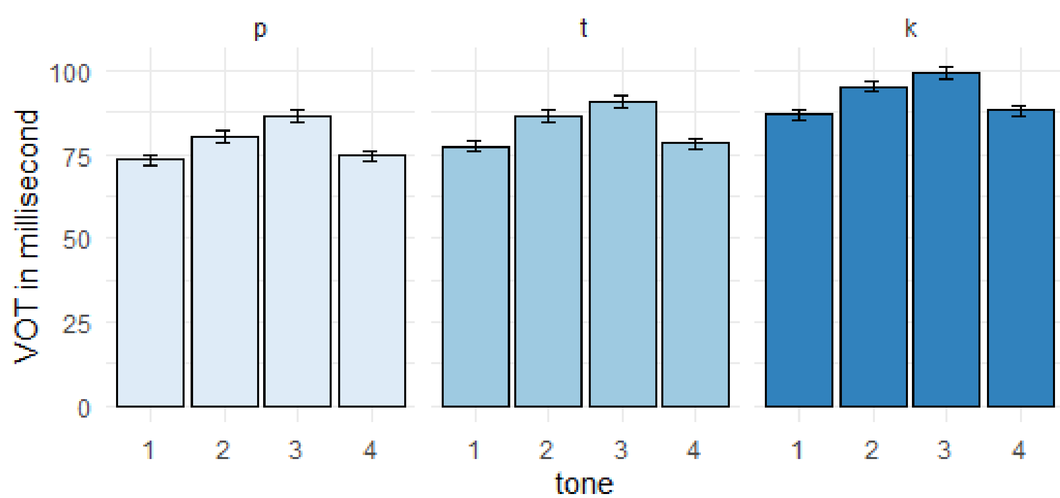

Figure 6. Overall VOT values by the lexical tone and by POA

3.6.2 Overall VOT in Places of Articulation

The overall VOT values by POA were <u>78.68 msec</u> (SD = 22.84) for /p/, <u>83.28 msec</u> (SD = 23.79) for /t/, and <u>92.42 msec</u> (SD = 24.88) for /k/. As shown in Figure 7, a consistent POA effect where the further back the place of articulation, the longer the VOT value was observed in each tone. Although the values were different between POAs, VOTs in T3 and T2 were still longer than those in T4 and T1 were.

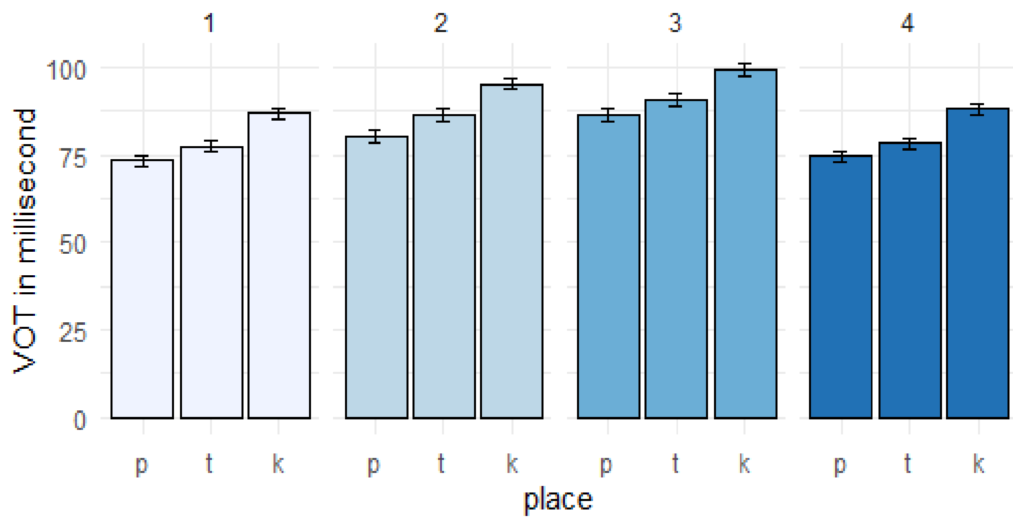

Figure 7. Overall VOT values by the POA and by the lexical tone

3.6.3 Overall VOT in Speech Rates

Figure 8 summarizes the average VOT values produced in three speech rates. Overall, the values by speech rate were <u>75.38 msec</u> (SD = 20.57), <u>84.04 msec</u> (SD = 22.65), and <u>95.11 msec</u> (SD = 25.99) for *fast*, *natural*, and *slow*, respectively. The general data showed that the faster the speech rate, the shorter the VOT.

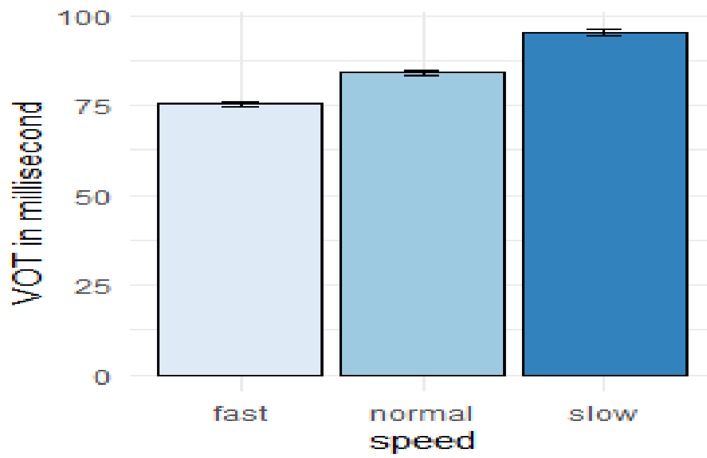

Figure 8. Overall VOT values in three speech rates

3.6.4 Overall VOT in Phrase-position

Figure 9 shows the overall VOT values in three phrase-positions. The mean VOT values were 83.86 msec (SD = 24.56), 84.40 msec (SD = 23.17) and 87.06 msec (SD = 25.63) at the utterance-*initial*, utterance-*medial*, and utterance-*final* positions. The descriptive data showed that the average values were similar in three contexts.

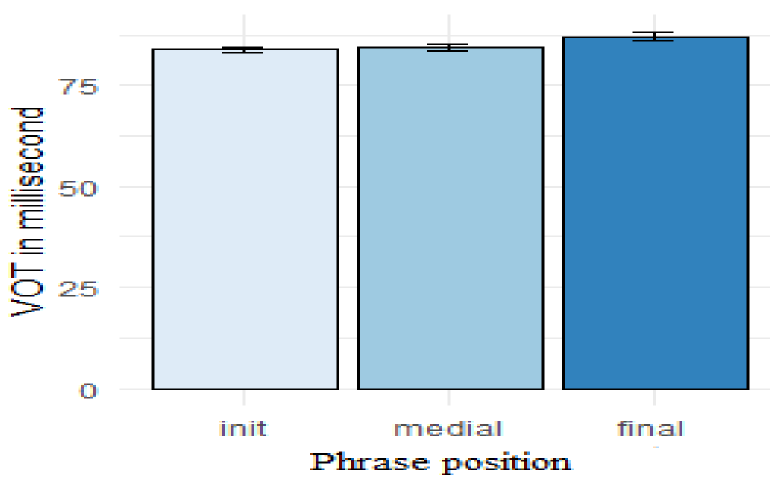

Figure 9. Overall VOT values by three phrase-positions

3.6.5 Overall VOT in Genders

Figure 10 shows the overall VOT by gender. The average values were 82.47 msec (SD = 24.27) for the males and 86.89 msec (SD = 24.57) for the female participants. Descriptively, females produced slightly longer VOT values than the males did from the given data.

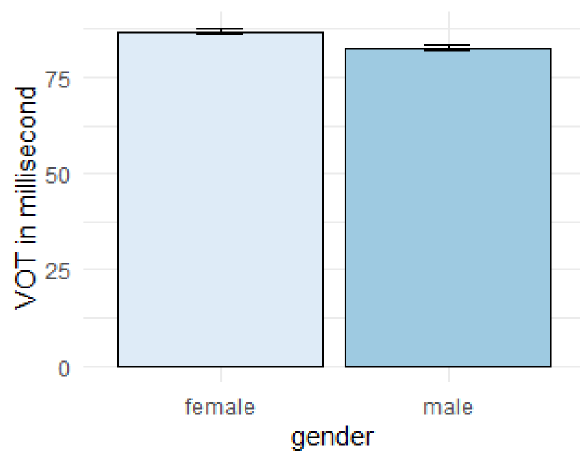

Figure 10. Overall VOT values by gender

3.7 The Statistical Model

Linear mixed-effects models were used to investigate the possible effects of tone, POA, speech rate, phrase-position, and gender on VOT duration. These five were the independent variables, and VOT was the dependent variable. The fitting of models was done following Winter (2013, 2015) using Likelihood ratio tests.

Before fitting the final model, a pre-test was conducted to test each fixed effect's predictive power. A null model (with only random effects of speaker-subject and worditem and no fixed effect) and five single mixed-effect models with each independent variable were built and compared. The results of the Likelihood comparison of each fixed effect are provided in Table 3.

As can be seen from the table, all mixed-effects under investigation were strong predictors, except for gender. All five fixed effects were incorporated, and the final model converged without any error. Full interactions of the four significant fixed effects were added one by one for as long as the model converged without an error message to keep the model maximal as permitted by the data. The interaction between gender and the other four significant predictors was only added up to two-way; adding the three-way erred for "*model matrix is rank deficient*," which means the model was over-fitted or the current data was deficient for the built.

Table 3. Results of Likelihood Ratio Tests of the fixed effects

Tone:	χ^2 = 17.721, p = 0.0005022 ***
POA:	χ^2 = 20.645, p = 3.288e-05 ***
Speed:	χ^2 = 49.853, p = 1.495e-11 ***
Phrase-position:	χ^2 = 37.096, p = 8.807e-09 ***
Gender:	χ^2 = 2.049, p = 0.1523

3.8 Statistical Modeling & Result Overview

This section provides statistical modeling approaches as a whole and an overview of the results. Detailed discussions of the results and tone effect are provided in the discussion section.

3.8.1 Statistical modeling

The current study used a series of mixed-effects linear regression models, implemented in the *lmerTest* package (Kuznetsova et al., 2017) in R (R Development Core Team, 2014) to test the change of VOT values as a function of lexical tone and other independent variables. Full interactions of the tone, POA, speech rates, and phrase-position fixed effects and two-way interactions of gender and the four main effects were also included. All fixed factors were coded using R default treatment (dummy) coding. The reference level for the intercept was set automatically to T1 for tone, /k/ for POA, *final* for phrase-position, *fast* for speech rate, and *female* for gender. Random intercepts and random slopes for items and subjects were included as maximally as permitted by the data (Barr, Levy, Scheepers, & Tily, 2013).

Thus, the final model for Experiment 1a included five independent variables, full interactions between four of the fixed effects, up to two levels of interactions between gender and the other effects, and the by-subject adjustments to the random slopes of tone, POA, and phrase-position, as well as random intercepts for by-subjects and by-items. The final model converged without any error messages. Pairwise comparisons were conducted using Tukey's HSD tests implemented in the *emmeans* (Lenth, 2020) package.

3.8.2 Result Overview

A summary of the Type III ANOVA test results' main effects with Satterthwaite's method is provided in Table 4. The results indicate that four of the five mixed-effects included had a significant influence on VOT. Most two-way interactions were significant, and none of the three- and four-way interactions was significant. Each factors' post-hoc results and significant interactions between factors of the native Mandarin are provided in the following sections.

Table 4. Summary of Type III ANOVA table with Satterthwaite's method of the mixed-effects and interactions between effects for the Taiwan group.

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	p.value
Tone (T)	43668	14556	3	72.2	57.8433	< 2.2e-16 ***
POA (P)	77172	38586	2	72.4	153.3361	< 2.2e-16 ***
Speed (S)	601290	300645	2	9184.0	1194.7244	< 2.2e-16 ***
Phrase Pos. (Pos)	2479	1239	2	68.0	4.9253	0.0100565 *
Gender (G)	746	746	1	68.0	2.9657	0.0895918 ˙
T:P	1380	230	6	9178.4	0.9139	0.4835052
T:S	4506	751	6	9190.9	2.9844	0.0065061 **
P:S	3907	977	4	9191.6	3.8815	0.0037445 **
T:Pos	9504	1584	6	9179.4	6.2948	1.286e-06 ***
P:Pos	4310	1077	4	9179.4	4.2817	0.0018380 **
S:Pos	5579	1395	4	9191.5	5.5422	0.0001876 ***
T:G	1124	375	3	67.7	1.4890	0.2253657
P:G	1112	556	2	68.1	2.2093	0.1175849
S:G	2321	1161	2	9181.5	4.6119	0.0099561 **
Pos:G	2486	1243	2	68.0	4.9398	0.0099297 **
T:P:S	3352	279	12	9178.4	1.1100	0.3463503
T:P:Pos	1551	129	12	9178.3	0.5137	0.9075131
T:S:Pos	1715	143	12	9179.4	0.5681	0.8694086
P:S:Pos	1562	195	8	9179.4	0.7761	0.6239072
T:P:S:Pos	3879	162	24	9178.3	0.6423	0.9078326

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

3.9 Results of the Native Taiwan group

The statistical results confirmed the expectation that three of the main effects strongly affect VOT in native Mandarin. The results showed that VOTs were found to be affected by tone (F (3, 72) = 57.843, p < 0.0001), as well as by POA (F (2, 72)

=153.336, $p < 0.0001$) and by speech rate (F (2, 9184) = 1194.72, $p < 0.0001$). The effect of phrase-position reached the statistical significance (F (2, 68) = 4.925, $p = 0.01005$), and gender did not (F (1, 68) = 2.965, $p = 0.0896$). These outcomes of tone and POA effects were in accord with most results reported in the literature in L1 Mandarin (Liu et al., 2008; Chen et al., 2009; Tseng, 2018). Detailed post-hoc pairwise comparison results are provided, and significant interactions between factors are discussed below in each subsection corresponding to each main factor.

3.9.1 Lexical Tone Effect in the Taiwan group

As expected, the model revealed a significant main effect of tone. The mean VOT values were 92.6 msec in T3 (SE = 1.87, DF = 68, lower.CL = 88.9, upper.CL = 96.4), 87.8 msec in T2 (SE = 1.66, DF = 68, lower.CL = 84.5, upper.CL = 91.1), 80.4 msec in T4 (SE = 1.52, DF = 68, lower.CL = 77.4, upper.CL = 83.4), and 79.2 msec in T1 (SE = 1.54, DF = 68, lower.CL = 76.2, upper.CL = 82.3). VOT values in T3 and T2 were significantly longer than those in T4 and T1 were.

The statistical model also revealed significant interactions between tone and speech rate (F (6, 9190) = 2.984, $p = 0.0065$) and tone and phrase-position (F (6, 9179) = 6.195, $p < 0.0001$). The clustered bar graph below shows general patterns of each tone across speech rates and phrase-positions (Figure 11). VOTs were longer in T3 and T2 across all three speech rates than those in T1 and T4 were. VOTs in T2 were significantly shorter than those in T3 in *slow* and *natural* speech rates; their difference was only marginal in the *fast* speech rate. All T2 and T3 were significantly longer than those in T1 and T4 were across three speech rates. No T1-T4 difference was significant across three speech rates

Moreover, the model revealed a significant interaction between tone and phrase-position. VOTs in T3 and T2 were significantly longer than those in T1 and T4 were in all conditions. T2-T3 differences were found significant in phrase-*initial*, phrase-*medial*, and phrase-*final*. T1-T4 differences were not significant across all three-phrase-positions. Table 5 provides detailed statistics of interactions and the significant differences of VOT in T2 and T3 in speech rates and phrase-positions.

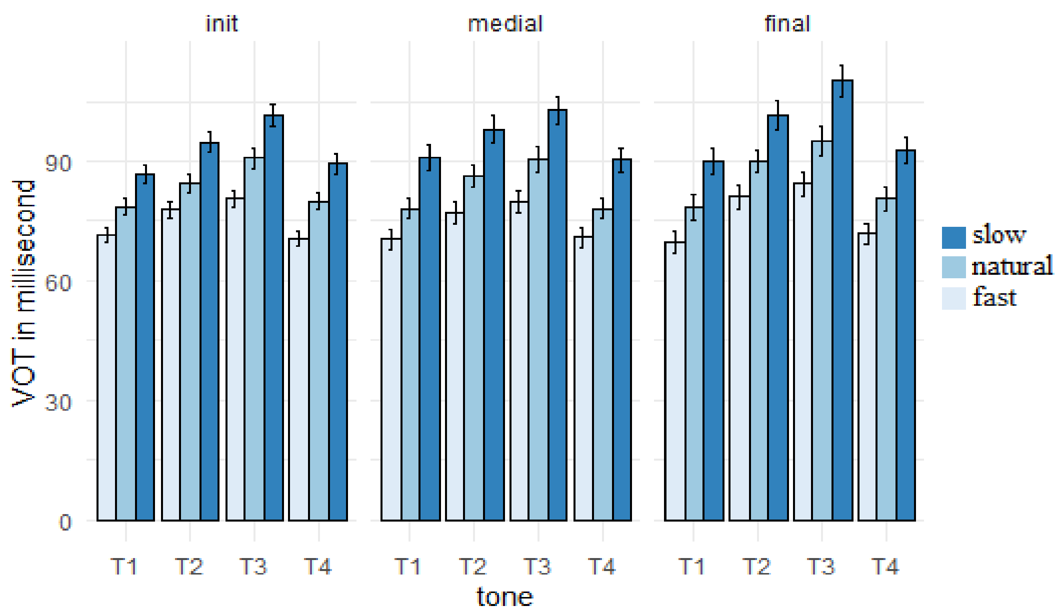

Figure 11. VOT patterns due to tone effect across speech rates and phrase-positions

As shown, these outcomes confirm a significant tone effect on VOT, where VOTs in T3 and T2 were significantly longer than those in T1 and T4 were regardless of the conditions of speech rate and phrase-position.

Table 5. Tukey HSD Post-hoc Analyses for significant interaction results between tones & speech rates and tones & phrase-positions for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
Tone : Speech Rate						
at *fast*						
T1 - T2	-8.26	1.05	209	-7.897	< .0001	***
T1 - T3	-11.03	1.30	131	-8.463	< .0001	***
T1 - T4	-0.68	1.07	199	-0.634	0.9209	
T2 - T3	-2.76	1.12	178	-2.469	0.0684	.
T2 - T4	7.58	1.23	144	6.145	< .0001	***
T3 - T4	10.35	1.30	132	7.973	< .0001	***
at *natural*						
T1 - T2	-8.51	1.05	211	-8.111	< .0001	***
T1 - T3	-13.64	1.31	132	-10.446	< .0001	***
T1 - T4	-1.17	1.08	200	-1.088	0.6973	
T2 - T3	-5.13	1.12	179	-4.569	0.0001	***
T2 - T4	7.34	1.24	145	5.937	< .0001	***
T3 - T4	12.47	1.30	133	9.592	< .0.001	***
at *slow*						
T1 - T2	-8.94	1.05	213	-8.495	< .0001	***
T1 - T3	-15.53	1.31	133	-11.871	< .0001	***
T1 - T4	-1.64	1.08	202	-1.524	0.4253	
T2 - T3	-6.59	1.13	181	-5.858	0.0001	***
T2 - T4	7.29	1.24	146	5.885	< .0001	***
T3 - T4	13.89	1.30	134	10.658	< .0001	***
Tone : Phrase-position						
in *initial*						
T1 - T2	-6.6875	0.908	119	-7.367	< .0001	***
T1 - T3	-11.945	1.195	93	-9.997	< .0001	***
T1 - T4	-1.0822	0.938	117	-1.153	0.6573	
T2 - T3	-5.2570	0.992	109	-5.301	< .0001	***
T2 - T4	5.6053	1.119	97	5.009	< .0001	***
T3 - T4	10.8623	1.189	93	9.137	<.0.001	***
in *medial*						
T1 - T2	-7.4405	1.112	266	-6.693	< .0001	***
T1 - T3	-11.177	1.356	154	-8.242	< .0001	***
T1 - T4	-0.0142	1.137	250	-0.012	1.0000	
T2 - T3	-3.7366	1.181	220	-3.163	0.0096	***
T2 - T4	7.4263	1.290	172	5.757	< .0001	***
T3 - T4	11.1629	1.351	155	8.262	<.0.001	***
in *final*						
T1 - T2	-11.579	1.114	267	-10.395	< .0001	***
T1 - T3	-17.071	1.358	155	-12.570	< .0001	***

T1 - T4	-2.3972	1.139	251	-2.105	0.1542	
T2 - T3	-5.4923	1.183	221	-4.641	< .0001	***
T2 - T4	9.1817	1.292	172	7.108	< .0001	***
T3 - T4	14.674	1.353	156	10.848	<.0.001	***

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

3.9.2 POA Effect in the Taiwan group

As expected, the universal effect of the place of articulation on VOT was shown in native Mandarin production. Their VOTs in / p, t, k / were 79.0 msec (SE = 1.48, DF = 68, lower.CL = 76.1, upper.CL = 82.0), 83.6 msec (SE = 1.57, DF = 68, lower.CL = 80.5, upper.CL = 86.8), and 92.4 msec (SE = 1.75, DF = 68, lower.CL = 88.9, upper.CL = 95.9), respectively. The model revealed two significant two-way interactions between POA & speech rate (F (4, 9192) = 3.8815, p = 0.0037) and POA & phrase-position (F (4, 9179 = 4.2817, p = 0.0018). Figure 12 shows the general VOT patterns due to POA effect across speech rates and phrase-positions.

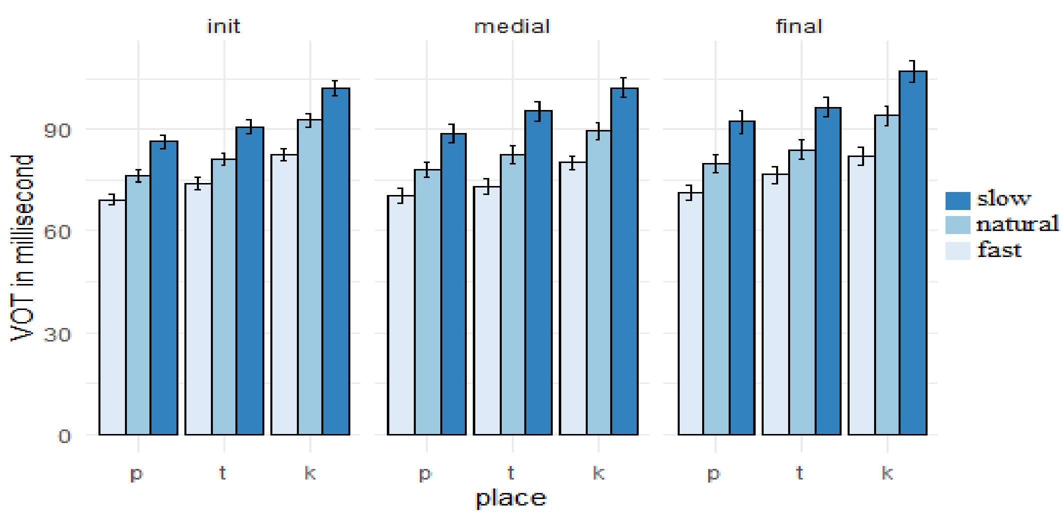

Figure 12. VOT patterns due to POA effect across speech rates and phrase-positions

As shown, at all three speech rates, VOTs in /k/ were significantly longer than those in /t/, which were also significantly longer than those in /p/. Likewise, across sentence-*initial*, -*medial*, and -*final* positions, VOTs in /k/ were significantly longer than those in /t/, which in turn were also significantly longer than those in /p/. In other words, regardless of the speech rate or phrase-position, the POA effect on VOT in Mandarin was confirmed. Table 6 provides detailed statistics of the interactions.

Table 6. Tukey HSD Post-hoc Analyses for significant interaction results between POA & speech rates and POA & phrase-positions for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
POA : Speech Rate						
at *fast*						
/k/ - /p/	11.26	0.968	178	11.626	<.0001	***
/k/ - /t/	6.99	0.964	180	7.251	<.0001	***
/p/ - /t/	-4.27	0.945	189	-4.514	<.0001	***
at *natural*						
/k/ - /p/	14.08	0.970	180	14.507	<.0001	***
/k/ - /t/	9.51	0.966	182	9.841	<.0001	***
/p/ - /t/	-4.57	0.947	191	-4.827	<.0001	***
at *slow*						
/k/ - /p/	14.80	0.973	181	15.203	<.0001	***
/k/ - /t/	9.81	0.969	183	10.120	<.0001	***
/p/ - /t/	-4.99	0.950	192	-5.249	<.0001	***
POA : Phrase-position						
in *initial*						
/k/ - /p/	15.13	0.857	110	17.658	<.0001	***
/k/ - /t/	10.47	0.852	110	12.281	<.0001	***
/p/ - /t/	-4.67	0.831	113	-5.617	<.0001	***
in *medial*						
/k/ - /p/	11.77	1.021	220	11.527	<.0001	***
/k/ - /t/	7.08	1.018	223	6.954	<.0001	***
/p/ - /t/	-4.70	1.000	236	-4.701	<.0001	***
in *final*						
/k/ - /p/	13.22	1.023	221	12.920	<.0001	***
/k/ - /t/	8.76	1.019	224	8.596	<.0001	***
/p/ - /t/	-4.46	1.001	238	-4.454	<.0001	***

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

3.9.3 Speech Rate Effect in the Taiwan group

As expected, the speech rate was also found to have a strong effect on VOT. VOT values in the *slow* speech rate were much longer than in the *fast* speech rate, about 20 milliseconds longer. The mean VOT values were 75.3 msec (SE = 1.56, DF = 71, lower.CL = 72.2, upper.CL = 78.4), 84.1 msec (SE = 1.56, DF = 71, lower.CL = 81.0, upper.CL = 87.2), and 95.7 msec (SE = 1.56, DF = 71, lower.CL = 92.5, upper.CL = 98.8), respectively.

In addition to tone & speech rate and POA & speech rate interactions discussed above, two more significant two-way interactions involved speech rate were found between speech rate & phrase-position (F (4, 9192) = 5.5422, p = 0.00019), and speech rate & gender (F (2, 9182) = 4.6119, p = 0.010). Figure 13 shows the general VOT patterns due to the speech rate effect across phrase-positions and genders.

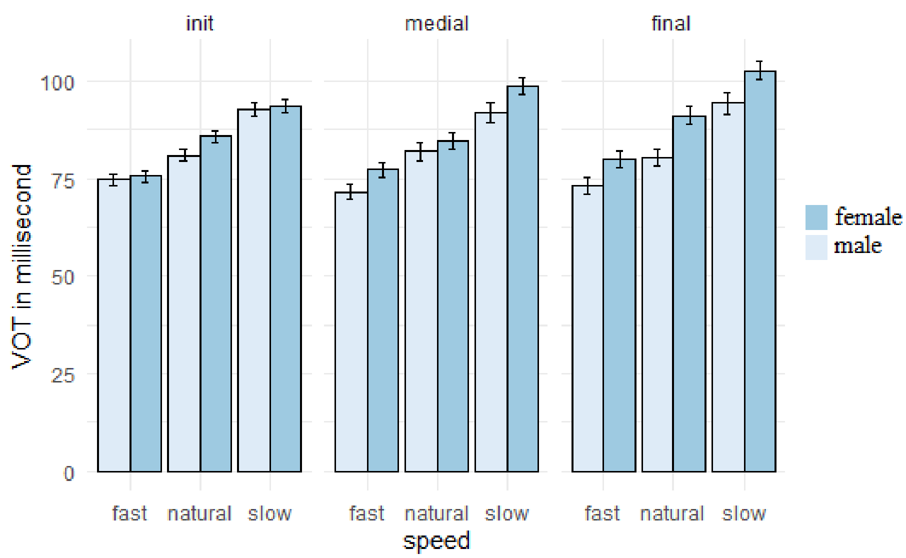

Figure 13. VOT patterns due to speech rate effect across phrase-positions and genders

As can be seen, the results showed that as the speakers' speech rate increased, VOT values significantly decreased across all factors. This indicates a strong speech rate effect on VOT in Mandarin regardless of conditions and genders; detailed statistics for each of these four two-way interactions are provided in Table 7 below.

Table 7. Post-hoc Analyses for significant interaction results between speech rates & tone, speech rate & POA, speech rate & phrase-position, and phrase-positions & gender for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
Speech Rate : Tone						
in T1						
fast - natural	-7.93	0.830	9185	-9.553	<.0001	***
fast - slow	-18.78	0.832	9200	-22.569	<.0001	***
natural - slow	-10.85	0.833	9185	-13.032	<.0001	***
in T2						
fast - natural	-8.17	0.830	1987	-9.848	<.0001	***
fast - slow	-19.46	0.832	9200	-23.380	<.0001	***
natural - slow	-11.28	0.833	9184	-13.547	<.0001	***
in T3						
fast - natural	-10.54	0.830	9185	-12.694	<.0001	***
fast - slow	-23.29	0.833	9196	-27.972	<.0001	***
natural - slow	-12.75	0.833	9183	-15.305	<.0001	***
in T4						
fast - natural	-8.42	0.829	9182	-10.151	<.0001	***
fast - slow	-19.75	0.832	9195	-23.725	<.0001	***
natural - slow	-11.33	0.833	9186	-13.604	<.0001	***
Speech Rate : POA						
in /p/						
fast - natural	-7.72	0.719	9187	-10.741	<.0001	***
fast - slow	-18.90	0.721	9200	-26.215	<.0001	***
natural - slow	-11.18	0.721	9185	-15.491	<.0001	***
in /t/						
fast - natural	-8.03	0.719	9183	-11.173	<.0001	***
fast - slow	-19.62	0.721	9195	-27.215	<.0001	***
natural - slow	-11.59	0.721	9186	-16.073	<.0001	***
in /k/						
fast - natural	-10.54	0.719	9186	-14.665	<.0001	***
fast - slow	-22.44	0.721	9197	-31.120	<.0001	***
natural - slow	-11.89	0.721	9184	-16.485	<.0001	***
Speech Rate : Phrase-position						

in *initial*						
fast - natural	-8.28	0.557	9180	-14.870	<.0001	***
fast - slow	-18.03	0.559	9188	-32.262	<.0001	***
natural - slow	-9.75	0.559	9181	-17.440	<.0001	***
in *medial*						
fast - natural	-8.68	0.786	9178	-11.043	<.0001	***
fast - slow	-20.85	0.790	9191	-26.408	<.0001	***
natural - slow	-12.18	0.790	9191	-15.419	<.0001	***
in *final*						
fast - natural	-9.34	0.790	9201	-11.821	<.0001	***
fast - slow	-22.08	0.790	9204	-27.931	<.0001	***
natural - slow	-12.74	0.791	9179	-16.101	<.0001	***
Speech Rate : Gender						
in *female*						
fast - natural	-9.93	0.558	9191	-17.803	<.0001	***
fast - slow	-20.68	0.562	9186	-36.817	<.0001	***
natural - slow	-10.75	0.562	9184	-19.129	<.0001	***
in *male*						
fast - natural	-7.60	0.587	9179	-12.934	<.0001	***
fast - slow	-19.96	0.587	9179	-33.982	<.0001	***
natural - slow	-12.36	0.587	9179	-21.049	<.0001	***

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

3.9.4 Phrase-position Effect in the Taiwan group

The mean VOTs by phrase-position were: 83.9 msec at the utterance-*initial* (SE = 1.69, DF = 68, lower.CL = 80.5, upper.CL = 87.2), 84.3 msec at the utterance-*medial* (SE = 1.57, DF = 68, lower.CL = 81.2, upper.CL = 87.4), and 86.9 msec at the utterance-*final* (SE = 1.68, DF = 68, lower.CL= 83.6, upper.CL = 90.3).

The VOT values were similar in different positions, but the main effect of phrase-position was significance (F (2, 68) = 4.9253, p = 0.01). The model also revealed four two-way interactions involved this effect: phrase-position & tone (F (6, 9179) = 6.2948, p < 0.0001), phrase-position & POA (F (4, 9179) = 4.2817, p = 0.0018), phrase-position & speech rate (F (4, 9192) = 5.5422, p = 0.0002), and phrase-position & gender

(*F* (2, 68) = 4.9398, *p* = 0.0099). The results showed that phrase-position effect was only significant in T2 and T3 between *final-initial* and *final-medial* positions. It only had a significant effect on VOT if the stimuli were /k/ between *final-medial*, /t/ between *final-initial*, and /p/ between *final-initial*. Concerning speech rate, the only significant phrase-position effect was at the *slow* speed between *final-initial* and *final-medial*. Lastly, it was only significant in *female* between *final-initial* and *final-medial*. See Table 8 for detailed statistics.

Lisker and Abramson's (1972) results, where VOT value in the utterance-*final* position is significantly greater than utterance-*medial* position, were only found in some conditions, not all in this study.

Table 8. Tukey HSD Post-hoc Analyses for significant interaction results between phrase-position & tone, phrase-position & POA, phrase-position & speech rate, and phrase-position & gender

Contrast	estimate	SE	df	t.ratio	p.value	
Phrase-position : Tone						
in T1						
final - initial	0.2201	1.24	141	0.178	0.9827	
final - medial	-0.4702	1.29	173	-0.365	0.9293	
initial - medial	-0.6903	1.15	160	-0.599	0.8211	
in T2						
final - initial	5.1114	1.24	141	4.132	0.0002	***
final - medial	3.6682	1.29	173	2.847	0.0136	*
initial - medial	-1.4433	1.15	160	-1.252	0.4247	
in T3						
final - initial	5.3467	1.24	141	4.322	0.0001	***
final - medial	5.4239	1.29	173	4.210	0.0001	***
initial - medial	0.0771	1.15	160	0.067	0.9975	
in T4						
final - initial	1.5351	1.24	141	1.241	0.4309	
final - medial	1.9128	1.29	173	1.485	0.3005	
initial - medial	0.3777	1.15	160	0.328	0.9426	
Phrase-position : POA						
in /p/						

final - initial	3.76	1.17	114	3.207	0.0049	**
final - medial	2.23	1.21	133	1.851	0.1571	
initial - medial	-1.53	1.08	125	-1.410	0.3387	
in /t/						
final - initial	3.55	1.17	114	3.031	0.0084	**
final - medial	1.99	1.21	133	1.653	0.2272	
initial - medial	-1.56	1.08	125	-1.440	0.3235	
in /k/						
final - initial	1.85	1.17	114	1.575	0.2605	
final - medial	3.68	1.21	133	3.051	0.0077	**
initial - medial	1.83	1.08	125	1.689	0.2135	
Phrase-position : Speech Rate						
at *fast*						
final - initial	1.353	1.17	113	1.156	0.4814	
final - medial	2.006	1.20	132	1.667	0.2216	
initial - medial	0.653	1.08	124	0.603	0.8186	
at *natural*						
final - initial	2.408	1.17	114	2.053	0.1045	
final - medial	2.666	1.21	133	2.211	0.0729	·
initial - medial	0.258	1.08	125	0.238	0.9693	
at *slow*						
final - initial	5.399	1.17	115	4.596	<.0001	***
final - medial	3.229	1.21	134	2.673	0.0229	*
initial - medial	-2.170	1.09	126	-1.996	0.1172	
Phrase-position : Gender						
in *female*						
final - initial	6.294	1.42	69	4.443	0.0001	***
final - medial	4.471	1.40	68	3.194	0.0060	**
initial - medial	-1.823	1.28	68	-1.428	0.3322	
in *male*						
final - initial	-0.187	1.50	68	-0.125	0.9914	
final - medial	0.797	1.48	68	0.538	0.8530	
initial - medial	0.984	1.35	68	0.728	0.7479	

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

3.9.5 Gender Effects in the Taiwan group

Finally, for the gender factor, the analysis showed no significant, but marginal difference, between females and males (F (1, 68) = 2.9657, p = 0.0896. The mean VOTs were 82.4 msec for males (SE = 2.25, DF = 68, lower.CL = 77.9, upper.CL = 86.9) and 87.7 msec for females (SE = 2.12, DF = 68, lower.CL= 83.5, upper.CL = 91.9).

As discussed earlier, the gender factor involved two two-way interactions with speech rate and phrase-position. From the given data, females had longer VOTs than males only at the *natural* speech rate and in the utterance-*final* position (see Table 9 for details).

Table 9. Tukey HSD Post-hoc Analyses for significant interaction results between gender & speech rates and gender & phrase-positions for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
Gender : Speech Rate						
at *fast*						
female - male	4.30	3.12	71	1.377	0.1728	
at *natural*						
female - male	6.64	3.12	71	2.126	0.0370	*
at *slow*						
female - male	5.02	3.12	71	1.607	0.1124	
Gender : Phrase-position						
in *initial*						
female - male	2.22	3.38	68	0.656	0.5137	
in *medial*						
female - male	5.03	3.14	68	1.604	0.1134	
in *final*						
female - male	8.70	3.36	68	2.591	0.0117	*

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

3.10 Result Summary & Preliminary Discussion

VOT values of the Taiwan Mandarin speakers were analyzed to explore the relationships between VOT variations as functions of tone, POA, speech rate, phrase-position, and gender in native Mandarin. It was found that tone affected VOT values in the L1 speech, where VOTs in T3 and T2 were significantly longer than those in T1 and

T4 across POAs, speech rates, and phrase-positions. VOT differences between T1 and T4 were all non-significant, so did those between T2 and T3, but T2 and T3 differences became significant in some speech rate and phrase-position conditions.

The POA and speech rate effects were shown to be strong factors across the board. VOTs were significantly longer in /k/ than those in /t/ and /p/, and those in *slow* speech rate were significantly shorter than those in the *natural* and *fast* speech rates. The effect of phrase position was shown to be a significant factor but in limited conditions. VOTs in words in utterance-*final* positions were more extended than in the utterance-*initial* positions. Lastly, the gender main effect was non-significant, but females did have significantly longer VOTs than males only at a *natural* speech rate.

In agreement with previous studies (Liu et al. 2008; Chen et al., 2009; Tseng, 2018), this study's general finding confirms a lexical tone effect on VOT. It provides supportive evidence of the effects of place-dependent and temporal-dependent VOT variations. There was no concrete evidence to support a phrase-position main effect on VOT, and neither was gender a significant main effect from the given data.

This experiment shows that VOT values were longer in T3 and T2 than in T4 and T1 in native Taiwan Mandarin within each factor's domain. That is, with other things being equal, VOT values varied between lexical tones. The reason may be the vocal fold tensity (Eshghi et al., 2016; Steven, 1999; Cho & Ladefoged, 1999).

Mandarin lexical tone is generally the manipulation of the pitch, which is the tension change of the vocal folds controlled by the cricoarytenoid muscles. In manipulating pitch for producing lexical tones, the movements subsequently change the vocal fold tension and shape. For instance, in producing T1 or T4, the muscle of the vocal folds contracts to change the elastic thinness to allow a high pitch onset at adduction; therefore, the quicker the reaction to the Bernoulli Effect can be with all other things being equal. The typical indication of T1 and T4 is the high pitch onset; thus, short VOT, and T2 and T3 exhibit low pitch onset, thus long VOT. This hypothesis successfully explains the results found in Experiment 1a that voicing delay was longer in T2 and T3 than in T1 and T4. To further test the pitch effect, this dissertation designs the second experiment to control tone (see Chapter 4). Chapter 7 gives a detailed overall discussion of the tone and other effects.

Furthermore, recall that one of the current study's hypotheses is that the effects found in L1 also exist in L2 production. From the L1 results, we show that VOT distribution from *long* to *short* markedly differed from: T3 > T2 > T4 > T1 for lexical tones, /k/ > /t/ > /p/ for POAs, and *slow* > *natural* > *fast* for speech rates. To test this hypothesis, this dissertation has conducted the same experiment with three groups of L2 learners of Mandarin: native English, Japanese and Spanish speakers. Chapter 5 provides detailed results of the tone effect on L2 Mandarin speech and the group comparison.

CHAPTER 4. PITCH EFFECT IN NATIVE SPEAKERS – EXPERIMENT 2A

4.1 Introduction

This chapter investigates the relationship between lexical tone and VOT by examining T1's F0 onset in various conditions in the same native Mandarin speakers. In Experiment 1a, we observed significant VOT variations between the T1-T4 pair and the T2-T3 pair. That is, VOTs in the mid-rising and falling-rising tones were significantly longer than those in the high-level and high-falling tones. A follow-up question is what specific component of the tone is responsible for the variations. As grouped above, two common properties of the T1-T4 and T2-T3 pairs are the high starting onset pitch for the former and the rising pitch of the latter at the end of the tones. This experiment focuses on the tone's onset pitch first.

We know from previous studies that F0 and vowel duration are the two primary components of tones (Yang, 2015; Lai et al., 2009; McCrea & Morris, 2005; Sun, 1998; Kessinger & Blumstein, 1997; Massaro et al., 1985; Port & Rotunno, 1979). Experiment 1a looks at how VOT acted in the conditions of lexical tone, place of articulation (POA), speech rate, phrase-position, and gender. Here, Experiment 2a controls the lexical tone and examines how VOT behaves only in T1 in the conditions of pitch register, post-stop vowel duration, as well as confirming POA, gender, and phrase-position. It addresses the relationship between the F0 onset pitch and VOT and compares the VOT difference with respect to post-stop vowel duration.

Recall that vowel duration was treated as a random effect in Experiment 1a since the speaker's speech rate was adopted. In Experiment 2a, the subjects recorded their speech naturally; therefore, the vowel duration is treated as an independent variable. Detailed vowel duration analysis is provided in section 4.7.

