PRODUCTION AND PERCEPTION OF LARYNGEAL CONTRASTS IN MANDARIN AND ENGLISH BY MANDARIN SPEAKERS

by

Yuting Guo
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: _____ Summer Semester 2020
George Mason University
Fairfax, VA

Production and perception of laryngeal contrasts in Mandarin and English by Mandarin speakers

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

by

Yuting Guo
Master of Arts
Syracuse University, 2015
Master of Arts
Hefei University of Technology, 2014

Director: Harim Kwon, Assistant Professor
Department of Linguistics

Summer Semester 2020
George Mason University
Fairfax, VA

DEDICATION

This is dedicated to my loving parents, husband, and daughter with LOVE.

ACKNOWLEDGEMENTS

I would like to take this special opportunity to publicly thank many people, without whom I could never have achieved so much.

My deepest gratitude and appreciation go to my advisor, Harim Kwon, for her generous support and unwavering dedication throughout the entire dissertation process. I am lucky enough to have her as my dissertation advisor. She has devoted her time and full attention to me whenever I asked for it, and she always gives me sound advice and thorough feedback. She has provided detailed comments on every draft and on every aspect of the dissertation from theoretical discussion to smallest typos. She is a great role model for me as a scholar, mentor, and teacher. I am especially grateful to Steven Weinberger for welcoming me to the GMU linguistics family and introducing me to the world of phonetics and phonology. His passion and love of phonetics and phonology has greatly influenced my research interests. He would always make time to meet with me when I needed and provide thoughtful comments at every stage of my dissertation. I am also grateful to Douglas Wulf for his expertise and insight. He helped me to clarify my thought with inspiring questions and feedback. I sincerely appreciate all the encouragement and support from my superb committee. This work would not have been possible without their invaluable insight and guidance. I am also grateful to Cynthia Lukyanenko for her help with the analysis of the data. I benefited a lot from her stats office hours and research methods class in terms of experiment design and data modeling. I also thank Tuuli Morrill for teaching me acoustics analysis and various experimental techniques.

I also would like to thank my MA advisors at Syracuse University: Amanda Brown and Bei Yu. Amanda is a great mentor. She introduced me to academia in the U.S., and she supported me in every possible way. Without her encouragement and help, I would not have been able to enter the field of linguistics and start this wonderful journey in my life. Bei opened the door of computational linguistics for me. She boosted my confidence to learn new things in a brand-new area and prepared me to learn R to analyze and visualize data. I also wish to thank my MA advisor at Hefei University of Technology, Jianghong Han, for his inspiration and guidance. I have been extremely fortunate to be surrounded by very caring and supportive teachers throughout my academic life. I am very grateful to all of them.

Many thanks to all my friends in the department. I thank Zhiyan Gao for generously sharing his learning experiences with me and cheering me up when I felt stressed. I thank

Omar Alkhonini, Sarah Alamri, and Sahar Almohareb for sharing the ups and downs of my graduate life. Special thanks to Chiu-ching Tseng and Yamei Wang for asking me out to chat and enjoy delicious food. I also thank Abdullah Alfaifi, Hussain Almalki, Ali Alelaiwi, Aseel Al-Ammar, and Ben Hunt for their friendship and support.

I owe my deepest gratitude to my parents, Shenglin Guo, Xiuying Chen and my older sister, Yumei Guo. I thank them for their unconditioned support over the years. The 12/13-hour time difference and the Pacific Ocean between us has never been an obstacle for their love. A very special thanks goes to my husband Zizhao for having faith in me, always being there for me, cooking tasty food for me, planning fun trips and making our life full of happiness. A super big thanks goes to my adorable daughter Rhyan. She has filled my life with joy. I love you all.

Finally, the experiments in this dissertation were supported by the linguistics program at George Mason University. I am grateful to Mason for supporting my Ph.D. studies by a Presidential scholarship, a summer fellowship grant and a dissertation completion grant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

PRODUCTION AND PERCEPTION OF LARYNGEAL CONTRASTS IN MANDARIN AND ENGLISH BY MANDARIN SPEAKERS

Yuting Guo, Ph.D

George Mason University, 2020

Dissertation Director: Harim Kwon

Word initial stop contrast is realized on many acoustic dimensions, and the acoustic realization of the laryngeal contrast is different across languages. The language-specific hybrid of acoustic cues for the laryngeal contrast shapes how speakers and listeners represent and identify the contrast. This dissertation explores how Mandarin speakers produce and perceive the laryngeal contrasts in their native language (L1) and the second language (L2), focusing on F0 (fundamental frequency) perturbation patterns. By doing so, this study aims to contribute to the theoretical discussion of the production-perception link, the L1-L2 interface, and the cross-linguistic comparison between tonal and non-tonal languages.

This dissertation features two sets of experiments: Mandarin speakers' L1 production and perception experiments (§3), and Mandarin speakers' L2 production and perception experiments (§4). Mandarin speakers' L1 experiments (§3) examine the

acoustic cues that Mandarin speakers use to produce and perceive the Mandarin aspiration contrast. The current study observes a high initial tone and low initial tone effect in producing the aspiration contrast in Mandarin. The F0 following aspirated stops (F0-aspirated stops) is found to be significantly higher than the F0 following the unaspirated stops (F0-unaspirated stops) in the high-level tone and the falling tone but is significantly lower than the F0-unaspirated stops in the rising and the dipping tone. The duration of significant F0 differences between the F0-aspirated stops and the F0-unaspirated stops is limited to the onset of the vowel, ranging from 11 ms to 75 ms. In perception, the voice onset time (VOT) is the primary cue for the aspiration judgement. Moreover, L1 Mandarin listeners are observed to extract both tonal and consonantal information from the post-onset F0. The listeners tend to associate high pitch with the aspirated stops and low pitch with the unaspirated stops across the four tonal contexts. The high initial tone and low initial tone effect is observed in the perception task as well. The low initial tones elicit significant more unaspirated responses than the high initial tones.

Mandarin speakers' L2 experiments (§4) investigate L1 Mandarin speakers' perception and production of the English voicing contrast with parallel tasks from the L1 experiments. Overall, the F0 following voiceless stops (F0-voiceless stops) is found to be produced significantly higher than the F0 following voiced stops (F0-voiced stops). The duration of significant F0 differences between the F0-voiceless stops and the F0-voiced stops is shorter than that produced by L1 English speakers. In perception, the L1 Mandarin listeners use VOT as a primary cue and pitch as a secondary cue for the

English voicing contrast. They tend to associate high F0 with the voiceless stops and low F0 with the voiced stops. The intrinsic F0 of the vowels also play a role and a perceptual compensation effect is observed in the English stop identification task. It seems that the listeners tend to attribute the high F0 they hear to the intrinsic F0 of the vowel rather than the voicelessness feature of the preceding stop when VOT is ambiguous.