The same 68 native Taiwan Mandarin speakers from Experiment 1a were tested. In Experiment 2a, each participant produced a new set of speech samples of 36 combinations of the same stimuli from Experiment 1a in their *natural* speech rate, but at three different pitch registers for acoustic measurement (pitch-level and pitch register are used interchangeably in this dissertation). The stimuli were the same, controlled by using three aspirated voiceless stops (/ p, t, k /) along with the low-central-unrounded vowel, /a/, in T1 in three utterance phrase-positions. The critical difference from Experiment 1a was adding the pitch-level variable and the control of the lexical tone variable. Thus, Experiment 2a's data analysis focuses on pitch-level and considers five primary mixed-effects: POA, pitch-level, phrase-position, vowel duration, and gender to keep the modeling as maximum as possible[6]. Vowel duration was converted and Z-normalized into categorical variables in three levels (see section 4.7).

The outcome revealed a strong effect of pitch on VOT in native Mandarin speech, suggesting that the tone effect found in Experiment 1a is due to lexical tone's F0 onset. The post-stop vowel duration effect on VOT was found non-significant from the given data. With the data collapsed, the findings were that as the pitch (or F0 onset) increased,

[6] Two new independent variables in Experiment 1b are the pitch-level and vowel-duration. Readopting POA, phrase-position, and gender variables is to confirm the inter-experiment consistency of their effects.

VOT values shortened, and the POA effect was consistent as the further back of the POA, the longer VOT values. Slightly differed from Experiment 1a, phrase-position affected VOTs only at the *high* pitch-level between *final* and *initial* utterance position and in /k/ instances. There were no gender and vowel duration main effects, but females had significantly longer VOTs than males only in /p/ and *long* vowel duration conditions. Similarly, there was no vowel duration main effect, but VOTs were longer in the *long* vowel duration than those in the *medium* vowel duration at the *low* pitch register.

This chapter starts with a brief review of the participants' background information, the stimuli, data measurement, and a descriptive overview of the results. It then summarizes the statistical results and provides a preliminary discussion of Experiment 2a for the native Mandarin data.

4.2 Participants (review)

The test subtests for Experiment 2a were the same Taiwan group of participants; i.e., 68 Taiwanese participants self-identified as native Taiwan Mandarin speakers. This Taiwan group had 36 females and 32 males. Their ages ranged from 18 to 68 years old, with an average age of 29.09 (SD = 13.19). They all reported having been born and raised in Taiwan. None of the participants self-reported any known history of a speech disorder or a hearing impairment.

Fifty-four participants recorded their speech production in the National Library of Public Information in Taichung city. Fourteen additional native speakers recorded their speech at three different universities - Fuguang University in Yilan city, Providence

University in Shalu city, and Tunghai University in Taichung city. Regardless of the recording sites, the recording equipment and the procedures were the same (discussed below).

4.3 Speech Samples (review)

Experiment 2a used the same test stimuli and the carrier phrase from Experiment 1a but further controlled the tone effect. Only T1 was used because we wanted to know whether different pitch registers in the same tone would influence VOT. Each participant produced 36 monosyllabic target stimuli made up of three articulation places (/p, t, k /), one open-unrounded vowel (/a/), one lexical tone (T1), three phrase-position conditions (utterance-*initial*, utterance-*medial*, & utterance-*final*) and three pitch registers (*normal*-, *high*-, & *low*-) for analyses. For producing these three pitch-levels, Experiment 2a adopted the method used in McCrea & Morris (2005) and Narayan & Bowden (2013), asked participants to speak at *normal*/*high*/*low* pitch registers without forsaking accuracy. The subjects first produced the samples at their *normal* tone of voice and then repeated the same task at a *high* pitch register, followed by another repetition of the task at a *low* pitch register. There was a self-paced break between each task.

For control of the lexical tone, only T1 was used so that every syllable and word in the stimulus sentence were all in T1, making it easier for the participants to keep the pitch-level consistent and avoid any possible tone *sandhi* being applied. It was also more comfortable for the subjects to reliably manipulate and produce the entire sentence in all three-pitch registers. For convenience, the next section provides a brief review of the stimuli design and an example of the exact carrier phrase from section 3.3.

4.3.1 Stimuli & Carrier Phrases

The stimuli were monosyllabic Mandarin words, which included the aspirated voiceless plosives /p, t, k/ followed by the low-central-unrounded-vowel /a/. Each stimulus appeared twice in two carrier phrases, once with a sentential ending classifier and once without the classifier to create open/closed-ending sentence conditions. Each sentence was produced three times at different pitch registers (at the *natural* speech rate).

Figure 4 is repeated in Figure 14 below from section 3.4, showing the stimuli and the carrier phrases. The change of phrase-position conditions was done by adding 七次 (qīcì); "*seven times*" after the second stimulus. The change of phrase-position conditions was done by adding the classifier 七次 (qīcì or [tɕʰītsʰì]); "*seven times*" after the second stimulus. The adding of the 七次, ([tɕʰītsʰì]) created a token-embedded condition for the second stimulus, and without it, the sentence created a token-ended condition. All Chinese characters in the carrier phrases were in T1 between the phrase-*initial* and the phrase-*medial* stimuli. Notice that 七次, [tɕʰītsʰì] enclosed the stimulus with a T1 character as well.

Figure 14. Screenshot examples of the visual presentation of stimulus sentence without (on the left) and with (on the right) the classifier (qīcì). The second stimulus without the classifier creates the utterance-*final* instance, and the utterance-*medial* instance is on the right.

The entire sentence was presented with a combination of p*inyin* and Chinese characters. For instance, **pā** 先生說 (xiān shēng shūo) **pā** 七次 (qīcì) had the target stimulus in pi*nyin*, but the carrier words were in Chinese characters.

The elicited data were controlled for the below conditions. Point 1, 2, and 5 were essentially the same as in Experiment 1a, and the only differences in Experiment 2a were Point 3 and 4:

1. Place of Articulation (POA): /p, t, k/
2. Post-stop vowel: /a/
3. One lexical tone: T1
4. Three pitch registers: *high*, *normal*, and *low*
5. Three types of phrase-position: utterance-i*nitial*, utterance-*medial*, and utterance-*final*

4.4 Data Elicitation

Experiment 2a had three blocks. The tasks were for the subjects to record their speech at *normal*, *high*, and *low* pitch-levels at the *natural* speech rate (Experiment 1a's blocks were at the *normal* pitch but in the *natural*, *slow,* and *fast* speech rates). The equipment, task procedures, and seating and setting were essentially the same as in Experiment 1a, but adding training to raise and lower pitches following a musical beep before the main recording tasks in the second and third blocks. The participants were trained with the same training stimuli composed of / s, l, n / onsets from Experiment 1a (see section 3.4).

For convenience, Experiment 1a's procedure is repeated:

1. Initial greeting and a minute of casual chatting.
2. Reading the consent form and the recording instructions.
3. Repeating the instructions orally by the experimenter to confirm the participants' understanding of the task.
4. Starting the training phase. This phase asked the participants to practice the task format with the training stimuli – the procedures were the same as the main task.
5. Beginning the main experiment phases.

After a self-paced break time from Experiment 1a, the participants first recorded the stimuli in their everyday pitch register and speech rate for Experiment 2a.

In the second block before the main task, they were trained to speak with a *high* pitch register by listening to a digitally created high-pitch piano F_4 note at 350Hz for the stimuli. The computer displayed the stimulus sentences with the training stimuli (with /s, l, n/) and replayed the guiding notes three times, about 330 milliseconds. They were instructed to raise their pitch in a self-regulated way, following the musical note they heard when producing the phrases.

Similarly, in the third block, they were trained to speak with a *low* pitch register by listening to a digitally made low-pitch piano G_1 note at 50Hz, for the training stimuli. In a self-regulated way, they lowered their pitch register to the lowest possible tone of voice without forsaking accuracy[7], following the triplet musical note, about 330 milliseconds, they heard when producing the phrases.

The participants were instructed to speak right after the guiding notes and practiced it for as much as they wanted until they felt comfortable and ready. They then informed the experimenter to start the second block and then the third block. The same high-/low-pitch piano tone was always replayed for each target stimulus in these blocks to guide them to produce the *high*-/*low*-pitch utterances.

4.5 Measurements

Acoustic measurements were obtained the same way as in Experiment 1a (also see section 3.5). VOT and vowel duration values were measured from /p, t, k/ in /a/ in T1 at three pitch-levels and three phrase-positions. Each stimulus response was digitally analyzed using Praat (Boersma & Weenink, 2018). The values of VOT and the following

[7] This was another reason to choose T1 instead of T3 for this experiment because when producing T3 at the lower pitch register, the lexical tone would normally be creaky voiced.

vowel duration for the stops were labeled and measured. The vowel duration was converted into categorical variables for statistical analysis, discussed in sections 4.7 and 4.8.5. Numeric vowel duration was also used to test a correlation using Pearson's correlation test.

Figure 5 (replotted in Figure 15 below for convenience) shows the waveform and spectrogram of an example and the labeling following the practice of previous studies. Four acoustic measurements for this experiment are defined below:

1. Voice onset time (VOT): the duration in millisecond (ms) from A to B; the stop release burst of the articulator to the beginning of the visible voice bar in the spectrogram and/or the first vertical spike in the waveform
2. Fundamental frequency (F0) onset before the vowel: the F0 at point B plus/minus 5 ms where frequency in Hertz (Hz) can be detected/measured
3. Post-stop vowel duration (vLength): from B to C; the start of aperiodic energy to the end of vowel's periodic energy
4. Vowel-end and F0 offset: the ending of the vowel's periodic energy or F0 offset at point C.

Following these criteria, each stimulus was manually edited to label for A, B, and C points one by one on the TextGrid file. In this dissertation, the VOT value is determined from point A to B, and the vowel duration is from B to C in the figure below. Acoustic measurement proceeds in this manner for all productions of the native speakers of Mandarin.

Several Praat scripts were used to extract the aforementioned acoustic measurements automatically. These values were entered into a spreadsheet alongside phonological variables and demographic information.

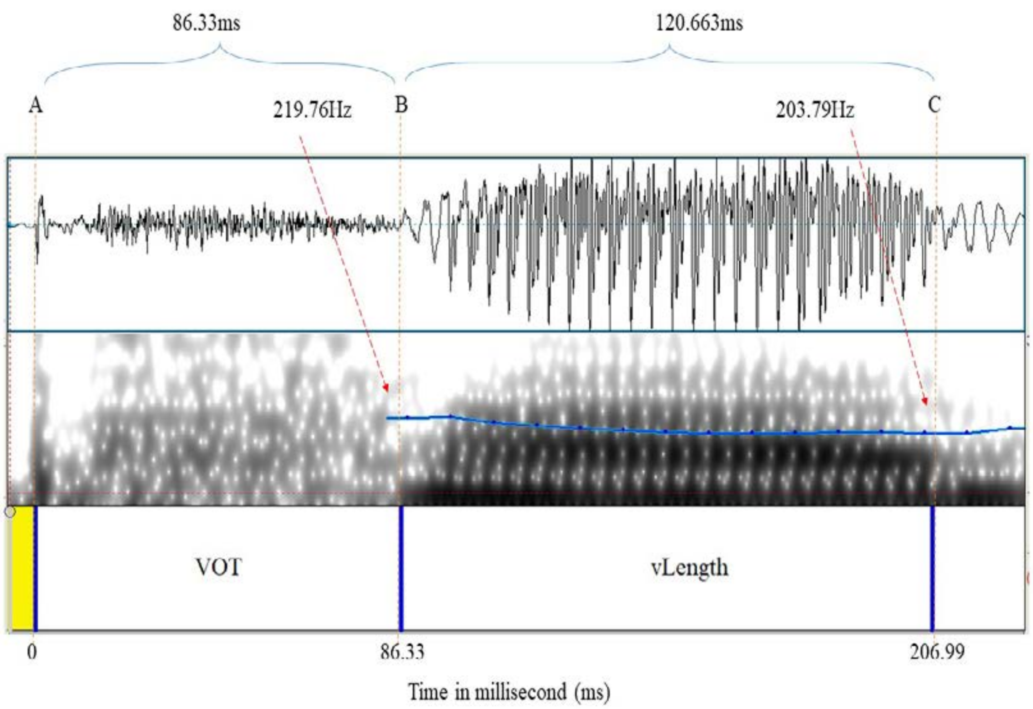

Figure 15. Waveform and spectrogram of an example of [pʰā] in T1. A is the stop release point; B is the start of voicing (or the aperiodic energy in the waveform); C is the end of F2 of the vowel (or the end of the periodic energy). The blue line depicts the overall pitch contour.

4.6 Descriptive Statistics Overview

This section provides descriptive and inferential statistics for Experiment 2a. Each of the categorical variables was dummy-coded accordingly, using R default. The final set of the data's relevant acoustic properties were measured from each stimulus, and the results collapsed across variables were graphed in the figures below. When collapsing the data, the overall T1's VOT and vowel duration values were 71.77 (SD = 21.72) and 202.82 (SD = 86.16) milliseconds, respectively. Recall that *speed* had a significant effect on VOT in Experiment 1a. The same participants' mean VOTs in T1 in Experiment 1a were 70.72, 78.46, and 88.59 milliseconds, and the vowel durations were 165.14, 201.17, and 236.41 milliseconds at *fast*, *natural*, and *slow* speeds, respectively. As we can see, although Experiment 2a's VOTs were similar to those at the *fast* speed from Experiment 1a, their vowel duration values seemed to match the ones at the *natural* speed, as they were supposed to. This observation may suggest a disassociation of the correlation between VOT and vowel duration.

The summaries of each data point of the measurements are provided. The data reported here were then analyzed for their significance using mixed-effect linear regression models discussed later.

4.6.1 Overall VOT in Pitch-levels

Figure 16 shows the overall VOT values by pitch-level. As can be seen, as the speaker's pitch-level raised, the VOT value decreased. VOTs at the *low*-pitch had the most extended average value at 76.10 milliseconds (SD = 24.15), *normal*-pitch the second at 73.28 milliseconds (SD = 19.31), and *high*-pitch the shortest at 65.92 milliseconds (SD = 20.14). When comparing to the *normal* pitch, VOT at the *high*-pitch register changed more than at the *low*-pitch.

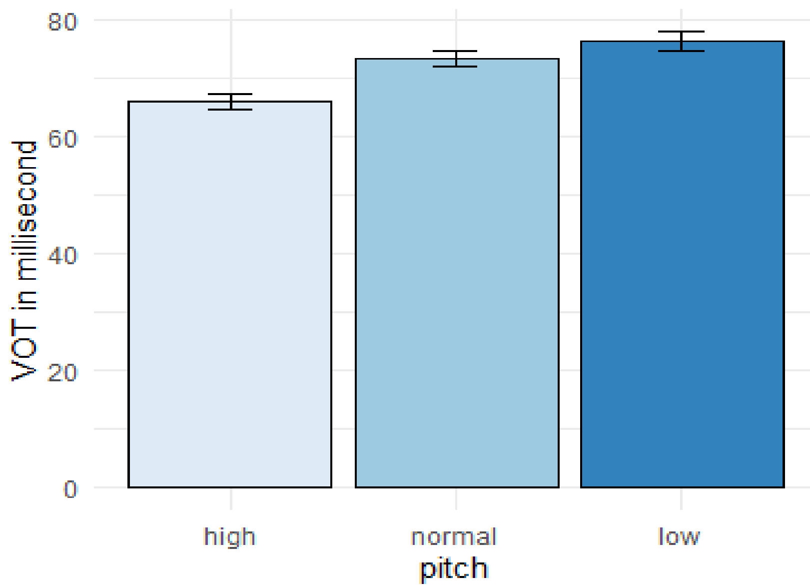

Figure 16. Overall VOT values by pitch-level across factors

4.6.2 Overall VOT in Places of Articulation

Figure 17 shows the overall VOT values by POA across three pitch-levels. As predicted, a consistent POA effect can be observed at each pitch-level, as well as the pitch effect in each place of articulation. The mean values were <u>66.63 milliseconds</u> (SD = 19.98) for /p/, <u>70.73 milliseconds</u> (SD = 22.43) for /t/, and <u>77.95 milliseconds</u> (SD = 21.18) for /k/. VOT in /k/ had the most change of value, except /p/ and /t/ at the *high* pitch-level.

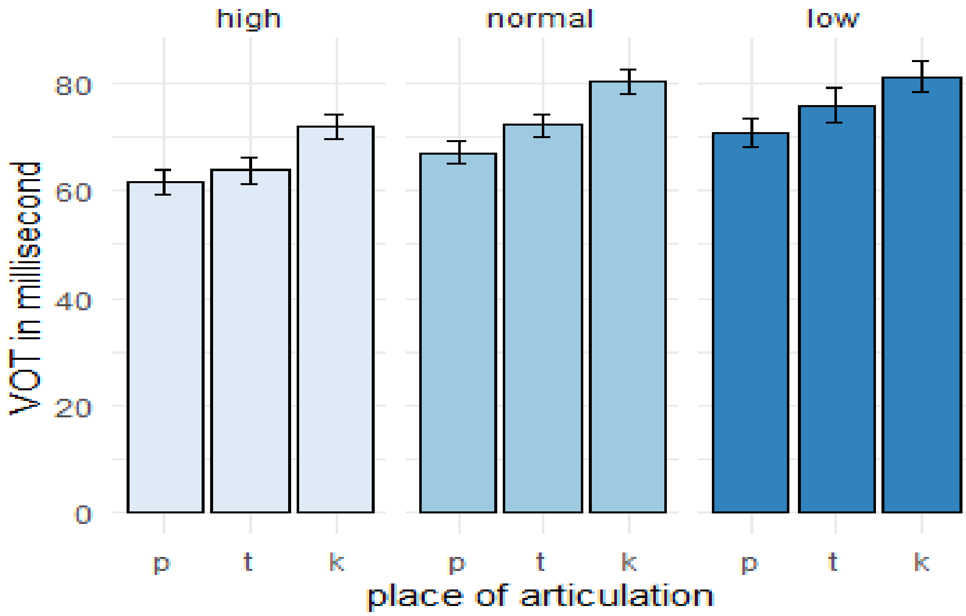

Figure 17. Overall VOT values by the POA across pitch-levels

4.6.3 Overall VOT in Phrase-position

Figure 18 shows the overall VOT values by pitch-level across three phrase-positions. The mean values were 72.37 milliseconds at the utterance-*initial* position (SD = 21.71), 69.96 milliseconds at the utterance-*medial* position (SD = 21.10), and 72.37 milliseconds at the utterance-*final* position (SD = 22.30). The descriptive data showed that the average values were similar in three different phrase-position, but the pitch effect was apparent in all three positions.

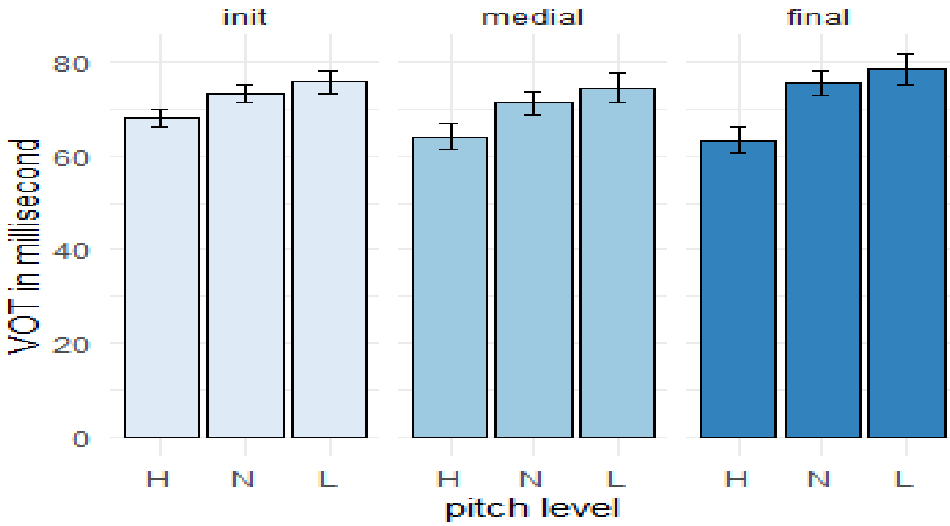

Figure 18. Overall VOT values by pitch-level across phrase-positions

4.6.4 Overall VOT in Genders

Figure 19 shows the overall VOT by gender across three pitch-levels. Descriptively, females produced slightly longer VOT values than the males did from the given data. This observation was interesting since females typically have higher F0. The mean values were 73.34 milliseconds (SD = 21.60) for the females and 70.00 milliseconds (SD = 21.73) for the males. As can be seen from the figure, the most observable gender difference was at the *normal* pitch-level; the difference was analyzed for the significance to be discussed in later sections.

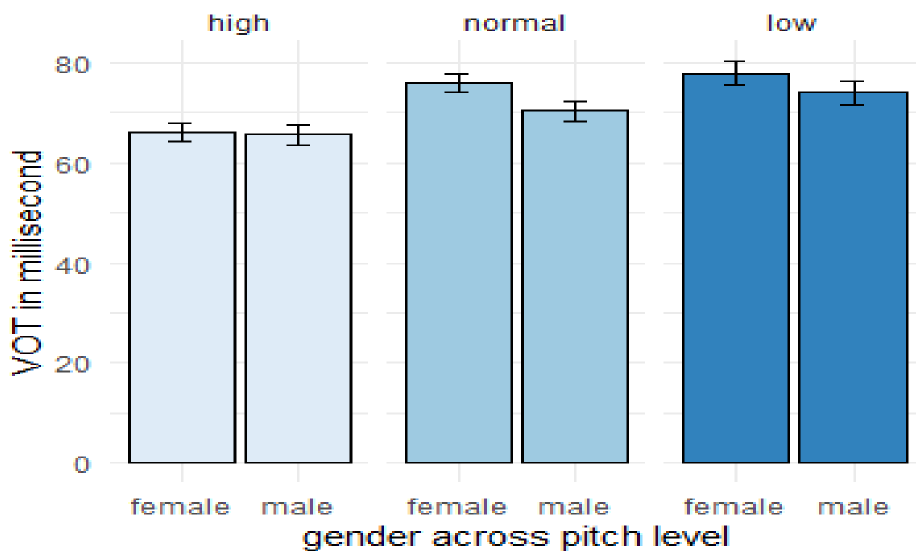

Figure 19. Overall VOT values by gender across pitch-level

4.6.5 Overall F0 Onset Frequency in Pitch-Levels

The averaged F0 onset frequency for the pitch-levels was analyzed. The first observation was that using a modified method from Narayan and Bowden's (2013) with the additional introduction of the precursing piano notes in this study successfully elicited participants' three different pitch-levels. In other words, the leading high and low piano notes successfully guided the participants to produce the stimulus sentence for their *high-* and *low*-pitch registers.

The averaged F0 (in Hertz) showed in Table 10 revealed that both genders could raise their pitch register from *normal* to *high* and lower their overall *low*-pitch register. Females ascended their pitch-level over 100 Hz from *low* to *high* pitch and about 70 Hz for the males; females' *low* pitch was still higher than males' *high* pitch.

Table 10. Overall F0 onset in Hertz (Hz) and vowel duration at each pitch-level in milliseconds (msec) between genders

	female	male
Pitch -level	F0 onset frequency	
at high	310.4 Hz (53.0)	187.9 Hz (32.8)
at normal	230.8 Hz (34.0)	127.0 Hz (19.5)
at low	207.8 Hz (31.3)	115.9 Hz (17.4)
	Vowel duration	
at high	212.8 msec (93.4)	215.4 msec (84.2)
at normal	170.3 msec (59.1)	159.1 msec (44.3)
at low	230.3 msec (106.1)	228.4 msec (84.1)

Concerning vowel duration, Table 10 also provides how each vowel duration in each pitch-level between genders. As can be seen, the averaged vowel duration was similar between genders and three pitch registers; they were able to speak faster at the *normal* pitch-level than *high* or *low* pitch-levels.

4.7 The Statistical Model

Experiment 2a used a series of mixed-effects linear regression models, implemented in the *lmerTest* package in R, to investigate the relationship between VOT and pitch-level as a function of POA, gender, phrase-position, and vowel duration. Five independent variables were included, and the primary focus was the pitch. The models' fitting was done the same way as in Experiment 1a by firstly testing each mixed-effects predictive power using the Likelihood Ratio test (LRT) implemented in afex-package in R (Singmann et al., 2015; Barr et al., 2013). The LRT comparison results of each fixed effect are provided in Table 11.

Table 11. Results of Likelihood Ratio Tests of the fixed effects

Pitch:	$\chi^2 = 52.65$	$p < 0.001$ ***
POA:	$\chi^2 = 45.47$	$p < 0.001$ ***
Phrase-position:	$\chi^2 = 6.760$	$p = 0.034$ *
Gender:	$\chi^2 = 3.030$	$p = 0.082$ ·
Vowel duration:	$\chi^2 = 0.190$	$p = 0.911$

As shown, three of the mixed-effects under investigation were seemingly good predictors. Gender was treated as one of the final model's main effects because the LRT test showed a marginal relationship (likewise, gender and pitch are intrinsically correlated; therefore, they are likely to involve meaningful interaction). Vowel duration was converted into three categorical variables for *short*, *medium*, and *long* speeds based on the distribution in the continuum. For example, in Experiment 2a, the vowel continuum was partitioned into three portions between 62.53 - 193.75, 193.76 - 296.02, and 296.03 - 857.14 milliseconds from the shortest vowel duration found to the longest one. The categorized vowel duration was then included in the final model for analysis. Thus, all five mixed effects were incorporated.

Full interactions of the pitch, POA, phrase-position, and gender predictors were added, and the interaction between vowel duration with others was added one by one for as long as no error message was shown (only up to two levels). Incorporating three-level or full interactions of it with others erred for "*model matrix is rank deficient*", which means either the model was over-fitted or the current data was deficient for the built. All fixed factors were coded using R default treatment. The reference level for the intercept was set automatically to *high* for pitch, /k/ for POA, *final* for phrase-position, *female* for gender, and *short* for vowel duration.

Thus, the final model included five independent variables, full interactions between four of the five, only up to two levels of vowel duration with others, and by-subject adjustments to the random slopes of pitch and POA; as well as random intercepts for by-subjects and by-items were included as maximally as permitted by the data (Barr et al., 2013). The final model converged without any error messages. Pairwise comparisons were conducted using Tukey's HSD tests implemented in the *emmeans* package.

4.7.1 Statistical Result Overview

A summary of the Type III ANOVA test result's main effects with Satterthwaite's method is provided in Table 12. The result indicates that three of the five mixed-effects included were significant effects.

The model revealed three significant main effects and six significant two-way interactions. The post-hoc results of each mixed effect are provided in the next section.

Table 12. Summary of Type III ANOVA table with Satterthwaite's method of the mixed-effects and interactions between the effects of the native Taiwan Mandarin speakers

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	p.value
Pitch (Ph)	9963.3	4981.6	2	135.42	23.3976	1.871e-09 ***
POA (P)	23187.3	11593.6	2	121.49	54.4526	< 2.2e-16 ***
Phrase Pos. (Pos)	1426.8	716.4	2	2233.12	3.3507	0.0352362 *
Gender (G)	466.6	466.6	1	75.36	2.1917	0.1429266
Vowel Dur. (V)	279.6	139.8	2	2331.66	0.6566	0.5187284
Ph:P	2076.4	519.0	4	2151.76	2.4381	0.0451377 *
Ph:Pos	3885.3	971.3	4	2115.94	4.5621	0.0011387 **
P:Pos	4539.1	1134.8	4	2021.03	5.3298	0.0002867 ***
Ph:G	602.9	301.5	2	75.49	1.4159	0.2490677
P:G	1740.3	870.2	2	77.11	4.0869	0.0205585 *
Pos:G	623.9	311.9	2	2190.03	1.4651	0.2312848

Ph:V	2174.6	543.7	4	1870.60	2.5534	0.0373259 *
P:V	1816.2	454.1	4	954.95	2.1326	0.0748443 ·
E:V	1305.7	326.4	4	2269.56	1.5331	0.1898649
G:V	1406.9	703.4	2	2311.67	3.3039	0.0369127 *
Ph:P:Pos	1936.2	242.0	8	2117.35	1.1367	0.3349834
Ph:P:G	629.5	157.4	4	2110.05	0.7391	0.5651289
Ph:Pos:G	1193.5	298.4	4	2122.84	1.4014	0.2310226
P:Pos:G	756.1	189.0	4	2115.69	0.8878	0.4703201
Ph:P:Pos:G	1102.6	137.8	8	2108.90	0.6473	0.7382075

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

4.8 Results of Experiment 2a

As shown in Table 12, the statistical modeling revealed that VOT values were significantly affected by pitch register, as well as by POA and phrase-position. There were six significant two-way interactions between pitch & POA, pitch & phrase-position, POA & phrase-position, POA & gender, pitch & vowel duration, and gender & vowel duration. Starting with pitch register, each main effect and the Post-hoc analyses of the significant interactions for the native Mandarin speakers are provided in the following subsections

4.8.1 Pitch register Effect in the Taiwan group

As expected, VOTs at the *high* pitch register were significantly shorter than those at *normal* and *low* pitch registers. The mean VOT values were 63.8 msec at *high* (SE = 1.83, DF = 80.5, lower.CL = 60.1, upper.CL = 67.4), 72.3 msec at *normal* (SE = 1.85, DF = 87.3, lower.CL = 68.6, upper.CL = 75.9), 76.1 msec at *low* (SE = 2.37, DF = 111.4, lower.CL = 71.4, upper.CL = 80.8). As shown in Figure 20, the general patterns were consistent, except for the interactions in the *medial* phrase-medial position at *high* and *low* pitch registers.

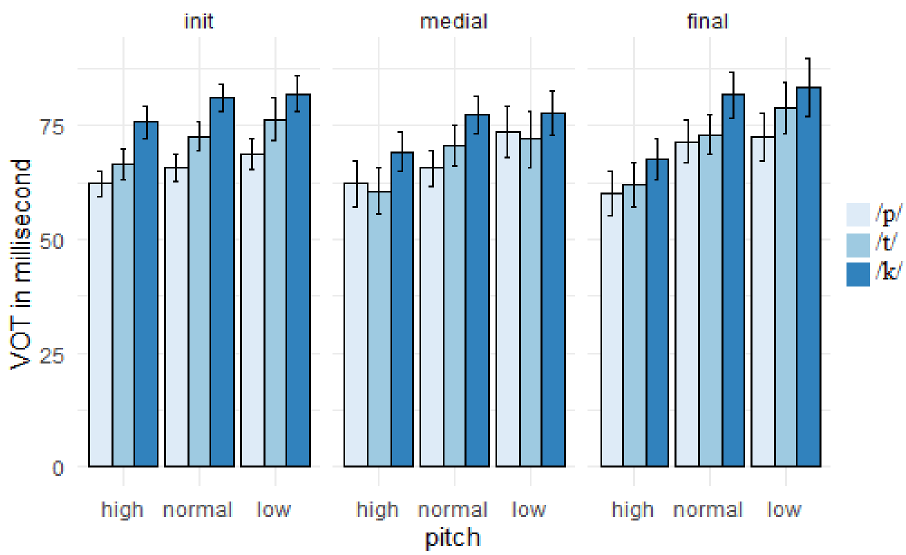

Figure 20. VOT patterns due to pitch register effect across POAs and phrase-positions

The post-hoc analysis of the two-way interaction revealed that *high* pitch register VOTs were significantly shorter than *low* ones in all three POAs; they were also shorter than those at *normal* pitch register, but not in /p/. VOTs at *low* pitch were longer than those at the *normal* pitch in /p/, but not in /t/ and /k/. Moreover, VOTs at the *high* pitch register were shorter than those at *low* in *the final* phrase-position; they were also shorter than those at the *normal* pitch register.

There was also a significant two-way interaction between pitch register and vowel duration. VOTs at *high* were significantly shorter than those at *low* and *normal* pitch registers, but not when the vowel was *long* at *normal* pitch. None of the differences between *low* and *normal* pitch registers was significant in all three categories of vowel duration (see Table 13 for detail).

Table 13. Tukey HSD Post-hoc Analyses for significant interaction results between pitch & POA, pitch & phrase-position, and pitch & vowel duration

Contrast	estimate	SE	df	t.ratio	p.value	
Pitch register : POA						
in /p/						
high - low	-10.51	1.99	13.2	-5.286	0.0004	***
high - normal	-3.55	2.50	32.0	-1.419	0.3435	
low - normal	6.96	2.48	31.2	2.807	0.0226	**
in /t/						
high - low	-12.84	1.97	12.7	-6.518	0.0001	***
high - normal	-7.17	2.47	30.6	-2.899	0.0183	*
low - normal	5.67	2.44	29.4	2.324	0.0680	.
in /k/						
high - low	-9.97	2.01	13.6	-4.967	0.0006	***
high - normal	-7.89	2.49	31.3	-3.172	0.0092	**
low - normal	2.08	2.45	29.9	0.849	0.6758	
Pitch register : Phrase-position						
in *initial*						
high - low	-7.34	1.91	62.5	-3.834	0.0009	***
high - normal	-3.01	2.58	176.7	-1.166	0.4753	
low - normal	4.34	2.49	166.1	1.741	0.1929	
in *medial*						
high - low	-10.47	2.41	50.5	-4.340	0.0002	***
high - normal	-4.37	2.95	116.4	-1.481	0.3039	
low - normal	6.10	2.88	112.4	2.118	0.0908	
in *final*						
high - low	-15.52	2.19	33.6	-7.083	< .0001	***
high - normal	-11.24	2.48	55.4	-4.531	0.0001	***
low - normal	4.28	2.44	53.6	1.75	0.1946	
Pitch register : Vowel duration						
at *short*						
high - low	-12.14	1.98	67.6	-6.125	< .0001	***
high - normal	-9.44	1.94	60.5	-4.875	< .0001	***
low - normal	2.70	1.91	59.8	1.413	0.3406	
at *medium*						
high - low	-7.76	1.98	73.2	-3.909	0.0006	***
high - normal	-6.51	2.33	122.4	-2.795	0.0165	**
low - normal	1.24	2.27	116.9	0.549	0.8471	
at *long*						

high - low	-13.43	2.86	218.2	-4.699	< .0001	***
high - normal	-2.66	4.84	1107.1	-0.549	0.8469	
low - normal	10.77	4.74	1087.8	2.274	0.0599	·

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

4.8.2 POA Effect in the Taiwan group

The consistent effect of POA was shown, where the overall VOTs in / p, t, k / were 66.9 msec (SE = 1.63, DF = 72.4, lower.CL = 63.6, upper.CL = 70.1), 68.9 msec (SE = 1.84, DF = 83.1, lower.CL = 65.3, upper.CL = 72.6), and 76.3 msec (SE = 1.98, DF = 91.9, lower.CL = 72.4, upper.CL = 80.3), respectively.

The analysis showed two two-way interactions between POA & phrase-position and POA & gender, as shown in Figure 21 below. The VOTs in /k/ were longer than those in /p/ and /t/ in all three phrase-positions. Those in /t/ were also longer than those in /p/ only in the *initial* phrase-position.

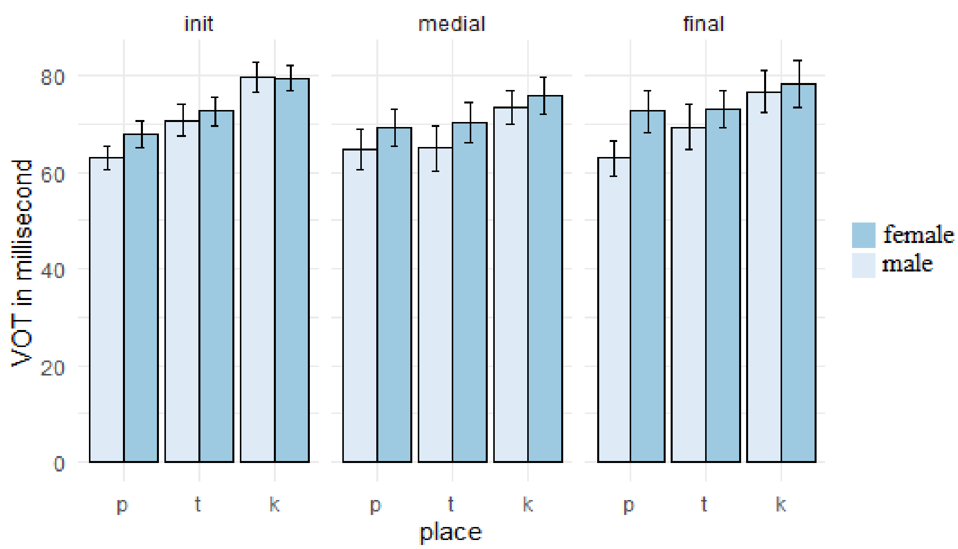

Figure 21. VOT patterns due to POA effect across phrase-positions and genders

Moreover, VOTs in /k/ were significantly longer than those in /t/, which in turn also longer than those in /p/ in males. For females, only /p/-/t/ differences were non-significant. Table 14 provides the post-hoc results of the interactions between the factors discussed.

Table 14. Tukey HSD Post-hoc Analyses for significant interaction results between POA & phrase-position and POA & gender for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
POA : phrase-position						
in *initial*						
/k/ - /p/	16.39	1.53	38.5	10.707	< .0001	***
/k/ - /t/	9.18	1.59	43.5	5.763	< .0001	***
/p/ - /t/	-7.21	1.53	38.3	-4.705	0.0001	***
in *medial*						
/k/ - /p/	10.06	2.08	29.7	4.840	0.0001	***
/k/ - /t/	8.35	2.13	32.5	3.920	0.0012	**
/p/ - /t/	-1.71	2.08	31.8	-0.820	0.6937	
in *final*						
/k/ - /p/	9.64	1.75	13.8	5.507	0.0002	***
/k/ - /t/	6.55	1.82	16.1	3.610	0.0062	**
/p/ - /t/	-3.09	1.78	15.1	-1.735	0.2250	
POA : Gender						
in *female*						
/k/ - /p/	9.32	1.48	28.9	6.299	< .0001	***
/k/ - /t/	6.83	1.62	36.5	4.210	0.0005	***
/p/ - /t/	-2.49	1.55	32.1	-1.612	0.2553	
in *male*						
/k/ - /p/	14.74	1.59	35.7	9.256	< .0001	***
/k/ - /t/	9.23	1.72	42.2	5.357	< .0001	***
/p/ - /t/	-5.51	1.65	38.1	-3.339	0.0052	**

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

4.8.3 Phrase-position Effect in the Taiwan group

The mean VOTs by phrase-position were: 72.3 msec at the utterance-*initial* position (SE = 1.62, DF = 74.8, lower.CL = 69.0, upper.CL = 75.5), 69.7 msec at the utterance-*medial* position (SE = 1.83, DF = 85.7, lower.CL = 66.1, upper.CL = 73.3), and 70.2 msec at the utterance-*final* position (SE = 2.32, DF = 183.0, lower.CL= 65.6, upper.CL = 74.7). There was a significant main effect (F (2, 2233) = 3.3507, p = 0.0352) and two two-way interactions between phrase-position & pitch (F (4, 2116) = 4.5621, p = 0.00114) and phrase-position & POA (F (4, 2021) = 5.3298, p = 0.00029). The only significant VOT difference in the two-way interaction between phrase-position and pitch register was those between utterance-*final* and utterance-*initial* in the *high* pitch register. Similarly, only for /k/, VOTs in the utterance-*final* position were longer than those in utterance-*initial* position, which in turn were longer than those in the utterance-*medial* position. See Table 15 for detailed statistics.

Table 15. Tukey HSD Post-hoc Analyses for significant interaction results between phrase-position & pitch register and phrase-position & POA for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
Phrase-position : Pitch Register						
at *high*						
final - initial	-7.638	1.80	279.2	-4.238	0.0001	***
final - medial	-3.103	2.64	74.1	-1.173	0.4727	
initial - medial	4.535	2.05	434.1	2.214	0.0700	.
at *normal*						
final - initial	0.593	1.86	405.1	0.319	0.9455	
final - medial	3.765	2.66	87.5	1.418	0.3364	
initial – medial	3.172	2.20	542.2	1.445	0.3186	
at *low*						
final - initial	0.535	1.66	221.5	0.323	0.9441	
final - medial	1.946	2.38	50.0	0.818	0.6936	
initial - medial	1.411	1.91	354.3	0.740	0.7398	
Phrase-position : POA						
in /p/						
final - initial	1.452	1.65	243.7	0.879	0.6540	
final - medial	0.546	2.49	62.8	0.219	0.9739	
initial - medial	-0.906	2.06	431.8	-0.440	0.8987	
in /t/						
final - initial	-2.669	1.60	200.2	-1.666	0.2209	

final - medial	1.930	2.41	52.6	0.800	0.7049	
initial - medial	4.599	2.01	408.5	2.287	0.0587	·
in /k/						
final - initial	-5.294	1.69	198.2	-3.138	0.0055	**
final - medial	0.131	2.48	53.5	0.053	0.9985	
initial - medial	5.425	2.05	440.8	2.650	0.0227	*

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

4.8.4 Gender Effect in the Taiwan group

The mean VOTs were 68.6 msec for males (SE = 2.33, DF = 82.0, lower.CL = 64.0, upper.CL = 73.3) and 72.8 msec for females (SE = 2.27, DF = 92.9, lower.CL= 68.3, upper.CL = 77.3). The ANOVA test result showed no significant main effect of gender (F (1, 75) = 2.1917, p = 0.1429). The model also revealed two two-way interactions between gender & POA (F (2, 77) = 4.0869, p = 0.0206) and gender & vowel duration (F (2, 2312) = 3.3039, p = 0.0369). However, it was mainly due to POA interaction; females had significantly longer VOTs than males only in /p/, as well as in the *long* vowel duration condition (see Table 16).

Table 16. Tukey HSD Post-hoc Analyses for significant interaction results between gender & POA and gender & vowel duration for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
Gender : POA						
in /p/						
female - male	7.40	3.03	80.2	2.445	0.0167	**
in /t/						
female - male	4.38	3.45	77.0	1.270	0.2080	
in /k/						
female - male	1.98	3.60	78.3	0.551	0.5834	
Gender : Vowel duration						
in *short*						
female - male	4.04	3.20	83.1	1.263	0.2103	
in *medium*						

female - male	1.72	3.27	91.1	0.526	0.6003	
in *long*						
female - male	8.00	4.04	198.6	1.981	0.0489	*

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

4.8.5 Vowel duration Effect in the Taiwan group

Finally, the mean VOT values at the three categorized vowel durations were <u>70.7 msec</u> (SE = 1.64, DF = 87.7, lower.CL = 67.5, upper.CL = 74.0), <u>71.7 msec</u> (SE = 1.67, DF = 95.4, lower.CL = 68.4, upper.CL = 75.1), and <u>72.7 msec</u> (SE = 2.70, DF = 545.6, lower.CL = 67.4, upper.CL = 78.0) for *short*, *medium*, and *long*, respectively. There was no significant main effect of vowel duration (F (2, 2332) = 0.6566, p = 0.5187), but as discussed earlier, the model showed two two-way interactions between pitch & vowel duration (F (4, 1870) = 1.4651, p = 0.037) and gender & vowel duration (F (2, 2312) = 3.3039, p = 0.0369). The only significant VOT difference was when the stops were produced at the *low* pitch-level and between *medium* and *long* vowel duration. The post-hoc analysis showed none of vowel duration was significant within each gender (see Table 17 below).

The overall vowel duration for *short*, *medium*, and *long* were <u>131.82</u> (SD = 19.99), <u>182.50</u> (SD = 14.55), and <u>294.20</u> (SD = 88.61) milliseconds, respectively. A Pearson correlation test revealed a weak positive correlation between VOT and vowel duration (r = 0.0262), but non-significant (t = 1.2978, DF = 2446, p = 0.1945). In other words, there was no vowel duration effect from the given dataset.

Table 17. Tukey HSD Post-hoc Analyses for significant interaction results between vowel duration & pitch register and vowel duration & gender for the Taiwan group.

Contrast	estimate	SE	df	t.ratio	p.value	
Vowel duration : Pitch Register						
at *high*						
short - medium	-3.431	1.73	1894	-1.986	0.1159	
short - long	-3.774	2.87	1763	-1.313	0.3879	
medium - long	-0.343	2.48	2291	-0.138	0.9895	
at *normal*						
short - medium	-0.505	1.82	1887	-0.277	0.9587	
short - long	3.009	4.70	2255	0.640	0.7981	
medium - long	3.513	4.47	2311	0.786	0.7115	
at *low*						
short - medium	0.950	1.56	2144	0.610	0.8144	
short - long	-5.065	2.71	1573	-1.872	0.1473	
medium - long	-6.015	2.44	1946	-2.461	0.0371	*
Vowel duration : Gender						
in *female*						
short - medium	0.1652	1.37	2393	0.121	0.9920	
short - long	-3.9223	2.85	2402	-1.376	0.3539	
medium - long	-4.0875	2.61	2359	-1.566	0.2605	
in *male*						
short - medium	-2.1554	1.43	2432	-1.508	0.2872	
short - long	0.0357	2.94	2365	0.012	0.9999	
medium - long	2.1911	2.73	2347	0.803	0.7013	

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

4.9 Result Summary

A series of mixed-effects linear regression analyses were conducted to explore the relationships between F0 onset and VOT as functions of speaker's pitch register, POA, phrase-position, gender, and vowel duration in Experiment 2a. The native Mandarin speakers' VOT values were analyzed, and three out of five fixed effects were significant.

The current survey hypothesizes that the higher F0 onset of the tone will lead to shorter VOT, and the tone effect found in Experiment 1a was due to the vocal fold tensity. Experiment 2a shows that when the speaker's vocal folds were stretched (for a higher pitch), they exhibited significantly shorter VOT values in the same lexical tone and POA; i.e., the higher pitch did lead to shorter VOT values.

The POA effect was consistent, where VOTs were significantly longer in /k/ than those in /t/ and /p/ across all three pitch-level conditions. There were no gender and vowel duration main effects shown in Experiment 2a from the ANOVA or Pearson correlation tests.

Mandarin's T1 is a level tone and generally exhibits a higher F0 onset value than other lexical tones. Therefore, it has been reported to have a shorter VOT value than, for instance, T2 or T3 of Mandarin in studies (Liu et al., 2008; Chen et al., 2009; Tseng, 2018). This dissertation proposes that pitch, which is the tension change of the vocal folds controlled by the cricoarytenoid muscles, is one of the main reasons due to the Myoelasic-aerodynamic effect (Ven dam Berg, 1958).

In agreement with previous studies (McCrea & Morris, 2005; Lai et al., 2009; Narayan & Bowden, 2013; Lou, 2018), this study's general finding is that pitch-level affects VOT values when factoring out other variables. We show that when the participants produced T1 at a *high* pitch-level, they bared even shorter VOT values than when they produced it at a *low* pitch-level in the same POAs. This finding indicates that the vocal folds contracted to change the elastic thinness to allow a high pitch onset at adduction had a quicker reaction to the Bernoulli Effect, thus shorter VOTs. Can this observation apply to L2 production from various language backgrounds? To answer this question, we conduct the same experiments with three L2 learner groups of Mandarin: native English, Japanese and Spanish speakers.

Recall that the third research question deals with L2 speech production. Experiment 1a and Experiment 2a reveal strong tone, POA, and pitch effects on VOT. Do we find the same effects in L2 learners of Mandarin Chinese? Does L2's native language matter? Do we find L1 VOTs for their L2 production? Chapter 5 reports Experiment 1b and Chapter 6 Experiment 2b for L2 speakers (replicating Experiment 1a and 2a for non-native Mandarin speakers).

CHAPTER 5. TONE EFFECT IN L2 SPEAKERS – EXPERIMENT 1B

5.1 Introduction

Recall that Chapter 3 reports Experiment 1a for native Mandarin speakers and addresses the VOT variation factors. To illuminate the phenomena further, the third research question asks whether the effects also exhibit in L2 Mandarin. This chapter investigates the same effects in L2 Mandarin by replicating the same experiment with three L2 groups of Mandarin learners. The hypothesis is that the effects observed in L1 Mandarin speech also exist in L2 Mandarin speech produced by English (ENG), Japanese (JPN), and Spanish (SPN) speakers. The reason for choosing these three language backgrounds, alongside Taiwan Mandarin (TMN), is that the L1 VOT continua compile agreeably with Cho and Ladefoged's (1999) cross-linguistic voiceless VOT taxonomy: unaspirated (SPN), weakly aspirated (JPN), aspirated (ENG), and strongly aspirated (TMN).

The experiment recruited 96 L2 participants (22 ENG, 40 JPN, & 34 SPN). Each participant produced speech samples of 72 combinations of stimuli for acoustic analysis. Stimuli and the carrier phrases were the same ones tested with the Taiwan group (See section 3.3). Data analysis was also approached the same way as the native Mandarin data, where we tested the data using a series of linear mixed-effects models that considered the main effects of tone, POA, speech rate, phrase-position, and gender. An additional fixed effect here was the native language (NL). For evaluating L2 speakers' Mandarin production, see section 5.4.

Similar to the modeling for the native speaker's analyses, the results with all language groups showed that VOT values differed significantly due to the lexical tone, POA, and speech rate. NL was also a strong significant factor; significant VOT differences between language groups were observed.

The general observed patterns of VOT values were similar from the longest to the shortest: T3 > T2 > T1 > T4 for tones, /k/ > /t/ > /p/ for POAs, *slow* > *natural* > *fast* for speeds, and ENG > Taiwan > SPN > JPN for NL backgrounds. These outcomes revealed that the testing method was considered reliable. An additional observation was that in a baseline VOT test, two groups of L2 learners used neither their native VOTs nor Mandarin VOTs (compared within the same L2 speaker groups).

This chapter starts with a report of the participants' background information, followed by the methodology section, which also describes the task procedures for the L2 speakers for their L2 and L1 speech productions. The approach that was taken for evaluating L2 tone production is provided in the method section. Section 5.5 reports descriptive statistics individually and a group comparison, including the Taiwan group for convenience. Finally, a result summary and a preliminary discussion conclude this chapter.

5.2 Participants

The subjects were 96 adult L2 speakers of Mandarin from three different language backgrounds. Each group's demographic information is provided in the following subsections. Fifty-four out of 96 L2 subjects recorded the speech in Taiwan at three different universities, 27 in Japan, and 11 in the US. At all recording sites, the recording equipment and the procedures were the same; none of the 96 L2 participants self-reported any known history of a speech disorder or a hearing impairment. Table 18 summarizes the demographic information of the L2 subjects, as well as for the Taiwan group (in Experiment 1a) for convenience.

Table 18. Demographic information of the L2 participants, plus the natives

Group	N (gender)	Mean age (sd)	Mean Length of L2 learning (range)
English	22 (11 female, 11 male)	33.64 (16.16)	3.68 (0.5yr - 13yr)
Japanese	40 (23 female, 17 male)	32.83 (12.61)	4.23 (0.5yr - 35yr)
Spanish	34 (19 female, 15 male)	23.03 (6.070)	1.88 (0.8yr - 10yr)
Taiwanese	68 (36 female, 32 male)	29.09 (13.19)	---

5.2.1 English L2 Group

The ENG group (22 subjects) is the smallest and most heterogeneous group. Twenty of them were from the US[8], one from Canada, and one from the UK. Their average age was 33.64, ranged from 18 to 76. Half of the 22 subjects recorded their speech production in Taiwan at the NPLI library (3 subjects), the Providence University (5 subjects), and the American School in Taichung city (3 subjects). They were university students and English teachers. The other 11 subjects participated in the US at George Mason University, and they were students there. Among all 22 participants, six self-reported that they had never lived in a Mandarin-speaking country, while 16 reported that they had.

[8] They were from states and counties of Fairfax, Falls Church. Washington DC, San Francisco, Radford, Macon, Colorado, Dunkirk, Clinton, Dallas, Houston, El Paso; Laval Canada; Blackpool UK.

5.2.2 Japanese L2 Group

The JPN group had 40 subjects. These speakers had an average age of 32.83, which ranged from 18 to 62 years old. Twenty-seven of them recorded their speech production in Tokyo, at the Arai-lab of Sophia University and the Nicchu Mandarin Institute of the Japan-China Friendship Center. They were students, businessmen/women, and airline company employees. The other 13 university students recorded their speech production in Taiwan at Providence University and Fuguang University. Among the 40 JPN participants, 16 of them self-reported that they had never lived in a Mandarin-speaking country, and 24 had.

5.2.3 Spanish L2 Group

All 34 native Spanish speakers recorded the speech production in Taiwan, at Fuguang University in Yilan city and Tunghai and Providence Universities in Taichung. Twenty-five of them were from Paraguay, six from Spain, one from Colombia, one from Peru, and one from Tunisia. The average age of the SPN group was 23.03 years old and ranged from 18 to 46. They were the youngest group and had the shortest average length of learning. This group self-reported that they had been living in Taiwan for more than a year at the time of the recording. They also reported that they had regularly used Mandarin outside the classroom. Therefore, we may predict that the SPN group might have the highest proficiency level in Mandarin.

5.3 Methodology

The test method, i.e., the stimuli, the carrier phrases, main task procedures, pieces of equipment, and data measurement, was precisely the same for the Taiwan group. The only differences were: a) the L2 subjects additionally recorded their L1 speech after the main experiments for a VOT baseline test and b) the L2 subjects filled out the questionnaire in their native languages. For convenience, the test stimuli, procedures, and data elicitation and measurement are reviewed here from Chapter 3.

5.3.1 Speech Samples (Review)

The stimuli were monosyllabic Mandarin words, which included the aspirated voiceless plosives /p, t, k/ followed by the low-central-unrounded-vowel /a/. Each stimulus appeared twice in two carrier phrases: once with a sentential ending classifier and once without the classifier to create an open/closed-ending sentence condition. Each sentence was produced three times at different speech rates.

The change of phrase-position conditions was done by adding the classifier 七次 (qīcì or [tɕ^{h}īts^{h}ì]); "*seven times*" after the second stimulus. The adding of the 七次, ([tɕ^{h}īts^{h}ì]) created a token-embedded condition for the second stimulus, and without it, the sentence created a token-ended condition. All Chinese characters in the carrier phrases were in T1 between the phrase-*initial* and the phrase-*medial* stimuli. Notice that 七次, [tɕ^{h}īts^{h}ì] enclosed the stimulus with a T1 character as well. Table 19 below (repeated from Table 2 in Chapter 3) lists the stimuli and the carrier phrases.

Table 19. **Monosyllabic stimuli & the carrier phrase plus the classifier**

Tone	POA		
T1	[pʰā]	[tʰā]	[kʰā]
T2	[pʰá]	[tʰá]	[kʰá]
T3	[pʰǎ]	[tʰǎ]	[kʰǎ]
T4	[pʰà]	[tʰà]	[kʰà]
Carrier phrase:	____先生說 (xiān shēng shūo) ____七次(qīcì); "*Mr.___ says seven times*"		

The entire sentence was presented with a combination of *pinyin*[9] and Chinese characters. For instance, **pā** 先生說 (xiān shēng shūo) **pā** 七次 (qīcì) had the target stimulus in *pinyin*, but the carrier words were in Chinese characters. There were two reasons for this. The first reason was that the monosyllabic word might not have a written character, which does not have a semantic meaning across all three articulatory gestures and four tones. For instance, neither /kī/ nor /kì/ has a linguistic meaning or a matched character in Mandarin. However, /kī/ and /kì/ are possible pronunciations for Mandarin speakers, which are accidental gaps, and /kì/ can be a surname for a foreigner when introduced (e.g., Key 先生 as *Mr. Key*). Secondly, the entire sentence was an attempt to create a meaningful interpretation (e.g., "*Mr. Key says Key seven times*") to provide more contextual support instead of producing [pʰā], or [tʰā], or [kʰā] monotonically (e.g., "*Pa, I now say Pa*").

[9] In Taiwan, the *pinyin* system is not officially taught in school, but the translation of street names and other foreigner-friendly signs are written in *pinyin.*

The elicited data were attended to the following conditions:

1. Place of Articulation (POA): the aspirated voiceless stops of Mandarin, /p, t, k/, were used.
2. Post-stop vowel: only the low-central-unrounded vowel /a/ was selected. Vowel duration control was done by adding 七次 (qīcì or [tɕʰītsʰì]); "*seven times*," which allowed for measurements of long and short vowel durations from the phrase-m*edial* token.
3. Four lexical tones: high-level (or T1), mid-rising (or T2), falling-rising (or T3), and high-falling (or T4).
4. Three speech rates: *slow*, *natural*, and *fast*. This study adapts Kessinger and Blumstein's (1998) method to ask the participants to speak as *natural*/*slow*/*fast* (in this order) as possible without forsaking accuracy. See the next section for detail.
5. Three types of phrase-positions: utterance-*initial* (i.e., the initial stimulus), utterance-*medial* (i.e., the medial stimulus due to the adding of 七次 (qīcì or [tɕʰītsʰì]); "sev*en times*," classifier), and utterance-*final* (i.e., the final stimulus without the classifier). The participants read the stimulus sentence once with the classifier, which generates two stimulus tokens of utterance-*initial* and utterance-*medial*, and once without the classifier, which generates two stimulus tokens of utterance-i*nitial* and utterance-*final*. There are twice as many *initial* tokens as *medial* and *final* tokens for analysis.

5.3.2 Procedures & Data Elicitation

Experiment 1b instructed the participants to read the same stimuli in their carrier phrases as the native speakers. The subjects' task was to read the sentences on the computer screen, using PsychoPy (version 1.83.04; Peirce 2009), which displayed the stimuli and the carrier phrases in a random order, as shown in Figure 22 (replotted from Figure 4).

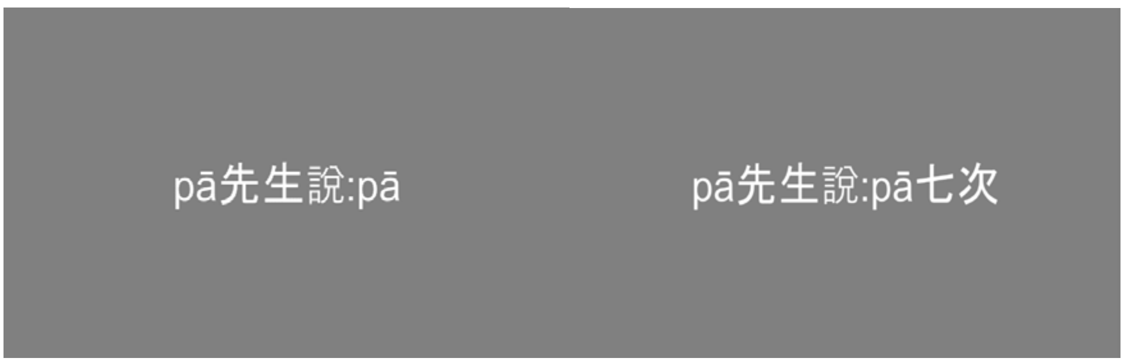

Figure 22. Screenshot examples of the visual presentation of stimulus sentence without (on the left) and with (on the right) the classifier (qīcì). The second stimulus without the classifier creates the utterance-*final* instance, and the utterance-*medial* instance is on the right.

This experiment's main task contained three blocks, which asked participants to record their speech at three different speeds. The first block required participants to read the sentences in their most comfortable speech rate (*natural*). The second block was to speak at a noticeably slower speed than his or her regular speech rate (*slow*), and the third block was to speak as *fast* as s/he could. A short self-paced break was allowed in between blocks. The entire procedure went as follows.

1. Initial greeting and a minute of casual chatting.

2. Reading the consent form and the recording instructions (in Mandarin or English of their choice; Appendixes A1-2 & C1-2).
3. Repeating the instructions orally in Mandarin (supported with English) by the experimenter to confirm the participants' understanding of the task.
4. Starting the training phase. This phase asked the participants to practice the task format with the training stimuli; the procedures were the same as the main task.
5. Beginning the main experiment blocks.

Several stimuli were used in the training phase to familiarize the participants with the task. The training stimuli consisted of /s, l, n/ as word onsets, rather than /p, t, k/ to avoid possible effects of training phase stimuli on the main experiment. In the training phase, the participants started a session to familiarize themselves with the experiment's task format, practicing it with 24 training stimuli in four tones. During the training phase, the L2 participant was given a chance to ask questions to confirm the pronunciation of the carrier phrases.

For the L2 speakers, there was 1-2 minutes of a casual conversation after the reading of the consent and the instruction forms. During this time, the conversation was always in Mandarin between the L2 subject and the experimenter. This attempt was to activate the participant's Mandarin system to ensure their production would be "more Mandarin-like" (see "gestural drift" for details in Sancier & Fowler, 1997).

During the tasks, the participants were seated in front of the computer and the microphone. The experimenter was seated on the left of the computer, controlling the keyboard. If the subject's utterance was perceptually unclear to the experimenter, the participant was asked to repeat the sentence before moving on to the next trial (there were a few requests for repeating, but it was not kept as a record).

The speech samples were counterbalanced to avoid the order-effects; stimulus presentation was randomized for each participant. Participants could ask questions or stop the experiment at any time, although none of the participants did so. The data collection was done using the same equipment as for the Taiwan group, an SSD laptop (Lenovo 110S) with a USB-connected Zoom Handy H2 recorder. All recordings were made using Audacity (version 2.1.3 of Audacity®, Audacity Team), stored as WAV files (44.100 kHz, 16 bit, mono).

In total, 23616 stimulus tokens were collected, and 397 tokens were excluded due to undetectable syllable boundaries and overlapping segments in the waveform and spectrogram. For the statistical analysis, the target stimuli were 23219, where the L2 groups contributed 3076 (ENG), 5587 (JPN), and 4836 (SPN), plus 9722 from TMN.

After Experiment 2b (see Chapter 6), the L2 participants filled out the questionnaire in their native languages (Appendixes B2-4). The L2 groups recorded a speech in their native language to elicit their native VOT values (section 5.12) after filling out the questionnaire in their native language, which would allow them approximately 2 minutes to switch back to their native language model (the conversation between the experimenter and the participant were then in their native language).

As mentioned above, the experiment was conducted in Taiwan, Japan, and the US. The entire experiment, included the baseline VOT recording for the L2 speakers, lasted about 20-30 minutes. All participants were paid 7 US dollars or the equivalent amount in local currency.

5.4 Evaluation of the L2 Tone Production

The current study evaluates L2 tone production by four native Mandarin speakers. After collecting all the recordings, four native Mandarin speakers in Taiwan (including the author) perceptually evaluated the samples. Samples of the 96 L2 speakers were partitioned into four quartiles and randomly assigned to the four listeners. Each listener examined 48 subjects (half of the population). The first person judged the samples from 1-48 subjects, the second from 24-72, the third from 48-96, and the last person from 72-96 plus 1-24—this method allowed each subject's samples to be heard twice by two different raters. Any L2 tone production that was perceived as unclear or inaccurate was marked for exclusion, and any stimulus token that had two marks was excluded from the analyses. Ninety-two from the ENG group, 173 from the JPN group, 62 from the SPN group (or about 2.5% in total) stimulus tokens were excluded from analyses. The raters were paid a small stipend and invited to a buffet dinner for their effort.

5.5 Descriptive Statistics Overview

This section provides descriptive statistics of Experiment 1b. It is a descriptive illustration of the data, which does not contain inferential analysis, such as ANOVA (the statistically analyzed results are provided in the next section). The results graphed below were collapsed across factors, and the native data were replotted here together for visual comparison.

5.5.1 Overall VOT by Lexical Tone

Figure 23 provides the overall VOT values by tone and by language groups. As can be seen from the graph, the ENG group had the longest mean VOT at 89.00 msec (SD = 32.51), the Taiwan group the second at 84.79 msec (SD = 24.53), the SPN group the third at 66.17 msec (SD = 48.86), and the JPN group the shortest at 63.56 msec (SD = 33.13). All groups showed that VOTs were longer in T3 (M = 85.63, SD = 39.30) and T2 (M = 81.60, SD = 36.00) than in T1 (M = 69.22, SD = 31.64) and T4 (M = 68.63, SD = 31.53). VOTs were also observed longer in the ENG and Taiwan groups than those in the JPN and SPN groups.

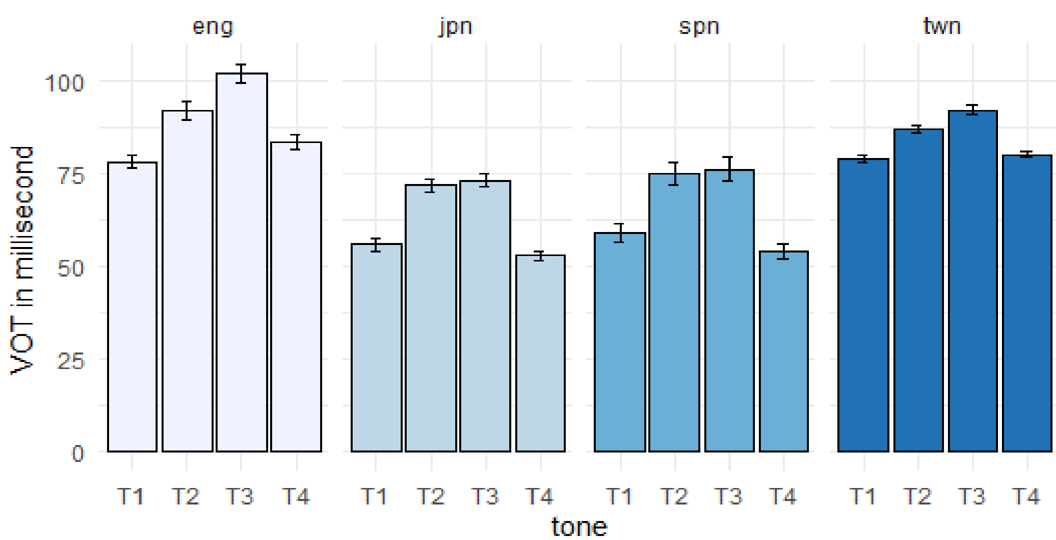

Figure 23. Overall VOT values by lexical tone and by language group

5.5.2 Overall VOT by Place of Articulation

The overall VOT values by POA were 68.99 msec (SD = 34.36) for /p/, 72.94 msec (SD = 34.87) for /t/, and 86.90 msec (SD = 34.96) for /k/. As shown in Figure 24, all groups showed that the further back the place of articulation, the longer the VOT value. Still, the ENG group and the Taiwan group showed longer VOT values than the JPN and SPN groups did. All four groups also displayed smaller VOT differences between bilabial and alveolar plosives than to the velar plosive.

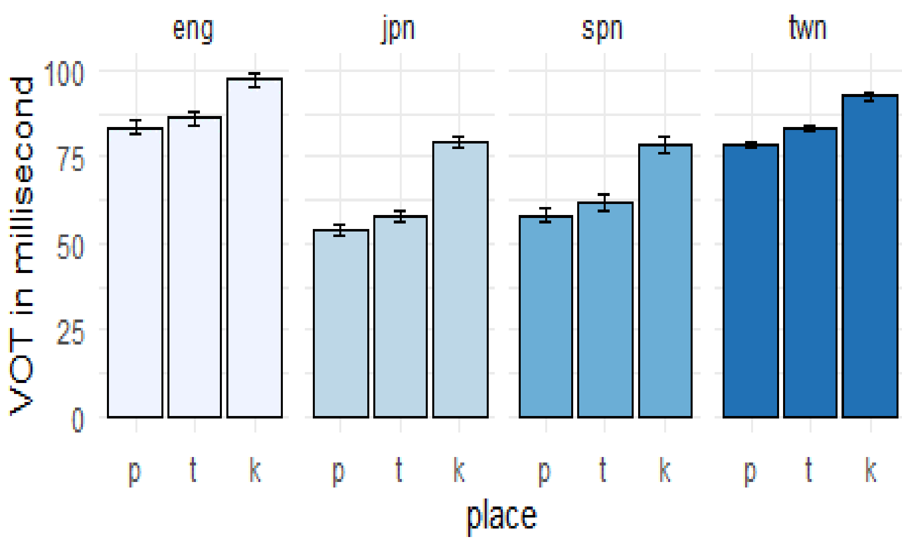

Figure 24. Overall VOT values by place and by language group

5.5.3 Overall VOT by Speech Rate

Figure 25 summarizes the average VOT values by speech rate. Overall, the values by speeds were 68.37 msec (SD = 30.99), 77.20 msec (SD = 34.72), and 83.21 msec (SD = 38.94) for *fast*, *natural*, and *slow*, respectively. The ENG and Taiwan groups exhibited longer VOTs than the JPN and SPN groups, and the faster speed led to the shorter VOT value in all groups (one exception in the SPN group, discussed in 5.9.3).

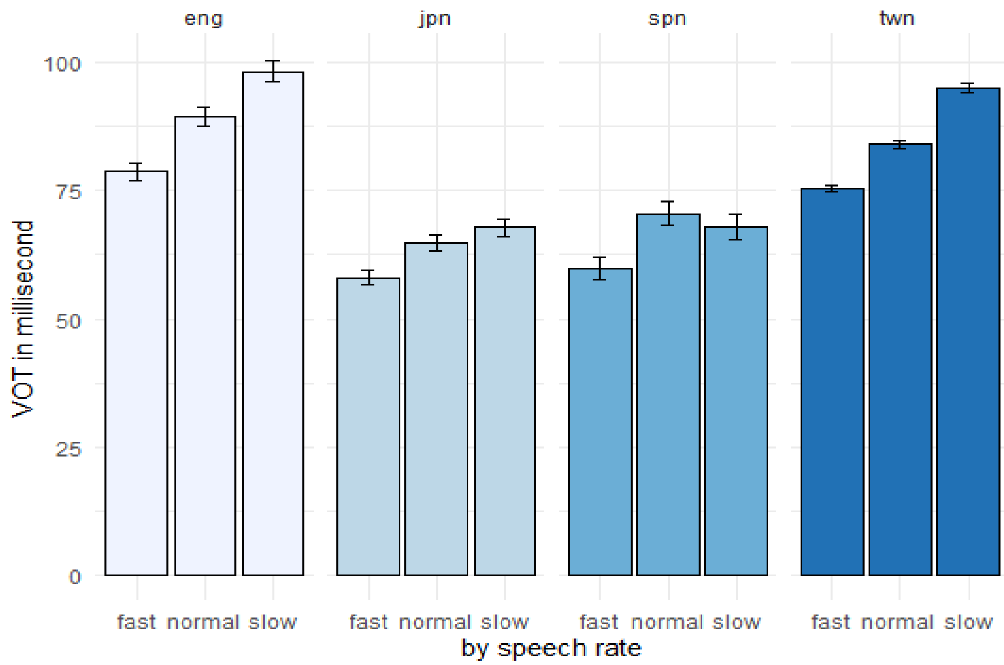

Figure 25. Overall VOT values by speech rate and by language group

5.5.4 Overall VOT by Phrase-position

Figure 26 summarizes the mean VOT values in three phrase-positions. The overall VOT values were <u>75.70 msec</u> (SD = 35.18), <u>75.34 msec</u> (SD = 34.72), and <u>78.31 msec</u> (SD = 37.06) at the utterance-*initial*, utterance-*medial*, and utterance-*final* positions. Stops in words at the utterance-*final* position, in general, showed longer VOT values than in the other two positions across all groups, except for the JPN group. The differences were trivial within each group concerning this effect.

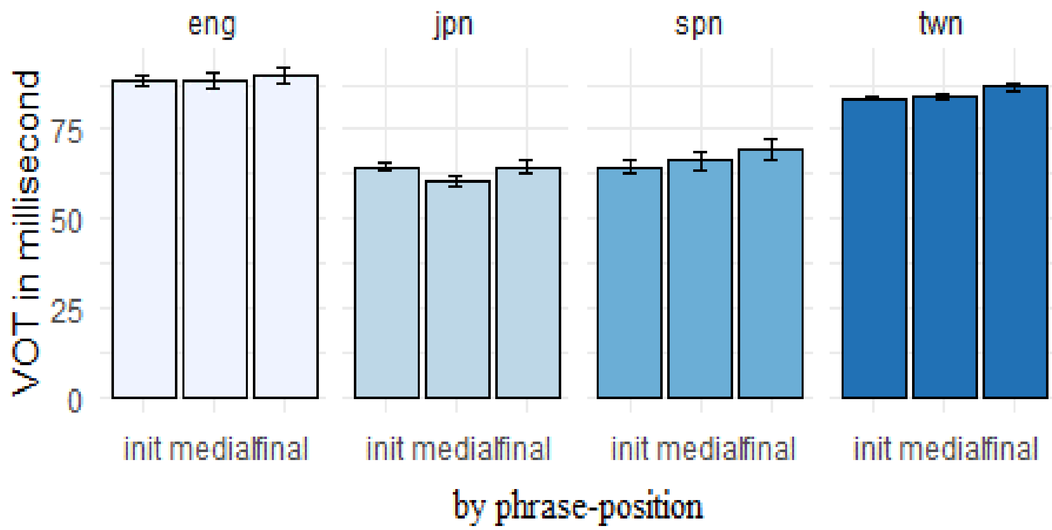

Figure 26. Overall VOT values by phrase-position and by language groups

5.5.5 Overall VOT by Gender

Finally, Figure 27 provides a summary of the VOT values by gender. The overall average values were <u>77.83 msec</u> for the males (SD = 33.56) and <u>74.94 msec</u> for the females (SD = 37.14). As can be seen, the VOT values were not consistent across language groups. The ENG and Taiwan groups showed longer VOTs than the JPN and SPN groups; however, males had longer VOTs than females in the ENG and JPN groups, but the opposite pattern in the SPN and Taiwan groups. The differences are analyzed and discussed in the later section and Chapter 7.

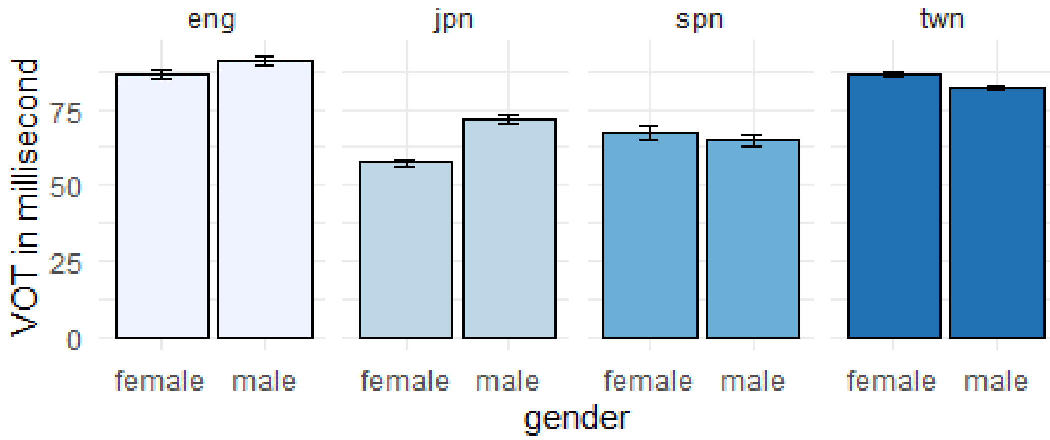

Figure 27. Overall VOT values by gender and by language group

5.6 The Statistical Model

The exact statistical modeling for the Taiwan group analysis was taken (see section 3.7 - 3.8), adding another fixed effect, native language (NL). This step was necessary to ensure that different language groups are significantly different before building separate mixed models for each language group. We could then a) decrease the complexity of the statistical model and b) look at the effects without the influence of the NL factor before investigating the main effects within each group.

The entire dataset included 164 test subjects (96 L2 speakers & 68 L1 speakers, 23616 data points). Each fixed effect was tested with the Likelihood Ratio tests for checking their predicting power. Table 20 shows the Likelihood Ratios test results, which suggest that tone, POA, speech rate, phrase-position, and NL were significant predictors; only the gender factor was not. Therefore, the group model included six fixed-effects and up to two-level interactions between gender and others. By-subject adjustments to the random slopes of tone and POA and random intercepts for by-subjects and by-items were also included. The group model did not err any messages. Pairwise comparisons were conducted using Tukey's HSD tests implemented in the *emmeans* (Lenth, 2020) package. This group model was also used for analyzing group comparison discussed in section 5.10.

Table 20. Results of Likelihood Ratio Tests of the fixed effects

Tone:	$\chi^2 = 31.343$, $p < 0.0001$***
POA:	$\chi^2 = 33.344$, $p < 0.0001$***
Speed:	$\chi^2 = 198.40$, $p < 0.0001$***
Native language:	$\chi^2 = 34.865$, $p < 0.0001$***
Phrase-position:	$\chi^2 = 18.252$, $p < 0.0001$***
Gender:	$\chi^2 = 0.5761$, $p = 0.4479$

5.6.1 Result Overview

The group model confirmed a strong NL effect (F (3, 164) = 12.6368, $p <$ 0.0001). Therefore, separate models were built and analyzed. Each group's analysis showed that the general results were similar between groups but slightly different in numeric values and significance of pairwise comparison by predictors. A result overview is given, and each group's results are provided individually.

Separate post-hoc result summaries of each group's main effects are provided in the order of the ENG, JPN, and SPN groups, and then the group comparison (adding the Taiwan group; for detailed L1 results, see section 3.8 - 3.9). In other words, the VOT differences between groups were significant. The post-hoc results of the ENG group are provided first.

5.7 Results of the English Group

As expected, the ENG group's analysis showed that the general results were similar to those of the Taiwan group, where tone, POA, and speech rate significantly affected VOT. None of the interactions was significant in this group, although the model revealed two marginal two-way interactions between POA & phrase-position and speech rate & phrase-position. These two marginal interactions are interpreted in each of the subsections of the factors involved. The post-hoc results and estimated mean values of each effect for the ENG group are provided in Table 21 and the following subsections.

Table 21. Summary of Type III ANOVA table with Satterthwaite's method of the mixed-effects and interactions between effects for the ENG group

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	p.value
Tone (T)	17963	5988	3	22.25	13.5734	3.008e-05 ***
POA (P)	26486	13243	2	34.39	30.0207	2.864e-08 ***
Speed (S)	165219	82609	2	2947.2	187.266	< 2.2e-16 ***
Phrase Pos. (Pos)	725	363	2	24.61	0.8219	0.45131
Gender (G)	149	149	1	22.01	0.3379	0.56694
T:P	2616	436	6	2943.11	0.9883	0.43132
T:S	4110	685	6	2952.68	1.5529	0.15694
P:S	558	139	4	2950.89	0.3160	0.86744
T:Pos	2110	352	6	2943.03	0.7974	0.57185
P:Pos	3682	920	4	2943.03	2.0867	0.07996 ·
S:Pos	3644	911	4	2946.91	2.0651	0.08279 ·
T:G	737	246	3	21.69	0.5567	0.64919
P:G	1349	675	2	31.74	1.5295	0.23217
S:G	998	499	2	2944.79	1.1315	0.32269
Pos:G	962	481	2	24.61	1.0906	0.35169
T:P:S	4793	399	12	2943.10	0.9054	0.54070
T:P:Pos	4743	395	12	2942.99	0.8959	0.55051
T:S:Pos	769	64	12	2943.03	0.1452	0.99971
P:S:Pos	1373	172	8	2934.03	0.3891	0.92700
T:P:S:Pos	4595	191	24	2943.00	0.4340	0.99249

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.7.1 Lexical Tone Effect in the English Group

The general pattern for VOTs in tones was T3 > T2 > T4 > T1; the mean values were: 102.7 msec (SE = 5.04, df = 22.0, lower.CL = 92.2, upper.CL = 113.1), 92.2 msec (SE = 4.26, df = 22.2, lower.CL = 83.4, upper.CL = 101.1), 83.2 msec (SE = 4.18, df = 22.2, lower.CL = 74.5, upper.CL = 91.8), and 78.5 msec (SE = 3.30, df = 22.1, lower.CL = 71.6, upper.CL = 85.3), respectively.

As shown from the ANOVA results, the tone had a significant effect on VOT in this group, too (F (3, 22) = 13.57, p < 0.0001). There was no significant interaction between tone and other factors. The post-hoc analysis showed that VOTs in T1 were significantly shorter than VOTs in T3 and T2, but not to those in T4. VOTs in T2 were significantly shorter than those in T3, and T4 VOTs were significantly shorter than those in T3, but only marginally different than those in T2 (see Table 22).

Table 22. Tukey HSD Post-hoc analysis for lexical tone main effect for the ENG group

Contrast	estimate	SE	df	t.ratio	p.value	
T1 - T2	-13.75	3.13	22.2	-4.392	0.0012	**
T1 - T3	-24.21	3.83	22.4	-6.329	< .0001	***
T1 - T4	-4.67	2.98	22.5	-1.568	0.4159	
T2 - T3	-10.46	3.12	21.6	-3.351	0.0144	*
T2 - T4	9.08	3.53	22.6	2.573	0.0753	·
T3 - T4	19.54	4.50	22.1	4.339	0.0014	**

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.7.2 POA Effect in the English Group

As expected, POA effect on VOT was found in the ENG group (F (2, 34) = 30.021, p < 0.0001). Their VOTs were the longest in /k/ (M = 97.3 msec; SE = 3.67, DF

= 22.1, lower.CL = 89.7, upper.CL = 104.9), the second in /t/ (M = 86.1 msec; SE = 3.86, DF = 22.1, lower.CL = 78.1, upper.CL = 94.1) and the third in /p/ (M = 84.1 msec; SE = 3.70, DF = 22.1, lower.CL = 76.4, upper.CL = 91.7).

The velar stop /k/ was significantly longer than /t/ and /p/, but the difference between /p/ - /t/ pair was not significant (t (70.2) = -2.02, p = 0.1523). The model also showed a marginal two-way interaction between POA and phrase-position (F (4, 2943) = 2.0867, p = 0.0799). The marginal interaction and the main effect were generally the same that VOTs were longer in /k/ than in /t/ and /p/ across three phrase-positions; VOT difference between /t/ - /p/ was not significant in the main or the two-way interaction, except in the *initial* phrase-position (see Table 23).

Table 23. Tukey HSD Post-hoc analyses for POA and the marginal two-way interaction results between POA & phrase-positions for the ENG group.

Contrast	estimate	SE	df	t.ratio	p.value	
/k/ - /p/	13.24	1.71	23.9	7.748	< .0001	***
/k/ - /t/	11.21	1.79	23.3	6.256	< .0001	***
/p/ - /t/	-2.02	1.08	70.2	-1.879	0.1523	
POA : Phrase-position						
in *initial*						
/k/ - /p/	15.70	1.92	37.9	8.181	< .0001	***
/k/ - /t/	12.32	1.99	35.7	6.177	< .0001	***
/p/ - /t/	-3.83	1.39	186.7	-2.436	0.0416	*
in *medial*						
/k/ - /p/	9.77	2.32	80.7	4.208	0.0002	***
/k/ - /t/	10.90	2.39	72.9	4.569	0.0001	***
/p/ - /t/	1.13	1.91	579.7	0.595	0.8229	
in *final*						
/k/ - /p/	14.24	2.33	81.5	6.118	<.0001	***
/k/ - /t/	10.42	2.39	73.1	4.361	0.0001	***
/p/ - /t/	-3.83	1.91	585.2	-2.001	0.1130	

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.7.3 Speech Rate Effect in the English Group

With respect to the speech rate, VOTs in words at *slow* speed were the longest (M = 98.7 msec; SE = 3.68, DF = 23.0, lower.CL = 91.1, upper.CL = 106.3), the second at *natural* speed (M = 88.9 msec; SE = 3.68, DF = 23.1, lower.CL = 81.3, upper.CL = 96.5), and the shortest at *fast* speed (M = 79.8 msec; SE = 3.68, DF = 23.1, lower.CL = 72.2, upper.CL = 87.4). There was no significant interaction, but a marginal two-way interaction between speech rate and phrase-position (F (4, 2947) = 2.0651, p = 0.0828). The model's main effect analysis revealed that VOT values at *fast* speech rate were significantly shorter than those at *slow* and *natural* speech rates. The post-hoc analysis of the interaction showed that all differences between all speech rates in all phrase-positions were significant (see Table 24). The temporal effect found in the Taiwan group was confirmed here from the ENG group.