The findings in this dissertation indicate that F0 perturbation effect is primarily an automatic effect. The features such as tone and vowel height that are realized with F0 can influence F0 perturbation patterns as different tones and vowels require different coordination of articulators. In addition, this study sheds light on why the F0 perturbation duration in tonal languages is shorter than that in non-tonal languages. A comparison between Mandarin speakers' L1 and L2 stop production suggests that speakers from a tonal language inhibit the F0 perturbation effect to keep the tonal information intact. In sum, the parallel studies of the laryngeal contrasts across languages and modalities in this dissertation offer insight into between-language (tonal vs. non-tonal) and within-language (production vs. perception) variations of how Mandarin speakers-listeners use different acoustic properties to contrast laryngeal features in their L1 and how they adapt the information of individual acoustic cues when they learn an L2. Along with such findings, this dissertation also provides a balanced corpus for testing models of the perception-production link as well as the L1-L2 interface.

CHAPTER 1. INTRODUCTION

1.1 Background

The difference between two L1 sounds is very straightforward to L1 listeners. Nonetheless, the process of integrating diverse information from speech signals is complex. This is because languages encode multiple acoustic cues to contrast speech sounds and the weight and function of each acoustic cue can differ across languages. For example, both English and Mandarin have a two-way laryngeal contrast (such as the /t/-/d/ distinction), but the phonetic realization of the laryngeal contrast could be language specific (see detailed discussion in §2.5).

The distinction between English stops is often considered as a voicing contrast and the two groups of stops are usually termed voiced stops and voiceless stops (the English stops are henceforth called voiced and voiceless stops). However, the voicing distinction is primarily a phonological one. The phonological voicing distinction, especially in syllable-initial positions is actually realized as a phonetic aspiration distinction between voiceless unaspirated stops and voiceless aspirated stops (e.g., Lisker & Abramson, 1964; Keating, 1984; Francis, Ciocca, Wong & Chan, 2006). Mandarin stops are typically classified as voiceless unaspirated stops and voiceless aspirated stops (Xu & Xu, 2003; Deterding & Nolan, 2007; Luo, 2018), with aspiration as the primary distinction (the Mandarin stops are henceforth called unaspirated and aspirated stops).

The role of F0 in English differs from that in Mandarin. In English, the F0 of the post-stop vowel is a secondary cue to the voicing contrast (e.g., Whalen, Abramson, Lisker & Mody, 1993). Voiceless stops tend to raise the F0 while voiced stops tend to lower the F0 of the following vowel in production (Hombert, Ohala & Ewan, 1979; House & Fairbanks, 1953; Lehiste & Peterson, 1961; Lea, 1973; Hombert, 1978; Ohde, 1984; Hansen, 2009). The phenomenon that the F0 pattern is correlated with the laryngeal feature of the preceding consonant is called *F0 perturbation*. Direction and duration are two commonly used measurements to describe F0 perturbation (Hombert et al., 1979; Luo, 2018). The direction of F0 perturbation measures whether the onset obstruents raise or lower the post-onset F0. The duration of F0 perturbation measures how long the F0 perturbation extends into the vowel. The effect of F0 perturbation in English can extend about 100 ms into the following vowel (e.g., Hombert et al., 1979). Studies have shown that L1 English listeners pay attention to the secondary cue (i.e., F0) in the perception of stop categories, not only when the primary cue (i.e., VOT) is ambiguous (e.g., Abramson & Lisker 1985) but also when it is not ambiguous (e.g., Whalen et al., 1993, see §2.3.2 for more detailed discussion of this issue).

Unlike English, Mandarin is a phonemic tonal language, in which tone can distinguish word meanings (e.g., Duanmu, 2007). There are four[1] tones in Mandarin: High level tone (T1), Rising tone (T2), Dipping tone (T3) and Falling tone (T4). F0 is the primary cue to convey the tonal information. Inconsistent results have been reported on F0 perturbation in Mandarin (Xu & Xu, 2003; Luo, 2018). Xu and Xu (2003) suggest that

[1] Mandarin also has a neutral tone and its tonal contour depends on the preceding tone. It is not considered in the current study.

aspiration is associated with low F0 while Luo (2018) argues that aspiration is associated with high F0. Despite the discrepancy of the direction of F0 perturbation in Mandarin, both studies have reported that the duration of F0 perturbation effects in Mandarin is limited to the onset of the following vowel. It is still unknown whether Mandarin listeners use F0 as a secondary cue to the laryngeal contrast, since F0 is the primary cue for tones and the F0 perturbation effect is very short in Mandarin productions.

L1 can influence L2 production and perception, which makes it interesting to examine the L1 influence on the acquisition of laryngeal contrasts in an L2, especially when the two languages differ in how they realize the laryngeal contrast. The learners may need to adjust the category definitions in their L1 in order to acquire the category mapping in their L2 to achieve target-like production and perception. For Mandarin speakers learning English as an L2, they may need to inhibit the link between F0 and tonal categories and associate F0 with the laryngeal contrast in order to recognize and produce speech sounds in a nativelike manner.

1.2 The current study

This study has been designed to contribute to the understanding of how speakers of a tonal language contrast the laryngeal features in production and identify the laryngeal features in perception, focusing on the role of VOT and F0. Along with the examination of native F0 perturbation patterns of Mandarin speakers, the current work also investigates how Mandarin speakers define the English laryngeal contrast in production and adjust their use of acoustic cues in perception.

For the Mandarin production experiment (§3.2), L1 Mandarin speakers were asked to read Mandarin words with alveolar stops in onset position in a carrier sentence. The production experiment thus attempts to reveal the influence of tone and intrinsic F0 of the vowel on the direction and duration of F0 perturbation in Mandarin. It seeks to provide additional empirical evidence as to whether the aspiration feature in Mandarin raises or lowers F0 at vowel onset, since previous studies (Xu & Xu, 2003; Luo 2018) have reported contradictory results. In order to explore the production-perception interface, this study also asks whether L1 Mandarin listeners use F0 as a cue for the laryngeal contrast in Mandarin (§3.3). A series of stops co-varying in VOT and F0 for each tone were created to test how VOT, F0, and tone influence L1 listeners' perception of Mandarin stops. The participants were asked to listen to Mandarin words and identify the aspiration feature by selecting the [t]-initial or [t^h]-initial words.

The same group of participants were asked to read English words with an alveolar stop as the onset in a carrier sentence to examine how Mandarin speakers produce the laryngeal contrast in English (§4.2). The participants' perception of the English stops (§4.3) was investigated using the same experimental design as the L1 perception experiment. The L2 experiments seek to investigate how L1 Mandarin speakers contrast the laryngeal feature in their non-tonal L2. The results from the two sets of parallel experiments thus explore the interface between L1and L2 in both production and perception. The results of the current study also have implications of Mandarin VOT categorical boundaries and English VOT categorical boundaries by L1 Mandarin speakers.

1.3 Organization of chapters

This dissertation proceeds as follows: Chapter 2 surveys the literature on the debate of the source of F0 perturbation, the production and perception of F0 perturbation in tonal and non-tonal languages, and F0 perturbation in L2 acquisition studies. Chapter 2 also provides a background review for Mandarin tones, and the laryngeal contrasts in Mandarin and English. Chapter 3 investigates the role of VOT and F0 in contrasting and identifying Mandarin stops by L1 speakers through a production and a perception experiment. Chapter 4 focuses on L1 Mandarin speakers' perception and production patterns of English stop contrasts. Chapter 5 summarizes the results, provides a general discussion the L1-L2 link, offers suggestions for future research and concludes the study.