Table 24. Tukey HSD Post-hoc analyses for speech rate and the marginal two-way interaction results between speech rate & phrase-positions for the ENG group.

Contrast	estimate	SE	df	t.ratio	p.value	
fast - natural	-9.15	0.997	2950	-9.183	< .0001	***
fast - slow	-18.97	0.981	2946	-19.336	< .0001	***
natural - slow	-9.82	0.979	2946	-10.032	< .0001	***
Speech Rate : Phrase-position						
in *initial*						
fast - natural	-10.33	1.33	2955	-7.737	< .0001	***
fast - slow	-18.46	1.32	2950	-14.036	< .0001	***
natural - slow	-8.13	1.31	2948	-6.201	< .0001	***
in *medial*						
fast - natural	-8.20	1.87	2944	-4.376	< .0001	***
fast - slow	-16.22	1.85	2945	-8.744	< .0001	***
natural - slow	-8.02	1.85	2945	-4.340	< .0001	***
in *final*						
fast - natural	-8.93	1.88	2942	-4.743	< .0001	***
fast - slow	-22.24	1.86	2947	-11.976	< .0001	***

natural - slow	-13.30	1.86	2946	-7.162	< .0001	***

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.7.4 Phrase-position Effect in the English-Group

The statistical model showed no main effect of phrase-position in all conditions. VOT values were 88.6 msec (SE = 3.69, DF = 22.1, lower.CL = 81.0, upper.CL = 96.3), 88.6 msec (SE = 3.94, DF = 22.0, lower.CL = 80.4, upper.CL = 96.8) and 90.2 msec (SE= 3.65, DF = 22.0, lower.CL = 82.6, upper.CL = 97.8) for utterance-*initial*, utterance-*medial* and utterance-*final* positions, respectively. None of the VOT differences shown in Figure 26 above was statistically significant for the ENG group.

5.7.5 Gender Effect in the English-Group

Lastly, females showed slightly shorter VOT values (M = 87.0 msec; SE= 5.15, DF = 22, lower.CL = 76.4, upper.CL = 97.7) than males (M = 91.3 msec; SE = 5.14, DF = 22, lower.CL = 80.6, upper.CL = 101.9). The difference between genders was non-significant (t (22) = -0.581, p = 0.5669). The non-significant gender effect was the same found in the Taiwan group.

5.8 Results of the Japanese Group

The general outcomes were similar to those of the Taiwan and ENG groups, but the JPN group was the only group that showed a significant gender effect[10]. ANOVA

[10] The possibility for this significant observation can be the F0 because we look at tones' F0 onset and females are generally have higher pitch; however the other groups did not show a significant effect. Another possibility is the sociolinguistic factor that we should discuss in Chapter 7.

results showed that tone, POA, speech rate, phrase-position, and gender were significant factors. The model also revealed five significant two-way interactions to be discussed later. No three-way nor four-way interactions were found. Table 25 provides the ANOVA results of this group.

Table 25. Summary of Type III ANOVA table with Satterthwaite's method of the mixed-effects and interactions between effects for the JPN group.

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	p.value
Tone (T)	30389	10230	3	41.6	31.2030	9.931e-11 ***
POA (P)	62124	31062	2	41.7	94.7455	3.126e-16 ***
Speed (S)	82745	41372	2	5439.2	126.1946	< 2.2e-16 ***
Phrase Pos. (Pos)	7208	3604	2	40.0	10.9932	0.0001568 ***
Gender (G)	1664	1664	1	40.0	5.0746	0.0298339 *
T:P	5114	852	6	5439.3	2.6000	0.0161741 *
T:S	2795	466	6	5439.3	1.4211	0.2022412
P:S	8671	2168	4	5439.2	6.6118	2.643e-05 ***
T:Pos	5731	955	6	5439.3	2.9135	0.0077352 **
P:Pos	6075	1519	4	5439.2	4.6322	0.0009848 ***
S:Pos	2259	565	4	5439.1	1.7225	0.1419788
T:G	417	139	3	40.0	0.4239	0.7369020
P:G	192	96	2	40.0	0.2921	0.7482867
S:G	4374	2187	2	5439.2	6.6712	0.0012772 **
Pos:G	205	102	2	40.0	0.3119	0.7337896
T:P:S	2666	222	12	5439.3	0.6776	0.7746879
T:P:Pos	2196	183	12	5439.2	0.5582	0.7967806
T:S:Pos	2611	218	12	5439.3	0.6638	0.7877288
P:S:Pos	1988	249	8	5439.2	0.7580	0.6400697
T:P:S:Pos	6459	269	24	5439.2	0.8209	0.7133261

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.8.1 Lexical Tone Effect in the Japanese Group

Their mean VOT values were 73.8 msec in T3 (SE = 3.86, DF = 40.1, lower.CL = 66.0, upper.CL = 81.5), 72.8 msec in T2 (SE = 3.62, DF = 40.1, lower.CL = 65.5, upper.CL = 80.1), 56.9 msec in T1 (SE = 3.26, DF = 40.1, lower.CL = 50.3, upper.CL = 63.4), and 53.9 msec in T4 (SE = 2.71, DF = 40.2, lower.CL = 48.4, upper.CL = 59.3), respectively.

The model revealed two two-way interactions between tone & POA (*F* (6, 5439) = 2.6000, *p* = 0.016) and tone & phrase-position (*F* (6, 5439) = 2.9135, *p* = 0.0077). Post-hoc analyses for the interactions showed that VOTs in T3 and T2 were significantly longer than those in T1 and T4 in all three POAs and all phrase-positions. The VOT differences between T2 & T3 and T1 & T4 were non-significant in all three POAs and phrase-positions. Table 26 provides detailed post-hoc results.

Table 26. Tukey HSD Post-hoc Analyses for significant interaction results between tone & POA and tone & phrase-positions for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
Tone : POA						
in /p/						
T1 - T2	-14.309	2.44	59.2	-5.871	< .0001	***
T1 - T3	-14.411	2.55	57.1	-5.649	< .0001	***
T1 - T4	2.993	1.74	94.9	1.724	0.3172	
T2 - T3	-0.102	1.78	90.7	-0.057	0.9999	
T2 - T4	17.302	2.25	64.0	7.685	< .0001	***
T3 - T4	17.403	2.49	58.2	6.989	< .0001	***
in /t/						
T1 - T2	-15.289	2.44	59.3	-6.271	< .0001	***
T1 - T3	-15.969	2.55	57.1	-6.260	< .0001	***
T1 - T4	4.135	1.74	95.3	2.379	0.0880	·
T2 - T3	-0.680	1.78	90.5	-0.383	0.9808	
T2 - T4	19.424	2.25	64.1	8.625	< .0001	***
T3 - T4	20.104	2.49	58.2	8.073	< .0.001	***
in /k/						
T1 - T2	-18.220	2.44	59.3	-7.474	< .0001	***
T1 - T3	-20.331	2.55	57.2	-7.966	< .0001	***
T1 - T4	1.820	1.74	95.3	1.048	0.7219	
T2 - T3	-2.111	1.78	90.9	-1.187	0.6368	
T2 - T4	20.040	2.25	64.1	8.899	< .0001	***

T3 - T4	22.151	2.49	58.3	8.891	< .0001	***
Tone : Phrase-position						
in *initial*						
T1 - T2	-16.957	2.31	47.8	-7.341	< .0001	***
T1 - T3	-19.683	2.43	47.0	-8.102	< .0001	***
T1 - T4	2.165	1.55	60.8	1.395	0.5076	
T2 - T3	-2.727	1.60	59.3	-1.706	0.3295	
T2 - T4	19.121	2.11	49.6	9.053	< .0001	***
T3 - T4	21.848	2.37	47.4	9.237	<.0.001	***
in *medial*						
T1 - T2	-13.591	2.50	65.5	-5.439	< .0001	***
T1 - T3	-13.766	2.61	62.5	-5.274	< .0001	***
T1 - T4	3.3281	1.82	115.1	1.826	0.2663	
T2 - T3	-0.1748	1.86	108.9	-0.094	0.9997	
T2 - T4	16.919	2.32	71.9	7.298	< .0001	***
T3 - T4	17.094	2.55	64.0	6.701	<.0.001	***
in *final*						
T1 - T2	-17.280	2.50	65.5	-6.910	< .0001	***
T1 - T3	-17.261	2.61	62.5	-6.613	< .0001	***
T1 - T4	3.4547	1.82	115.2	1.895	0.2358	
T2 - T3	0.0088	1.86	108.8	0.005	1.0000	
T2 - T4	20.725	2.32	72.0	8.940	< .0001	***
T3 - T4	20.716	2.55	64.1	8.121	<.0.001	***

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.8.2 POA Effect in the Japanese Group

The POA-dependent VOT values in /p, t, k / were 54.6 msec (SE = 3.11, DF = 40.1, lower.CL = 48.3, upper.CL= 60.9), 58.3 msec (SE = 3.13, DF = 40.1, lower.CL = 51.9, upper.CL = 64.6), and 80.1 msec (SE = 3.68, DF = 40.1, lower.CL = 72.7, upper.CL = 87.6), respectively.

The model showed three significant two-way interactions between POA and tone (F (6, 5439) = 2.6000, p = 0.016), speech rate (F (4, 5439) = 6.6118, p < 0.0001), and phrase-position (F (4, 5439) = 4.6322, p = 0.0010). VOTs in /k/ were longer than those in

/t/ and /p/ in all four tone, except those in between /t/ and /p/ in T1 and T4. All POA-dependent VOT differences in the *fast* speech rate were significant, except for those in between /t/ and /p/ in *normal* and *slow* speech rates. Similarly, VOTs in /k/ were longer than those in /t/ and /p/ in all three phrase-positions, except for the VOT differences between /t/ and /p/ in *final* phrase-position. See Table 27 for detailed post-hoc results.

Table 27. Tukey HSD Post-hoc Analyses for significant interaction results between POA & tone, POA & speech rate, and POA & phrase-positions for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
POA : Tone						
in T1						
/k/ - /p/	22.80	2.16	71.4	10.559	< .0001	***
/k/ - /t/	19.47	2.34	64.7	8.306	< .0001	***
/p/ - /t/	-3.33	1.53	159	-2.174	0.0788	·
in T2						
/k/ - /p/	26.71	2.16	71.5	12.366	< .0001	***
/k/ - /t/	22.40	2.34	64.7	9.557	< .0001	***
/p/ - /t/	-4.31	1.53	159.4	-2.812	0.0152	*
in T3						
/k/ - /p/	28.72	2.16	71.6	13.292	< .0001	***
/k/ - /t/	23.83	2.34	64.8	10.166	< .0001	***
/p/ - /t/	-4.89	1.53	159.1	-3.191	0.0048	**
in T4						
/k/ - /p/	23.97	2.16	71.5	11.095	< .0001	***
/k/ - /t/	21.79	2.35	64.9	9.288	< .0001	***
/p/ - /t/	-2.19	1.53	159.9	-1.426	0.3299	
POA : Speech Rate						
in *fast*						
/k/ - /p/	23.83	2.07	60.3	11.512	< .0001	***
/k/ - /t/	18.03	2.26	56.2	7.968	< .0001	***
/p/ - /t/	-5.81	1.40	112.4	-4.141	0.0002	***
in *natural*						
/k/ - /p/	24.47	2.07	60.4	11.818	< .0001	***
/k/ - /t/	22.47	2.26	56.2	9.934	< .0001	***
/p/ - /t/	-2.00	1.40	112.5	-1.426	0.3311	
in *slow*						
/k/ - /p/	28.35	2.07	60.3	13.692	< .0001	***
/k/ - /t/	25.12	2.26	56.1	11.107	< .0001	***
/p/ - /t/	-3.23	1.40	112.5	-2.299	0.0601	·

POA : Phrase-position						
in *initial*						
/k/ - /p/	24.38	1.96	48.2	12.456	< .0001	***
/k/ - /t/	20.05	2.16	46.6	9.289	< .0001	***
/p/ - /t/	-4.33	1.23	66.7	-3.517	0.0023	**
in *medial*						
/k/ - /p/	24.15	2.12	66.9	11.367	< .0001	***
/k/ - /t/	19.87	2.31	61.2	8.596	< .0001	***
/p/ - /t/	-4.28	1.48	139.6	-2.886	0.0125	*
in *final*						
/k/ - /p/	28.12	2.12	67.0	13.235	< .0001	***
/k/ - /t/	25.69	2.31	61.2	11.113	< .0001	***
/p/ - /t/	-2.43	1.48	139.8	-1.641	0.2321	

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.8.3 Speech Rate Effect in the Japanese Group

As expected, the speech rate had a strong effect on VOT. The mean values for speech rates were <u>58.9 msec</u> (SE = 3.18, DF = 41, lower.CL = 52.5, upper.CL = 65.3), <u>65.5 msec</u> (SE = 3.18, DF = 41, lower.CL = 59.1, upper.CL = 71.9), and <u>68.6 msec</u> (SE = 3.18, DF = 41, lower.CL = 62.1, upper.CL = 75.0) for *fast*, *natural*, and *slow*, respectively.

There were two two-way significant interactions between speech rate & POA (F (4, 5439) = 6.6118, p < 0.0001) and speech rate & gender (F (2, 5439) = 6.6712, p = 0.0013). The post-hoc results showed that all VOT differences between speech rates were significant, except for those in /p/ and in *females* between *natural* and *slow* speech rates (see Table 28).

Table 28. Tukey HSD Post-hoc Analyses for significant interaction results between speech rate & POA and speech rate & gender for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
Speech Rate : POA						
in /p/						
fast - natural	-7.66	1.07	5439	-7.161	< .0001	***
fast - slow	-9.03	1.07	5439	-8.441	< .0001	***
natural - slow	-1.37	1.07	5439	-1.279	0.4068	
in /t/						
fast - natural	-3.85	1.07	5439	-3.600	0.0009	***
fast - slow	-6.45	1.07	5439	-6.023	< .0001	***
natural - slow	-2.59	1.07	5439	-2.423	0.0408	*
in /k/						
fast - natural	-8.30	1.07	5439	-7.740	< .0001	***
fast - slow	-13.54	1.07	5439	-12.641	< .0001	***
natural - slow	-5.25	1.07	5439	-4.896	< .0001	***
POA : Gender						
in *female*						
fast - natural	-8.69	0.795	5439	-10.936	< .0001	***
fast - slow	-10.25	0.795	5439	-12.887	< .0001	***
natural - slow	-1.55	0.795	5439	-1.951	0.1247	
in *male*						
fast - natural	-4.52	0.918	5439	-4.921	< .0001	***
fast - slow	-9.10	0.917	5439	-9.925	< .0001	***
natural - slow	-4.59	0.918	5439	-5.000	< .0001	***

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.8.4 Phrase-position Effect in the Japanese Group

The mean VOT values were 65.6 msec (SE = 3.14, DF = 40, lower.CL = 55.4, upper.CL = 68.1), 61.8 msec (SE = 3.11, DF =40, lower.CL = 59.3, upper.CL = 71.9), and 65.5 msec (SE = 3.47, DF = 40, lower.CL = 58.5, upper.CL = 72.6) for utterance-*initial*, utterance-*medial*, and utterance-*final* positions, respectively.

The phrase-position main effect was significant in this group (F (2, 40) = 10.9932, p = 0.0002), and the model revealed two significant two-way interactions between tone & phrase-position (F (6, 5439) = 2.9135, p = 0.0077) and POA & phrase-position (F (4, 5439) = 4.6322, p = 0.0010). The post-hoc results showed that VOTs in the *medial* phrase-position were significantly shorter than those in the *final* and *initial* phrase-positions only in T2 and T3. Similarly, only VOTs in the phrase *medial* position tend to be significantly shorter than those in the *final* and *initial* phrase-positions only in /k/. VOTs in /t/ and /p/ in the *medial* position were also shorter than those in the *initial* phrase-position (see Table 29).

Table 29. Tukey HSD Post-hoc Analyses for significant interaction results between phrase-position & tone and phrase-position & POA for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
Phrase-position : Tone						
in T1						
final - initial	0.761	1.85	81.0	0.410	0.9114	
final - medial	1.981	1.63	165.7	1.212	0.4477	
initial - medial	1.220	1.49	135.5	0.817	0.6933	
in T2						
final - initial	1.075	1.85	81.0	0.580	0.8314	
final - medial	5.661	1.63	165.9	3.462	0.0020	**
initial - medial	4.586	1.49	135.9	3.069	0.0073	**
in T3						
final - initial	-1.660	1.85	81.0	-0.895	0.6448	
final - medial	5.447	1.64	166.0	3.350	0.0029	**
initial - medial	7.137	1.49	135.9	4.776	< .0001	***
in T4						
final - initial	-0.529	1.86	81.2	-0.285	0.9562	
final - medial	1.854	1.64	166.6	1.133	0.4952	
initial - medial	2.382	1.50	136.2	1.594	0.2517	
Phrase-position : POA						
in /p/						
final - initial	-0.705	1.76	65.8	-0.401	0.9154	
final - medial	3.033	1.49	114.56	2.037	0.1079	
initial - medial	3.738	1.37	97.6	2.720	0.0209	*
in /t/						
final - initial	-2.600	1.76	65.8	-1.477	0.3084	
final - medial	1.189	1.76	65.8	-1.477	0.3084	
initial - medial	3.788	1.38	97.8	2.755	0.0190	*

in /k/						
final - initial	3.040	1.76	65.9	1.727	0.2030	
final - medial	7.007	1.49	115.0	4.702	< .0001	***
initial - medial	3.968	1.38	97.9	2.884	0.0133	*

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.8.5 Gender Effect in the Japanese Group

The mean VOTs were 57.2 msec for the females (SE = 4.12, DF = 40; lower.CL = 48.9, upper.CL = 65.5) and 71.4 msec for the males (SE = 4.79, DF = 40, lower.CL = 61.7, upper.CL = 81.1). The JPN group was the only group that showed a main gender effect on VOT (*F* (1, 40) = 10.9932, *p* = 0.0298), but the model also revealed a significant two-way interaction between speech rate and gender (*F* (2, 5439) = 6.6712, *p* = 0.0012). The post-hoc analysis revealed that VOTs in females were significantly shorter than those in males at both *fast* and *slow* speech rates; at the *natural* speech rate, the difference became marginal (*t* (41) = -1.1861, *p* = 0.0699). Table 30 shows the post-hoc analysis results. What might have caused the significant gender effect is discussed in Chapter 7.

Table 30. Tukey HSD Post-hoc Analyses for significant interaction results between speech rate & gender for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
Gender Main Effect						
female - male	-14.2	6.32	40	-2.253	0.0298	*
Gender : Speech rate						
in *fast*						
female - male	-16.0	6.36	40.9	-2.518	0.0158	*
in *natural*						
female - male	-11.8	6.36	40.9	-1.861	0.069	·
in *slow*						

female - male	-14.9	6.36	40.9	-2.339	0.0243

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.9 Results of the Spanish Group

Although the overall VOT values were much shorter in the SPN group than those in the ENG and Taiwan groups, the general results were similar for each factor. ANOVA results showed the significant main effects of tone, POA, speech rate, and phrase-position. The model also revealed were four significant and one marginal two-way interactions to be discussed later. No three-way nor four-way interactions were found. Table 31 provides the ANOVA results of this group.

Table 31. Summary of Type III ANOVA table with Satterthwaite's method of the mixed-effects and interactions between effects for the SPN group.

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	p.value
Tone (T)	26834	8945	3	35.2	12.8454	7.938e-06 ***
POA (P)	55811	27905	2	35.6	40.0750	7.726e-10 ***
Speed (S)	127160	63580	2	4567.6	91.3082	< 2.2e-16 ***
Phrase Pos. (Pos)	9926	4963	2	34.2	7.1275	0.0025866 **
Gender (G)	2	2	1	34.0	0.0026	0.9597623
T:P	3394	566	6	4563.4	0.8125	0.5600407
T:S	8685	1447	6	4571.9	2.0787	0.0524422 ·
P:S	4796	1199	4	4574.0	1.7218	0.1421632
T:Pos	13594	2266	6	4563.4	3.2537	0.0034124 **
P:Pos	15979	3995	4	4563.4	5.7368	0.0001329 ***
S:Pos	2667	667	4	4579.0	0.9574	0.4296675
T:G	2140	713	3	34.1	1.0244	0.3939881
P:G	336	168	2	33.9	0.2410	0.7872052
S:G	19123	9561	2	4564.6	13.7312	1.134e-06 ***
Pos:G	5653	2826	2	34.2	4.0591	0.0262001 *
T:P:S	8914	743	12	4563.4	1.0667	0.3839779
T:P:Pos	8961	747	12	4563.3	1.0725	0.3788970
T:S:Pos	7340	612	12	4563.4	0.8785	0.5685993
P:S:Pos	1944	243	8	4563.4	0.3489	0.9466964

T:P:S:Pos	11169	564	24	4563.3	0.6683	0.8862442

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.9.1 Lexical Tone Effect in the Spanish Group

The mean VOT values, in T3 > T2 > T1 > T4, were 76.5 msec (SE = 7.98, DF = 34.0, lower.CL = 60.2, upper.CL = 92.7), 75.2 msec (SE = 7.24, DF = 34.1, lower.CL = 60.5, upper.CL = 89.9), 59.4 msec (SE = 5.53, DF = 34.1 lower.CL = 48.2, upper.CL= 70.6) 53.8 msec (SE = 5.18, DF = 34.1, lower.CL = 43.3, upper.CL = 64.4), respectively. There were a significant two-way interaction between tone and phrase-position and one marginal two-way interaction between tone and speech rate, as shown in Table 31. The post-hoc analyses revealed that in all three phrase-positions, VOTs in T3 and T2 were significantly longer than those in T1 and T4. The differences between those in T2 and T3 was not significant, so were not for those between T1 and T4; except for those in T1 in the *final* phrase-position were significantly longer than those in T4 (see Table 32).

Table 32. Tukey HSD Post-hoc Analyses for significant interaction results between tone & phrase-position and the marginal interaction between tone & speech rate for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
Tone : Phrase-position						
in *initial*						
T1 - T2	-15.265	3.08	44.1	-4.957	0.0001	***
T1 - T3	-15.641	4.03	39.4	-3.878	0.0021	***
T1 - T4	3.367	2.79	47.1	1.208	0.6251	
T2 - T3	-0.375	2.95	45.3	-0.127	0.9993	
T2 - T4	18.632	3.83	40.1	4.865	0.0001	***
T3 - T4	19.007	4.34	38.5	4.379	0.0005	***
in *medial*						
T1 - T2	-18.118	3.43	67.9	-5.281	< .0001	***
T1 - T3	-15.184	4.31	51.2	-3.525	0.0048	**
T1 - T4	4.696	3.17	78.8	1.480	0.4542	

T2 - T3	2.935	3.32	72.1	0.885	0.8128	
T2 - T4	22.814	4.12	53.5	5.540	< .0001	***
T3 - T4	19.879	4.60	48.5	4.324	0.0004	***
in *final*						
T1 - T2	-13.945	3.44	68.4	-4.057	0.0007	***
T1 - T3	-20.376	4.31	51.5	-4.725	0.0001	***
T1 - T4	8.660	3.18	79.3	2.725	0.0388	*
T2 - T3	-6.431	3.32	72.5	-1.936	0.2224	
T2 - T4	22.604	4.12	53.7	5.485	< .0001	***
T3 - T4	29.035	4.60	48.6	6.312	< .0001	***
Tone : Speech Rate (marginal)						
in *fast*						
T1 - T2	-13.06	3.33	60.3	-3.919	0.0013	**
T1 - T3	-15.41	4.23	47.6	-3.645	0.0036	**
T1 - T4	4.63	3.06	68.5	1.511	0.4369	
T2 - T3	-2.36	3.22	63.6	-0.733	0.8835	
T2 - T4	17.69	4.04	49.4	4.380	0.0003	***
T3 - T4	20.04	4.52	45.5	4.430	0.0003	***
in *natural*						
T1 - T2	-18.19	3.31	59.1	-5.491	< .0001	***
T1 - T3	-16.05	4.21	47.0	-3.808	0.0022	**
T1 - T4	7.79	3.04	66.8	2.559	0.0601	·
T2 - T3	2.14	3.19	62.1	0.669	0.9082	
T2 - T4	25.98	4.02	48.6	6.464	< .0001	***
T3 - T4	23.84	4.51	44.9	5.287	<.0.001	***
in *slow*						
T1 - T2	-16.08	3.32	59.6	-4.843	0.0001	***
T1 - T3	-19.73	4.22	47.1	-4.679	0.0001	***
T1 - T4	4.30	3.05	67.4	1.411	0.4971	
T2 - T3	-3.65	3.20	62.4	-1.142	0.6650	
T2 - T4	20.39	4.02	48.8	5.066	< .0001	***
T3 - T4	24.04	4.51	44.9	5.329	<.0.001	***

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.9.2 POA Effect in the Spanish Group

The mean VOT values were 79.0 msec in /k/ (SE = 6.30, DF = 34, lowe.CL = 66.2, upper.CL = 91.8), 61.2 msec in /t/ (SE = 6.42, DF = 34, lowe.CL = 48.2, upper.CL = 74.3), and 58.4 msec in /p/ (SE = 6.40, DF = 34, lowe.CL = 45.4, upper.CL = 71.4). The post-hoc of the main POA effect showed that /k/ had significantly longer VOTs than

/t/ and /p/, but VOTs in /p/ - /t/ were not (*t* (36) = -1.639, *p* = 0.2425). The model also revealed a significant two-way interaction between POA and phrase-position (*t* (4563) = 5.7368, *p* = 0.0001). The post-hoc analyses of the interaction showed that the same pattern that VOTs in /k/ were significantly longer than those in /t/ and /p/ at all three speech rates. See Table 33 for detail.

Table 33. Tukey HSD Post-hoc Analyses for the significant interaction results between POA & speech rate for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
POA : Speech Rate						
in *fast*						
/k/ - /p/	20.928	2.72	65.5	7.695	< .0001	***
/k/ - /t/	16.133	2.77	64.0	5.823	< .0001	***
/p/ - /t/	-4.795	2.21	101.2	-2.171	0.0811	·
in *natural*						
/k/ - /p/	18.798	2.70	64.1	6.955	< .0001	***
/k/ - /t/	16.126	2.75	62.6	5.855	< .0001	***
/p/ - /t/	-2.672	2.19	97.9	-1.223	0.4427	
in *slow*						
/k/ - /p/	22.046	2.71	64.6	8.141	< .0001	***
/k/ - /t/	21.147	2.76	63.0	7.667	< .0001	***
/p/ - /t/	-0.899	2.19	98.7	-0.410	0.9115	

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.9.3 Speech Rate Effect in the Spanish Group

For this group, the mean VOTs in three speech rates were <u>58.7 msec</u> (SE = 6.28, DF = 34.6, lowe.CL = 46.0, upper.CL = 71.5), <u>71.6 msec</u> (SE = 6.27, DF = 34.5, lowe.CL = 58.9, upper.CL = 84.4), and <u>68.3 msec</u> (SE = 6.27, DF = 34.6, lowe.CL = 55.5, upper.CL = 81.0) for *fast*, *natural*, and *slow* speeds, respectively.

In the post-hoc analyses for the main effect, VOTs in the *fast* speech rate were significantly shorter than those in *natural* (*t* (4567) = -13.057, *p* < 0.0001) and *slow* (*t* (4570) = -9.620, *p* < 0.0001). Those in *natural* were also significantly longer than those in *slow* (*t* (4566) = 3.431, *p* = 0.0018). It was unexpected to find that VOT values in the *natural* speed be longer than those in the *slow* speed. The preliminary suspicion was that the SPN group's utterance speed in the *natural* speech rate was slower than in the *slow* speech rate. However, a closer examination showed that the mean vowel durations in these two speech rates were 329.43 milliseconds in the *natural* speed vs. 350.49 milliseconds in the *slow* speed (the mean syllable durations were 399.96 milliseconds vs. 418.51 milliseconds). The vowel duration difference between the two was statistically significant (*t* (212) = -9.140. *p* < 0.0001). In other words, the vowel duration was not the cause of the longer VOT in the *natural* speed since the vowel duration increased from *natural* speed to *slow* speed, but VOT decreased in the same setting. The cause might be from other factors.

There was also a significant two-way interaction between speech rate and gender (*F* (2, 4565) = 13.7312, *p* < 0.0001). The post-hoc analyses showed that only VOTs between *natural* and *slow* speech rates in females were non-significant (see Table 34).

Table 34. Tukey HSD Post-hoc Analyses for the significant interaction results between speech rate & gender for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
Speech Rate : Gender						
in *female*						
fast - natural	-8.695	1.30	4567	-6.691	< .0001	***
fast - slow	-9.583	1.30	4567	-7.381	< .0001	***
natural - slow	-0.888	1.28	4564	-0.696	0.7657	

in *male*						
fast - natural	-17.132	1.43	4565	-6.627	< .0001	***
fast - slow	-9.495	1.43	4569	-6.627	< .0001	***
natural - slow	7.637	1.43	4567	5.332	< .0001	***

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.9.4 Phrase-position Effect in the Spanish Group

The mean VOT values were 63.8 msec at utterance-*initial* (SE = 6.06, DF = 34, lowe.CL = 51.5 upper.CL = 76.1), 65.6 msec at the utterance-*medial* (SE = 6.19, DF = 34, lowe.CL = 53.1, upper.CL = 78.2), and 69.2 msec at utterance-*final* (SE = 6.74, DF = 34, lowe.CL = 55.5, upper.CL = 82.9). The model revealed a significant main effect (F (2, 4565) = 13.7312, p < 0.0001) and two significant two-way interaction between POA & phrase-position (F (4, 4563) = 5.7368, p = 0.0001) and phrase-position & gender (F (2, 34) = 4.0591, p = 0.026). The post-hoc analyses showed that only in /k/, VOTs in *final* phrase-position were significantly longer those in *initial* phrase-position, which in turn shorter than those in the *medial* phrase-position; VOTs in *final* phrase-position were significantly longer those in *initial* phrase-position in /p/ (see Table 35). Moreover, VOTs in *final* phrase-position were also significantly longer those in *initial* phrase-position only in male. This finding agrees with Lisker and Abramson's (1967) report that VOT in words at the utterance-*final* position would have longer VOT than at the utterance-*initial* position.

Table 35. Tukey HSD Post-hoc Analyses for the significant interaction results between phrase-position & POA and phrase-position & gender for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value
Phrase-position : POA					
in /p/					

final - initial	6.34	2.06	96.9	3.071	0.0077	**
final - medial	3.07	2.76	70.3	1.115	0.5078	
initial - medial	-3.26	2.01	103.9	-1.624	0.2401	
in /t/						
final - initial	0.49	2.06	96.8	0.238	0.9693	
final - medial	3.43	2.76	70.2	1.246	0.4306	
initial - medial	2.94	2.01	103.7	1.467	0.3111	
in /k/						
final - initial	9.55	2.07	97.5	4.624	< .0001	***
final - medial	4.27	2.76	70.5	1.547	0.2753	
initial - medial	-5.28	2.01	104.4	-2.630	0.0264	*
Phrase-position : Gender						
in *female*						
final - initial	1.261	2.11	354.4	0.597	0.8229	
final - medial	0.621	3.06	34.3	0.203	0.9775	
initial - medial	-0.639	2.02	34.4	-0.316	0.9464	
in *male*						
final - initial	9.659	2.37	34.2	4.068	0.0008	**
final - medial	6.564	3.44	34.1	1.909	0.1515	
initial - medial	-3.095	2.27	34.0	-1.366	0.3698	

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.9.5 Gender Effect in the Spanish Group

Lastly, the mean VOT values were 66.5 msec for the females (SE = 8.30, DF = 34, lower.CL = 49.7, upper.CL = 83.4) and 65.9 msec for the males (SE = 9.34, DF = 34, lower.CL = 46.9, upper.CL = 84.9). The analyses showed that the VOT difference concerning gender was not significant (F (2, 34) = 0.0026, p = 0.9597), but the model also showed two significant two-way interaction between speech rate & gender (F (2, 4565) = 13.731, p < 0.0001), and phrase-position & gender (F (2, 34) = 4.0591, p = 0.026). The post-hoc analyses revealed that the significant interactions found were due to speech rate and phrase-position discussed above; gender difference were all non-significant across all speech rates and phrase-positions (see Table 34, Table 35, & Table 36).

Table 36. Tukey HSD Post-hoc Analyses for the interactions between speech rate & gender and phrase-position & gender for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value
Gender : Speech Rate					
in *fast*					
female - male	3.42	12.5	34.5	0.272	0.7869
in *natural*					
female - male	-5.02	12.5	34.5	-0.400	0.6915
in *slow*					
female - male	3.51	12.5	34.5	0.279	0.7815
Gender : Phrase-position					
in *initial*					
female - male	4.25	12.1	34	0.351	0.7279
in *medial*					
female - male	1.80	12.4	34	0.145	0.8854
in *final*					
female - male	-4.14	13.5	34	-0.307	0.7605

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.10 Language Background Comparison

As predicted, the ENG group and the Taiwan group showed significantly longer VOT values than the SPN and JPN groups. The overall VOTs across factors were: 89.1 msec for the ENG group (SE = 4.65, df = 164, lower.CL = 80.0, upper.CL = 98.3), 85.1 msec for the Taiwan group (SE = 2.65, df = 164, lower.CL = 79.9, upper.CL = 90.3), 66.3 msec for the SPN group (SE = 3.76, df = 165, lower.CL = 59.0, upper.CL = 73.7), 64.2 msec for the JPN group (SE = 4.63, df = 164, lower.CL = 57.4, upper.CL = 71.1).

For the NL effect, ANOVA revealed that NL was a significant predictor (F (3, 164) = 12.6368, $p < 0.0001$). The group model also revealed four significant two-way interactions between tone & NL (F (9, 169.5) = 4.9389, $p < 0.0001$), between POA & NL (F (6, 173.6) = 9.1672, $p < 0.0001$), between speech rate & NL (F (6, 22409) = 60.4706, $p < 0.0001$), and between NL & phrase-position (F (6, 22404) = 8.9097, $p < 0.0001$). Overall, the model revealed the same patterns of the main effect and the two-way interactions. As shown in Figure 28 and Figure 29, the ENG and Taiwan groups' VOT values were significantly longer than those of the SPN and JPN groups across all four tones, three POAs, and three speech rates. Those in between the ENG and Taiwan groups and in between the JPN and SPN groups were non-significant. In other words, the VOT values of the ENG and Taiwan groups were similar, and so were the SPN and JPN groups.

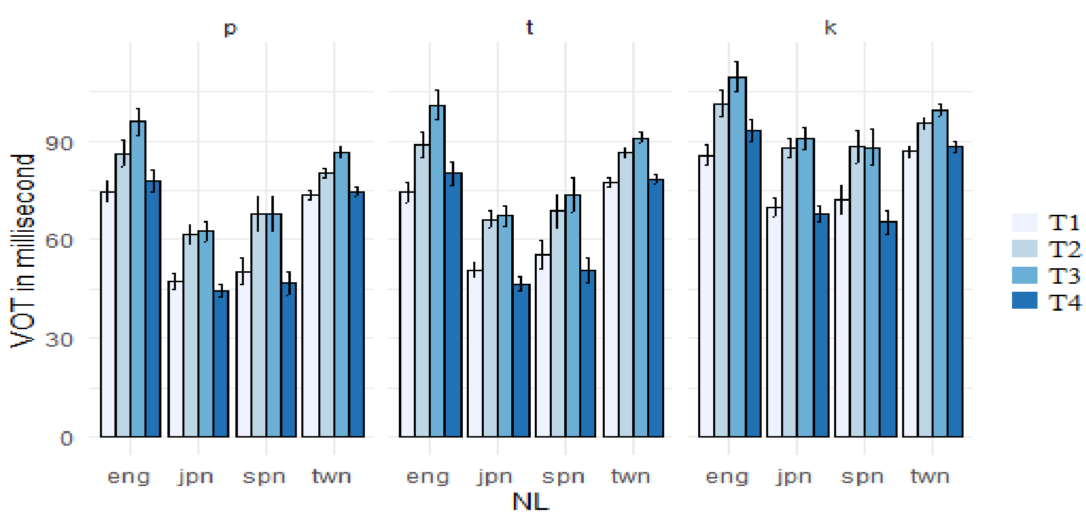

Figure 28. VOT patterns due to NL effect across lexical tones and POAs

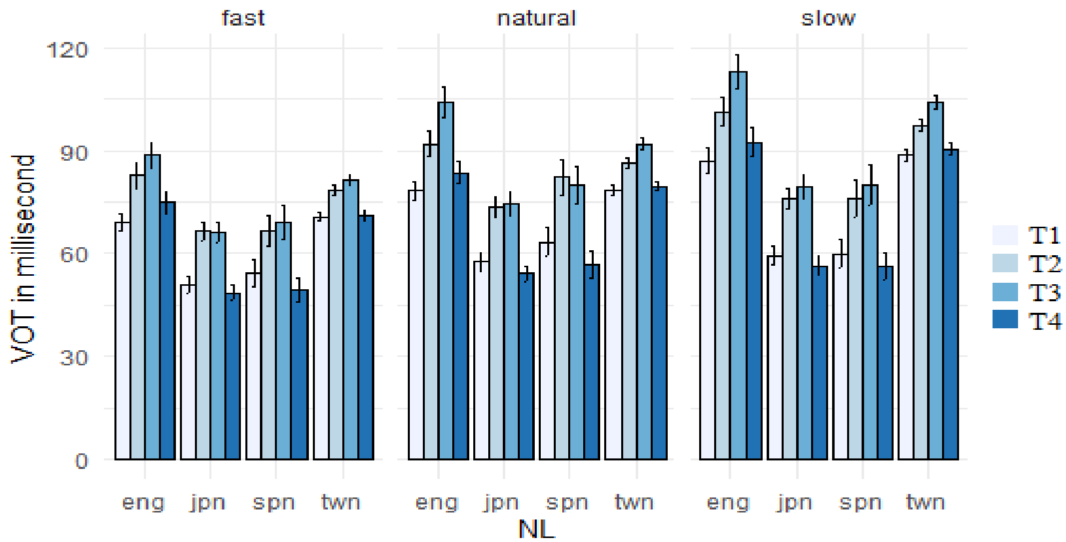

Figure 29. VOT patterns due to NL effect across lexical tones and speech rates

There were also three three-way interactions between tone & NL & phrase-position (F (18, 22405) = 4.7580, p = 0.0028), between POA & NL & phrase-position (F (12, 22405) = 4.7580, p < 0.0001), and between speech rate & NL & phrase-position (F (12, 22404) = 1.7812, p = 0.0452). The post-hoc analyses for the three-way interactions showed the patterns were the same as the two-way interactions with a few marginal exceptions in T2 between NLs and phrase-positions, in /k/ between phrase-positions, and in *natural* speech rate in *final* phrase-position between the groups of ENG & SPN and SPN & Taiwan. Table 37 reports the post-hoc results of the main NL effect, and lists these unexpected marginal results concerning interactions.

Table 37. Tukey HSD Post-hoc Analyses for the NL main effect and significant interactions for the group comparison.

Contrast	estimate	SE	DF	t.ratio	p.value	
Main effect by Language						
ENG - JPN	24.9	2.78	164	4.3111	0.0002	***

ENG - SPN	22.8	5.95	164	3.833	0.0010	***
ENG - Taiwan	4.1	5.32	164	0.770	0.8679	
JPN - SPN	-2.1	5.09	165	-0.412	0.9763	
JPN - Taiwan	-20.8	4.35	165	-4.785	< .0001	***
SPN - Taiwan	-18.7	4.57	164	-4.089	0.0004	***
NL : Tone : Phrase-position						
In T2 in *initial*						
ENG - JPN	16.638	6.67	170	2.493	0.0646	·
JPN - Taiwan	-11.638	5.02	171	-2.319	0.0976	·
SPN - Taiwan	-13.181	5.28	170	-2.496	0.0642	·
In T2 in *medial*						
ENG - SPN	15.856	7.00	183	2.265	0.1101	
SPN - Taiwan	-10.449	5.38	183	-1.944	0.2136	
In T2 in *final*						
ENG - SPN	17.177	7.00	183	2.454	0.0709	·
In T3 in *final*						
SPN - Taiwan	-13.023	5.92	179	-2.199	0.1275	
NL : POA: Phrase-position						
In /k/ in *initial*						
JPN - Taiwan	-12.022	4.66	169	-2.582	0.0517	·
In /k/ in *medial*						
ENG - SPN	15.852	6.48	180	2.448	0.0720	·
SPN - Taiwan	-10.919	4.98	180	-2.194	0.1288	
In /k/ in *final*						
ENG - JPN	15.052	6.29	180	2.393	0.0820	·
ENG - SPN	14.840	6.48	180	2.291	0.1040	
JPN - Taiwan	-10.866	4.73	180	-2.298	0.1023	
SPN - Taiwan	-10.653	4.98	180	-2.140	0.1445	
NL : Speech Rate: Phrase-position						
In *natural* in *final*						
ENG - SPN	13.150	6.15	187	2.139	0.1446	
SPN - Taiwan	-10.261	4.72	187	-2.173	0.1346	

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

5.11 Result Summary

Three L2 groups of speakers participated in this study. Their VOT values were analyzed to explore the relationships between VOT & lexical tone, VOT & POA, VOT & speech rate, VOT & phrase-position, VOT & gender, and VOT & native language background in their Mandarin production. A series of mixed-effects linear regression analyses were performed for a full analysis, and five out of six of the mixed-effects were found to be significant factors.

In general, all groups exhibited similar VOT distribution, where the patterns were from long to short: T3 > T2 > T1 > T4 for lexical tones, /k/ > /t/ > /p/ for POA, and *slow* > *natural* > *fast* for the speech rate. VOTs at the utterance-*final* were the longest, and languages feature aspirated voiceless stops in their L1 had longer VOTs than those who do not. Table 38 below provides the averaged VOT values for all groups' five main effects (included the Taiwan group for convenience).

Table 38. Averaged VOT values for all groups of participants for five main effects: in milliseconds

	Taiwanese	English	Japanese	Spanish
Tone 1	79.2 (22.1)	78.2 (25.8)	55.9 (29.5)	59.2 (42.9)
Tone 2	87.8 (24.5)	92.1 (32.3)	71.9 (32.9)	74.9 (53.1)
Tone 3	92.6 (26.4)	102.0 (36.7)	73.4 (37.1)	76.4 (55.2)
Tone 4	80.3 (22.4)	83.7 (29.2)	53.0 (26.9)	54.2 (38.4)
/p/	78.7 (22.8)	83.6 (31.5)	53.9 (29.4)	58.2 (47.9)
/t/	83.3 (23.8)	86.0 (32.3)	57.7 (30.0)	62.0 (47.9)
/k/	92.4 (24.9)	97.4 (32.2)	79.1 (34.1)	78.4 (48.6)
fast	75.4 (20.6)	78.8 (27.4)	58.0 (48.6)	59.8 (43.6)
natural	84.0 (22.6)	89.4 (30.7)	64.8 (33.0)	70.5 (49.7)
slow	95.1 (26.0)	98.3 (35.7)	67.8 (35.3)	68.0 (52.2)
initial	83.9 (24.6)	88.5 (33.6)	64.5 (33.0)	64.5 (47.4)
medial	84.4 (23.2)	88.6 (30.7)	60.7 (31.7)	66.3 (48.5)
final	87.1 (25.6)	90.3 (32.1)	64.6 (34.6)	69.4 (51.9)
male	82.5 (24.3)	91.0 (34.2)	71.9 (35.8)	64.8 (41.1)
female	86.9 (24.6)	86.8(30.5)	57.4 (29.5)	67.2 (54.3)

Our research question asks whether lexical tone affects VOT in L2 Mandarin. Figure 30 below shows that, overall, VOT values in T3 were the most prolonged and T2 the second across all language groups. This suggests that the tone effect exhibits in L2 speech production and that all speakers had the same tone production mechanism. Furthermore, the post-hoc results showed that even with the tone effect, these L2 VOTs were remarkably different from the native Mandarin speakers (except for the ENG group; t (163) = 0.757, p = 0.8738). This finding indicates an L1 influence; the SPN and JPN groups' VOT values were significantly shorter than the ENG group, as they are in their L1 systems, reported in the literature. Interestingly, these L2 participants' native VOTs were not found in their L2 production. Two groups extended their L2 VOT production significantly (it may not be the ENG group case; see next section).

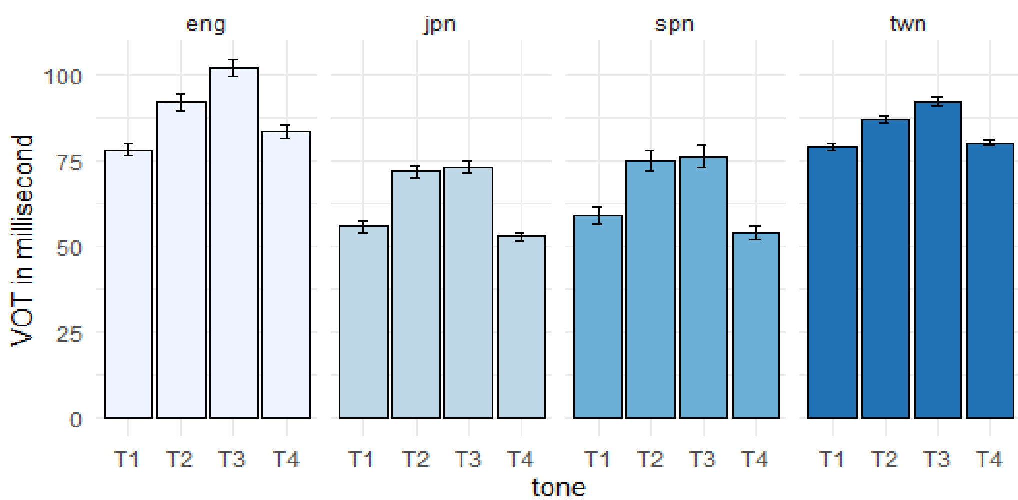

Figure 30. Overall VOT values by lexical tone and by language group

In sum, we report a strong tone effect on VOT in L2 Mandarin, and the POA and speech rate effects were consistent. The phrase-position effect was significant in some conditions, not all. Lastly, this study did not support the gender effect, although Li (2013) and Peng et al. (2014) reported significant VOT variation in Mandarin Chinese due to gender. Future studies are needed for understanding phrase-position and gender effects.

Secondly, one of the current study's hypotheses is whether we find the same effects found in L1 in L2 learners of Mandarin. In other words, does NL matter? Our L2 results confirmed that it does, and within the same conditions, VOTs were significantly longer produced by the ENG and Taiwan groups than by the JPN and SPN groups. Thus, native language background affects L2 VOT duration. The latter two groups also showed extended VOT values compared to their native VOTs (see section 5.12). These findings suggest an influence of L1 phonology in L2 production, and there was a different system that extended the JPN and SPN groups' Mandarin VOT production.

As mentioned in Chapter 2, while there are inconsistent results reported in the literature concerning the lexical tone effect, this study offers supportive evidence from L1 and L2 Mandarin productions.

Furthermore, this study offers comparative results and shows that tone affected VOT in L2 production from three language backgrounds. The significantly short VOT found in the JPN and SPN groups may be due to other reasons, but it is apparent that the tone effect was observed within each group. Thus, this outcome suggests the relationship between VOT and lexical tones in the L2 speakers. The L2 results further support the proposition that the vocal fold tensity may be one reason for the tone effect. Experiment 2b further controls the tone to test the vocal fold tensity proposition with the L2 groups.

5.12 Elicitation of Native VOTs of the L2 speakers – A VOT Baseline Test

The L2 groups completed an additional task for eliciting their native VOTs. This VOT baseline test was conducted to offer more comparative VOT values from the same L2 groups of participants in some regards. The intentions were twofold: a) to provide more comparable values to native VOT values, and b) to glance at our participant's phonetic process of their inter-language (Gass, 2013; Major, 2001).

In general, for monolingual speakers, voiceless stop VOTs have been reported around 58 to 80 milliseconds in English (Cho & Ladefoged, 1999), around 30 to 60 milliseconds in Japanese (Shimizu, 1990; Cho & Ladefoged, 1999), and around 10 to 30 milliseconds in Spanish (Abramson & Lisker, 1972; Cho & Ladefoged, 1999). These values are intended to give a sense of the possible VOT ranges rather than guide actual VOT values in these languages.

For a given place of articulation, most language contrasts two or three phonemes, but the laryngeal contrasts are different in each case. Since not all languages in question are tonal languages, the comparison cannot provide certainty because studies do not typically consider lexical tone effect in these languages. However, this comparison can nonetheless provide us some degree of L2 learner's phonetic and phonological processes.

5.12.1 The Stimuli

The stimuli for eliciting their native VOTs were three short passages that preserved the translated meaning from the Mandarin stimulus sentences into English, Japanese, and Spanish. The voiceless stops under investigation appeared throughout the passages. Table 39 below presents examples of the passages. There were six target stimuli in the English translation. Three of them were enclosed by other stops or fricatives, similar to those in the Mandarin stimuli closed by /s/ and /tʃ/. Others were open phrasal endings separated by a comma or a period indicating an ending or a pause, which would allow for final vowel lengthening. These settings yielded similar conditions to the Mandarin carrier phrases for the utterance-*medial* and utterance-*final* positions. The L2 participants recorded the passage with the voiceless bilabial stops. Then /p/s were replaced with the voiceless alveolar stops and then the voiceless velar stops. The Japanese and Spanish translations also had similar target stimuli and the attached meaning.

Table 39. Examples of short passages in each language

English	Please call Mr. **Pa**, ask Mr. **Pa** to say **Pa**. When Mr. **Pa** say(s) **Pa** ten times, ask him to say **Pa** for the last time.
Japanese	パさんに電話して、パさんに“パ”を言わせ、パさんは“パ“と言ったら、パさんはパです”と言わせます。 “**Pa**-san-ni denwashite, **Pa**-san-ni “**Pa**”-o iwase, **Pa**-san-wa “**Pa**”-toittara, **Pa**-san-wa “**Pa**” desu-to iwasemasu.”
Spanish	Por favor llame al Sr. **Pa**, pregunte al Sr. **Pa** que diga **Pa** cuando el Sr. **Pa** diga **Pa** tantas veces, Dile que diga **Pa** por una última vez.
- The bilabial stops were replaced by the alveolar or velar stops for repetitions.	

5.12.2 The Procedure & *VOT Measurement*

L2 participants recorded their native speech after they had finished the Mandarin tasks. After filling out the demographic questionnaire in their native language, the participant was orally instructed in his/her native language to read the passage on the computer screen silently. After receiving an indication of the participant's readiness, the author started the recorder and asked him/her to read it clearly and comfortably into the microphone. Each passage was repeated three times with different voiceless stops. Each of the ENG and SPN groups produced 18 (6 x 3) instances of each of the voiceless stops and 21 (7 x 3) by the JPN group. The measurement criteria were the same as the Mandarin stimuli described in section 3.5. After excluding a few defective, unmeasurable stimuli, 1830 data points from the three groups were analyzed.

5.12.3 The Results

The results reported here are a descriptive illustration of the data and do not contain inferential analysis, such as ANOVA, since investigating L2 inter-language is not the focus of this study. The intention here is to provide a glimpse of the VOT values from the same participants without making any inferential analysis.

The first observation was that the POA effect consistently showed in their L1s' VOTs; i.e., those in /k/ were longer than those in /p/ and /t/, and secondly, VOTs were much shorter in their L1 production than those in their Mandarin production for the SPN and JPN groups.

Figure 31 below shows the visual comparison of their L1 and L2 VOT productions. The average native VOT values for the SPN, JPN, and ENG groups were <u>26.33</u> (SD = 19.31), <u>37.34</u> (SD = 22.09), and <u>84.18</u> (SD = 22.35) milliseconds, respectively. In comparison to their Mandarin productions, their mean L2 VOT values were 66.18 (SD = 48.86), 63.56 (SD = 33.13), and 89.00 (SD = 32.51) milliseconds for the SPN. JPN and ENG groups, respectively.

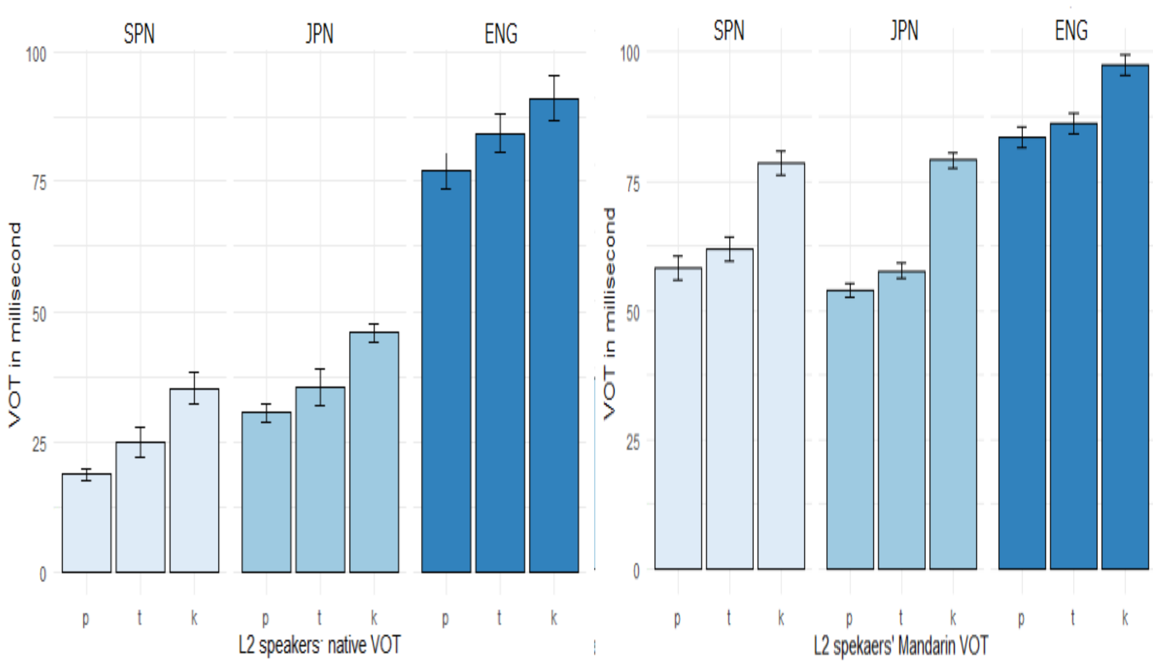

Figure 31. Comparison of L2 groups' native & Mandarin VOTs by language group & by POA

As we can see, these native VOT numbers were within the VOT values reported above from the literature (Cho & Ladefoged, 1999; Shimizu, 1990; Abramson & Lisker, 1972). Therefore, for this VOT baseline test, it is safe to say that our participants recorded the speech "natively" in their languages[11].

[11] While the JPN- and SPN- groups' L1 VOTs were within the range reported in the literature, the ENG-group's 84.18 milliseconds was slightly longer than that was reported by Abramson & Lisker (1964).

CHAPTER 6. PITCH EFFECT IN L2 SPEAKERS – EXPERIMENT 2B

6.1 Introduction

Chapter 6 presents Experiment 2b of the same L2 groups tested in Experiment 1b for the pitch register effect on VOT. Recall that Experiment 1b looks at how VOT behaves with contrasting lexical tones, places of articulation (POA), speech rates, phrase-positions, and genders. Experiment 2b further controls the lexical tone variable and examines how VOT behaves in T1 at three different pitch registers.

In Chapter 3, we report a tone effect on VOT and show significant VOT variations in different POAs, speech rates, and phrase-positions in native Taiwan Mandarin. In Chapter 4, we further show that the pitch-level significantly affected VOT values in Mandarin T1; thus, we propose that the tone effect on VOT might be due to vocal cord tensity. In Chapter 5, we bring attention to L2 speech production and show the tone effect in three groups of L2 learners of Mandarin. Chapter 6 aims to broaden our reports of the relationship between lexical tone and VOT by replicating the second experiment with the same L2 speakers.

This chapter addresses the relationship between the F0 onset pitch and VOT in T1 and compares the VOT difference with respect to post-stop vowel duration in L2 production. The essential differences from Experiment 1b were the additions of the pitch-level variable and the control of the lexical tone variable. Thus, Experiment 2b's data analysis focused on pitch-level and considers six primary mixed-effects: POA, pitch-level, phrase-position, vowel duration, native language (NL), and gender. Vowel duration was converted into categorical variables at three levels. Recall that vowel duration was treated as a random effect in Experiment 1b because the speaker's speech rate was adopted. In Experiment 2b, the subjects recorded their speech at a *natural* rate; therefore, the vowel duration is treated as an independent variable. Detailed vowel duration analysis is provided in the later subsections of each language group.

The key findings of Experiment 2b for L2 speakers were similar to that of the L1 speakers, where strong significant pitch-level, POA, and NL effects were found. There were no significant effects of vowel duration, phrase-position, and gender on VOT.

With the data collapsed, it shows that a) as the pitch (or F0 onset) increased, VOT values shortened, b) as the POA moved further back, VOT values prolonged, and c) the ENG and Taiwan groups had significantly longer VOTs than the SPN and JPN groups. Unlike the native Mandarin data results, the phrase-position effect was only significant if it interacted with NL and vowel duration or in a three-way interaction. Gender and vowel duration alone were not significant factors unless they interacted with other fixed effects. The outcome revealed a strong effect of pitch on VOT in L2 Mandarin speech, suggesting that the tone effect found in Experiment 1b is due to the lexical tone's F0 onset.

For convenience, this chapter starts with a brief review of the participants' background information from Chapter 5 and the stimuli design and data measurement from Chapter 3 and Chapter 4. A review of L2 T1 production at three pitch-levels is provided, followed by a descriptive overview of L2 participants' Experiment 2b results. It then summarizes the statistical analyses and provides a preliminary discussion for the

L2 speakers; native Taiwan Mandarin data from Chapter 4 is included for overall comparison.

6.2 Participants (review)

Experiment 2b's test subtests were the same L2 speaker groups of participants, i.e., English (22 ENG), Japanese (40 JPN), and Spanish (34 SPN) learners of Mandarin. 54 out of 96 L2 subjects recorded the speech in Taiwan at three different universities, 27 in Japan, and 11 in the US. At all recording sites, the recording equipment and the procedures were the same—none of the 96 L2 participants self-reported any known history of a speech disorder or a hearing impairment. Table 40 summarizes the demographic information of the L2 subjects. In all, the total number of subjects was 164 subjects, including the Taiwan group.

Table 40. Demographic information of the L2 participants, plus the natives

Group	N (gender)	Mean age (sd)	Mean Length of L2 learning (range)
English	22 (11 female, 11 male)	33.64 (16.16)	3.68 (0.5yr - 13yr)
Japanese	40 (23 female, 17 male)	32.83 (12.61)	4.23 (0.5yr - 35yr)
Spanish	34 (19 female, 15 male)	23.03 (6.07)	1.88 (0.8yr - 10yr)
Taiwanese	68 (36 female, 32 male)	29.09 (13.19)	---

6.2.1 English L2 Group

Twenty of the 22 ENG subjects were from the US, one from Canada, and one from the UK. Their average age was 33.64, ranged from 18 to 76. Half of the 22 subjects recorded their speech production in Taiwan at the NPLI library (3 subjects), at the

Providence University (5 subjects), and at the American School in Taichung city (3 subjects). They were university students and English teachers. The other 11 subjects participated in the US at George Mason University, and they were students there. Among all 22 participants, six self-reported that they had never lived in a Mandarin-speaking country, while 16 reported that they had.

6.2.2 Japanese L2 Group

The JPN group had 40 subjects. These speakers had an average age of 32.83, which ranged from 18 to 62 years old. Twenty-seven of them recorded their speech production in Tokyo—at the Arai lab of Sophia University and the Nicchu Mandarin Institute of the Japan-China Friendship Center. They were students, businessmen/women, and airline company employees. The other 13 university students recorded their speech production in Taiwan at Providence University and Fuguang University. Among the 40 JPN participants, 16 self-reported that they had never lived in a Mandarin-speaking country, and 24 reported that they had.

6.2.3 Spanish L2 Group

All 34 native Spanish speakers recorded the speech production in Taiwan at three different universities. Twenty-five of them were from Paraguay, six from Spain, one from Colombia, one from Peru, and one from Tunisia. The average age of this group was 23.03 years old and ranged from 18 to 46. They were the youngest group and had the shortest average length of learning. This group self-reported that they had been living in Taiwan for more than a year. They also reported that they had regularly used Mandarin outside the classroom.

6.3 Methodology

The test method, i.e., the stimuli, the carrier phrase, main task procedures, pieces of equipment, and data measurement, was precisely the same as for the Taiwan group. The only differences were: a) the L2 subjects additionally recorded their L1 speech after Experiment 2b, and b) the L2 subjects filled out the questionnaire in their native languages. For convenience, the test stimuli, procedures, and data elicitation and measurement are repeated here from Chapters 3, 4, and 5.

6.3.1 Speech Samples (Review)

Experiment 2b used the same test stimuli and the carrier phrases as Experiment 1b but further controlled the lexical tone. Each participant produced 36 monosyllabic target stimuli made up of /p, t, k /, one open-unrounded vowel (/a/), one lexical tone (T1), three phrase-positions (utterance-*initial*, utterance-*medial*, & utterance-*final*), and three pitch registers (*high*-, *normal*-, & *low*-) for analyses. For producing these three pitch registers, the second experiment adopted the method used in McCrea and Morris (2005) and Narayan & Bowden (2013), which asked participants to speak at *normal*/*high*/*low* pitch registers without forsaking accuracy. There were digitally made piano *high*/*low* notes to guide the subjects. The subjects firstly produced the samples at their *normal* tone of voice and then repeated the same task at a *high* pitch register, followed by another repetition of the task at a *low* pitch register. For the control of the lexical tone, only T1 was used so that every syllable and word in the stimulus sentence were all in T1, making it easier for the participants to keep the pitch-level consistent and avoid any possible tone *sandhi* being applied. It was also more comfortable for the subjects to reliably manipulate and produce the entire sentence in all three-pitch registers. For convenience, the next section repeats the stimuli design and an example of the same carrier phrase from section 3.3.

6.3.2 Stimuli & Carrier Phrase (Review)

The stimuli were monosyllabic Mandarin words. Each stimulus appeared twice in the carrier phrases, and each stimulus sentence was spoken twice, once with the classifier and once without the classifier, to create an open/closed-ending sentence condition. The stimulus sentence was also produced three times at different pitch registers (at the *natural* speech rate). Figure 32 below is repeated here from section 3.4, showing the stimuli and the carrier phrases. The change of phrase-position conditions was done by adding 七次 (qīcì); "*seven times*" after the second stimulus. The adding of the "七次 (qīcì)" created a token-embedded condition for the second stimulus, and without it, the sentence created a token-ended condition. All Chinese characters in the carrier phrases were in T1 between the phrase-*initial* and the phrase-*medial* stimuli. Notice that "七次/qīcì/" enclosed the stimulus with a T1 character as well.

Figure 32. Screenshot examples of the visual presentation of stimulus sentence without (on the left) and with (on the right) the classifier (qīcì). The second stimulus without the classifier creates the utterance-*final* instance, and the utterance-*medial* instance is on the right.

The entire sentence was presented with a combination of *pinyin* and Chinese characters. For instance, **pā** 先生說 (xiān shēng shūo) **pā** 七次 (qīcì) had the target stimulus in p*inyin*, but the carrier words were in Chinese characters. The elicited data were controlled for the below conditions. Point 1, 2, and 5 were essentially the same as in Experiment 1b, and the only differences in Experiment 2b were Point 3 and 4:

1. Place of Articulation (POA): the aspirated voiceless stops of Mandarin, /p, t, k/, were used.
2. Post-stop vowel: only the low-central-unrounded vowel /a/ was selected. Vowel duration control was done by adding 七次 (qīcì); "sev*en times*," which allowed for measurements of long and short vowel durations from the phrase-m*edial* token.
3. One lexical tone: high-level (or T1).
4. Three pitch registers: *high*, *normal*, and *low*.
5. Three types of phrase-position: utterance-i*nitial*, utterance-*medial*, and utterance-*final*. The participants read once the stimulus sentence without the classifier,

which generates two stimulus tokens of utterance-i*nitial* and utterance-*medial*, and once without the classifier, which generates two tokens of utterance-*initial* and utterance-*final*. This setting creates twice as many initial tokens as medial and final tokens for analysis.

6.3.3 Data Elicitation (Review)

Experiment 2b had three blocks. The subjects' tasks were to record their speech at *normal*, *high,* and *low* pitch-levels at the *natural* speech rate (Recall that there were also three blocks in Experiment 1b at the *normal* pitch but in the *natural*, *slow,* and *fast* speech rates). The equipment, task procedures, and seating and setting were the same as in Experiment 1b, but adding training for raising and lowering pitches following a musical beep before the main recording task in the second and third phases, in which participants were trained with the same training stimuli composed of / s, l, n / onsets.

For convenience, Experiment 1b's procedure is repeated here from Chapter 5.

1. Initial greeting and a minute of casual chatting.
2. Reading the consent form and the recording instructions.
3. Repeating the instructions orally by the experimenter to confirm the participants' understanding of the task.
4. Starting the training phase. This phase asked the participants to practice the task format with the training stimuli – the procedures were the same as the main task.
5. Beginning the main experiment blocks.

After a self-paced break time from Experiment 1b, the participants first recorded the stimuli in their everyday pitch register and speech rate for Experiment 2b.

In the second block before the main task, they were trained to speak with a *high* pitch register by listening to a digitally created high-pitch piano F_4 note at 350Hz for the stimuli. The computer displayed the stimulus sentences with the training stimuli and replayed the guiding notes three times, about 330 milliseconds. They were instructed to raise their pitch in a self-regulated way, following the musical note they heard when producing the phrases.

Similarly, in the third block, they were trained to speak with a *low* pitch register by listening to a digitally made low-pitch piano G_1 note at 50Hz, for the training stimuli. In a self-regulated way, they lowered their pitch register to the lowest possible tone of voice without forsaking accuracy, following the triplet musical note, about 330 milliseconds, they heard when producing the phrases.

The participants were instructed to speak right after the guiding notes and practiced it for as much as they wanted until they felt comfortable and ready. They then informed the experimenter to start the second block and then the third block. The same high-/low-pitch piano tone was always replayed for each target stimulus in these tasks to guide them to produce the *high*-/*low*-pitch utterance.

6.3.4 Measurements (Review)

Acoustic measurements were obtained the same way as in Experiment 1a (also see section 3.5). VOT and vowel duration values were measured from /p, t, k/ in /a/ in T1 at three pitch-levels and three phrase positions.

Each stimulus response was digitally analyzed using Praat. The VOT and the following vowel duration values for the stops were labeled and measured. The vowel duration was converted into categorical variables and used to test a correlation using Pearson's correlation coefficient and statistical analysis.

Figure 33 shows the waveform and spectrogram of an example and the labeling following the practice of previous studies. Four acoustic measurements for this experiment are defined below:

1. Voice onset time (VOT): the duration in millisecond (ms) from A to B; the stop release burst of the articulator to the beginning of the visible voice bar in the spectrogram and/or the first vertical spike in the waveform
2. Fundamental frequency (F0) onset before the vowel: the F0 at point B plus/minus 5 ms where frequency in Hertz (Hz) can be detected/measured
3. Post-stop vowel duration (vLength): from B to C; the start of aperiodic energy to the end of vowel's periodic energy
4. Vowel-end and F0 offset: the ending of the vowel's periodic energy or F0 offset at point C.

Following these criteria, each stimulus was manually edited to label for A, B, and C points one by one on the TextGrid file. In this dissertation, the VOT value is determined from point A to B, and the vowel duration is from B to C, shown in the figure below. Acoustic measurement proceeds in this manner for all productions of L2 speakers of Mandarin. Acoustic measurement proceeded in this manner for all stimuli.

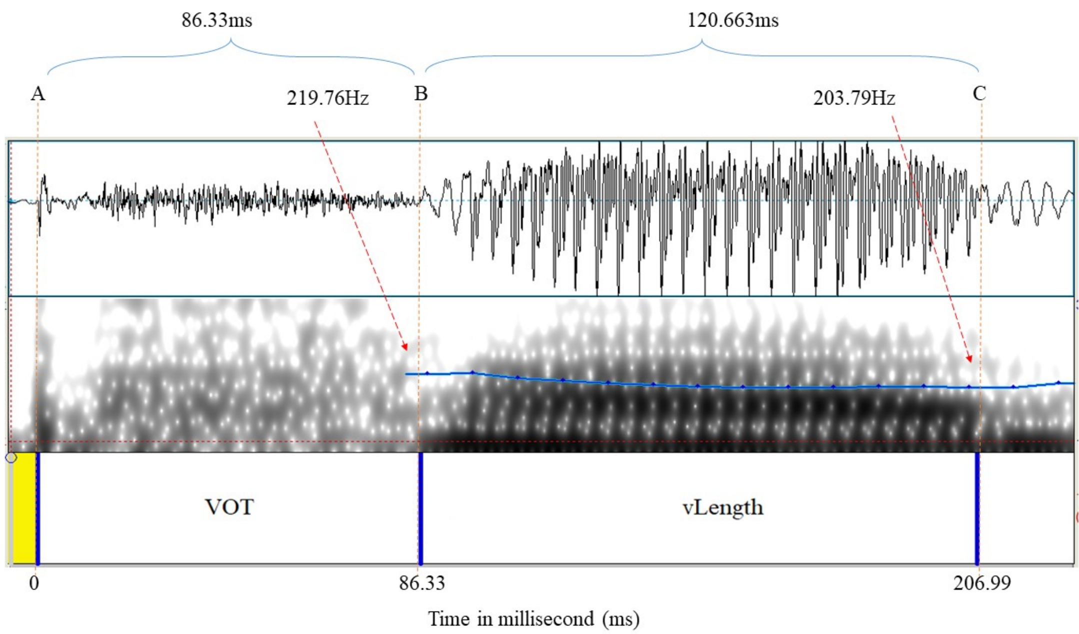

Figure 33. Waveform and spectrogram of an example of [pʰɑ̃] in T1. A is the stop release point; B is the start of voicing (or the aperiodic energy in the waveform); C is the end of F2 of the vowel (or the end of the periodic energy). The blue line depicts the overall pitch contour.

Several Praat scripts were used to extract the aforementioned acoustic measurements automatically. These values were entered into a spreadsheet alongside phonological variables and demographic information. Overall, 5904 stimulus tokens were collected from the 96 L2 Mandarin speakers and imported into R (R Development Core Team, 2014) for further analysis (792 from ENG, 1440 from JPN, and 1224 from SPN; see section 6.5; 2448 data points from the native Mandarin speakers were also included).

6.4 Descriptive Statistics Overview

This section provides the L2 groups' descriptive statistics for Experiment 2b. VOT was measured in milliseconds (msec), and each of the categorical variables was dummy-coded accordingly. It is a descriptive illustration of the data, which does not contain inferential analysis, such as ANOVA (the statistically analyzed results are provided in the next section). When collapsing the data for all variables, the overall T1's VOT values were 70.87 msec (SD = 24.8) for ENG, 49.74 msec (SD = 32.6) for JPN, and 55.84 msec (SD = 43.0) for SPN. The averaged vowel durations in T1 were 312.13 msec (SD = 95.5) for ENG, 305.31 msec (SD = 118.9) for JPN, and 291.31 msec (SD = 96.5) for SPN.

The data's relevant acoustic properties were measured from each stimulus and recorded in a spreadsheet. The results graphed below were collapsed across factors, and the native data were also replotted here together for visual comparison.

6.4.1 Overall VOT in Pitch-levels

Figure 34 shows the overall VOT values by pitch-level. VOTs at the *low*-pitch had the most extended values at 76.78 msec (SD = 26.3) for ENG, 57.23 msec (SD = 36.8) for JPN, and 59.99 msec (SD = 46.7) for SPN. At the *normal*-pitch, VOTs were at 70.51 msec (SD = 24.6) for ENG, 53.10 msec (SD = 32.6) for JPN, and 58.62 msec (SD = 42.8) for SPN. VOTs were the shortest at the *high* pitch register for all groups, where they were 65.34 msec (SD = 22.0) for ENG, 38.89 msec (SD = 24.2) for JPN, and 48.92 msec (SD = 38.5) for SPN.

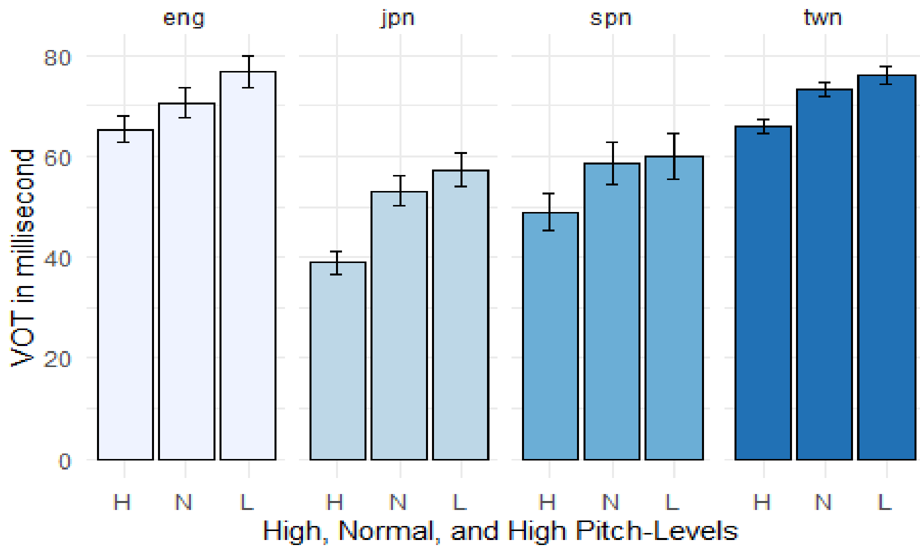

Figure 34. Overall VOT values by pitch-level and by language group across factors

6.4.2 Overall VOT in Places of Articulation

Figure 35 shows the overall VOT values by POA across all factors. The ENG group had the longest VOTs, and the JPN groups had the shortest values. The mean values in /p/ for ENG, JPN and SPN were 64.82 msec (SD = 22.8), 43.43 msec (SD = 30.5)/, and 46.97 msec (SD = 41.6), respectively. VOTs in /t/ were 69.66 msec (SD = 24.1) for ENG, 43.62 msec (SD = 29.3) for JPN, and 51.43 msec (SD = 42.0) for SPN. VOT values in /k/ were the longest across all groups concerning POA: 78.14 msec (SD = 22.8) for ENG, 62.18 msec (SD = 34.2) for JPN, and 69.12 msec (SD = 42.2) for SPN. Overall, VOT in /k/ at the *low* pitch-level was the longest, except for the SPN group in /k/.

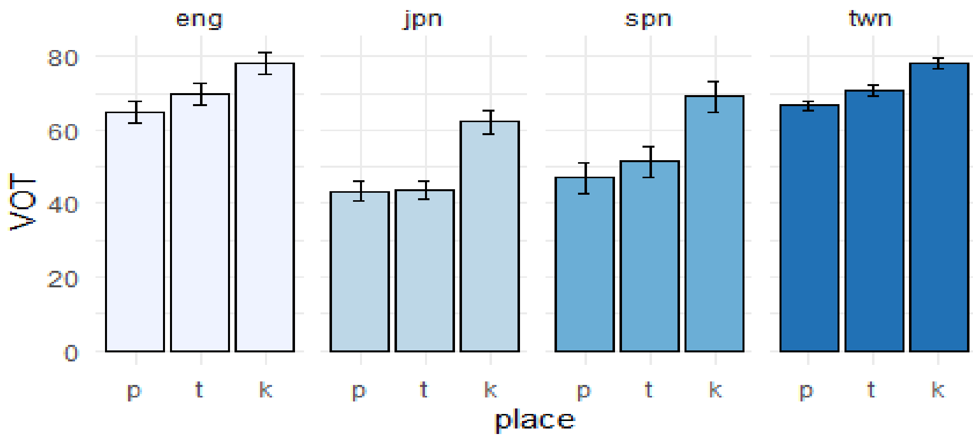

Figure 35. Overall VOT values by the POA and by language group across factors

6.4.3 Overall VOT in Phrase-Positions

Figure 36 shows the overall VOT values of the L2 groups for three phrase-positions. The mean values for ENG, JPN, and SPN at the utterance-*initial* position were 72.49 msec (SD = 25.4), 50.79 msec (SD = 31.5), and 54.24 msec (SD = 42.5), respectively. VOTs at the utterance-*medial* position were 70.16 msec (SD = 24.7) for ENG, 49.18 msec (SD = 33.6) for JPN, and 54.93 msec (SD = 41.6) for SPN. VOTs at the utterance-*final* position were 68.36 msec (SD = 23.4), 48.21 msec (SD = 33.7), and 59.96 msec (SD = 45.2) for ENG, JPN, and SPN, respectively. The descriptive data showed an inconsistent pattern of VOT distributions across groups for the phrase-positions. Statistical analyses of these variations are provided in the next sections.

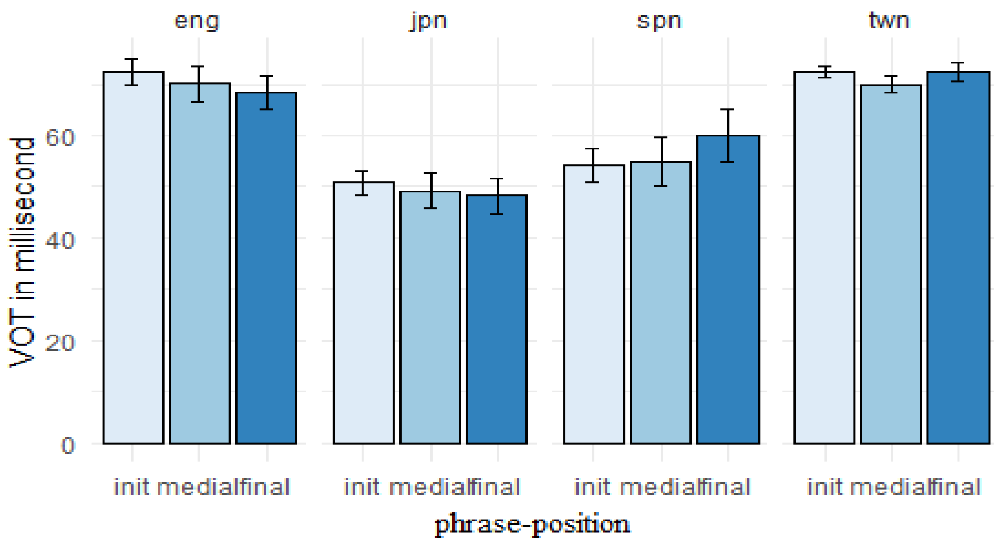

Figure 36. Overall VOT values at initial, medial, and final phrase-positions across factors

6.4.4 Overall VOT in Genders

Descriptively from the given data, ENG and JPN males produced slightly longer VOT values than the females did, but SPN females produced longer VOT values than the males. The ENG mean VOT values were 67.80 msec (SD = 24.8) for females and 73.95 msec (SD = 24.4) for males. The JPN mean VOT values were 46.69 msec (SD = 29.9) for the female and 53.9 msec (SD = 24.4) for the male. As for the SPN group, the VOT values were 58.37 msec (SD = 47.9) for the female and 52.64 msec (SD = 35.8) for the male. The SPN group's observation was interesting since females typically have higher F0; therefore, we would expect a shorter VOT value due to the pitch effect. Figure 37 shows the overall VOT by gender, and the differences were analyzed and reported in the following section.

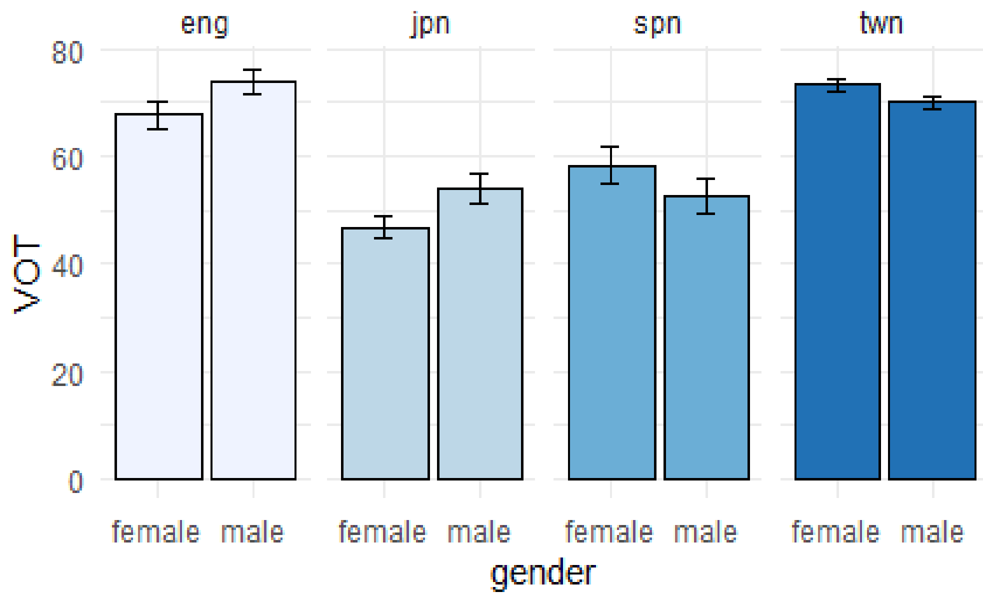

Figure 37. Overall VOT values by gender across pitch-level

6.4.5 Overall F0 Onset Frequency in Pitch-Level

The averaged F0 onset frequency for the pitch-levels was measured and analyzed. This study adopted Narayan and Bowden's (2013) method to elicit participants' three different pitch-levels but introduced the preceding piano notes. The first observation was that the leading *high* and *low* piano notes used in Experiment 2b successfully and consistently guided the participants to produce the stimulus sentence for their *high*- and *low*-pitch registers.