CHAPTER 2. LITERATURE REVIEW

F0 perturbation has been widely attested in different languages across the world, such as Cantonese (Francis et al., 2006; Luo, 2018), Danish (Jeel, 1975; Reinholt Petersen, 1983), English (House & Fairbanks, 1953; Lehiste & Peterson, 1961; Lea, 1973; Hombert, 1978; Hombert et al., 1979; Ohde, 1984; Hanson, 2009), French (Kirby & Ladd, 2016), German (Kohler, 1982), Japanese (Gao & Arai, 2018), Korean (Han & Weitzman, 1970; Jun 1996), Mandarin (Xu & Xu, 2003; Luo, 2018), Russian (Mohr, 1971), Spanish (Dmitrieva, Llanos, Shultz & Francis 2015), Southern British English (Silverman, 1984), Thai (Gandour, 1974; Ewan, 1976), Xhosa (Jessen & Roux, 2002), and Yoruba (Hombert, 1977; Hombert, 1978). Despite the universality of the phenomenon, there are controversial patterns within and across languages, which could be due to the complexity of the phenomenon itself.

Although all the languages listed above have a laryngeal contrast, the acoustic realization of the contrast differs across languages. Specifically, the languages differ in terms of whether they are tonal, non-tonal, or pitch-accent languages, whether they are true voicing or aspiration languages, and the number of stops they have along the laryngeal contrast. Tonal languages use F0 to express different lexical or grammatical meanings while non-tonal languages do not. As to pitch-accent languages, F0 is essential to determine the meaning of a word in some tonal minimal pairs (Gao & Arai, 2018). It is

still unclear whether F0 can carry both consonantal and tonal information in the tonal or pitch-accent languages. Aspiration languages, such as Mandarin, have two series of voiceless stops (voiceless aspirated and voiceless unaspirated), and aspiration is the phonemic feature contrasting the two groups of stops. In contrast, in true voicing languages, such as Spanish (Dmitrieva et al., 2015), the phonologically voiced stops are phonetically voiced (with voicing during stop closure) and the phonologically voiceless stops are phonetically voiceless. Voicing is the phonemic feature contrasting the two groups of stops in true voicing languages. Languages like English, Mandarin, French, and Italian have a two-way stop contrast in which they have two-member-pairs for the laryngeal contrast. There are also languages such as Korean in which they have three-member-groups for the laryngeal contrast. Even within one language, various acoustic properties are used to realize the laryngeal contrast. For example, Lisker (1986) listed as many as 16 acoustic patterns involved in a /p/-/b/ contrast between *rabid* and *rapid*.

In addition to these various cross-linguistic differences related to the laryngeal contrast, different research procedures and measurements could be another potential reason for the contradictory results on F0 perturbation, such as the direction of the F0 contours throughout the vowel. Some studies (e.g., House & Fairbanks, 1953) have measured average F0 over the vowels following the onset consonants, some studies (e.g., Lehiste & Peterson, 1961) have measured the maximum F0 in the vowel, and some studies (e.g., Luo, 2018) have measured F0 from certain equidistant points throughout the vowel. As to the stimuli, some studies have used real words (e.g., Lehiste & Peterson, 1961), whereas others have used nonsense words (e.g., Lea, 1973). In terms of syllable

structure, some have used one syllable words (e.g., House & Fairbanks, 1953), some bi-syllabic words (e.g., Lea, 1973; Xu & Xu, 2003), and some tri-syllabic words (e.g., Silverman, 1984). Some have recorded words in isolation (e.g., Lehiste & Peterson, 1961), while other studies have embedded the target syllables in carrier phrases (e.g., Luo, 2018). More cross-linguistic evidence is needed to provide a better understanding of this universally observed phenomenon.

The current study focuses on how L1 Mandarin speakers distinguish the laryngeal contrast in production and identify the stops in perception, both in their L1 and L2. The following sections survey literature on the source of F0 perturbation (§2.1), F0 perturbation in English (§2.2) and tonal languages (§2.3) in both production and perception tasks, the laryngeal contrast in L2 acquisition studies (§2.4) and the background of the two testing languages in the present study (§2.5).

2.1 The source of F0 perturbation

2.1.1 F0 perturbation

As introduced above, obstruents influence the F0 of adjacent vowels. The phenomenon that the F0 in a vowel is correlated with the laryngeal feature of the preceding consonant is called *F0 perturbation* (Hombert et al., 1979; Kirby & Ladd, 2016; Hanson, 2009; Luo, 2018). Direction and duration are two common measurements to define F0 perturbation (Hombert et al., 1979; Hanson, 2009; Luo, 2018). The direction of F0 perturbation measures whether the onset obstruents raise or lower the post-stop F0. The duration of F0 perturbation measures how long F0 perturbation extends into the vowel.

Studies have collected data from various languages to understand the source of F0 perturbation. Two major approaches have been proposed to account for this universally attested phenomenon: the phonetic approach and the phonological approach.

2.1.2 The phonetic approach

The phonetic approach claims that F0 perturbation is intrinsic to the voicing contrast of the preceding consonant and it is restricted by physiological and/or aerodynamic factors (e.g., Ladefoged, 1967; Slis, 1970; Halle & Stevens, 1971; Ohala & Ohala, 1972; Löfqvist, 1975; Hombert et al., 1979; Kohler, 1984). Various articulatory and aerodynamic mechanisms have been proposed to elucidate the nature of this phenomenon.

One such possible mechanism is *larynx height*. That is, the larynx is lowered to facilitate voicing by increasing the size of the oral cavity. A larger oral cavity helps to maintain the necessary trans-glottal pressure for voicing. The lowered larynx leads to a decrease in F0 at the vowel onset, which has been observed in many previous studies (Moeller & Fischer, 1904; Ohala, 1972; Ewan, 1976).

Another proposed mechanism is the *cricothyroid (CT) muscles and vocal cord tension.* Halle and Stevens' (1971) claim that the vocal cord (a.k.a., vocal fold) tension differs in the course of producing voiced and voiceless stops. During the production of the voiceless stops, the CT muscles contract to inhibit voicing, which stretches the vocal cords. Thus, the vocal cords become stiff. In contrast, during the production of voiced stops, the vocal cords are slack. The effect of the different vocal cord tension during the

stop closure spreads to the following vowel. Tense or stiff vocal folds lead to high F0 and slack vocal folds lead to low F0.

Trans-glottal pressure has also been examined. Trans-glottal pressure refers to the pressure difference between the subglottal pressure and the oral pressure. During stop closure, oral pressure builds up. And upon the release of the stop, the oral pressure drops. However, the pressure accumulated during the closure period is different for voiced and voiceless stops. The oral pressure builds up quickly for the voiceless stops since the glottis is widely open while it builds up gradually for the voiced stops (Ladefoged, 1971). As a result, the oral pressure is higher for the voiceless stops than for the voiced stops. The increase of the oral pressure during the closure of a voiced stop reduces the trans-glottal pressure, which lowers the F0 at the vowel onset. Voiceless stops, with high oral pressure, induce faster trans-glottal airflow at the vowel onset and raise the F0 at the beginning of the following vowel (Ladefoged, 1967; Kohler, 1984).