With all data collapsed, the overall F0 frequency (in Hertz) was 252.45 Hz (SD = 61.6) for the females and 155.59 Hz (SD = 47.0) for the males. The average *high* pitch F0 was 303.25 Hz (SD = 53.9) for the females and 198.92 Hz (SD = 42.2) for the males. The average *normal* pitch F0 was 211.40 Hz (SD = 41.6) for the females and 122.12 Hz (SD = 25.2) for the males. Lastly, the average *low* pitch F0 was 242.74 Hz (SD = 48.8) for the females and 145.74 Hz (SD = 33.6) for the males. Table 41 further separated the F0 onset frequency by language groups; vowel duration data were also provided.

As shown in the table, both genders could raise and lower their pitch registers without overlapping; the female exhibited about 100 Hz higher than the male across all three pitch-levels. F0 onset at the *high* pitch register was about 61 Hz higher than that at the *normal*, which in terms about 31 Hz higher than that at the *low* pitch register for females. The male-*high* F0 was about 53 Hz higher than the male-*normal* F0, which was about 23 Hz higher than the male-*low* F0. In general, both JPN males and females had higher pitch frequencies than the ENG and SPN groups.

Table 41. L2 groups' Overall F0 onset in Hertz (Hz) & vowel duration in milliseconds (msec) at each pitch-level between genders across factors

	female			**male**		
	F0 onset frequency by L2 groups					
Pitch -level	ENG	JPN	SPN	ENG	JPN	SPN
at *high*	289.13 Hz (59.5)	317.81 Hz (44.4)	280.21 Hz (53.8)	216.99 Hz (42.5)	224.02 Hz (41.7)	180.79 Hz (42.4)
at *normal*	251.46 Hz (56.3)	266.75 Hz (40.6)	231.19 Hz (63.3)	155.01 Hz (33.32)	163.06 Hz (34.2)	159.23 Hz (37.1)
at *low*	194.36 Hz (42.1)	215.34 Hz (46.0)	223.41 Hz (48.4)	125.89 Hz (24.1)	121.02 Hz (30.3)	133.82 Hz (29.2)
	Vowel duration					
at *high*	306.5msec (95.0)	327.2 msec (108.6)	308.5 msec (110.1)	332.6 msec (98.5)	254.3 msec (96.3)	282.2 msec (81.1)
at *normal*	292.5 msec (77.8)	317.2 msec (85.3)	307.0 msec (95.2)	279.5 msec (88.2)	264.7 msec (128.0)	256.2 msec (75.4)
at *low*	321.2 msec (100.6)	368.9 msec (133.5)	296.1 msec (106.0)	340.5 msec (98.0)	265.3 msec (111.0)	287.9 msec (90.0)

Concerning vowel duration, Table 41 provides an overview of how each vowel duration behaved in each pitch-level. The female's vowel duration was longer than that of the males. JPN females had the most extended vowel duration, but in males, the ENG group was the longest across all pitch-levels. The participants' vowel duration in the *normal*-pitch utterance was the shortest (237.7 msec, SD = 102.1), the second in the *high*-pitch utterance (265.4 msec, SD = 107.0), and the longest in the *low*-pitch utterance (279.4 msec, SD = 116.4).

The statistics reported here were combined in a spreadsheet and imported into R to statistically evaluate their significance using a series of linear mixed-effects models to be discussed in the following section.

6.5 The Statistical Model

The statistical analysis was conducted in two parts, using a series of mixed-effects linear regression models implemented in the *lmerTest* package in R. In part A, each language was modeled separately the same way as in the Taiwan group analysis, considered five main effects (pitch, POA, phrase-position, gender, and vowel duration). Vowel duration was converted into three categorical variables for *short*, *medium*, and *long* speeds based on the distribution in the continuum; see section 4.7 for detail. In part B, all groups' datasets were combined. The native language (NL) fixed-effect was then added to the final model to investigate the pitch-level effect on VOT in each mixed-effects.

The datasets were initially tested with the Likelihood Ratio tests for checking their predicting power (included 22 ENG, 40 JPN, and 34 SPN L2 speakers & 68 native Mandarin speakers). Table 42 shows the Likelihood Ratios Test results, which suggest that pitch, POA, and NL were significant predictors; phrase-position, gender, and vowel duration were not.

Table 42. Results of Likelihood Ratio Tests of the fixed effects

Place of Articulation:	$\chi^2 = 69.50$	$p < 0.0001$***
Pitch-Level:	$\chi^2 = 61.32$	$p < 0.0001$***
Native Language:	$\chi^2 = 28.18$	$p < 0.0001$***
Phrase-position:	$\chi^2 = 2.34$	$p = 0.310$
Vowel duration:	$\chi^2 = 0.61$	$p = 0.737$
Gender:	$\chi^2 = 0.04$	$p = 0.846$

The linear regression modeling was approached similarly as the L1 modeling to keep the model maximal as permitted by the data (Barr, Levy, Scheepers, & Tily, 2013); phrase-position[12], gender, and vowel duration factors were included for modeling consistency. Thus for each language group, all five mixed-effects were incorporated, as in the native model. Full interactions of pitch, POA, phrase-position, gender, and up to two-level vowel duration interactions with the other main effects were added. Post-hoc pairwise comparisons were conducted using Tukey's HSD tests implemented in the *emmeans* (Lenth, 2020) package.

Each group's analysis showed that the general results were similar between groups but slightly different in numeric values and significance of pairwise comparison

[12] Phrase-position was shown a significant predictor in the Taiwan group analysis.

by predictors. Starting with the ENG group's analysis, each group's detailed results are provided individually in the following sections.

6.6 Results of the English Group

The ANOVA results showed significant effects of pitch register (F (2, 772) = 7.8266, p = 0.0004), POA (F (2, 61) = 17.0402, p < 0.0001), and vowel duration (F (2, 787) = 7.1662, p = 0.0008). The general patterns were that the higher the pitch, the shorter the VOT, the further back of the POA, the longer the VOT, and the longer vowel duration, the longer VOT. The model also showed a significant two-way interaction between pitch and gender (F (2, 773) = 9.1506, p = 0.0001). Tukey HSD post-hoc analyses are provided in the following subsections for each fixed effect.

6.6.1 Pitch-Level Effect in the English Group

As expected, the general pattern for VOTs in pitch-levels was *low* > *normal* > *high*; the mean values were 72.7 msec (SE = 4.08, DF = 64, lower.CL = 64.5, upper.CL = 80.8), 69.7 msec (SE = 3.76, DF = 44, lower.CL = 62.2, upper.CL = 77.3), and 63.4 msec (SE = 3.96, DF = 56, lower.CL = 55.5, upper.CL = 71.3), respectively. The model revealed that only VOTs at the *high* pitch-level were significantly shorter than those at the *low* pitch-level. This outcome was slightly different from the Taiwan group. As shown in Table 41 above, ENG participants increased their pitch frequency by about 100 Hz from *low* to *high* and about 50 Hz proportionally between *normal* to *high* and *normal* to *low,* and VOT difference was only significant between *low*–*high*; the differences between *high–normal* and *low–normal* pairs were non-significant. This might suggest a range of frequency for a pitch-level effect on VOT. The post-hoc analyses for the significant two-way interaction between pitch and gender showed that VOTs in *female-high* pitch register were significantly shorter than those in *female-normal* pitch register, and those in *male-high* were shorter than those in *male-low*, which in turn were longer than those in *male-normal* pitch register (see Table 43 detailed statistics).

Table 43. Tukey HSD Post-hoc Analyses for the significant pitch effect and interaction between pitch & gender for the ENG group.

Contrast	estimate	SE	DF	t.ratio	p.value	
in Pitch						
high - low	-9.26	3.26	145	-2.842	0.0141	**
high - normal	-6.33	3.10	106	-2.042	0.1071	
low - normal	2.93	3.28	151	0.893	0.6457	
Pitch : Gender						
in *female*						
high - low	-7.04	3.65	262	-1.928	0.1328	
high - normal	-11.66	3.56	234	-3.273	0.0035	**
low - normal	-4.61	3.71	281	-1.243	0.4286	
in *male*						
high - low	-11.48	3.83	320	-3.001	0.0081	**
high - normal	-1.01	3.69	276	-0.272	0.9600	
low - normal	10.48	3.86	331	2.717	0.0190	*

˙ significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.6.2 POA Effect in the English Group

POA effect on VOT was significant between all POAs. Their VOTs were the longest in /k/ (M = 77.1 msec; SE = 3.80, DF = 47, lower.CL = 69.5, upper.CL = 84.8), the second in /t/ (M = 68.5 msec; SE = 3.90, DF = 52, lower.CL = 60.7, upper.CL = 76.3) and the third in /p/ (M = <u>60.2 msec</u>; SE = 4.03, DF = 60, lower.CL = 52.1, upper.CL = 68.2).

The statistical model showed a significant main effect of POA (F (2, 770) = 28.1556, $p < 0.0001$) and no significant interaction. The post-hoc analyses of the main effect showed that the differences between POAs were all statistically significant. While the /p/ - /t/ pair difference was significant here (t (131) = -2.596, $p = 0.0282$), it was not significant in their Experiment 1b's T1 instances (t (70) = -1.879, $p = 0.1523$). The peculiar observation of inconsistent results for the /p/ - /t/ pair is unclear; however, it may be due to the tongue's placement of the test subjects during the articulation. Since the stop consonant closure affects voicing delay, the significance might have differed if alveolar /t/ or dental alveolar /t/ was employed. This motivates a future study to investigate this phenomenon.

6.6.3 Vowel duration Effect in the English Group

The mean VOT values for stops preceding *short*, *medium*, and *long* vowel durations were <u>65.8 msec</u> (SE = 5.19, DF = 196.3, lower.CL = 55.5, upper.CL = 76.0), <u>66.3 msec</u> (SE = 3.42, DF = 38.4, lower.CL = 59.4, upper.CL = 73.3), and <u>73.7 msec</u> (SE = 3.39, DF = 36.8, lower.CL = 66.5, upper.CL = 80.6), respectively. ANOVA Type III result revealed a significant main effect of vowel duration and no significant interaction involved vowel duration. The post-hoc analyses showed that only VOTs accompanied with *medium* vowel duration were significantly shorter than those with *long* vowel duration (t (844) = -3.559, $p = 0.0011$).

The mean vowel duration values for *short*, *medium*, and *long* were analyzed: 171.4 (SD = 16.1), 251.5 (SD = 27.2), and 386.8 (SD = 71.2) milliseconds, respectively. A Pearson correlation test showed a non-significant weak positive correlation between VOT and vowel duration (r = 0.00899; t = 0.25272, DF = 790, p = 0.8006) from this group.

6.6.4 Phrase-position Effect in the English-Group

The mean VOT values for utterance-*initial*, utterance-*medial* and utterance-*final* positions were 70.0 msec (SE = 3.41, DF = 37.8, lower.CL = 63.1, upper.CL = 76.9), 70.2 msec (SE = 3.69, DF = 46.6, lower.CL = 62.8, upper.CL = 77.6), and 65.6 msec (SE= 4.88, DF = 150.9, lower.CL = 55.9, upper.CL = 75.2), respectively. None of the VOT differences concerning phrase-position was significant in the ENG group; there was no significant interaction either.

6.6.5 Gender Effect in the English-Group

Lastly, females showed slightly shorter VOT values (M = 64.8 msec; SE= 4.64, DF = 31.5, lower.CL = 55.4, upper.CL = 74.3) than males (M = 72.4 msec; SE = 4.86, DF = 38.0, lower.CL = 62.5, upper.CL = 82.2). The difference between two genders was non-significant (F (1, 25) = 1.4860, p = 0.2340); however, the model revealed a significant two-way interaction between pitch and gender (F (2, 773) = 9.1506, p = 0.0001). The post-hoc analyses showed that only in the *low* pitch register, females had significantly shorter VOTs than males (t (36.6) = -2.030, p = 0.0496).

6.7 Results of the Japanese Group

The general outcomes were that VOTs were significantly affected by pitch-level (F (2, 1400) = 92.2329, $p < 0.0001$) and POA (F (2, 1399) = 107.5270, $p < 0.0001$). The model also revealed four significant two-way interactions between pitch & POA, POA & gender, pitch & vowel duration, and phrase-position & vowel duration. The post-hoc results and estimated mean values of each fixed effect for the JPN group are provided in the following subsections.

6.7.1 Pitch-Level Effect in the Japanese Group

The overall pattern for VOTs in pitch-levels was *low* > *normal* > *high*; the mean values were 57.7 msec (SE = 3.80, DF = 46.8, lower.CL = 50.0, upper.CL = 65.3), 54.3 msec (SE = 3.77, DF = 45.5, lower.CL = 46.7, upper.CL = 61.9), and 38.9 msec (SE = 3.77, DF = 45.4, lower.CL = 31.3, upper.CL = 46.5), respectively[13]. The analyses showed that high pitch (or F0 onset) led to short VOT value. The model also revealed two two-way interaction between pitch & POA (F (4, 1400) = 2.6316, $p = 0.0328$) and between pitch & vowel duration (F (4, 1403) = 3.3220, $p = 0.0102$). The post-hoc analyses showed that VOTs at the *high* pitch register were significantly shorter than those at *low* and *normal* pitch registers in all three POAs and in all three-vowel durations, except for those in between *low-normal* in all three POAs, and those in between *short* and *medium* long vowel durations. Detailed post-hoc analyses are provided in Table 44.

[13] These VOT values were noticeably much shorter than the ENG-group and Taiwan groups.

The overall F0 onset frequency for the JPN participants increased by about 100 Hz from *low* to *high* and about 50 Hz in between *normal* to *high* or *normal* to *low* (see Table 41). Although the frequency was proportionally distributed between pitch-levels, the VOT was not.

Table 44. Tukey HSD Post-hoc Analyses for the significant pitch effect and interactions between pitch & POA and pitch & vowel duration for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
in Pitch						
high - low	-18.75	1.53	4.21	-12.254	0.0004	***
high - normal	-15.41	1.50	3.89	-10.281	0.0013	**
low - normal	3.34	1.54	4.38	2.163	0.1832	
Pitch : POA						
in /p/						
high - low	-20.825	2.52	3.47	-8.257	0.0045	**
high - normal	-12.926	2.51	3.42	-5.143	0.0214	*
low - normal	7.899	2.52	3.47	3.131	0.0867	·
in /t/						
high - low	-15.657	2.53	3.49	-6.196	0.0113	*
high - normal	-12.848	2.50	3.36	-5.137	0.0225	*
low - normal	2.809	253	3.51	1.111	0.5636	
in /k/						
high - low	-19.770	2.52	3.48	-7.833	0.0053	**
high - normal	-20.453	2.49	3.31	-8.205	0.0055	**
low - normal	-0.682	2.53	3.50	-0.270	0.9611	
Pitch : Vowel Duration						
in *short*						
high - low	-24.78	3.30	78.7	-7.512	< .0001	***
high - normal	-18.90	3.17	71.2	-5.957	< .0001	***
low - normal	5.88	3.24	74.8	1.813	0.1722	
in *medium*						
high - low	-14.56	2.62	35.9	-5.560	< .0001	***
high - normal	-16.88	2.53	31.3	-6.680	< .0001	***
low - normal	-2.32	2.84	49.1	-0.816	0.6949	
in *long*						
high - low	-16.91	2.09	14.1	-8.109	< .0001	***
high - normal	-10.45	2.09	14.3	-5.001	0.0005	***
low - normal	6.47	2.02	12.4	3.208	0.0185	*

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.7.2 POA Effect in the Japanese Group

The POA-dependent VOT differences were highly significant (F (2, 1399) = 107.5270, $p < 0.0001$). The values in /p, t, k / were 43.5 msec (SE = 3.78, DF = 45.8, lower.CL = 35.9, upper.CL= 51.1), 44.5 msec (SE = 3.78, DF = 45.8, lower.CL = 36.9, upper.CL = 52.1), and 62.9 msec (SE = 3.78, DF = 46.0, lower.CL = 55.3, upper.CL = 70.5), respectively. The model revealed two two-way significant interactions between pitch & POA (F (4, 1400) = 2.6316, $p = 0.03287$) and between POA & gender (F (2, 1399) = 4.0272, $p = 0.0180$). The post-hoc analyses showed that VOTs in /k/ were significantly longer than those in /p/ and /t/ across all three-pitch registers and both genders (see Table 45).

Table 45. Tukey HSD Post-hoc Analyses for the significant POA effect and interactions between POA & pitch and pitch & gender for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
in POA						
/k/ - /p/	19.4	1.52	4.16	12.716	0.0004	***
/k/ - /t/	18.4	1.52	4.12	12.093	0.0005	***
/p/ - /t/	-1.0	1.51	3.99	-0.663	0.7956	
POA : pitch						
at *high*						
/k/ - /p/	17.23	2.51	3.41	6.864	0.0088	**
/k/ - /t/	14.48	2.50	3.35	5.795	0.0157	*
/p/ - /t/	-2.75	2.49	3.30	-1.104	0.5699	
at *normal*						
/k/ - /p/	24.76	2.51	3.38	9.884	0.0028	**
/k/ - /t/	22.09	2.50	3.36	8.827	0.0041	**
/p/ - /t/	-2.67	2.51	3.39	-1.066	0.5877	
at *low*						
/k/ - /p/	16.18	2.54	3.57	6.368	0.0097	**
/k/ - /t/	18.60	2.54	3.59	7.308	0.0060	**
/p/ - /t/	2.42	2.53	3.48	0.957	0.6423	
POA : Gender						
in *female*						
/k/ - /p/	15.62	2.12	15.4	7.368	< .0001	***

/k/ - /t/	18.07	213	15.6	8.504	< .0001	***
/p/ - /t/	2.46	2.14	15.8	1.148	0.5002	
in *male*						
/k/ - /p/	23.17	2.18	17.4	10.634	< .0001	***
/k/ - /t/	18.71	2.18	17.3	8.600	< .0001	***
/p/ - /t/	-4.46	2.14	16.2	-2.083	0.1245	

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.7.3 Vowel duration Effect in the Japanese Group

ANOVA Type III results showed that there was no main vowel duration effect (*F* (2, 1430) = 1.9814, *p* = 0.1382), but two two-way significant interactions between pitch & vowel duration (*F* (4, 1403) = 3.3220, *p* = 0.0108) and between phrase-position & vowel duration (*F* (4, 1404) = 3.3004, *p* = 0.0106).

Mean VOT values for stops preceding *short*, *medium*, and *long* vowel durations were 53.3 msec (SE = 4.13, DF = 68.9, lower.CL = 45.0, upper.CL = 61.5), 49.2 msec (SE = 3.79, DF = 49.8, lower.CL = 41.6, upper.CL = 56.8), and 48.5 msec (SE = 3.76, DF = 47.8, lower.CL = 40.9, upper.CL = 56.0), respectively. The post-hoc analyses revealed that only VOTs with *short* vowel duration were longer than those with *medium* vowel duration in the *low* pitch register. Those with the *short* vowel duration were also longer than those with *long* vowel duration in the *final* phrase-position (see Table 46).

An additional analysis was conducted to investigate this reversed VOT and vowel duration distribution. The overall mean vowel duration values for *short*, *medium*, and *long* were analyzed: 156.57 (SD = 27.6), 243.64 (SD = 29.7), and 394.96 (SD = 90.0) milliseconds, respectively. A Pearson correlation test showed a significant weak negative correlation between VOT and vowel duration (*r* = -0.2.072; *t* = -8.0325, DF = 1438, $p <$ 0.0001), which suggests the longer vowel duration would lead to a shorter VOT value.

Table 46. Tukey HSD Post-hoc Analyses for the vowel duration effect and interactions between vowel & pitch and vowel & phrase-position for the JPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
By Vowel duration						
short - medium	4.107	2.33	1500	1.763	0.1824	
short - long	4.799	2.60	1397	1.846	0.1550	
medium - long	0.692	1.74	1485	0.398	0.9165	
Vowel : Pitch						
at *high*						
short - medium	0.0253	3.27	1490	0.008	1.0000	
short - long	-0.6426	3.60	1387	-0.179	0.9826	
medium - long	-0.6679	2.52	1479	-0.265	0.9620	
at *normal*						
short - medium	2.0514	3.39	1496	0.606	0.8171	
short - long	7.8139	3.38	1461	2.315	0.0541	·
medium - long	5.7625	2.58	1488	2.229	0.0666	·
at *low*						
short - medium	10.2429	3.40	1475	3.016	0.0073	**
short - long	7.2249	3.45	1391	2.095	0.0912	·
medium - long	-3.0180	2.69	1488	-1.122	0.5008	
Vowel : Phrase-position						
in *initial*						
short - medium	-1.997	2.88	1429	-0.693	0.7674	
short - long	-2.819	3.08	1048	-0.915	0.6308	
medium - long	-0.823	2.12	1414	-0.388	0.9205	
in *medial*						
short - medium	4.778	3.34	1490	1.430	0.3254	
short - long	6.851	3.42	1504	2.009	0.1112	
medium - long	2.073	2.90	1482	0.714	0.7550	
in *final*						
short - medium	9.538	4.07	1479	2.342	0.0505	·
short - long	10.364	4.15	1498	2.497	0.0338	*
medium - long	0.826	2.98	1481	0.278	0.9584	

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.7.4 Phrase-position Effect in the Japanese Group

The mean VOT values for utterance-*initial*, utterance-*medial* and utterance-*final* positions were 50.6 msec (SE = 3.72, DF = 46.2, lower.CL = 43.1, upper.CL = 58.1), 49.8 msec (SE = 3.80, DF = 47.7, lower.CL = 42.1, upper.CL = 57.4), and 50.5 msec (SE= 3.90, DF = 52.3, lower.CL = 42.7, upper.CL = 58.3), respectively. The model revealed no significant main effect (F (2, 1402) = 0.1931, p = 0.8244), and no significant interaction either.

6.7.5 Gender Effect in the Japanese Group

The mean VOTs were 46.9 msec for the females (SE = 4.86, DF = 46.6; lower.CL = 37.1, upper.CL = 56.7) and 53.7 msec for the males (SE = 5.51, DF = 42.3, lower.CL = 42.6, upper.CL = 64.8). The statistical modeling revealed no significant gender main effect (F (1, 41) = 0.9113, p = 0.3453) in Experiment 2b, but a significant two-way interaction between POA and gender effect (F (2, 1400) = 4.0272, p = 0.018). Post-hoc analyses showed that there were no significant interactions between genders in all three POAs. In other words, and the significance was not due to gender, but within each gender between different POAs, as shown in Table 45 above.

6.8 Results of the Spanish Group

The outcomes of the SPN group, in a way, were similar to the JPN group. The general results were that pitch register (F (2, 1191) = 20.9050 p < 0.0001) and POA (F (2, 1190) = 65.7553, p < 0.0001) significantly affected VOT. The main effect of vowel duration (*F* (2, 1214) = 3.0235, *p* = 0.0490) also reached at the borderline 95% confidence level. The model also revealed two significant two-way interactions between POA & gender and phrase & gender, and three marginal two-way interactions between vowel duration & pitch, phrase-position, and gender. Post-hoc analyses and factors to be discussed are provided in the following subsections for the SPN group.

6.8.1 Pitch-Level Effect in the Spanish Group

The general pitch main effect on VOT was consistently exhibited in this group. VOTs were found the shortest at the *high* pitch register. The mean VOT values were <u>47.5 msec</u> at *high* (SE = 6.50, DF = 46.2, lower.CL = 34.4, upper.CL = 60.6), <u>58.7 msec</u> at *normal* (SE = 6.50, DF = 46.2, lower.CL = 45.6, upper.CL = 71.8), and <u>59.3 msec</u> at *low* (SE = 6.51, DF = 46.6, lower.CL = 46.2, upper.CL = 72.4) pitch levels.

There were a significant main effect and a marginal two-way interaction between pitch register and vowel duration (*F* (4, 1192) = 2.2586, *p* = 0.0609). Post-hoc analyses showed that VOT values at the *high* pitch-level were significantly shorter than those at the *low* and *normal*. The VOT differences between those at l*ow* and *normal* were non-significant (*t* (231) = 0.162, *p* = 0.9856); the reason might be the degree of lowering their pitch-level. As shown in Table 41, female-*high* F0 onset frequency was 280.2 Hz (SD = 53.8), which was about 57 Hz higher than female-*normal*, which was in turn only about 8 Hz higher than female-*low*. For males, *male-high*, *male-normal*, and *male-low* were 180.8 Hz (SD = 42.4), 159.2 Hz (SD = 37.0), and 133.8 Hz (SD = 29.2). This motivates a future study to investigate whether VOT's pitch-effect can occur only at the *high* pitch register and not at the *low* pitch register.

The marginal two-way interaction between pitch and vowel duration was looked at. VOTs in the *high* pitch register were significantly shorter than those in *normal* and *low* in both *short* and *medium* vowel durations; between *low* and *normal* pitch-levels, differences were non-significant in *short* and *medium* vowel duration. None of the differences between the three pitch registers in the *long* vowel duration was significant (see Table 47).

Table 47. Tukey HSD Post-hoc Analyses for the significant main pitch effect and the marginal interaction between pitch & vowel duration for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
in Pitch						
high - low	-11.764	3.65	207	-3.3223	0.0042	**
high - normal	-11.165	3.66	215	-3.049	0.0073	**
low - normal	0.599	3.69	231	0.162	0.9856	
Pitch : Vowel duration (only marginal)						
in *short*						
high - low	-14.19	5.61	1270	-2.531	0.0308	*
high - normal	-17.88	5.75	1289	-3.110	0.0054	**
low - normal	-3.68	5.60	1304	-0.624	0.8068	
in *medium*						
high - low	-13.57	4.09	553	-3.316	0.0028	**
high - normal	-11.99	4.14	529	-2.894	0.0110	*
low - normal	1.58	4.05	512	0.390	0.9198	
in *long*						
high - low	-7.53	4.08	540	-1.847	0.1556	**
high - normal	-3.63	4.08	542	-0.888	0.6480	**
low - normal	3.90	4.14	594	0.942	0.6140	

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.8.2 POA Effect in the Spanish Group

The mean VOT values in /p, t, k / were 46.2 msec (SE = 6.51, DF = 46.6, lower.CL = 33.1, upper.CL= 59.3), 50.9 msec (SE = 6.49, DF = 46.1, lower.CL = 37.9, upper.CL = 64.0), and 68.3 msec (SE = 6.49, DF = 46.1, lower.CL = 55.3, upper.CL = 81.4), respectively. The model showed a significant main POA effect (*F* (2, 1190) = 65.7553, *p* < 0.0001) and a significant two-way interaction between POA and gender (*F* (2, 1190) = 3.0578, *p* = 0.0473). Post-hoc analyses revealed that VOTs in /k/ were significantly longer than those in /p/ and /t/ in both genders and the differences between those in /p/ and /t/ were non-significant (see Table 48).

Table 48. Tukey HSD Post-hoc Analyses for the significant POA effect and interaction between POA & gender for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value	
in POA						
/k/ - /p/	22.09	3.64	202	6.067	< .0001	***
/k/ - /t/	17.40	3.64	202	4.777	< .0001	***
/p/ - /t/	-4.69	3.67	221	-1.278	0.4093	
POA : Gender						
in *female*						
/k/ - /p/	19.752	4.00	473	4.944	< .0001	***
/k/ - /t/	19.591	4.00	479	4.893	< .0001	***
/p/ - /t/	-0.161	4.02	494	-0.040	0.9991	
in *male*						
/k/ - /p/	24.429	4.19	637	5.829	< .0001	***
/k/ - /t/	15.211	4.19	634	3.632	0.0009	***
/p/ - /t/	-9.219	4.22	660	-2.185	0.0745	·

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.8.3 Vowel duration Effect in the Spanish Group

As mentioned earlier, vowel duration effect on VOT reached 95% confidence-level with the given data (*F* (2, 1214) = 3.0235, *p* = 0.0490). The mean VOT values in three categorized vowel durations were 53.9 msec (SE = 6.75, DF = 63.0, lower.CL = 40.4, upper.CL = 67.4), 58.2 msec (SE = 6.19, DF = 44.2, lower.CL = 45.7, upper.CL = 70.7), and 53.4 msec (SE = 6.27, DF = 46.5, lower.CL = 40.8, upper.CL = 66.0) for *short*, *medium*, and *long*, respectively. There were three marginal interactions between vowel duration and pitch (F (4, 1191) = 2.2586, p = 0.0609), phrase-position (F (4, 11192) = 2.2999, p = 0.0.0569), and gender (F (2, 1215) = 2.7858, p = 00.0621). Post-hoc analyses were conducted to look at these marginal interactions. The results showed that there were no significant VOT difference in all three vowel duration conditions, but when they were at the *normal* pitch register, VOTs in *medium* vowel duration were significantly longer than those in *long* (t (1256) = 2.500, p = 0.0355) and those in *medium* vowel duration were significantly longer than those in *male* (t (1262) = 2.694, p = 0.0196). See Table 49 for detailed results.

The overall mean vowel duration values for *short*, *medium*, and *long* were analyzed: 165.49 (SD = 21.2), 242.04 (SD = 28.0), and 380.12 (SD = 69.8) milliseconds, respectively. The current data analysis showed that as their vowel duration increased, the VOT decreased. A Pearson correlation test revealed a significant weak negative correlation between VOT and vowel duration (r = -0.1778; t = -6.3161, DF = 1222, p < 0.0001).

Table 49. Tukey HSD Post-hoc Analyses for the vowel duration effect and interactions between vowel & pitch, vowel & phrase-position, and vowel duration & gender for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value
By Vowel duration					
short - medium	-4.256	3.15	1243	-1.350	0.3680
short - long	0.524	3.81	1261	0.137	0.9896

medium - long	4.770	2.33	1262	2.049	0.1012	
Vowel : pitch						
at *high*						
short - medium	-6.4139	4.52	1237	-1.419	0.3312	
short - long	-6.4449	4.93	1254	-1.307	0.3914	
medium - long	-0.0309	3.25	1248	-0.010	1.0000	
at *normal*						
short - medium	-0.5306	4.48	1240	-0.118	0.9923	
short - long	7.8024	4.97	1252	1.569	0.2594	
medium - long	8.3330	3.33	1256	2.500	0.335	*
at *low*						
short - medium	-5.7924	4.62	1238	-1.254	0.4215	
short - long	0.2143	5.09	1251	0.042	0.9990	
medium - long	6.0067	3.20	1246	1.876	0.1461	
Vowel : Phrase-position						
in *initial*						
short - medium	-7.46	3.40	1249	-2.192	0.0729	·
short - long	-1.16	3.92	1271	-0.296	0.9528	
medium - long	6.30	2.85	1267	2.211	0.0697	·
in *medial*						
short - medium	6.08	4.46	1237	1.363	0.3609	
short - long	7.40	5.14	1250	1.438	0.3217	
medium - long	1.32	3.78	1241	0.349	0.9351	
in *final*						
short - medium	-11.35	6.60	1228	-1.720	0.1980	
short - long	-4.66	6.84	1237	-0.682	0.7739	
medium - long	6.69	3.42	1240	1.954	0.1242	
Vowel : Gender						
in *female*						
short - medium	-1.932	3.58	1236	-0.540	0.8518	
short - long	-2.106	4.14	1256	-0.509	0.8669	
medium - long	-0.174	2.84	1259	-0.061	0.9979	
in *male*						
short - medium	-6.559	4.74	1250	-1.385	0.3492	
short - long	3.154	6.03	1262	0.523	0.8601	
medium - long	9.713	3.61	1262	2.694	0.0196	*

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.8.4 Phrase-Position Effect in the Spanish Group

For this group, the mean VOT values were 52.8 msec at utterance-*initial* (SE = 6.17 DF = 43.5, lowe.CL = 40.4 upper.CL = 65.2), 55.7 msec at the utterance-*medial* (SE

= 6.33, DF = 46.3, lowe.CL = 42.9, upper.CL = 68.4), and 57.0 msec at utterance-*final* (SE = 6.45, DF = 53.3, lowe.CL = 43.9, upper.CL = 70.1). Lisker and Abramson's (1967) report that VOT in words at the utterance-*final* position would have longer VOT than at the utterance-*initial* position due to sentential final stressing. Our model revealed no significant phrase-position main effect on VOT from this group (F (2, 1191) = 2.1227, p = 0.1202), but a significant two-way interaction between phrase-position and gender (F (2, 1192) = 5.2376, p = 0.0054) and a marginal two-way interaction between phrase-position and vowel duration (F (2, 1192) = 2.2999, p = 0.0569). It was shown in the post-hoc analyses that only in males, VOTs in the *initial* phrase-position were significantly shorter than those in the *medial* position (t (1331) = -2.930, p = 0.0097). The results also showed that VOTs were significantly shorter when they were in the *initial* phrase-position than those in the *medial* in the *short* vowel duration condition. Those in *final* phrase-position were significantly longer than those in the *medial* phrase-position in the *medium* vowel duration condition. Table 50 provides detailed statistics.

Table 50. Tukey HSD Post-hoc Analyses for the phrase-position effect and significant and marginal interactions between phrase-position & gender & phrase-position & vowel for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value
By Phrase-position					
final - initial	4.22	2.68	1334	1577	0.2557
final - medial	1.37	3.29	859	0.416	0.9092
initial - medial	-2.85	2.10	1280	-1.356	0.3644
Phrase-position : Gender					
in *female*					
final - initial	1.94	2.98	1332	0.651	0.7919
final - medial	4.77	3.62	1073	1.319	0.3851
initial - medial	2.84	2.63	1334	1.080	0.5268
in *male*					
final - initial	6.50	3.48	1320	1.869	0.1483
final - medial	-2.04	4.20	1257	-0.486	0.8780

initial - medial	-8.55	2.92	1331	-2.930	0.0097	**
Phrase-position : Vowel duration						
in *short*						
final - initial	1.76	6.62	1264	0.265	0.9619	
final - medial	-8.47	7.19	1314	-1.177	0.4672	
initial - medial	-10.22	4.32	1278	-2.368	0.0473	*
in *medium*						
final - initial	5.65	2.96	1335	1.910	0.1363	
final - medial	8.97	3.58	1070	2.504	0.0333	*
initial - medial	3.32	2.84	1334	1.169	0.4719	
in *long*						
final - initial	5.26	2.76	1333	1.908	0.1368	
final - medial	3.60	3.77	1157	0.954	0.6064	
initial - medial	-1.66	3.11	1332	-0.534	0.8544	

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.8.5 Gender Effect in the Spanish Group

Lastly, females had slightly longer VOT than males, but the model revealed that the gender main effect was non-significant (F (1, 34) = 0.2141, p = 0.6464). The average VOT values were 57.9 msec for the females (SE = 8.08, DF = 40.0, lower.CL = 41.5, upper.CL = 74.2) and 52.5 msec for the males (SE = 9.09, DF = 39.6, lower.CL = 34.1, upper.CL = 70.8). The model also showed two significant two-way interactions between POA & gender (F (2, 1190) = 3.0578, p = 0.0474) and phrase-position & gender (F (2, 1190) = 5.2376, p = 0.0005) and a marginal two-way interaction (F (2, 1215) = 2.7858, p = 0.062) between gender & vowel duration. Post-hoc analyses showed that the interaction between gender and POA was mainly due to POA because all three POAs between genders were non-significant. Similarly, all gender difference between three phrase-position were also non-significant; so were the interaction between gender & vowel duration. In other words, the significances shown were mainly due to other effects, not gender. Table 51 shows the post-hoc results of the gender main effect and interactions.

Table 51. Tukey HSD Post-hoc Analyses for the gender effect and significant and marginal interactions between gender & POA & gender & phrase-position, and the marginal interaction between gender & vowel duration for the SPN group.

Contrast	estimate	SE	DF	t.ratio	p.value
By Gender					
female - male	5.4	12	36.7	0.449	0.6559
Gender : POA					
in /p/					
female - male	9.980	12.2	39.3	0.816	0.4192
in /t/					
female - male	0.922	12.2	39.3	0.075	0.9403
in /k/					
female - male	5.302	12.2	39.0	0.435	0.6662
Gender : Phrase-position					
in *initial*					
female - male	10.717	12.1	37.5	0.887	0.3808
in *medial*					
female - male	-0.664	12.3	39.7	-0.0554	0.9571
in *final*					
female - male	6.150	12.3	40.8	0.498	0.6209
Gender : Vowel duration (marginal)					
in *short*					
female - male	5.190	13.0	49.9	0.399	0.6912
in *medium*					
female - male	0.563	12.1	38.2	0.046	0.9633
in *long*					
female - male	10.451	12.3	39.9	0.851	0.3999

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.9 Group Comparison

For the group modeling, the best-fit model included six fixed effects (the individual model's main effects plus the NL effect) and incorporated full interactions of the four predictors of pitch, POA, NL, and phrase-position. Interactions between gender and vowel duration with others were added one by one for as long as no error message was shown (only up to two-level). Incorporating the three-level interaction erred for "*model matrix is rank deficient*", which means the model was over-fitted or the current data was deficient for the built. By-subject adjustments to the random slopes of pitch and

POA, as well as random intercepts for by-subjects and by-items were included. All fixed factors were coded using R default treatment (dummy coding). The reference level for the intercept was set automatically to *high* for pitch, /k/ for POA, ENG for NL, final for phrase-position, *female* for gender, and *short* for vowel duration.

A summary of the Type III ANOVA test results is provided in Table 52. The model showed that VOTs were found to be affected by pitch-level, POA, and NL. Six significant two-way interactions and two significant three-way interactions were found. Post-hoc analyses for the significant main effects and interactions are provided in the following subsections.

Table 52. Summary of Type III ANOVA table with Satterthwaite's method of the mixed-effects and interactions between main effects of the group analyses

Fixed effect	Sum Sq	Mean Sq	NumDF	DenDF	F value	p.value
Pitch (Ph)	31841	1592	2	177.9	53.6496	<2.2e-16 ***
POA (P)	74906	37453	2	187.2	126.2130	<2.2e-16 ***
NL	9247	3082	3	170.4	10.3876	2.587e-06 ***
Phrase-Pos. (Pos)	1301	650	2	5170.7	2.1913	0.1118705
Gender (G)	9	9	1	164.6	0.0308	0.8608627
Vowel-Dur. (V)	428	214	2	5585.1	0.7204	0.4865972
Ph:P	4968	1242	4	5113.6	4.1857	0.0021921 **
Ph:NL	3131	522	6	187.1	1.7585	0.1097890
P:NL	9062	1510	6	194.9	5.0899	7.061e-05 ***
Ph:Pos	2640	660	4	5172.5	2.2245	0.0638514 .
P:Pos	419	105	4	5163.6	0.3527	0.8422981
NL:Pos	13291	2215	6	51808	7.4648	5.586e-08 ***
Ph:G	766	383	2	173.1	1.2911	0.2776051
P:G	3528	1764	2	182.2	5.9445	0.0031556 **
NL:G	631	210	3	164.7	0.7085	0.5482235
Pos:G	1680	840	2	5092.1	2.8303	0.0590899 .
Ph:V	2339	585	4	2961.0	1.9703	0.0963314 .
P:V	1272	318	4	2214.0	1.0720	0.3687003
Pos:V	6917	1729	4	5369.4	5.8274	0.0001122 ***
NL:V	8176	1363	6	5567.1	4.5919	0.0001170 ***

Ph:P:NL	3763	314	12	5099.2	1.0567	0.3929029
Ph:P:Pos	1271	159	8	5088.1	0.5353	0.8307018
Ph:NL:Pos	5550	463	12	5115.3	1.5586	0.0963660 ·
P:NL:Pos	7194	599	12	5110.0	2.0202	0.0190509 *
Ph:P:G	1327	332	4	5088.5	1.1179	0.3460665
Ph:NL:G	2315	386	6	172.2	1.3001	0.2595658
P:NL:G	1637	273	6	182.2	0.9192	0.4823152
Ph:Pos:G	862	216	4	5084.3	0.7265	0.5737181
P:Pos:G	1886	471	4	5086.4	1.5886	0.1743690
NL:Pos:G	7003	1167	6	5095.5	3.9332	0.0006306 ***
Ph:P:NL:Pos	10568	440	24	5086.4	1.4838	0.0604590 ·
Ph:P:NL:G	2007	167	12	5084.0	0.5637	0.8726584
Ph:P:Pos:G	1086	136	8	5084.2	0.4577	0.8862466
Ph:NL:Pos:G	3569	297	12	5087.4	1.0023	0.4436607
P:NL:Pos:G	2508	209	12	5090.8	0.7044	0.7486845
Ph:P:NL:Pos:G	6001	250	24	5085.0	0.8426	0.6838034

· significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.9.1 Pitch Effect in the Group Comparison

As expected, the model showed that pitch was a strong factor in VOT variations, and the higher the pitch register, the shorter the VOT. The mean VOT values were 54.6 msec at *high* (SE = 1.66, DF = 178, lower.CL = 51.4, upper.CL = 57.9), 64.2 msec at *normal* (SE = 1.92, DF = 172, lower.CL = 60.4, upper.CL = 68.0), and 67.6 msec at *low* (SE = 2.22, DF = 173, lower.CL = 63.3, upper.CL = 72.0). It revealed that there was a two-way interaction between pitch and POA. Post-hoc analyses showed that all interactions between the two were significant, except for those between *low* pitch register and *normal* pitch register in the /k/ (t (375) = 0.419, p = 0.9076) from the given data. As shown in the figure below, each group exhibited the pitch effect across all POAs, as the low the pitch register, the longer the VOT (the Taiwan group was added for convenience).