Lastly, *subglottal pressure* has been considered. During stop closure, a constant subglottal pressure is retained (Slis, 1970; Ohala & Ohala, 1972; Löfqvist, 1975). Upon stop release, a higher rate of air flows out of the glottis for aspirated stops than for unaspirated stops, which leads to a greater decrease of the subglottal pressure for aspirated stops than for unaspirated stops. The F0 at vowel onset is positively correlated with the remaining subglottal pressure after stop release. Therefore, the F0-aspirated stops is lower than the F0-unaspirated stops (Shi, 1998; Xu & Xu, 2003; Francis et al., 2006).

Most of the mechanism mentioned above are about phonetically voiceless and voiced obstruents. Only the last one, the subglottal pressure mechanism, focuses on the differences between voiceless aspirated and voiceless unaspirated stops. If the F0 perturbation is an automatic effect, the mechanism activated for voiced obstruents is expected to be different for the voiceless unaspirated stops, although they could belong to the same phonological voicing category.

2.1.3 The phonological approach

The phonetic approach suggests that F0 perturbation is a physiological or aerodynamic artifact of how obstruents are produced (e.g., Ladefoged, 1967; Halle & Stevens, 1971; Hombert et al., 1979). However, Kingston and Diehl (1994) propose that F0 perturbation is 'controlled' rather than 'automatic.' It is worthwhile to note that the 'controlled' hypothesis does not mean that the speakers are doing it consciously. It is engrained as a part of their subconscious phonological knowledge, which is language specific. The aim of F0 perturbation is to enhance the perceptual salience of the contrastive laryngeal features. The F0 perturbation pattern is related to the phonological status of the obstruents rather than the specific phonetic realization of the obstruents with different degrees of prevoicing or delayed onset of voicing. In English, the F0-voiced stops is lowered in both the syllable initial positions and the intervocalic positions although voicing during the closure typically only occurs in the intervocalic positions (Kingston, 1985; Kingston & Diehl, 1994). Cross-linguistically, English phonologically voiced (phonetically voiceless unaspirated) and French phonologically voiced (phonetically voiced) stops both lower the F0 of the following vowel (Hombert, 1978).

Based on the observed patterns, Kingston and Diehl (1994) argued that the F0 perturbation was independent of the specific phonetic realizations of the laryngeal contrast and it was maintained to reinforce a phonological contrast.

To sum up, the phonetic approach and the phonological approach provide explanations of F0 perturbation from different perspectives with the former focusing on production and the latter focusing on perception. However, they do not directly contradict to one another. They have different predictions for production patterns, but shares some predictions for perceptual patterns. Specifically, the phonetic approach predicts that all human languages show the same pattern as long as the obstruents are phonetically equivalent regardless of the phonological contrasts in each individual language. The phenomenon is expected to be universal rather than language specific. It does not, arguably, predict a direct link between the production and perception. The L1 listeners may or may not use F0 as cue for the laryngeal contrast.

On the other hand, the phonological approach suggests the F0 perturbation is related to the phonological contrast in a given language regardless of the specific phonetic realization of the laryngeal contrasts. It predicts languages exhibit the same pattern when they have the same phonological contrast. It predicts that the production is closely related to the perception as the purpose of F0 perturbation is to enhance a phonological contrast for listeners. L1 listeners are predicted to be able to use F0 as a cue for the laryngeal contrast.

2.2 F0 perturbation in English

2.2.1 Contrasting English voicing in production

The F0 perturbation effect in English is well documented in the literature (e.g., House & Fairbanks, 1953; Lehiste & Peterson, 1961; Lee, 1973; Hombert, 1975; Hanson, 2009). In general, studies have found that the F0-voiceless stops in English is higher than the F0-voiced stops. The duration of the perturbation can extend 100 ms into the vowel (e.g. Lehiste & Peterson, 1961; Hombert et al., 1979). Despite the different experimental designs and acoustic measurements across studies, the general pattern of F0 perturbation is robust both in its direction and duration.

The study of House and Fairbanks (1953) measures average F0 over the vowels in CVC syllables and finds that the average F0 in vowels following the voiceless consonants was higher than that in vowels following the voiced consonants in English. Examining F0 contours of a subset of data, they note that the F0 perturbation affects only the initial period after voicing rather than throughout the vowel. Lehiste and Peterson (1961) measure Maximum F0 and the F0 contour of English CVC syllables, reporting that F0 is higher when the prevocalic consonant is voiceless than when it is voiced. They maintain that the effect of F0 perturbation can extend to mid-vowel, with maximum F0 occurring immediately after the initial voiceless consonant while occurring at about the middle of the vowel after voiced consonants.

Lea (1973) uses bisyllabic nonsense words and real English words. The structure of the nonsense words is /həCVC/ with stress on the second syllable. The study used target words recorded in isolation. He reports that the F0 following voiceless obstruents is

higher than that following voiced obstruents. For the nonsense words, he notes that the F0 contour following voiceless obstruents descends from the onset of voicing to mid-vowel while the F0 contour following voiced obstruents ascends from voicing onset to mid-vowel. The general finding is that stress and the phonological voicing contrast indeed influence the F0 contour. Lea's (1973) study can be compared to Silverman (1984), who uses three-syllable nonsense words embedded in carrier phrases in Southern British English. In Silverman's (1984) study, the consonant voicing is found to influence the F0 of vowels preceding and following the consonant with stressed syllables exhibiting a stronger effect of F0 perturbation than the unstressed syllables. Both Lea (1973) and Silverman (1984) suggest word-level stress influences the F0 perturbation.

Hombert (1975) also examines the F0 contour in English by asking the participants to read '*Say C[i] again*.' He finds that voiceless stops raise the F0 of the following vowel and voiced stops lower the F0 of the following vowel with the greatest F0 difference at the vowel onset. In addition, he observes inter-speaker differences in terms of the F0 contour during the post-onset vowels, suggesting the specific phonetic realization of the voicing contrast in English could be slightly different for different speakers.

In addition to the 'microprosody' (word-level stress), Jun (1996) observes that the 'macroprosody' (sentence-level intonation) influences the F0 perturbation. She uses data from speakers of American English, Korean and French. In her study, the target CV syllables are placed in positions varying in their prosodic contexts. For American English, she finds that the F0 perturbation effect varies by intonation environments with

the focused syllable showing the greatest F0 difference between the voiced and voiceless stops. In a similar study, Hanson (2009) places /CVm/ syllables in carrier sentences and asks the participants to produce the syllables in high, low and neutral pitch environments. In the high pitch environment, the F0 following voiceless obstruents is significantly higher than the nasal baseline and the F0 following the voiced obstruents closely traces the baseline. In the low pitch environments, the F0 following both groups of obstruents is slightly higher than the baseline. She observes a conflict between the segmental level F0 perturbation and the sentence level intonation. When the two were in opposite directions, the segmental level perturbation would be weakened. This finding that stress and sentence level intonation influences F0 perturbation potentially indicates that F0 perturbation is an automatic effect. When other linguistic information needs to be realized with F0, the F0 perturbation pattern is influenced due to the possible restriction of human articulators.

2.2.2 Identifying English voicing in perception

Given the robust correlation between the post-stop F0 and the voicing feature of the stop, studies (Abramson & Lisker, 1985; Haggard, Ambler & Callow, 1969; Fujimura, 1971; Whalen, Abramson, Lisker & Mody, 1990; Whalen et al. 1993) have investigated whether L1 listeners use F0 information for the voicing identification. The phonological approach (§2.1.3) predicts that L1 listeners will use F0 as a cue for a voicing distinction as it argues that the aim of F0 perturbation is to enhance the voicing contrast. The phonetic approach (§2.1.3) does not directly predict whether the F0 information will be used in perception. However, it is not incompatible with the idea that

listeners will use the information. Since the F0 perturbation pattern in English is consistent, the L1 listeners may learn the correlation between F0 and the voicing features of the preceding consonant, which as a result, helps them to identify stops in perception. Studies confirmed the predictions (Abramson & Lisker, 1985; Haggard et al., 1969; Fujimura, 1971; Whalen et al., 1990, Whalen et al., 1993). VOT has been reported as the primary cue for the voicing feature of syllable initial stops for English listeners (e.g., Lisker & Abramson, 1964; Abramson & Lisker, 1985). F0 has been reported as a secondary cue (Abramson & Lisker, 1985; Haggard et al., 1969; Fujimura, 1971; Whalen et al., 1990; Whalen et al., 1993). Abramson and Lisker (1985) create a set of synthesized labial stops followed by the low back vowel [a]. The stimuli covary with four steps of VOT (5 ms, 20 ms, 35 ms, 50 ms), four steps of onset F0 (98 Hz, 108 Hz, 120 Hz, 130 Hz) and three levels of F0 perturbation durations (50 ms, 100 ms, 150 ms). They ask L1 English listeners to identify whether the stimuli had /b/ or /p/ as the onset. They find that pitch has a heavier influence on the stimuli with ambiguous VOT than on the stimuli with unambiguous VOT. Based on these findings, they conclude that VOT is the dominant cue while F0 has a modest effect on the consonant voicing judgement. These results are in line with the study conducted by Whalen et al. (1990), which focuses on the effect of sentence level intonation on the perception of the stop voicing contrasts in syllable-initial positions. They have determined that the F0 onset values rather than the intonation contour contributed to the voicing judgement. Whalen et al. (1990) also report that the onset F0 results in a gradient effect on the voicing judgement although the stimuli with lower onset F0 differ from that with higher onset F0. For the stimuli with higher F0 onset,

the higher the F0, the more voiceless responses. However, for lower F0 onsets, there is only a marginal statistical significance of the gradient effect, i.e. the lower the F0 onsets, the more voiced responses.

Whalen et al. (1993) further explore the role of F0 in voicing judgements by collecting both categorical responses of voicing and the reaction time of the perceptual judgements. They use stimuli covaried with seven steps of VOT (5 ms, 10 ms, 15 ms, 20 ms, 25 ms, 35 ms, 50 ms) and five steps of F0 onset values (98 Hz, 108 Hz, 114 Hz, 120 Hz, 130 Hz). Whalen et al. (1993) conduct two experiments. In experiment 1, the F0 perturbation duration in the stimuli is 50 ms and in experiment 2, the duration is the entire vowel duration. They confirm that F0 influenced listeners' judgement when VOT is ambiguous. Further, they find that when the VOT and F0 information are incongruent (i.e. stimuli with long VOT and low F0 or short VOT and high F0), it takes listeners a longer time to give a response than when the two cues are congruent (i.e. stimuli with long VOT and high F0 or short VOT and low F0). Based on this finding, they suggest that listeners use F0 information even when VOT is not ambiguous. Their results indicate that listeners take both VOT and F0 information into consideration when making voicing judgements.

A majority of the studies on English F0 perturbation in production and perception have examined the pattern with L1 English participants. Cross linguistic evidence is needed to enable a better understanding of the source of F0 perturbation. If F0 perturbation is a pure phonetic effect due to the constriction of human articulator mechanisms, the L2 learners are expected to show the same pattern as the native

speakers. If the F0 perturbation is a phonological effect which is language specific, the L2 learners may not be able to produce the native-like pattern especially when their L1 and the target language use different acoustic cues to signal the phonological contrast. Chapter 4 of this dissertation examines F0 perturbation in English by L2 learners, focusing on how L1 Mandarin speakers produce and perceive the English voicing contrasts.

2.3 F0 perturbation in Mandarin and other tonal languages

2.3.1 Contrasting the laryngeal feature in production

Only a few studies have examined F0 perturbation in tonal languages. Studies that have examined tonal languages agree that the duration of F0 perturbation is shorter in tonal languages than in non-tonal languages, but report contradictory results in terms of the direction of F0 perturbation. See Table 1 for a summary of F0 perturbation direction in tonal languages reported by previous studies.

Table 1. F0 perturbation direction in tonal languages

	F0-aspirated stops >F0-unaspirated stops	**F0-aspirated stops <F0-unaspirated stops**
Thai	Ewan (1976)	Gandour (1974)
Mandarin	Luo (2018)	Xu & Xu (2003), Howie (1974)
Cantonese	Zee (1980), Luo (2018)	Francis et al. (2006)
Taiwanese	Lai et al. (2009)	NA

A handful of studies have examined F0 perturbation in Mandarin and the results have been inconsistent. Howie (1974) measures the F0 contours of the Mandarin citation syllables. He visually illustrates tone variation between aspirated and unaspirated syllables, showing that F0-unaspirated stops is lower than F0-aspirated stops. However, Howie (1974) provides no statistical analysis to compare the F0 values of the vowel onset.

Xu and Xu (2003) examine the effect of aspiration on tones and used dissyllabic words to test the effect of tonal co-articulation. The target syllable is either the first or the second syllable. They report that the F0-unaspirated stops is higher than the F0-aspirated stops in all the four tones. The pattern is different from what has been observed in non-tonal languages, such as English (e.g., Hombert et al., 1979). Based on aerodynamic conditions, they suggest that the subglottal pressure is the main factor causing the F0 differences between aspirated and unaspirated stops (see §2.1.2 for detailed discussion). Xu & Xu (2003) also suggest that tone of the target syllable and tone of the preceding

syllable play a role in F0 perturbation in Mandarin. As to the perturbation effect, Xu & Xu (2003) suggest it is greater in T2 and T3 than in T1 and T4. In addition, the perturbation effect is greater when the preceding tones are T1 or T2, which are the tones end at a high F0 range, than when the preceding tones end at a low F0 range. They do not report the duration of the perturbation effect on F0.

Luo (2018) reports a study on F0 perturbation patterns in Mandarin and Cantonese, but her findings are quite different from the results reported by Xu and Xu (2003). She finds the F0-aspirated stops to be higher than the F0-unaspirated stops and the F0-sonorants (the F0 of the vowel following sonorants) in Mandarin and Cantonese. She finds that perturbation durations of 35 ms in T1, 5 ms in T3 and 30 ms in T4. She observes no significant F0 perturbation effect in T2. She proposes that the F0 perturbation effect tends to be longer in tones with high onset pitch than with low onset pitch. She argues that the high initial tones are more salient than low initial tones; therefore, they permit more F0 variability.

Both Xu and Xu (2003) and Luo (2018) have examined F0 perturbation in Mandarin. They both suggest the duration of F0 perturbation is limited to the onset of the vowel. In addition, they both find that the lexical tone influences the F0 perturbation and the tones with high onset pitch behave differently from the tones with low onset pitch. However, they provide opposite results in terms of the direction of F0 perturbation. As to the influence of lexical tones, Xu and Xu (2003) report the perturbation effect is greater in tones with low onset pitch than in tones with high onset pitch, while Luo (2018) reports the opposite pattern. More studies are needed to explain the contradictory results

reported by the two previous studies (Xu & Xu, 2003; Luo, 2018). Chapter 3 of this dissertation investigates the F0 perturbation in Mandarin and aims to enrich the empirical evidence of understanding the F0 perturbation in tonal languages.