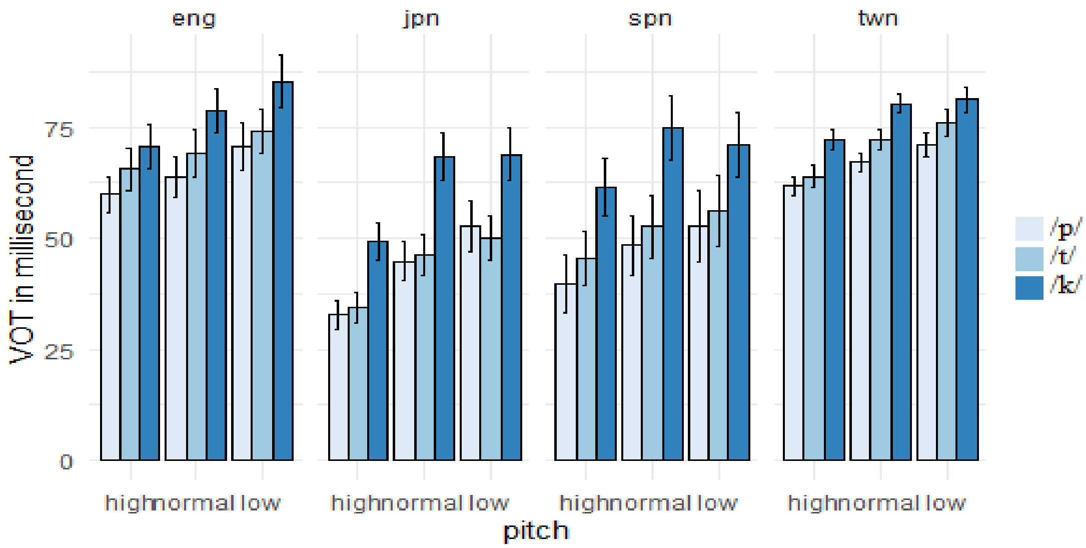

Figure 38. VOT patterns due to pitch register effect across language groups and POAs

Table 53 provides the post-hoc results concerning the pitch effect. It indicates a strong pitch effect, and it confirms the reports from the literature, where VOT duration decreases as speakers' pitch increases (Ewan, 1976; McCrea & Morris, 2005; Lai et al., 2009; Narayan & Bowden, 2013; Lou, 2018).

Table 53. Tukey HSD Post-hoc Analyses for the pitch effect and significant interaction between pitch & POA for the group comparison.

Contrast	estimate	SE	DF	t.ratio	p.value	
in Pitch						
high - low	-13.00	1.37	179	-9.491	< .0001	***
high - normal	-9.59	1.18	177	-8.108	< .0001	***
low - normal	3.41	1.33	178	2.570	0.0294	*
Pitch : POA						
in /p/						
high - low	-13.098	1.64	363	-7.989	< .0001	***
high - normal	-7.764	1.49	438	-5.215	< .0001	***
low - normal	5.334	1.61	384	3.309	0.0029	*
in /t/						

high - low	-12.221	1.64	360	-7.470	< .0001	***
high - normal	-7.998	1.49	434	-5.384	< .0001	***
low - normal	4.223	1.60	373	2.641	0.0234	*
in /k/						
high - low	-13.685	1.63	358	-8.392	< .0001	***
high - normal	-13.014	1.48	429	-8.797	< .0001	***
low - normal	0.671	1.60	375	0.419	0.9076	

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.9.2 POA Effect in the Group Comparison

In the group analyses, the final model showed that POA was a strong factor in VOT variations and involved many interactions. The mean VOT values were 55.4 msec in /p/ (SE = 1.86, DF = 174, lower.CL = 51.8, upper.CL = 59.1), 59.1 msec in /t/ (SE = 1.84, DF = 174, lower.CL = 55.4, upper.CL = 62.7), and 72.0 msec in /k/ (SE = 1.95, DF = 173, lower.CL = 68.1, upper.CL = 75.8). Figure 39 provides a clustered bar chart of the effect across pitch registers and language groups. The model revealed three two-way interactions between POA & pitch, POA & NL, and POA & gender and a three-way interaction between POA & NL & gender. Post-hoc analyses showed that a) VOT differences between POAs were all significant in all three-pitch registers, b) the differences between POAs were all significant, except for the differences between /p/ and /t/ in the JPN group, and c) similarly, all differences between POAs were all significant, except for the differences between /p/ and /t/ in females.

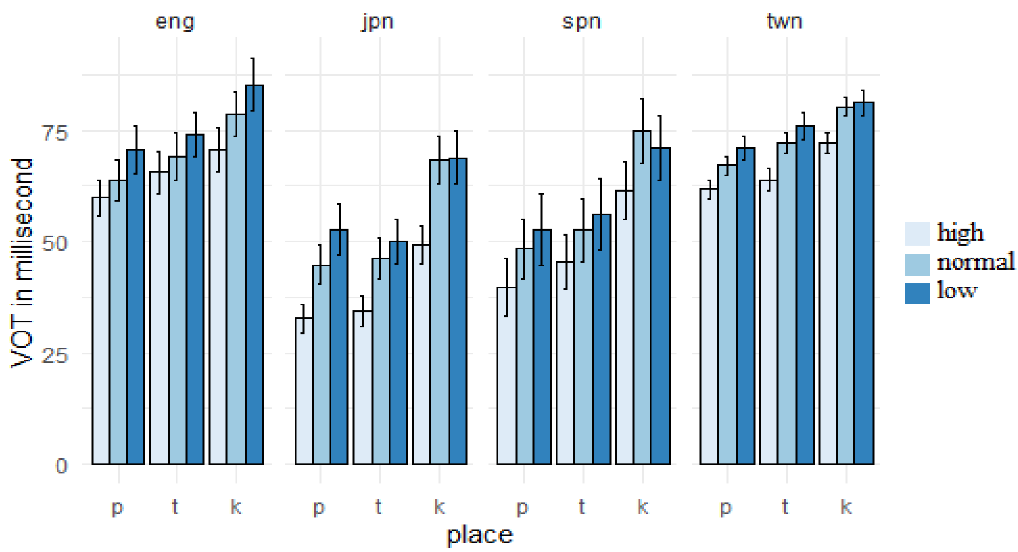

Figure 39. VOT patterns due to the POA effect across language groups and pitch registers

The post-hoc analyses for the three-way interaction showed that most differences between POAs were significant, except for those differences between /p/ and /t/ in JPN and ENG males, ENG, Taiwan, JPN, and SPN females and between /k/ and /t in ENG females (see detailed statistics in Table 54).

Table 54. Tukey HSD Post-hoc Analyses for the POA effect and significant interactions involved POA for the group comparison.

Contrast	estimate	SE	DF	t.ratio	p.value	
in POA						
/k/ - /p/	16.55	1.042	183	15.886	< .0001	***
/k/ - /t/	12.90	1.116	180	11.565	< .0001	***
/p/ - /t/	-3.65	0.814	196	-4.477	< .0001	***
POA : Pitch						
in *high*						
/k/ - /p/	14.60	1.38	541	10.617	< .0001	***
/k/ - /t/	10.74	1.43	480	7.501	< .0001	***
/p/ - /t/	-3.86	1.21	890	-3.185	0.0043	**
in *normal*						
/k/ - /p/	19.85	1.36	531	14.561	< .0001	***
/k/ - /t/	15.76	1.42	472	11.080	< .0001	***

/p/ - /t/	-4.09	1.20	892	-3.411	0.0020	**
in *low*						
/k/ - /p/	15.19	1.39	551	10.949	< .0001	***
/k/ - /t/	12.21	1.44	485	8.485	< .0001	***
/p/ - /t/	-2.98	1.22	899	-2.440	0.394	*
POA : NL						
in ENG						
/k/ - /p/	14.070	2.63	185	5.360	< .0001	***
/k/ - /t/	8.122	2.81	183	2.889	0.0120	*
/p/ - /t/	-5.949	2.05	198	-2.898	0.0116	*
in JPN						
/k/ - /p/	18.535	1.97	184	9.432	< .0001	***
/k/ - /t/	17.729	2.10	181	8.432	< .0001	***
/p/ - /t/	-0.807	1.53	194	-0.528	0.8577	
in SPN						
/k/ - /p/	22.721	2.12	184	10.708	< .0001	***
/k/ - /t/	18.261	2.27	182	8.039	< .0001	***
/p/ - /t/	-4.460	1.66	197	-2.692	0.0210	*
in Taiwan						
/k/ - /p/	10.871	1.55	204	7.021	< .0001	***
/k/ - /t/	7.505	1.65	200	4.545	< .0001	***
/p/ - /t/	-3.66	1.21	214	-2.782	0.0162	*
POA : Gender						
in *female*						
/k/ - /p/	13.73	1.42	184	9.640	< .0001	***
/k/ - /t/	12.34	1.52	181	8.098	< .0001	***
/p/ - /t/	-1.38	1.11	197	-1.242	0.4297	
in *male*						
/k/ - /p/	19.37	1.51	178	12.821	< .0001	***
/k/ - /t/	13.46	1.62	176	8.313	< .0001	***
/p/ - /t/	-5.91	1.17	188	-5.040	< .0001	***
POA : NL: Gender						
in ENG *female*						
/k/ - /p/	11.486	3.68	180	3.122	0.0059	**
/k/ - /t/	5.314	3.94	178	1.347	0.3710	
/p/ - /t/	-6.172	2.86	192	-2.156	0.0816	·
in JPN *female*						
/k/ - /p/	14.197	2.60	191	5.468	< .0001	***
/k/ - /t/	17.023	2.78	188	6.131	< .0001	***
/p/ - /t/	2.826	2.04	204	1.386	0.3504	
in SPN *female*						
/k/ - /p/	21.006	2.81	182	7.479	< .0001	***
/k/ - /t/	20.761	3.01	180	6.902	< .0001	***
/p/ - /t/	-0.245	2.19	193	-0.112	0.9931	
in Taiwan ***female***						

/k/ - /p/	8.222	2.07	189	3.967	0.0003	***
/k/ - /t/	6.275	2.21	187	2.833	0.0141	*
/p/ - /t/	-1.947	1.61	199	-1.210	0.4486	
in ENG *male*						
/k/ - /p/	16.655	3.69	182	4.512	< .0001	***
/k/ - /t/	10.930	3.95	179	2.766	0.0172	*
/p/ - /t/	-5.725	2.87	193	-1.992	0.1169	
in JPN *male*						
/k/ - /p/	22.874	2.95	178	7.762	< .0001	***
/k/ - /t/	18.434	3.16	176	5.837	< .0001	***
/p/ - /t/	-4.440	2.28	186	-1.949	0.1279	
in SPN *male*						
/k/ - /p/	24.436	3.15	180	7.761	< .0001	***
/k/ - /t/	15.761	3.37	178	4.673	< .0001	***
/p/ - /t/	-8.675	2.45	192	-3.542	0.0014	**
in Taiwan ***male***						
/k/ - /p/	13.519	2.21	192	6.130	< .0001	***
/k/ - /t/	8.734	2.35	188	3.712	0.0008	***
/p/ - /t/	-4.785	1.72	202	-2.785	0.0161	*

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

6.9.3 NL Effect in the Group Comparison

Our model revealed a strong NL effect on VOT (F (3, 170) = 10.3876, $p <$ 0.0001). The mean VOTs were 70.6 msec for the ENG group (SE = 4.50, DF = 173, lower.CL = 61.7, upper.CL = 79.5), 50.9 msec for the JPN group (SE = 3.34, DF = 167, lower.CL = 44.3, upper.CL = 57.5), 55.6 msec for the SPN group (SE = 3.62, DF = 168, lower.CL = 48.5, upper.CL = 62.7), and 71.6 msec for the Taiwan group (SE = 2.58, DF = 178, lower.CL = 66.5, upper.CL = 76.6). The model also revealed three two-way and two three-way interactions (see Table 52). The post-hoc pairwise comparisons are reported in Table 55.

In general, the Taiwan and ENG groups exhibited longer VOTs than the SPN and JPN groups. The differences between the Taiwan and ENG groups were non-significant across all three POAs, so were those between the SPN and JPN groups. The group differences in /k/ were mainly non-significant except those in JPN and Taiwan groups. Concerning the interaction between NL and phrase-position, the general patterns were the same: Taiwan and ENG groups had longer VOTs than those of the JPN and SPN groups, except for the VOT differences between the ENG and SPN groups in the *final* phrase-position condition. Similarly, all VOT differences were significant concerning the interaction between NL and vowel duration, except for those between the ENG and SPN groups in the *medium* vowel duration. See Table 55 for detail.   In addition to Figure 38 and Figure 39 above, Figure 40 also provides the clustered bar chart for the group comparison across phrase-positions and genders.

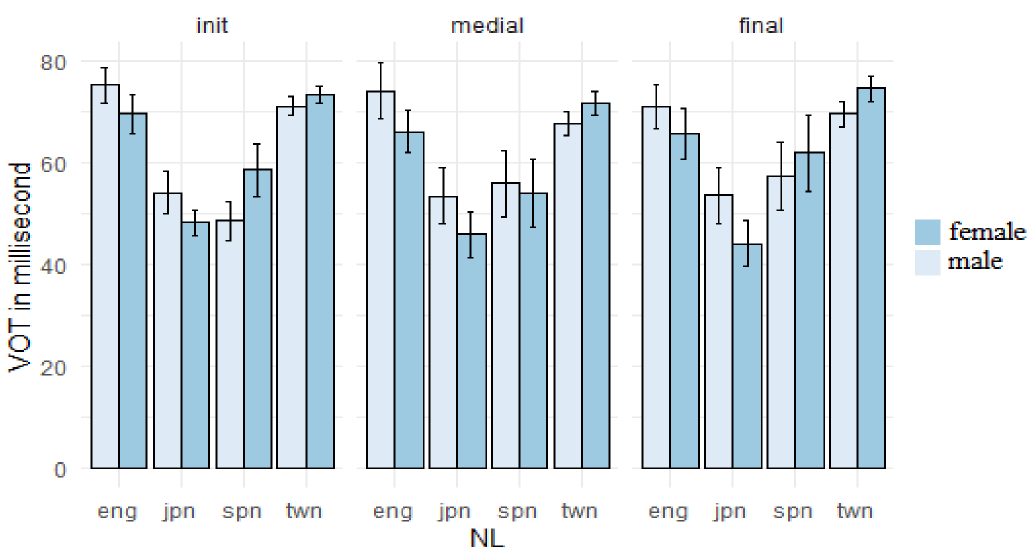

Figure 40. VOT patterns due to the NL effect across phrase-positions and genders

There were two three-way interactions between POA & NL & phrase-position and between NL & phrase-position & gender. The general patterns were that the Taiwan and ENG groups had longer VOTs than the SPN and JPN groups did in all conditions, except for a few exceptions in /k/ in all three phrase-positions. For the second three-way interaction, only nine significant differences were found. Detailed post-hoc results are provided in the table below (only significant results are reported).

Table 55. Tukey HSD Post-hoc Analyses for the NL main effect and significant interactions involved NL for the group comparison.

Contrast	estimate	SE	DF	t.ratio	p.value	
in NL						
ENG - JPN	19.688	5.60	171	3.514	0.0031	**
ENG - SPN	15.001	5.77	171	2.600	0.0493	*
ENG - Taiwan	-0.951	5.18	173	-0.184	0.9978	
JPN - SPN	-4.687	4.92	167	-0.952	0.7767	
JPN - Taiwan	-20.639	4.22	170	-4.894	< .0001	***
SPN - Taiwan	-15.952	4.44	170	-3.596	0.0024	**
NL : POA						
in /p/						
ENG - JPN	19.462	5.80	172	3.357	0.0053	**
ENG - SPN	17.388	5.97	172	2.913	0.0209	*
ENG - Taiwan	-2.879	5.38	177	-0.535	0.9503	
JPN - SPN	-2.074	5.09	168	-0.407	0.9771	
JPN - Taiwan	-22.341	4.38	173	-5.101	< .0001	***
SPN - Taiwan	-20.267	4.61	174	-4.398	0.0001	***
in /t/						
ENG - JPN	24.604	5.77	172	4.265	0.0002	***
ENG - SPN	18.877	5.94	172	3.178	0.0094	**
ENG - Taiwan	-0.296	5.35	177	-0.055	0.9999	
JPN - SPN	-5.727	5.07	169	1.131	0.6712	
JPN - Taiwan	-24.900	4.35	173	-5.722	< .0001	***
SPN - Taiwan	-19.173	4.58	174	-4.186	0.0003	***
in /k/						
ENG - JPN	14.997	6.08	170	2.468	0.0687	.
ENG - SPN	8.737	6.26	170	1.397	0.5032	
ENG - Taiwan	0.321	5.64	175	0.057	0.9999	

JPN - SPN	-6.260	5.35	167	-1.170	0.6466	
JPN - Taiwan	-14.676	4.61	173	-3.185	0.0092	**
SPN - Taiwan	-8.417	4.84	173	-1.738	0.3073	
NL : Phrase-position						
in *initial*						
ENG - JPN	19.936	5.64	175	3.534	0.0029	**
ENG - SPN	19.029	5.81	175	3.277	0.0069	**
ENG - Taiwan	-2.129	5.24	181	-0.406	0.9773	
JPN - SPN	-0.907	4.96	173	-0.183	0.9978	
JPN - Taiwan	-22.065	4.29	181	-5.149	< .0001	***
SPN - Taiwan	-21.158	4.50	180	-4.705	< .0001	***
in *medial*						
ENG - JPN	20.988	5.73	187	3.663	0.0018	**
ENG - SPN	16.388	5.90	187	2.778	0.0305	*
ENG - Taiwan	2.288	5.38	202	0.425	0.9741	
JPN - SPN	-4.600	5.04	184	-0.912	0.7985	
JPN - Taiwan	-18.701	4.41	203	-4.242	0.0002	***
SPN - Taiwan	-14.101	4.64	203	-3.041	0.0141	*
in *final*						
ENG - JPN	18.138	5.81	197	3.125	0.0110	*
ENG - SPN	9.585	5.98	196	1.604	0.3788	
ENG - Taiwan	-3.013	5.35	198	-0.563	0.9429	
JPN - SPN	-8.554	5.09	191	-1.681	0.3365	
JPN - Taiwan	-21.152	4.33	189	-4.885	< .0001	***
SPN - Taiwan	-12.598	4.56	190	-2.762	0.0318	*
NL : Vowel duration						
in *short*						
ENG - JPN	16.680	6.35	278	2.627	0.0447	*
ENG - SPN	18.376	6.55	281	2.803	0.0276	*
ENG - Taiwan	0.756	5.79	268	0.131	0.9992	
JPN - SPN	1.696	5.43	245	0.312	0.9894	
JPN - Taiwan	-15.924	4.48	214	-3.557	0.0026	**
SPN - Taiwan	-17.620	4.77	227	-3.690	0.0016	**
in *medium*						
ENG - JPN	18.060	5.72	185	3.159	0.0099	**
ENG - SPN	9.403	5.85	181	1.607	0.3772	
ENG - Taiwan	-3.497	5.26	184	-0.665	0.9101	
JPN - SPN	-8.657	5.04	184	-1.717	0.3177	
JPN - Taiwan	-21.557	4.34	190	-4.968	< .0001	***
SPN - Taiwan	-12.901	4.51	182	-2.859	0.0243	*
in *long*						
ENG - JPN	24.323	5.69	181	4.277	0.0002	***
ENG - SPN	17.223	5.86	182	2.937	0.0194	*
ENG - Taiwan	-0.113	5.43	210	-0.021	1.0000	

JPN - SPN	-7.100	5.05	185	-1.406	0.4971	
JPN - Taiwan	-24.436	4.55	230	-5.366	< .0001	***
SPN - Taiwan	-17.336	4.77	226	-3.637	0.0019	**
NL: POA : Phrase-position (only significant results are reported)						
in /p/ in *initial*						
ENG - JPN	22.617	5.92	187	3.818	0.0010	**
ENG - SPN	23.589	6.10	187	3.870	0.0009	***
JPN - Taiwan	-22.131	4.53	198	-4.887	< .0001	***
SPN - Taiwan	-23.104	4.75	195	-4.868	< .0001	***
in /t/ in *initial*						
ENG - JPN	24.038	5.89	188	4.080	0.0004	***
ENG - SPN	20.184	6.06	187	3.329	0.0057	**
JPN - Taiwan	-27.409	4.50	197	-6.091	< .0001	***
SPN - Taiwan	-23.555	4.72	195	-4.995	< .0001	***
in /k/ in *initial*						
JPN - Taiwan	-16.654	4.74	194	-3.514	0.0031	**
SPN - Taiwan	-16.815	4.98	194	-3.376	0.0049	**
in /p/ in *medial*						
ENG - JPN	17.558	6.19	223	2.836	0.0255	*
ENG - SPN	16.746	6.38	224	2.627	0.0453	*
JPN - Taiwan	-21.442	4.80	249	-4.470	0.0001	***
SPN - Taiwan	-20.631	5.03	247	-4.099	0.0003	***
in /t/ in *medial*						
ENG - JPN	26.939	6.19	228	4.356	0.0001	***
ENG - SPN	20.340	6.37	228	3.195	0.0086	**
JPN - Taiwan	-21.761	4.76	246	-4.573	< .0001	***
SPN - Taiwan	-15.162	5.02	250	-3.021	0.0147	*
in /k/ in *medial*						
ENG - JPN	18.468	6.46	217	2.857	0.0241	*
JPN - Taiwan	-12.899	4.98	236	-2.591	0.0496	*
in /p/ in *final*						
ENG - JPN	18.212	6.28	236	2.902	0.0210	*
JPN - Taiwan	-23.449	4.67	224	-5.025	< .0001	***
SPN - Taiwan	-17.066	4.93	228	-3.459	0.0036	**
in /t/ in *final*						
ENG - JPN	22.834	6.23	234	3.665	0.0017	**
JPN - Taiwan	-25.530	4.65	226	-5.489	< .0001	***
SPN - Taiwan	-18.802	4.90	227	-3.839	0.0009	***
in /k/ in *final*						
JPN - Taiwan	-14.475	4.94	228	-2.932	0.0193	*
NL: Phrase-Position : Gender (only significant results are reported)						
in *initial* in *female*						
JPN - Taiwan	-25.70	5.69	180	-4.520	0.0001	***
SPN - Taiwan	-17.06	6.01	176	-2.839	0.0258	*

in *medial* in female						
JPN - Taiwan	-23.26	5.83	199	-3.987	0.0005	***
SPN - Taiwan	-16.72	6.17	195	-2.712	0.0363	*
in *final* in *female*						
JPN - Taiwan	-26.33	6.09	187	-2.023	< .0001	***
in *initial* in *male*						
ENG - SPN	26.91	8.36	172	3.218	0.0083	**
JPN - Taiwan	-18.43	6.33	173	-2.912	0.0210	*
SPN - Taiwan	-25.26	6.61	174	-3.824	0.0010	**
in *medial* in *male*						
ENG - JPN	22.09	8.30	185	2.663	0.0416	*

. significance codes: 0.1, * significance codes: 0.05, **significance codes: 0.01, ***significance codes: 0.001

Recall that we hypothesized that vocal fold tensity, which is due to pitch change, can affect VOT values. Both individual and all-inclusive statistical models bore out this prediction. Regardless of their L1, all groups of participants exhibited this effect. Thus, the tone effect on VOT in Mandarin found in Experiment 1a and 2a might be a universal tendency due to the physiology of the vocal tract rather than due to language-specific phonology.

However, the significant VOT differences found between groups also suggest a degree of L1 influence. Cho and Ladefoged (1999) report that English VOT is typically around 58-80 milliseconds. Our ENG group might have used their English VOT for their Mandarin production because it was about the suggested range. Similarly, Shimizu (1990) reports that Japanese VOT can be around 30-60 milliseconds. Our JPN group's Mandarin VOT value was 49.74 msec (SD = 32.6) milliseconds, which was within the range; therefore, it might be the case that they had used their L1 Japanese VOT. However, in the VOT baseline test, our JPN group showed an average Japanese VOT value of 37.34 (SD = 22.09) milliseconds, which was different from their L2 VOT.

Studies have shown native Spanish VOT to be around 10-30 milliseconds (Abramson & Lisker, 1972; Cho & Ladefoged, 1999). Our SPN group had an average L2 VOT of <u>55.84 msec</u> (SD = 43.0; see section 6.4). As a result, they had used non-Spanish or L2 VOT for their Mandarin production (also see Section 5.12 for L1 VOT elicitation task and results for the current study L2 speaker's VOT).

Concerning the vowel duration effect, the overall averaged vowel duration values for *short*, *medium*, and *long* were analyzed: <u>153.79</u> (sd = 27.2), <u>240.77</u> (sd = 29.0), and <u>387.97</u> (sd = 81.8) milliseconds, and the overall averaged VOT values were: <u>69.77</u> (sd = 25.6), <u>64.31</u> (sd = 33.5), and <u>54.85</u> (sd = 34.0) milliseconds, respectively.

A Pearson correlation test showed an overall significantly negative correlation between VOT and vowel duration (r = -1.9165; t = -15.001, DF = 5902, p-value < 2.2e-16). This result suggests an inverse relationship that the longer vowel duration would lead to shorter VOT values.

6.10 Result Summary

A series of mixed-effects linear regression analyses were performed to explore the relationships between F0 onset and VOT in various conditions: pitch register, POA, NL, phrase-position, gender, and vowel duration in L2 Mandarin. Three out of six of the mixed-effects were significant factors, and three were non-significant.

Mandarin's T1 is a level tone and generally exhibits a higher F0 onset value than other lexical tones. Therefore, it has been reported to have a shorter VOT value than, for instance, T2 or T3 of Mandarin in studies of tone effect in Mandarin Chinese (Liu et al., 2008; Chen et al., 2009; Tseng, 2018). Experiment 2b hypothesizes that pitch, which is the tension change of the vocal folds controlled by the cricoarytenoid muscles, is one of the main reasons for the shortened VOT due to the Myoelasic-aerodynamic effect (Ven dam Berg, 1958).

Our native Mandarin data showed that when the participants produced T1 at a *high* pitch register exposed even shorter VOT values than when they produced it at a *low* pitch register. The vocal folds contracted to change the elastic thinness for a high pitch onset at the cord adduction had a quicker reaction to the Bernoulli Effect, thus shorter VOTs. We propose that this observation would apply to L2 Mandarin production from various language backgrounds. Experiment 2b revealed just that.

In general, the findings were that:

- The higher the F0 onset frequency leads to shorter VOT.
- The further back of the place of articulation leads to longer VOT.
- The speakers of languages with aspirated stops produce more Mandarin like VOT

A detailed discussion of the general findings is provided in Chapter 7.

CHAPTER 7. GENERAL DISCUSSION AND CONCLUSION

7.1 Introduction

This dissertation asked three research questions to investigate VOT variations in L1 and L2 Mandarin. It surveyed VOT variations affected by phonetic and phonological properties, i.e., lexical tone, POA, speech rate, and pitch register. While there are disagreements on the lexical tone-VOT relationship, this study provided empirical data to support the tone effect on VOT. Moreover, this study explored pitch register and vowel duration, two of the acoustic properties of lexical tone, and revealed that the pitch-level (or F0 onset) was central for the tone effect, whereas vowel duration was not.

Based on empirical observations, the current study offered evidence that the lexical tone affects VOT within different places of articulation, pitch registers, speech rates, language groups, genders, vowel durations, and phrase-positions; i.e., in most cases, VOTs in T1 and T4 will be shorter than those in T2 and T3 across all conditions.

Non-native Mandarin production was also investigated. The question was whether L2 learners of Mandarin from different language backgrounds do not feature aspirated voiceless stops (e.g., Spanish) or only a set of weak aspirated voiceless stops (e.g., Japanese) would also show the same effects. The results revealed that the VOT influencers found in native Mandarin showed in all three L2 groups. Two of the L2 groups also showed non-native VOTs nor Mandarin-like VOT for their Mandarin production.

We report that the most significant finding of this dissertation is that the effect of lexical tone on VOT in Mandarin is not due to language-specific phonology. Instead, it is due to the physiology of the vocal tract, and thus we propose that it is universal as it is exhibited in the L2 learners regardless of their L1.

The following section reviews the research questions and summarizes the study results. A summary table of experiments between groups is also provided, followed by discussion sections of each investigated factor. A conclusion and implications for future studies are also provided.

7.2 Study Questions & Result Summary

This dissertation features two experiments with the native speakers (denoted as Experiment 1a & Experiment 2a) and the same two experiments (denoted as Experiment 1b & Experiment 2b) with the L2 speakers to investigate VOT and lexical tone relationship under various phonetic and phonological conditions in L1 and L2 Mandarin speech production. L2 speakers additionally participated in a VOT baseline test[14] for measuring their native VOTs. While studies have inspected Mandarin prevocalic stop VOT's tone effect, very few have attended to multiple VOT influencers and looked at which acoustic property is responsible for the tone effect.

[14] The VOT baseline test was the elicitation of the L2 speakers' native VOT values. We did not statistically test the difference between their L1 and L2 VOTS because their L1s are non-tonal languages.

The current study examines 164 test subjects from four language backgrounds and looks at six effects on the prevocalic stop VOTs. We ask the following three research questions:

1. What are the lexical roles of tone, POA, speech rate, phrase-position, pitch register, vowel duration, and gender factors in VOT variation in Mandarin?
2. What is the specific component of lexical tone responsible for the tone effect on VOT?
3. Do we find the same effects in L2 learners of Mandarin Chinese?
    a. Does L2's native language matter? In other words, do we find L1 VOTs for their L2 production?

Chapter 3 presented Experiment 1a to investigate how VOT behaved in native Mandarin when POA, vowel duration, speech rate, phrase-position, and gender were attended to. Thus, for research question 1, the general findings showed that tone, POA, and speech rate significantly affected VOT, and gender and phrase-position were non-significant (see subsections below). Based on the empirical observation, we confirmed the tone effect on VOT.

Chapter 4 further investigated two of the essential acoustic properties of tone responsible for the tone effect (i.e., pitch-level and vowel duration). In Experiment 2a, the lexical tone was controlled, and POA, pitch register, phrase-position, and gender were attended to. Thus, for research question 2, the general findings revealed that the speaker's pitch-level (or tone's F0 onset) and POA significantly affected VOT, whereas gender, vowel duration, and phrase-position were non-significant. The crucial finding in

Experiment 2a was that VOT values varied significantly within the same conditions of the factors mentioned above. Therefore, we reported a pitch-level effect on VOT, and it was mainly due to the vocal fold tensity in the pitch change.

Chapter 5 extended Experiment 1a, as Experiment 1b, to L2 Mandarin learners to answer research question 3 in L2 speech. The findings of the effects were consistent in native English, Japanese, and Spanish speaking learners of Mandarin, where tone, POA, and speech rate significantly affected VOT. The results of gender and phrase-position effects were mixed, where the former was significant only in the JPN group but not the other two groups, and the latter was significant in the JPN and SPN groups, but not in the ENG group. See more discussions in sections 7.2.5 and 7.2.7.

Chapter 6 reported on the replication of Experiment 2a, as Experiment 2b, with the L2 groups to answer research question 3. Pitch-level and POA effects were consistently found in the L2 Mandarin. Gender and phrase-position effects were non-significant across all L2 groups. Vowel duration effect was non-significant in the ENG and SPN group s but was significant in the JPN group. Thus, for research questions 2 and 3, we reported a pitch effect on VOT across all language groups.

The results pooled from all test subjects and experiments revealed three apparent significant effects and three less-clear effects. As we can see from an overview provided in Table 56, a significant and robust POA effect was found in both experiments, and tone, speech rate, and pitch-level were significant across all groups. These three main effects consistently exhibited in L1 and L2 Mandarin across all fixed conditions, except for a few outliers discussed earlier. The other three factors were not as consistent in most conditions, as discussed in each relevant chapter and section. Overall, gender, vowel duration, and phrase-position main effects were found non-significant, but not all when interactions are considered in both Experiment1b and Experiment 2b.

Table 56. An overview of the study's significant findings of the main effects proposed in two experiments and across language groups

Effects	Lexical Tone		Speech Rate		Pitch Level		POA		Gender		Phrase-Position		Vowel-Duration	
Experiments / Groups	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Taiwan	√	---	√	---	---	√	√	√	×	×	√	√	---	×
ENG	√	---	√	---	---	√	√	√	×	×	×	×	---	√
JPN	√	---	√	---	---	√	√	√	√	×	√	×	---	×
SPN	√	---	√	---	---	√	√	√	×	×	√	×	---	√

√ : Significant effect
---: Not Applicable
×: Non-significant effect

Overall averaged VOT values pooled from all test subjects across the experiments are provided in Table 57. In general, for significant findings, the lower the F0 onset of a tone leads to longer VOT. VOT is also longer as the place of articulation moves further back of the mouth. In other words, velar plosive in T3 would be the longest. The slower the speech rate also leads to longer VOT, and the high pitch register causes a shorter VOT value. Lastly, languages that feature aspirated stops exhibit longer VOTs than the ones that do not.

Table 57. Mean VOT values pooled from all test subjects and factors in two experiments; VOT measured in milliseconds

Factors	Experiments	Conditions			
Lexical Tone	1	In T1: 69.22 (31.6)	In T2: 81.60 (36.0)	In T3: 85.63 (39.3)	In T4: 68.63 (31.5)
	2	In T1: 62.98 (31.8)	---	---	---
POA	1	In /p/: 68.99 (34.4)	In /t/: 72.94 (34.8)	In /k/: 86.90 (35.0)	
	2	In /p/: 56.65 (30.5)	In /t/: 59.97 (31.7)	In /k/: 72.30 (31.1)	
Speech Rate	1	In *fast*: 68.37 (31.0)	In *natural*: 77.20 (34.7)	In *slow*: 83.21 (38.9)	
	2	---	---	---	
Pitch-level	1	---	---	---	
	2	In *high*: 55.73 (28.6)	In *normal*: 64.94 (30.8)	In *low*: 68.25 (34.5)	
Vowel duration	1	---	---	---	
	2	In *short*: 69.77 (25.6)	In *medium*: 64.31 (33.5)	In *long*: 54.85 (34.0)	
Phrase-position	1	In *initial*: 75.70 (35.2)	In *medial*: 75.38 (34.7)	In *final*: 78.31 (37.1)	
	2	In *initial*: 63.37 (31.6)	In *medial*: 61.81 (31.4)	In *final*: 63.37 (32.8)	
Gender	1	In *female*: 74.94 (37.1)	In *male*: 77.8 (33.6)		
	2	In *female*: 62.57 (33.2)	In *male*: 63.46 (30.1)		
Native Language	1	In Taiwan: 84.79 (24.5)	In ENG: 89.00 (32.5)	In JPN: 63.56 (33.1)	In SPN: 66.18 (48.9)
	2	In Taiwan: 71.77 (21.7)	In ENG: 70.87 (24.8)	In JPN: 49.74 (32.6)	In SPN: 55.84 (43.0)

---: Not Applicable

Concerning L2 speakers' interlanguage, in both Experiments 1b and 2b, non-L1 or L2 VOTs were observed in the JPN and SPN groups. Their L2 VOT values were significantly shorter than the ENG and Taiwan groups. In other words, even though the

Japanese and Spanish speakers were approximating their target VOTs, the values were still not close to the typical native Mandarin VOTs.

L2 test subjects' L1 VOTs were compared to their L2 VOTs, descriptively in the VOT baseline test, reported in section 5.12. As reported, our Spanish and Japanese speakers showed averaged L1 VOTs of 26.33 (SD = 19.31) and 37.34 (SD = 22.09) milliseconds and L2 VOTs of 66.18 (SD = 48.86) and 63.56 (SD = 33.13) milliseconds, respectively. The SPN group had a slightly more extended L2 VOT than the JPN group, but the difference was not statistically significant. The crucial observation was that the L2 speakers' Mandarin VOTs were nevertheless affected by the significant factors found in native Mandarin speech. These findings indicate an NL influence (see detailed discussion in 7.2.8).

7.2.1 Tone Effect due to Pitch Register

The production study in Experiment 1a and Experiment 2b investigated the effect of lexical tone on aspirated stop VOT production in L1 and L2 Mandarin when several factors were individually attended to. In agreement with previous studies (Liu et al. 2008; Chen et al., 2009; Tseng, 2018), this study's general finding was that lexical tone significantly affects VOT values between the T1 and T4 pair and the T2 and T3 pair in three POAs and three speech rates in all participant groups. That is, VOTs in the mid-rising and falling-rising tones were significantly longer than those in the high-level and high-falling tones.

We attempted to understand the tone effect upon VOT value within the physiological, aerodynamic, and temporal accounts (Cho and Ladefoged (1999) and add the Myoelastic-aerodynamic theory (Van dem Berg, 1958) into the discussion. While place-dependent VOT variation can be explained by physiological and aerodynamic theories, they alone do not account for variations due to the tone effect because when we attended to the place of the articulation, VOT still varied between four lexical tones. Thus, we proposed that vocal fold tensity might be the leading cause of the tone effect.

Two of the primary acoustic properties of the tones are F0 and vowel duration. Experiment 2a and Experiment 2b showed that VOT values in T1 exhibited significant differences in three pitch-levels and in both L1 and L2 groups when these two properties were scrutinized. As a result, when the speaker's pitch settings (and presumably vocal folds' tensity) changed, VOT was affected. Therefore, we propose that the tone effect is due to the vocal fold tensity. Before discussing this proposition, a quick review of the VOT difference caused by POA is needed.

As reviewed in Chapter 2, VOT is a product of many things. The Bernoulli Effect explains how the vocal fold oscillation is possible, but it does not account for the voicing delay. Our results concerning the POA effect support the hypothesis that the higher the intraoral air pressure, the longer the voicing delay (Eshghi et al., 201; Steven, 1999; Cho & Ladefoged, 1999). However, in increasing the pitch level to produce a lexical tone, particularly for high onset T1 and T4 pitches, the larynx must elevate to create vertical tension for a high pitch (see vertical tension hypothesis in Homert, Ohala, & Ewan, 1979). Doing so decreases the supraglottal cavity volume, which increases the air pressure in the cavity, and the higher air pressure should lead to a longer voicing delay, analogous to the instance of the velar stop VOT. However, this study found that T1 and T4 had significantly shorter VOTs than those in T2 and T3.

Aerodynamically, when comparing voiceless alveolar fricative and voiceless alveolar aspirated plosive productions, the latter has a blockage of airflow. For example, in producing /sa/, there is no articulator closure (only vocal fold adduction) between segments, so the intraoral air pressure is constant. When the vocal folds are adducted together, the phonation airstream pressure is sufficient enough to initiate the oscillation, but there is a pressure change in producing /ta/ because of the pushed back transglottal air pressure between supraglottal and subglottal (Eshghi et al., 2016). As a result, some milliseconds are needed to resume a sufficient equilibrium air pressure for /ta/. This illustration of change of intraoral air pressure provides a sounding explanation for prevocalic stop’s voicing delay due to POA; however, it fails to explain the effect of lexical tone when we attended to the same articulation place. Hence, aerodynamics and laryngeal devoicing gesture theories seem to be insufficient to account for the lexical tone effect; the theory of Myoelastic-aerodynamics (Van dem Berg, 1958) is needed.

As reviewed, lexical tone is generally the manipulation of the pitch, which is the tension change of the vocal folds controlled by the cricoarytenoid muscles. In manipulating pitch for producing lexical tones, the movements subsequently change the vocal fold tension and shape. For instance, in producing T1 or T4, the muscle of the vocal folds contracts to change the elastic thinness to allow a high pitch onset at adduction; therefore, the quicker the reaction to the Bernoulli Effect when all other things are kept constant. The typical indication of T1 and T4 are high pitch onset, and T2 and T3 are low; thus, significant VOT differences can be found between the pairs. This hypothesis successfully explains the results found in this study that voicing delay was longer in T3 and T2 than in T1 and T4. To further test the pitch effect, this dissertation designed the second experiment to control the tone and observed a VOT's pitch effect in native Mandarin and L2 Mandarin.

In Experiment 2a and 2b, our test subjects produced the same stimuli in three pitch-levels in one level lexical tone. They showed that the higher pitch-level led to the shorter VOT values. This provided us first-hand data for the vocal fold tensity hypothesis in Mandarin tone effect on VOT. Thus, we propose that the tone effect on VOT is mainly due to vocal fold tensity.

The vocal folds are long and loose when producing low F0 so that they oscillate slower than when they are short and tensed (Reetz & Jongman, 2014, p. 77). Therefore, VOT duration in aspirated stops decreases as speakers' pitch increases (Ewan, 1976; McCrea & Morris, 2005; Lai et al., 2009; Narayan & Bowden, 2013; Lou, 2018). In other words, the delay of voicing should be longer with non-tensed vocal folds than tensed ones. A follow-up task is to test the pitch register effect in Mandarin T4, T2, and T3 and other languages for future studies.

Furthermore, there is a suggestion about the tone's F0 offset on VOT. Tseng (2018) reports that the long T2 and T3 VOT found in Mandarin may be due to the tones' rising parts (c.f., Liu et al., 2008). However, in this study, tone one and tone four, which have a similar F0-onset pitch, but relatively different F0-offset pitch, showed non-significant VOT differences. In other words, if long T2 and T3 VOTs were due to tone-ending pitch, our results would have shown significant T1 and T4 VOT differences. The current study did not statistically test the F0-offset effect due to the number of independent variables that had already been included in the model; however, intuitively, we remain skeptical about the F0-offset effect. Future studies might consider it.

7.2.2 POA Effect

Adding to the previous studies (Lisker & Abramson, 1964; Abramson & Lisker, 1972; Shimizu, 1990; Cho & Ladefoged, 1999; Lai et al., 2004; Pearce 2005; Liu et al., 2008; Narayan & Bowden, 2013; Chen et al., 2009; Tseng, 2018), the current study confirms the POA effect, by adding empirical data from Mandarin speakers in Taiwan.

As discussed above, place-dependent VOT variations can generally be explained by physiological and aerodynamic accounts. That is that VOT depends on the places of constriction's contact and occlusion area plus the intraoral air pressure (Cho & Ladefoged, 1999 and Eshghi et al., 2016).