2.3.2 Identifying the laryngeal feature in perception

Given that F0 is the primary cue for lexical tones in tonal languages and that the duration of F0 perturbation is relatively limited to the onset of the vowel, it is interesting to investigate whether listeners of a tonal language use F0 at the vowel onset as a cue for the laryngeal contrast at syllable initial positions. Few studies have tested perception of the laryngeal contrast by listeners from a tonal language. One exception, to the author's knowledge, is a study conducted by Francis et al. (2006). They explore whether L1 Cantonese speakers use F0 as a cue for aspiration. They select naturally produced [p^ha] with high level tone as their base token. They use sixteen syllables synthesized from the base token by fully crossing four levels of onset F0 (127 Hz, 137 Hz, 147 Hz, 57 Hz) and four levels of F0 perturbation duration (10 ms, 20 ms, 40 ms, 80 ms). In their study, listeners select between two buttons labeled with Chinese characters, constituting the aspiration pair in Cantonese. The study does not produce straightforward results. With a perturbation duration of 10 ms, no significant responses difference is found among the four onset levels, suggesting a 10 ms perturbation duration is too short for the listeners to hear a difference. For the other three F0 perturbation duration levels, syllables with 157 Hz onsets are identified as aspirated words significantly more often than the lower onset F0 onsets levels, suggesting listeners tend to associate the high onset F0 with aspirated stops. When the F0 perturbation duration is above 40 ms, the onset 147 Hz syllables are

identified as aspirated stops significantly more often than the onset 127 Hz and 137 Hz syllables. The 137 Hz syllables are significantly more aspirated than the 127 Hz syllables only when the perturbation duration is 80 ms. There seems to be an interaction between the onset F0 differences and the F0 perturbation duration. Both high onset F0 and long perturbation duration contribute to the identification of aspiration. When the onset F0 is high (157 Hz), the listeners give aspirated responses despite the short perturbation duration. When the onset F0 is relatively low, the perturbation duration has to be long enough for the listeners to give aspirated responses.

There is a disconnection between the results of Francis et al.'s (2006) production and perception experiments. In their production experiment with a different group of individuals, they find the F0-unaspirated stops to be higher than the F0-aspirated stops and the perturbation duration to be about 10 ms for the high level tone stimuli. Taking the results from both of their experiments together, it indicates the listeners use F0 information in a way that is not consistent with the production patterns. They propose that the strategies that the Cantonese listeners use in the aspiration judgement task are from their exposure of English rather than their L1 experience, although they acknowledge the English that the participants exposed to is mostly Hong Kong English. Their participants were recruited from the University of Hong Kong and were familiar with Hong Kong English at the time of testing. However, it remains unclear whether Hong Kong English, whose speakers are mostly bilinguals of Cantonese and Hong Kong English, shows the F0 perturbation patterns similar to what has been reported in other varieties of English spoken by monolingual English speakers.

Luo (2018) suggests that F0 is a weak cue for aspiration in Mandarin. However, no study so far has examined whether L1 Mandarin listeners use F0 as a cue for the identification of aspiration in Mandarin. Francis et al. (2006) find an intriguing disconnection between production and perception in Cantonese. This might be due to that they use different groups of individuals for their production and perception experiments. In addition, they do not examine the potential influence of tones on F0 perturbation. Chapter 3 of this dissertation seeks to fill the gap and to explore the effect of tone, onset F0 and VOT on consonant aspiration judgement in both production and perception tasks completed by the same group of participants.

2.4 The laryngeal contrast in second language studies

2.4.1 Factors influencing the laryngeal contrast in L2

The L1 influence on L2 perception and production is well-known in the literature. However, the majority of the studies on L2 production and perception have focused on how non-native speakers produce and perceive speech sounds that are not used in their L1, such as the English /l/-/ɹ/ contrast for Japanese leaners (e.g., Miyawaki, Jenkins, Strange, Liberman, Verbrugge & Fujimura, 1975) and the /i/-/ɪ/ and /æ/-/ɛ/ distinction for L2 English learners of German, Spanish, Mandarin and Korean (e.g., Flege, Bohn & Jang, 1997). There are also studies that have explored how learners acquire non-native speech contrast between aspiration and true voicing languages, such as the Russian stop voicing contrast by Mandarin learners (Yang, Chen & Xiao, 2020), the English aspiration contrast by Dutch learners (Simon, 2009), the Spanish stop voicing contrast by English learners (Flege & Eefting, 1987a), and the Dutch stop voicing contrast by English

learners (Flege & Eefting, 1987b). However, most of the above-mentioned studies on laryngeal contrast focus only on VOT, the primary acoustic property that distinguishes the contrast, rather than F0. In addition, cases in which the L1 and L2 contrasts assign different weights to the corresponding cues in the two languages and cases in which the same acoustic cues are used for different purposes in the learners' L1 and L2 are under-studied. In these cases, the learners need to adjust or switch their attention to appropriate cues when they acquire L2 speech sounds.

Among the few studies investigating language pairs in which L1 and L2 differ in the weight of corresponding acoustic cues, Schertz, Cho, Lotto and Warner (2015) examine Korean speakers' L1 and L2 (English) stop contrasts in production and perception. Seoul Korean stops contrast in VOT, F0 at vowel onset, and closure duration. Both VOT and F0 serve as the primary cues (e.g., Lee & Jongman, 2012; Lee, Politzer-Ahles & Jongman, 2013) as Korean has a three-way stop contrast. The English stop contrast is realized primarily in VOT. F0 at vowel onset and closure duration are considered to be secondary cues (e.g., Francis, Kaganovich & Driscoll-Huber, 2008). Schertz et al. (2015) first examine L1 Korean speakers' productions of word-initial stops in their L1 and L2. They use a set of stops covarying in seven steps of VOT, seven steps of F0 and three steps of closure duration to examine the influence of the three acoustics cues (i.e., VOT, onset F0 and closure duration) on Korean listeners' perception of the stop contrast in their L1 and L2. Overall, they determine that both VOT and F0 are important in L1 Korean speakers' representations of the L2 English stop contrasts due to the fact that VOT and F0 are important cues in the L1 Korean stop contrast. The weight of the

acoustic cues in L1 influences the weight distribution in L2 production and perception tasks, and it is challenging for the learners to adapt the weight of acoustic cues in their L1 to the target language.

In addition to the influence of L1, L2 proficiency and L2 learning experience (e.g., Flege & Eefting, 1987a), age and sex (Yu, De Nil & Pang, 2015) may contribute to non-target like patterns in L2 production and perception. L2 proficiency and L2 learning experience are reported to be positively correlated with target-like patterns. In general, high proficiency groups and groups with more authentic language exposure are shown to be more likely to show target-like patterns (e.g., Flege et al., 1997). Age of onset negatively correlates with native-like patterns. The earlier the learner starts to learn the L2, the more likely the learner successfully acquires the target patterns (e.g., Jia, 2006). The L2 proficiency, L2 learning experience, age and gender, which could potentially influence the L2 patterns, are carefully controlled in this study, in order to focus on the L1 influence.

Although the L1 influence on L2 production has been well studied in the literature, research on the role of F0 in the L2 voicing contrast by learners from a tonal language remains sparse. Chapter 4 of this dissertation attempts to understand how L1 Mandarin speakers demonstrate English voicing contrast in production and perception focusing on the role of F0, which carries different information in Mandarin and English.

2.4.2 The interface between production and perception in L2

In L1 production, VOT is a reliable indicator of the laryngeal contrast across languages (Lisker & Abramson, 1964) while F0 at vowel onset exhibits different patterns

across languages and even within the same language (Xu & Xu, 2003; Luo, 2018). In perception, L1 listeners attend to various acoustic cues when judging stop voicing categories (e.g., Whalen et al., 1990). The use of these acoustic cues in the learners' L1s can be fully or partially reflected in their L2 production and perception (Flege & Eefting, 1987a; Schertz et al., 2015). Studies have showed that learners use various acoustic cues for L2 stop contrasts in both production and perception tasks (Schertz et al., 2015; Flege & Eefting, 1987a). Flege and Eefting (1987a) examine how monolingual L1 Spanish speakers and native Spanish speakers learning English as an L2 produce and identify English and Spanish stops in word initial positions. Unlike English, Spanish is a pre-voicing language. Instead of contrasting short lag VOT and long lag VOT, the true voicing languages such as Spanish contrast voicing lead with short lag VOT. They find that the participants produce significantly shorter VOTs for Spanish stops than for English stops. In perception, the mean VOT boundary for the /d/-/t/ continuum for the adult participants is significantly shorter than that for the monolingual English speakers, suggesting that the adult participants' perception of the English voicing contrast is influenced by the Spanish pattern in their L1.

Schertz et al. (2015) find VOT and F0 are important cues in the Korean speakers' L2 production and perception at the group level. At the individual level, all the Korean speakers in the study consistently use VOT and F0 to contrast English stops in production, although F0 is a weak cue for the English contrast. However, the perception results are more variable than the production results. Listeners in the study employ different categorization strategies. Some mainly rely on VOT (consistent with the L2

pattern), whereas some primarily depend on F0 (consistent with their L1 pattern) and some attend to both cues equally to distinguish the L2 stop contrast. They (Schertz et al., 2015) suggest that the different perception patterns are not directly linked to the production patterns in either L1 or L2. The lack of connection between production and perception is still unclear.

Chapter 4 of this study examines an under-studied case in which the learners' L1 and L2 share the same laryngeal contrast (i.e. phonetic aspiration), but one of the acoustic cues involved in the contrast has different functions in the two languages in both production and perception: F0 is a primary cue for tone in Mandarin (a tonal language) and a secondary cue for stop contrast in English (a non-tonal language). By examining this, this study can contribute to the understanding of whether L2 learners can adjust the function of acoustic cues in their L1 to the L2 norm by inhibiting certain acoustic information.

2.5 Background of the target languages

2.5.1 Mandarin tones

Mandarin is a phonemic tonal language in which tone expresses word meanings (e.g., Duanmu, 2007). There are four[2] tones in Mandarin: High level tone, Rising tone, Dipping tone and Falling tone. They are usually represented by T1, T2, T3 and T4 respectively. The tones in Mandarin are commonly transcribed with numbers introduced by Chao (1930). It is a numeric translation of the tonal contours, where 5 represents the highest pitch and 1 represents the lowest pitch. Each tone is represented by a starting

[2] Mandarin also has a neutral tone and its tonal contour depends on the preceding tone. It is not considered in the current study.

pitch and an ending pitch, with an optional number in the middle representing the mid pitch (Chao, 1930; Duanmu, 2007). Table 2 represents the tone inventory in Mandarin.

Table 2. Tone inventory in Mandarin

Tones	**High level**	**Rising**	**Dipping**	**Falling**
Tone labels	T1	T2	T3	T4
Pitch number (Chao, 1930)	55	35	214	51
Examples	tu1 (55)	tu2 (35)	tu3 (214)	tu4 (51)

According to Xu (1997), T1 (55) begins with a high F0 and maintains the same level though the entire vowel; T2 (35) starts with a low F0, then falls slightly until 20% into the vowel before rising throughout the rest of the vowel; T3 (214) begins with a low F0, falls to the lowest F0 at the midpoint of the vowel, then rises sharply to the end of the syllable; and T4 (51) starts with the highest F0 and drops sharply from the 20% of the vowel until the end of the syllable. See Figure 1 for a visualization of the tonal contours. The four tonal contours are statistically different from each other (Luo, 2018). A number of studies have reported the relationship between post-stop F0 and the physiological correlates. Moisik, Lin and Esling (2014) have examined the laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound. They find that larynx height positively correlated with F0 in the production of Mandarin tones. This pattern is also reported in previous electromyographic research by Sagart, Pierre,

Benedicte & Catherine (1986) and Hallé (1994). Overall, the previous studies report that the larynx is higher in T1 and T4 than T2 and T3.

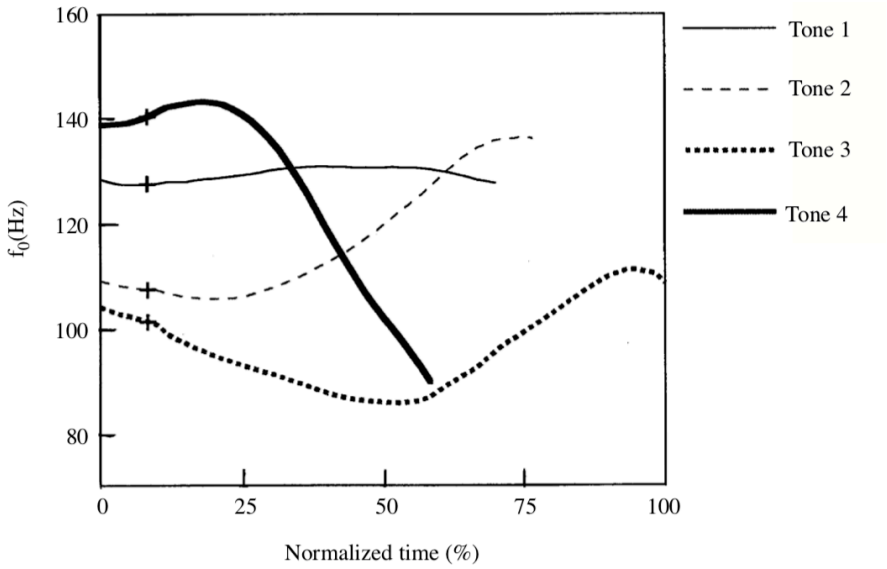


Figure 1. Mean F0 contours of Mandarin tones in monosyllable /ma/ produced in insolation. Time is normalized, with all tones plotted with their average duration proportional to the average duration of T3 (from Xu, 1997).