A review note on the POA effect is that our overall results showed significant VOT differences between all /p/ - /t/, /p/ - /k/, and /t/ - /k/ pairs of POAs in both experiments; however, the individual analyses also showed non-significant /p/ - /t/ pair in the SPN and ENG groups in Experiment 1b and the SPN and JPN groups in Experiment 2b. This peculiar observation of inconsistent results for the /p/ - /t/ pair is unclear; however, it may be due to the tongue's placement of the test subjects during the articulation. Since the stop consonant closure affects voicing delay, the non-significant results might have been due to the employment of the dental alveolar /t/ instead of the Mandarin alveolar /t/.

Spanish /t/ has been reported to be dental or more dentalized (Maddieson, 1984; Hualde et al., 2001). Our Spanish participants might have used their dental /t/ for the L2 production. Since producing /p/ and dental /t/ would result in similar oral cavity sizes, it is reasonable that the/p/ - /t/ pair's VOT difference would be non-significant, as shown in the SPN group in both experiments. The present study did not control the type of alveolar plosive, which motivates a further study to investigate this phenomenon.

7.2.3 Speech Rate Effect

Studies have suggested a temporal account for VOT alterations, which proposes that a stop closure's temporal adjustment during a *fast* or *slow* speech rate can affect vocal folds' oscillation (Cho and Ladefoged, 1999; Port and Rotunno, 1979). The motivation for a speech rate effect is that the airflow is more potent and in higher air pressure during a fast speech rate, resulting in a quicker Bernoulli Effect (Ohala & Ewan, 1979; Eshghi et al., 2016).

Our results showed that the faster release of a stop closure led to a shorter VOT and a shorter voicing delay since the air pressure is weaker when the stop closure's relative timing is short (Eshghi et al., 2016). The speech rate effect was confirmed in both our L1 and L2 Mandarin. Thus, the current study adds experimental data of the speech rate effect on VOT from Mandarin.

7.2.4 Vowel duration Effect in L1

Recall that the vowel duration effect was non-significant when data were pooled together. As shown in Figure 41 and Figure 42 below, an observation of vowel duration showed that vowels in T3 were significantly shorter than those in other tones, but VOT values were nonetheless significantly longer. Vowel durations in /k/ were also shorter than those in /p/ and /t/, which was also an inverse pattern to the significant VOT pattern concerning POA.

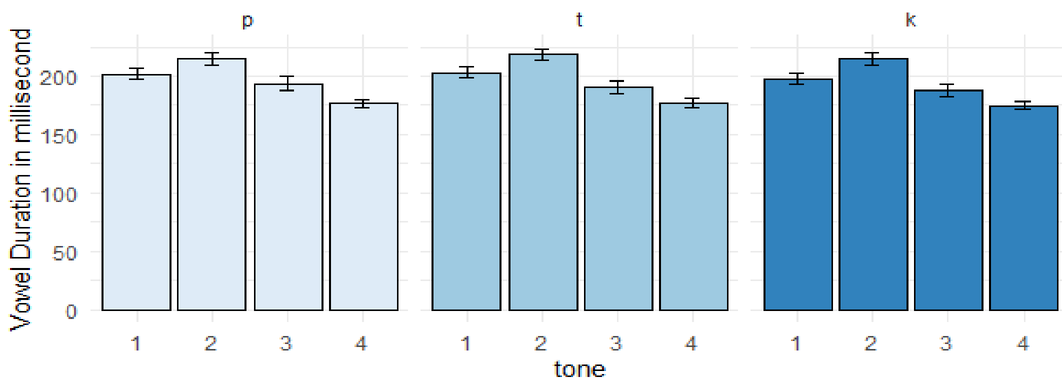

Figure 41. **An overview of Taiwan Mandarin vowel duration by lexical tone and by POA**

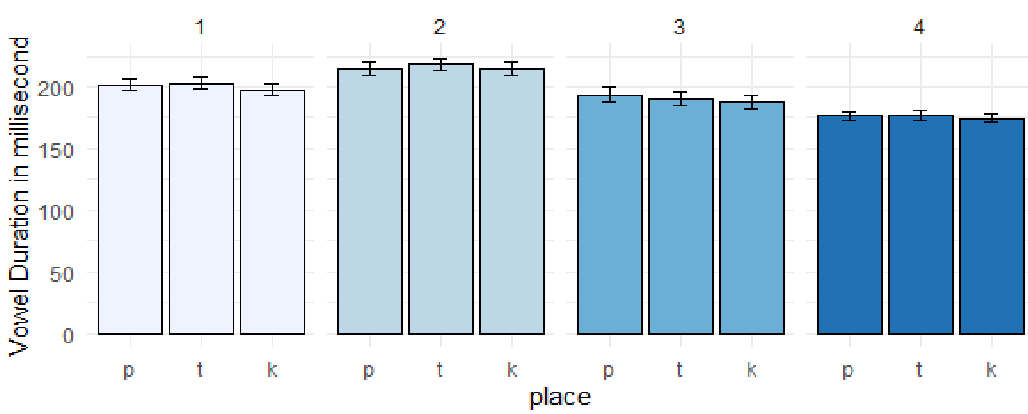

Figure 42. **An overview of Taiwan Mandarin vowel duration by POA and by lexical tone**

It is commonly known that Mandarin T3 is impressionistically longer in duration than the other three tones. However, the present study observed a different pattern where the vowel duration in T3 was significantly shorter than T1 and T2. A closer examination showed that the reason might be due to the third tone *sandhi* (Yang 2015).

Despite the short T3 vowel duration, VOT values in the native Mandarin T3 were still significantly the longest amongst all tones. For the Taiwan group, the mean vowel duration was 200.7 (SD = 69.8), 216.0 (SD = 73.0), 190.9 (SD =77.1)[15], and 176.1 (SD = 54.8) milliseconds whereas the mean VOT values were 79.2 (SD = 22.1), 87.3 (SD = 24.5), 92.3 (SD = 26.4), and 80.3 (SD = 224) milliseconds for T1, T2, T3 and T4, respectively. In this study, the mean vowel duration of the Taiwan group was almost 50% shorter than those of the L2 groups, yet their VOT values were still significantly longer than those of the SPN and JPN groups (see Chapter 5). This result led us to speculate about the VOT difference due to the vowel duration effect (e.g., Tseng, 1979).

From the given data, the overall Pearson correlation test reached the significance (t (9720) = 23.391, $p < 0.0001$), but a weak negative correlation of $r = -0.23$. In other words, VOT and vowel durations were inversely correlated, and vowels and VOTs did not change proportionally according to speaking rate. This finding is in agreement with the study of Allen and Miller (1999).

Additionally, Experiment 2a showed a confirmation of a non-significant vowel duration effect (F (2, 1509) = 0.8551, $p = 0.4254$). Thus, the results suggest that speech rate affects both VOT and vowel durations with both experiments, but vowel duration

[15] The largest standard deviation in T3 might indicate the instances of tone sandhi.

does not affect VOT. A more detailed study is needed to provide a definitive answer to this discussion. Nonetheless, the current results discussed here present a challenge to the proposals of the vowel duration effect on VOT (Port & Rotunno, 1979; Tseng, 1979).

7.2.5 Vowel duration Effect in L2

In comparison to the native speakers, L2 speakers' averaged vowel durations across all factors were 315.75 (SD = 132.6), 303.93 (SD = 118.4), and 320.35 (SD = 121.80) for the ENG, JPN, and SPN groups, respectively. As shown in Figure 43, L2 vowel durations in T3 were the longest across all L2 groups. Secondly, tone *sandhi* was not observed in the L2 groups; they probably have not yet acquired the *sandhi* rule. Therefore, they carefully produced the stimuli in a full T3.

From the given data, the overall Pearson correlation test reached the borderline significance (t (23389) = -2.027, p = 0.043), but a weak negative correlation of r = -0.0132 between VOT and the vowel duration. The Pearson *R-values* for the ENG, JPN, and SPN groups were 0.29, -0.04, and 0.04, respectively. None of the groups separately had a strong correlation concerning VOT and vowel duration. Since we have found that speech rate affected VOT duration, so it is possible that the SPN and JPN groups' extended L2 VOTs might be due to the speech rate in comparison to their L1 VOTs since they did have a faster speech rate in their L1s than L2 (about 150 msec faster). The current study was not set to investigate the L1-L2 VOT differences in these two groups; thus, this motivates future study.

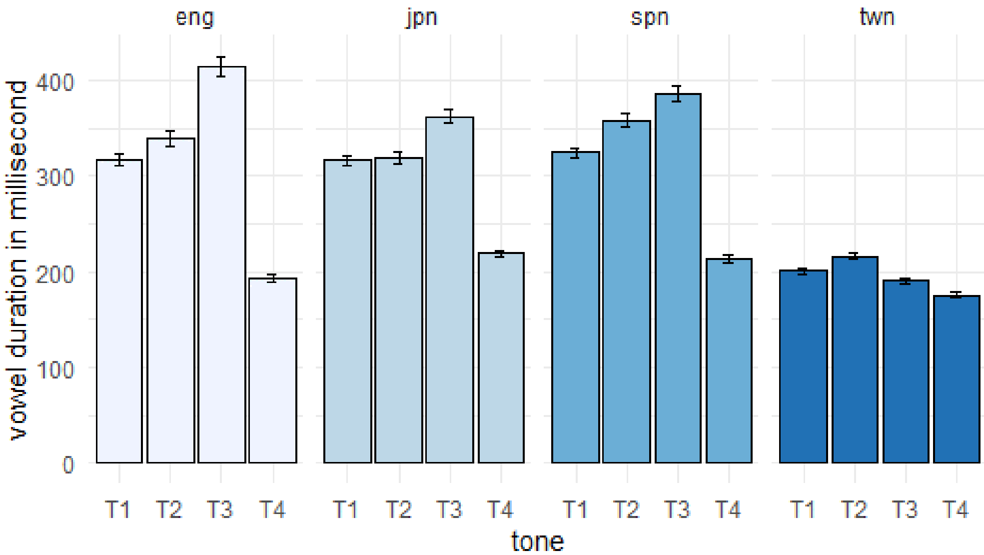

Figure 43. Visual comparison of L2 speakers' vowel durations values in four lexical tones across factors; native Taiwan Mandarin values are added for convenience.

7.2.6 Phrase-Position Effect

Recall that we observed mixed results concerning the phrase-position effect. Lisker and Abramson (1967) propose voicing-lag in syllables at the sentence-*final* would exhibit extended VOT values due to final sentence stress over those at the sentence-*initial* or sentence-*medial* positions. Our results cannot fully support this proposition because in Experiment 1a and 1b, while the SPN, JPN, and Taiwan groups' data showed significant results, the ENG groups did not, and none of the groups showed a significant phrase-position effect in Experiment 2a and 2b. A more carefully designed study is needed for investigating VOT variations due to this effect.

7.2.7 Gender Effect.

Intuitively, females would exhibit shorter VOTs than males because of the cavity size and the high pitch frequency governed by physiological constraints. In the current study, the gender effect was not significant in all groups between experiments, except the JPN group in Experiment 1b; they showed significantly shorter VOT values in females than males. Our challenge here is to explain why other groups did not show the same gender effect and why the overall non-significance.

While our JPN group showed significantly longer VOTs in males than females, Li's (2013) Mandarin study showed that females produced significantly longer VOTs than males. In contrast, from our native data, we observed that our Taiwan group did show longer VOTs in females than males, although non-significant.

Furthermore, in Mandarin Chinese, Li (2010) reports that females tend to adopt more formal language styles than males, which might lead to a slower speech rate (Kendall, 2013). Our finding was that the slower speech rate led to a significantly longer VOT in the L1 Mandarin. Thus, it might be the case that our native Taiwanese males recorded the stimuli informally, and females did so too carefully. This study treated sociolinguistic factors as random effects; therefore, future studies should further control such possible influence and instruct test subjects to provide formal or informal recordings.

On the contrary, Shuju et al. (2016) report that Japanese speakers tend to slow their speech rate to signal their politeness. Therefore, our male Japanese participants probably had recorded the stimuli "politely", thus had significantly longer VOTs than females. Data discussed here seems to suggest a sociolinguistic factor. Future studies will need to account for politeness in speech for VOT influences.

7.2.8 Native Language Effect

The motivation for an NL effect for the current study is the L1 transfer theory (Flege, 1995; Major, 2001; Gass, 2013). Speakers of languages with aspirated stops would produce more Mandarin-like VOT, whereas those without aspirated stops would produce less Mandarin-like VOT (Eckman, 1977). In the present study, we observed a significant NL effect and that the ENG group's VOTs were more native Mandarin-like than the JPN and SPN groups. The preponderance of the evidence does indicate the advantage for the ENG group over the latter two groups, who have to use their newly learned L2 VOT.

Not only did we find an NL effect, but we also observed non-L1 VOT used in L2 Mandarin. As shown in Figure 31 in section 5.12, while the SPN group's L1 voiceless VOTs were not nearly close to the English speakers or the native Mandarin speakers, they showed slightly more Mandarin native-like VOT values than the Japanese participants did. Our Spanish participants had an average of 1.88 years of learning, Japanese 3.68 years, and English 4.23 years (see Table 39). Second language acquisition theories would predict a more native-like production with a longer length of learning (Major, 2001; Gass, 2031). Our results showed that the English participants had the more native-like VOTs (might be due to L1 transfer), Spanish the second, and Japanese the third with respect to VOT length. However, when considering language use, the Spanish participants have lived in Taiwan and reported that they had used it daily by the time of the recording. They showed that with more experience, their VOTs were more native-like than the Japanese participants. Still, their VOTs were not as native-like as the English participants, who might have transferred their L1 VOT instead of learning the Mandarin VOT. Future studies may need to be more selective for test subjects concerning the length of learning and language use.

Even though the specific VOT values differed depending upon the NL were observed, the effects of tone, POA, pitch-level, and speech rate on VOT did occur. These observations provide evidence for responding to our research question 3 that a) the effects found in L1 Mandarin did occur in L2 speech, b) L2 native language did matter in L2 production, and c) we did find both L1 VOTs (i.e., English speakers) and non-L1 VOTs (i.e., Japanese and Spanish speakers) in three different groups of L2 speakers' Mandarin.

7.3 Conclusion and Implication

This dissertation provides empirical evidence that an isolated acoustic property involved complex phonological categories. The present study attempts to investigate VOT with several effects and in various languages. It discusses how the effects operate within phonetics and phonology, which may benefit future studies concerning the experiment's design. Our results strongly suggest that lexical tone and suprasegment features must be considered when examining VOT variations. In our view, the most crucial finding is that the effect of lexical tone on VOT in Mandarin is not due to the language-specific phonological grammar but due to the physiology of the vocal tract and thus be universal.

In addition, we attempted to compare our L2 learners' L1 and L2 VOTs directly. Since not all languages in questions are tonal languages, this direct comparison might be considered a shortcoming. However, if the F0 onset of the post-stop vowel duration can be controlled, the comparison can be more validated.

This cross-linguistic survey also offers multi-dimensional comparisons and confirmation of the interlanguage process as relevant to the field of Second Language Acquisition. The observed phenomena may provide insights to Mandarin instructors about L2 accent variations for L2 English, Japanese, and Spanish learners.

In sum, VOT is more complicated than we thought, and there are many effects on VOT. We showed that lexical tone is one of them, and we showed that there are tone components that specifically affect VOT.

APPENDIXES

Appendix A1: Informed Consent Form, Chinese Version

題目: 漢語第一(L1)和第二(L2)語音中詞彙音調對VOT影響的調□
(An investigation of the effect of lexical tone on VOT in L1 and L2 speakers of Mandarin)

研究程序:
此為研究進行為一項調查 L2 普通話學習之研究。 如果您同意參加，您將被要求提供有關您年齡，性別，語言背景和熟練程度的信息以做統計。 在任務期間，您將被要求在計算機屏幕上大聲朗讀單詞和句子列表。 您的回復將被記錄下來。 完成大約需要 15-20 分鐘。 您的所有錄音及一切有關資料都將被保密。在查問捲和錄音中都不會在調詢問您的姓名。 您參與該研究的是基於自願的，但會有 7 美元或 200 台幣的現金獎勵。

風險:
參與本研究並無任何可預見的風險。

優點:
除了加強您的普通話學習經驗或是幫助理解語言成分的相互關係之外，作為參與者，您並不會獲得第定的任何好處。

保密:
本研究中的數據將保密。 音頻文件不會包含您的姓名。 研究人員只會使用識別碼。 音頻文件將保存在主要調查員辦公室的受密碼保護的大學計算機上。 錄音將保密，並在 5 年後妥善銷毀。 此外，調查問卷將在主要調查員辦公室保存 5 年，並在此之後妥善銷毀。

- 雖然沒有任何計算機傳輸可以完全安全，但我們將做出合理的努力來保護您的傳輸機密性。
- 我們將持有未經參與者的額外同意，可以將去使用此次的數據用於未來的研究。
- 只有相關之研究人員方可使用錄音紀錄。

參與:
您的參與是自願的，但您將獲得 7 美元美元或 200 台幣的現金獎勵，以補償您的參與。

您必須是:

- 台灣出身

- 會□國語
- 有清晰正常之ㄩ語與聽力

您可以隨時以任何理由退出研究。 如果您決定不參加或退出研究，則不會受到任何懲罰或您有權獲得的福利損失。 您無須需支付任何費用或責任。

聯繫:
這項研究由喬治梅森大學語言學部門的 Steven Weinberger 博士進行。 可致電 (703)993-1188 或發送電子郵件至 weinberg@gmu.edu 獲取問題或報告研究相關問題。 如果您對作為研究參與者的權利有任何問題或意見，請聯繫喬治梅森大學，機構審查委員會（IRB）辦公室，電話 703-993-4121。
本研究已根據喬治梅森大學的程序進行了審查，該程序管理您參與本研究。喬治梅森大學機構審□委員會已放棄在此同意書上簽字的要求。 IRBnet＃:1446423-1

承諾書: 如果您已□讀此表並且所有問題都已得到解答，請ㄩ頭告知研究人員您是否同意參加本研究。

Appendix A2: Informed Consent Form, English version
Title: An investigation of the effect of lexical tone on VOT in L1 and L2 speakers of Mandarin

RESEARCH PROCEDURES
This research is being conducted to investigate second language (L2) Mandarin lexical tone learning. If you agree to participate, you will be asked to provide demographic information about your (age, gender, language background, and proficiency). During the task, you will be asked to read aloud a list of words and sentences on a computer screen. Your responses will be recorded. It will take approximately 15 minutes to complete. All your responses will be confidential. You will be asked for your name neither on the questionnaire nor in the recording. Your participation in the study is VOLUNTARY-based, although you will receive US$7.00 (or NTD$200 / JPY¥700 / EUR€6.00) compensation for your time.

RISKS
There are no foreseeable risks for participating in this research.

BENEFITS
There are no benefits to you as participants other than reinforcing your Mandarin learning experience and understanding linguistic components' interrelationships.

CONFIDENTIALITY
The data in this study will be confidential. The audio files will not include your name. The researcher will use identification codes instead. The audio files will be maintained on a password-protected university computer in the principal investigator's office. The recordings will remain confidential and will be destroyed appropriately after five years. In addition, the questionnaire will be kept for five years in the principal investigator's office and will be destroyed appropriately after that time.

- While it is understood that no computer transmission can be perfectly secure, reasonable efforts will be made to protect your transmission's confidentiality.
- The de-identified data could be used for future research without additional consent from participants.
- Only the researchers will have access to the audio recordings.

PARTICIPATION
Your participation is voluntary-based, but you will receive US$7 (or NTD$200 / JPY¥700 / EUR€6.00) compensation for your participation. You must:

- Be a native speaker of English, Japanese or Spanish.
- Have studied Chinese Mandarin for at least six months.
- Can produce lexical tones of Mandarin.
- Have normal hearing and normal speaking speech production.

You may withdraw from the study at any time and for any reason. If you decide not to participate or withdraw from the study, there is no penalty or loss of benefits you are otherwise entitled to. There are no costs to you or any other party.

CONTACT
This research is being conducted by Dr. Steven Weinberger in the linguistics program at George Mason University. He may be reached at (703) 993-1188 or weinberg@gmu.edu for questions or to report a research-related problem. You may contact the George Mason University Institutional Review Board (IRB) Office at 703-993-4121 if you have questions or comments regarding your rights as a participant in the research.
This research has been reviewed according to George Mason University procedures governing your participation in this research.

The George Mason University, Institutional Review Board has waived the requirement for a signature on this consent form. IRBnet#: 1446423-1

CONSENT: If you have read this form and all of your questions have been answered, please verbally inform the researcher(s) if you consent to participate in this study or not.

Appendix B1: demographic questionnaire for the Mandarin natives

Case#:____________
Date: _____________
Gender:____________

1. 您的母語是哪種語言？ ____________________

2. 您有沒有聽力及說話能力上的問題？ Y/N________________________________

3. 您的出身及成長地區？城市 ________________(省份_______________)

4. 請問您的年紀? _____________________________

5. 除了國語，您還會哪些方言？ ______________________________________

6. 您還會說哪些外語？ ____

學習多久了？ _______________ 幾歲開始學習的？ _______________

7. 除了英語您還會哪些外語?___________________________________

8. 有沒有學過音樂? Y/N_____________

9. 國語使用程度? ___國語為主 ____一半一半____方言居多

____跟家人說方言跟外人說國語 or _____________________________

10. 有沒有上過___語言學___音律學___音位學___語音學? No______

11. 您覺得您說話的速度 ___快 ___普通 ___慢?

Principal Investigator: Dr. Steven Weinberger, English Department
Institutional Review Board IRBnet#: 1446423-1

Appendix B2: demographic questionnaire (English Speakers)

Case#:____________
Date: _____________
Gender:___________

1. Do you have normal hearing and speech production? YES/NO___________
2. What is your native Language? ____
3. How old are you? _________________________
4. Where were you born?

   city______________(state_______) country___________
5. What other foreign language(s) besides Mandarin do you know?

   ____
6. How old were you when you started learning Chinese Mandarin?

   ____
7. How long have you studied Mandarin?_________________
8. How did you learn Mandarin? (academically or

   naturalistically)_______________
9. What are your reasons for learning Mandarin?

   ____
10. Do you use Mandarin outside of the classroom? YES/NO___________
11. Have you ever lived in a Mandarin speaking country? YES/NO___________

Principal Investigator: Dr. Steven Weinberger, English Department
Institutional Review Board IRBnet#: 1446423-1

Appendix B3: demographic questionnaire (Japanese Speakers)

Case#:____________
Date: _____________
Gender:___________

1. 会話や聴力など特に問題はありますか? YES/NO___________
2. あたなの母国語はなんですか? ____________________________________
3. 年齢は? _____________________________
4. 出身地は?

   市町村_____________(県_______) 国:___________

5. 母国語以外に話せる外国語はありますか?

   ____

6. いつ頃から中国語勉強し始めたでしょうか? _______________
7. どのぐらい中国語を習いましたか?________________
8. 中国語を習うきっかけ? (カルチャースクール又はその他)______________
9. 中国語を習う理由は? ____
10. 日常生活上、中国語を使う機会ありますか? YES/NO___________
11. 中国語圏などの国で暮らした経験ありますか? YES/NO___________

Principal Investigator: Dr. Steven Weinberger, English Department
Institutional Review Board IRBnet#: 1446423-1

Appendix B4: demographic questionnaire (Spanish Speaker)

Caso#:___________
Fecha: ____________
Género: ___________

1. ¿Tiene una audición y producción de habla normal? SI/NO___________
2. ¿Cuál es su lengua nativa? _______________________________________
3. ¿Cuantos años tiene? ___________________________
4. ¿Lugar de nacimiento?

   ciudad_____________(estado_______) país___________
5. ¿Habla otras lenguas extranjeras aparte de chino-mandarín?____
6. ¿Cuántos años tenía cuando comenzó a estudiar chino-mandarín?

   ____
7. ¿Por cuánto tiempo ha estudiado chino-mandarín?

   ____
8. ¿Cómo aprendió chino-mandarín? (académicamente-de forma natural)

   ____
9. ¿Cuáles son las razones por las que decidió aprender chino-mandarín?

   ____
10. ¿Utiliza el chino-mandarín fuera del aula?SI/NO___________
11. ¿Ha vivido en algún país en donde se habla chino-mandarín? SI/NO___________

Principal Investigator: Dr. Steven Weinberger, English Department
Institutional Review Board IRBnet#: 1446423-1

Appendix C2: 錄音程序

讀錄任務

此任務分為兩部分。您無須戴耳機。

實驗一:您首先會坐在一個電腦前面，顯示著”按空白鍵開始”。開始時會有八題練習題。電腦上顯示的每一個句子都會是混合著國字跟拼音。請您大聲唸出。

當您練習結束後，請告知研究人員開啟錄音機。開始之後請大聲往麥克風念出您在螢幕上看到的句子。如果有任何您不確定的句子，請在腦海裡練習幾次在念。請您用您最自然平常的音調及音量。每按空白鍵前往下一個句子。

當實驗一結束時，您會看到”謝謝”兩個字。 請休息一分鐘。實驗一大約需要四至五分鐘完成。

實驗二:錄音程序及方式同上，但是句子不同，而且會要求您使用不同的音調。請跟著電腦銀幕的指示提高或降低您的音調。請練習到您覺得可以了再開始。實驗二大約需要五分鐘完成。

您如果有任何問題，可以現在或是在實驗途中問研究人員。

您可以在任何時候中止任務，並停止實驗程序。

指導教授: Dr. Steven Weinberger, English Department, Linguistics program, GMU
Institutional Review Board IRBnet#: 1446423-1

Appendix C2: Recording Instruction

Reading aloud task

There are two main tasks. You will NOT be wearing a headphone for either experiment.

For the first task, you will first see the phrase "*press space-key to begin*" on the computer screen. There will be eight practice sentences before the main task. You will see the stimulus sentence displayed in pinyin and Chinese characters on the computer screen one by one. Please read it aloud.

When you are ready, please inform the researcher to start the recorder. When you see the stimuli printed on the screen, you will read it aloud to the microphone in front of you. If you are unsure about the word or the phrase, you can repeat it a time or two in your head before reading it aloud. Please speak in a voice you find most natural and comfortable but loud enough to be recordable. Once you have read it aloud, press the space key to move to the next word or phrase.

When it reaches the end of the task, the computer will show "*Thank you!*" on the screen. Please take a 1-minute break. The first task will take about 4 to 5 minutes to complete.

For the second experiment, the recording procedures are the same as for the first experiment, except for different stimulus sentences, and you will be required to speak in different voice pitches. Please follow the computer screen instructions and practice as many times as you want until you feel comfortable proceeding. The second experiment will take about 5 minutes as well.

If you have any questions, you can ask the research now or during the task.

You can withdraw from the experiments at any time for any reason.

Principal Investigator: Dr. Steven Weinberger, English Department
Institutional Review Board IRBnet#: 1446423-1

REFERENCES

Allen, J. S., & Miller, J. L. (1999). *Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words*. The Journal of the Acoustical Society of America, 106(4), 2031-2039.

Ashby, M., & Maidment, J. (2005). *Introducing phonetic science*. Cambridge University Press.

Abramson, A. S., & Lisker, L. (1973). *Voice-timing perception in Spanish word-initial stops*. Journal of Phonetics, 1(1), 1-8.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). *Random effects structure for confirmatory hypothesis testing: Keep it maximal*. Journal of memory and language, 68(3), 255-278.Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. Journal of Phonetics 29, 155-190.

Boersma, P. (1998). *Functional phonology*. Netherlands Graduate School of Linguistics.

Boersma, P., & Weenink, D. (2018):*Praat: doing phonetics by computer [Computer program]. Version 6.0.37*, retrieved 14 March 2018 from http://www.praat.org/

Browman, C. P., & Goldstein, L. (1984). *Dynamic modeling of phonetic structure*. Status Report on Speech Research, Haskins Laboratories, New Haven, 79(80), 1-17.

Chang, S. S., Ohala, J. J., Hansson, G., James, B., Lewis, J., Liaw, L., & Van Bik, K. (1999). *Vowel-dependent VOT variation: An experimental study.* The Journal of the Acoustical Society of America, 105(2), 1400-1400.

Chao, K. Y., Khattab, G., Chen, L. M. (2006, January). *Comparison of VOT patterns in Mandarin Chinese and in English*. In Proceedings of the 4th Annual Hawaii International Conference on Arts and Humanities (Vol.840, p. 859).

Chao, Y.R. 1930. *A system of 'tone letters'*. Le Maitre Phonetique 45: 24–27.

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

Chen, K., Tsay, J., & Hong, G. (1998). *Duration of initials in Mandarin: Fundamental acoustic research and its clinical significance*. Journal of Speech Language-hearing Association, 13, 138-149.

Chen, L. M., Peng, J. F., & Chao, K. Y. (2009, December). *The effect of lexical tones on voice onset time*. In 2009 11th IEEE International Symposium on Multimedia (pp. 552-557). IEEE.

Ding, H., Jokisch, O., & Hoffmann, R. (2010, May). *Perception and production of Mandarin tones by German speakers*. In Proc. of 5th Conference on Speech Prosody, Chicago, USA.

Eckman, F. R. (1977). *Markedness and the contrastive analysis hypothesis*. Language learning, 27(2), 315-330.

Eckman, F. R. (2004). *From phonemic differences to constraint rankings: Research on second language phonology*. Studies in Second Language Acquisition, 513-549.

Escudero, P., & Boersma, P. (2004). *Bridging the gap between L2 speech perception research and phonological theory*. Studies in Second Language Acquisition, 551-585.

Eshghi, M., Alemi, M., & Zajac, D. J. (2016). *Aerodynamic and laryngeal characteristics of place-dependent voice onset time differences*. Folia Phoniatrica et Logopaedica, 68(5), 239-246.

Ewan, W. G. (1976). *Laryngeal Behavior in Speech*. Doctoral Dissertation, University of California, Berkeley.

Francis, A.L., Ciocca, V., Wong, V.K.M., and Chan, J.K.L. (2006). *"Is fundamental frequency a cue to aspiration in initial stops?"* Journal of Acoustical Society of America 120, 2884-2895

Flege, J. E. (1991). *Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language*. The Journal of the Acoustical Society of America, 89(1), 395-411.

Flege, J. E. (1995). *Second language speech learning: Theory, findings, and problems*. Speech perception and linguistic experience: Issues in cross-language research, 92, 233-277.

Gass, S. M. (2013). *Second language acquisition: An introductory course*. Routledge.

Geissler, C. A. (2018). *VOT in Tibetan is conditioned by tone*. The Journal of the Acoustical Society of America, 143(3), 1756-1756.

Goldinger, S. D. (1998). *Echoes of echoes? An episodic theory of lexical access*. Psychological Review, 105(2), 251.

Hirata, Y. (2004). *Computer-assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts*. Computer Assisted Language Learning, 17(3-4), 357-376.

Hoit, J. D., Solomon, N. P., & Hixon, T. J. (1993). *Effect of lung volume on voice onset time (VOT)*. Journal of Speech, Language, and Hearing Research, 36(3), 516-520.

Hombert, J. M., & Ladefoged, P. (1977). *frequency of the following vowel*. Working Papers in Phonetics, 33.

Hualde, J. I., Olarrea, A., & Escobar, A. M. (2001). *Introducción a la lingüística hispánica*. Cambridge: Cambridge University Press.

Ioup, G., & Weinberger, S. (1987). *Interlanguage phonology: The acquisition of a second language sound system (Vol. 8)*. Newbury House.

Jaeger, J. J. (1978, September). *Speech aerodynamics and phonological universals*. In Annual Meeting of the Berkeley Linguistics Society (Vol. 4, pp. 312-329).

Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch* (Doctoral dissertation, University Library Groningen] [Host]).

Kalveram, K. T., & Jancke, L. (1989). *Vowel duration and voice onset time for stressed and nonstressed syllables in stutterers under delayed auditory feedback condition*. Folia Phoniatrica, 41(1), 30-42.

Kendall, T. (2013). *Speech rate, pause and sociolinguistic variation: studies in corpus sociophonetics*. Springer.

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

Kessinger, R. H., & Blumstein, S. E. (1998). *Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies*. Journal of Phonetics, 26(2), 117-128.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). *lmerTest package: tests in linear mixed-effects models*. Journal of statistical software, 82(13), 1-26.

Kwon, H. (2015). *Cue primacy and spontaneous imitation: Is imitation phonetic or phonological?* (Doctoral dissertation).

Ladefoged, P., & JOHNSON, K. (2010). *A course in phonetics*. Boston: Thomson Wadsworth.

Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). *The raising effect of aspirated prevocalic consonants on F0 in Taiwanese*. In Proceedings of the 2nd International Conference on East Asian Linguistics (Vol. 2). Simon Fraser University Vancouver, Canada.

Lai, Y., & Jongman, A. (2004). *The effect of aspiration on fundamental frequency in Taiwanese syllables*. Unpublished MA Thesis. University of Kansas.

Lam, C. L. *Effect of tones on voice onset time (VOT) in Cantonese aspirated*. Journal of Phonetics, 27, 207-229.

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means.* R package version 1.4.5. <https://CRAN.R-project.org/package=emmeans>.

Luo, Q. (2018). *Consonantal effects on F0 in tonal languages* (Doctoral dissertation). Michigan State University, East Lansing.

Li, F. (2013). *The effect of speakers' sex on voice onset time in Mandarin stops*. The Journal of the Acoustical Society of America, 133(2), EL142-EL147.

Li, X. (2010). *Sociolinguistic variation in the speech of learners of Chinese as a second language*. Language Learning, 60(2), 366-408.

Lisker, L., & Abramson, A. S. (1964). *A cross-language study of voicing in initial stops: Acoustical measurements.* Word, *20*(3), 384-422.

Lisker, L., & Abramson, A. S. (1967). *Some effects of context on voice onset time in English stops*. Language and Speech, 10(1), 1-28.

Liu, H., Ng, M. L., Wan, M., Wang, S., & Zhang, Y. (2008). *The effect of tonal changes on voice onset time in Mandarin esophageal speech*. Journal of Voice, 22(2), 210-218.

Ma, J., Chen, X., Wu, Y., & Zhang, L. (2018). *Effects of age and sex on voice onset time: Evidence from Mandarin voiceless stops*. Logopedics Phoniatrics Vocology, 43(2), 56-62.Maddieson. (1984). Patterns of Sounds. Cambridge: Cambridge University Press.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Routledge.

Massaro, D. W., Cohen, M. M., & Tseng, C. Y. (1985). *The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese*. Journal of Chinese Linguistics, 267-289.

McCrea, C. R., & Morris, R. J. (2005). *The effects of fundamental frequency level on voice onset time in normal adult male speakers*.

Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). *Voice onset time differences between adult males and females: Isolated syllables*. Journal of Phonetics, 36(2), 308-317.

Miller, J. L., Green, K. P. & Reeves, A. (1986) *Speaking rate and segments: a look at the relation between speech production and speech perception for the voicing contrast*, Phonetica, 43, 106-115

Narayan, C., & Bowden, M. (2013, June). *Pitch affects voice onset time (VOT): A cross-linguistic study*. In Proceedings of Meetings on Acoustics ICA2013 (Vol. 19, No. 1, p. 060095). Acoustical Society of America.

Nearey, T. M., & Rochet, B. L. (1994). *Effects of place of articulation and vowel context on VOT production and perception for French and English stops*. Journal of the International Phonetic Association, 24(1), 1-18.

Oh, E. (2011). *Effects of speaker gender on voice onset time in Korean stops*. Journal of Phonetics, *39*(1), 59-67.

Ohala, J. (1983). *The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (editor)*, The Production of Speech, 189-216. New York: Springer-Verlag.

Pearce, M. (2005). *Kera tone and voicin*g. University College London working papers in linguistics, 17, 61.

Peng, J. F., Chen, L. M., & Lee, C. C. (2009, December). *Tonal effects on voice onset time*. In International Journal of Computational Linguistics & Chinese Language Processing, Volume 14, Number 4, December 2009.

Peng, J. F., Chen, L. M., & Lee, C. C. (2014). *Voice onset time of initial stops in Mandarin and Hakka: Effect of gender*. Taiwan Journal of Linguistics, 12(1), 63-79.

Peirce, J. W. (2009). *Generating stimuli for neuroscience using PsychoPy*. Frontiers in neuroinformatics, 2, 10.

Pind, J. (1995) *Speaking rate, voice-onset time, and quantity: the search for higher-order invariants for two Icelandic speech cues*, Perception & Psychophysics, 57, 291-304

Port, R. F., & Rotunno, R. (1979). *Relation between voice-onset time and vowel duration*. The Journal of the Acoustical Society of America, 66(3), 654-662.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.URL https://www.R-project.org/.

Ran, Q. B. (2005). *Experimental studies on Chinese obstruent consonants: with the emphasis on standard Chinese* (Doctoral dissertation, Ph. D. dissertation, Nankai University, Mainland China).

Reetz, H., & Jongman, A. (2014). *Phonetics: Transcription, production, acoustics, and perception*. John Wiley & Sons.

Rochet, B.L. and Fei, Y. (1991). *Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception*. Canadian Acoustics, vol. 19, no. 4, 1991, pp. 105-106.

Sancier, M. L., & Fowler, C. A. (1997). *Gestural drift in a bilingual speaker of Brazilian Portuguese and English*. Journal of phonetics, 25(4), 421-436.

Shimizu, K. (1990). *Cross-Language study of voicing contrasts of stop consonants in Asian languages*. Annexe Thesis Digitisation Project 2016 Block 7.

Shuju, S., Chiharu, T., Xiaoli, F., Jinsong, Z., & Nobuaki, M. (2016, October). *Acoustic correlates and gender effects in production and perception of Japanese polite speech*. In 2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP) (pp. 1-5). IEEE.

Silva, D. J. (1993). *The phonetics and phonology of stop lenition in Korean*. Ph.D. Dissertation, Cornell University

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2015). *Package 'afex'*.

Stevens, K. N. (2000). *Acoustic phonetics* (Vol. 30). MIT press.

Stölten, K., & Engstrand, O. (2002). Effects of sex and age in the Arjeplog dialect: a listening test and measurements of preaspiration and VOT. In *Proc. of Fonetik* (Vol. 44, pp. 29-32).

Sun, S. H. (1998). *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese* (No. 16). University of Hawaii Press.

Swartz, B. L. (1992). *Gender difference in voice onset time*. Perceptual and motor skills, 75(3), 983-992.

Titze, I. R. (1980). *Comments on the Myoelastic-aerodynamic theory of phonation*. Journal of Speech, Language, and Hearing Research, *23*(3), 495-510.

Tse, H. (2005). *The Phonetics of VOT and Tone Interaction in Cantonese* (Doctoral dissertation, University of Chicago).

Tseng, C. C. (2018). *The Effect of Lexical Tones on Voice Onset Time in L2 Mandarin Production by English Speakers*. In Proc. ISAPh 2018 International Symposium on Applied Phonetics (pp. 120-125).

Tseng, C. Y. (1979). *Production and perception of Mandarin tones with respect to vowel duration*. The Journal of the Acoustical Society of America, *65*(S1), S32-S33.

Van den Berg, J. (1958). *Myoelastic-aerodynamic theory of voice production*. Journal of speech and hearing research, 1(3), 227-244.

Whiteside, S. P., & Irving, C. J. (1997). *Speakers'sex differences in voice onset time: Some preliminary findings*. Perceptual and motor skills, 85(2), 459-463E.

Winter, B. (2013). *Linear models and linear mixed-effects models in R: Tutorial 11*. arXiv preprint arXiv: 1308.5499.

Winter, B. (2015). *The F distribution and the basic principle behind ANOVAs*. Author, Birmingham.

Wu, Z. (1997). *Cong Shengdiao yu Yueli de Guanxi Tichu Putonghua Yudiao Chuli de Xinfangfa*. Shangwu Yinshuguan.

Wu, X. (2005). *(Non-)categorical perception of Mandarin tones: A comparison between speakers of tone and non-tone languages*. M.A. thesis, Victoria: University of Victoria.

Xu, G. X, and Y. Xu (2003*). "Effects of consonant aspiration on Mandarin tones."* Journal of the International Phonetic Association 33, 165-181.

Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners*. Berlin: Springer Berlin Heidelberg.

Yavas, M. S. (2011). *Applied English phonology (2nd ed).* Oxford; Malden, MA: Wiley-Blackwell.

BIOGRAPHY

Chiu-ching Tseng, a.k.a. Ken, graduated from CUNY-Queens college with a Bachelor's degree in Computer Science. Ken received his first Master in Science in Foreign Language Education - NCATE accredited TESOL track and his second Master of Arts in General Linguistics from Florida International University. Ken accepted a four-year TAship and started his Ph.D. training in Theoretical Linguistics at GMU, where he also served as a TA for the English department and a graduate lecturer/substitute teacher at INTO Mason, Academic English program.

Ken's teaching career starts in 2010 in Tokyo. He has taught English, Japanese and Chinese commercially and academically. He has also taught General Linguistics, Modern English Grammar, Phonetics, Applied Linguistics, and TESL Practicum courses. Ken has joined the Defense Language Institute, Foreign Language Center in spring 2021.

Ken is interested in Phonetics, Phonology, and Second Language Acquisition. He has researched these fields in production and perception cross-linguistically. His teaching approach is fundamentally Task-Based (TBLT) and utilizing other methods to target specific learning objectives because he visualizes that learning should be microscopic. Ken believes in Skill Acquisition Theory (DeKeyser, 2000) and an Object-Orientated-Learning approach. Thus, his idea of teaching style is to be as micro, adaptable, and eclectic as possible, and he devotes himself to helping students to develop skills for approaching language learning both holistically and analytically.