F0 height and F0 contour has been proposed as the primary acoustic cues to Mandarin tones in production studies (Liang & van Heuven, 2004; Alexander, 2010). In perception, F0 height, overall F0 contours (Howie, 1976; Xu, 1997), amplitude contour (Whalen & Xu, 1992), voice quality (Garding, Kratochvil, Svantesson & Zhang, 1986) and vowel duration (Blicher, Diehl & Cohen, 1990) are important cues for listeners to identify tones. Among all the acoustic cues, the F0 contour has been reported as the most

salient cue for L1 speakers to distinguish different tones (Gandour & Harshman, 1978; Massaro, Cohen & Tseng, 1985).

2.5.2 Mandarin stops

There are six stops in Mandarin with three different places of articulation, which are bilabial stops /p^h/-/p/, alveolar stops /t^h/-t/, and velar stops /k^h/-/k/. All Mandarin stops are voiceless and aspiration is the primary feature of the laryngeal contrast. The Mandarin stops are typically termed as voiceless unaspirated and voiceless aspirated stops (Xu & Xu, 2003; Duanmu, 2007; Luo, 2018). The mean VOTs of Mandarin stops from three previous studies are summarized in Table 3.

Table 3. Mean VOTs of Mandarin stops from three previous studies

	Rochet & Fei[3] (1991)	**Liu et al. (2008)**	**Li (2013)**	
Gender	Not reported	Male	Female	Male
/p^h/	99.6 ms	96.5 ms	NA	NA
/t^h/	98.7 ms	97.8 ms	93.3 ms	77.9 ms
/k^h/	110.3 ms	104.6 ms	90.8 ms	78.8 ms
/p/	13.0 ms	NA	NA	NA
/t/	13.7 ms		12.5 ms	17.5 ms
/k/	26.3 ms		22.5 ms	29.5 ms

[3] The mean VOT of the unaspirated group was calculated based on a VOT Figure in Rochet & Fei (1991).

2.5.3 English stops

English stops are often considered as a contrast in voicing (e.g., Lisker, 1986; Francis et al., 2006). However, the distinction is mainly a phonological one especially for syllable-initial stops. The voicing contrast in onset position is primarily realized with aspiration, so the stop distinction is between voiceless unaspirated stops and voiceless aspirated stops (Lisker & Abramson, 1964; Zlatin, 1974; Keating, 1984; Francis et al., 2006). In intervocalic position, voicing tends to occur during stop closure of voiced stops, which makes it a true voicing contrast (Deterding & Nolan, 2007). Mean VOTs of English stops from Lisker and Abramson (1964) are summarized in Table 4.

Table 4. American English VOT means and ranges (Lisker & Abramson, 1964)

Stops	**/p/**	**/t/**	**/k/**	**/b/**	**/d/**	**/g/**
Mean VOT	58 ms	70 ms	80 ms	1 ms	5 ms	21 ms
Range	20–120 ms	30–105 ms	50–135 ms	0–5 ms	0–25 ms	0–35 ms

2.5.4 Comparing the English and Mandarin stops

The laryngeal contrast in syllable-initial position in both English and Mandarin is realized as a phonetic distinction between voiceless unaspirated and voiceless aspirated stops. However, studies have provided different views whether Mandarin and English should be categorized in the same group along the VOT continuum. Chao and Chen (2008) note that Mandarin aspirated stops have longer VOT values than English aspirated

stops, and they suggest that Mandarin should be categorized into the highly aspirated group and English into the aspirated group according to the framework proposed by Cho and Ladefoged (1999). Cho and Ladefoged (1999) summarize the VOT of velar stops from 18 languages and argued that languages can be divided into four different phonetic categories along the VOT continuum: the unaspirated stops with VOT around 30 ms, the slightly aspirated stops with VOT around 50 ms, the aspirated stops with VOT around 90 ms and the highly aspirated stops with VOT above 100[4] ms. However, Lisker and Abramson (1964) suggest Mandarin and English belong to the same two-way contrast language group with voiceless unaspirated stops ranging from 0 to 25 ms along the VOT continuum and 60 to 100 ms for voiceless aspirated stops. Deterding and Nolan (2007), having asked L1 Mandarin and British English speakers to produce isolated stops in their L1, identify no significant VOT differences between the corresponding syllable initial stops in the two languages. This suggests that Mandarin and English should be categorized into the same group. However, Deterding and Nolan (2007) do find a significant difference of voicing between English and Mandarin when the stop with a short lag VOT is in the intervocalic position. The English stop with a short lag VOT exhibits a significantly longer voicing duration than the Mandarin stops with a short lag VOT during stop closure. The voicelessness of Mandarin unaspirated stops in both syllable-initial and intervocalic positions indicates that Mandarin is an aspiration language (Jessen, 2001; Luo, 2018). However, the different behavior of English stops in syllable-initial and intervocalic positions makes it a controversial case. Following other

[4] In their paper, Cho and Ladefoged (1999) did not specify a specific VOT value for the highly aspirated stops. The number is inferred from Figure 9 in their paper on page 223.

studies, the Mandarin stops in this study will be referred to as unaspirated stops and aspirated stops (e.g., Xu & Xu, 2003; Luo, 2018) and the English stops will be referred to as voiced and voiceless stops (e.g., Lisker & Abramson, 1964).

This study examines how L1 Mandarin speakers produce and perceive Mandarin and English stop contrasts, primarily focusing on the role of F0 at vowel onset. If the F0 perturbation is a pure automatic effect, the L1 Mandarin speakers are expected to produce the English stop contrast in a target-like manner. Because in syllable-initial position, the phonetic distinction of the English and Mandarin stops is realized as one between voiceless unaspirated and voiceless aspirated stops. Phonetically, the stops in both languages are realized in a similar way as to the laryngeal contrast. If the F0 perturbation pattern produced by the L1 Mandarin speakers is different between English and Mandarin, it is likely resulted from the influence of the lexical tones in Mandarin. If the F0 perturbation is a phonological effect, the L1 Mandarin speakers may not be able to produce the English stops in a native-like way. Because the phonological approach indicates that the phonological contrast could be language specific. The phonological contrast in Mandarin and English could potentially be represented by different features. Some studies (Keating, 1984; Kingston & Diehl, 1994) proposed that the laryngeal feature in both true voicing and aspiration languages was [voice], while some other studies (Jessen & Ringen, 2002; Beckman, Jessen & Ringen, 2009) argued that the laryngeal contrast in aspiration languages should be represented by the feature [spread glottis] rather than [voice]. Examining Mandarin speakers' L2 behaviors can help to understand the source of F0 perturbation and comparing Mandarin speakers' L1 and L2

behaviors can contribute to the understanding of why F0 perturbation in tonal languages are shorter than that in non-tonal languages. It could be the speakers from non-tonal languages enhance the stop contrast by making it longer or the speakers from tonal languages inhibit the effect of F0 perturbation to avoid it distorts the onset of the tonal contours. In tonal languages, F0 perturbation and tonal information are both represented by pitch. Therefore, the enhancement of F0 perturbation could potentially influence the perception of the tonal contours. If the Mandarin speakers produce longer F0 perturbation duration in English than in Mandarin, it could indicate that maintaining the full tonal information will inhibit the effect of F0 perturbation. If the Mandarin speakers produce short F0 perturbation in English, it may suggest that the production of the tones is not the primary reason of the short perturbation duration in tonal languages.

CHAPTER 3. F0 PERTURBATION AND STOP ASPIRATION IDENTIFICATION IN MANDARIN

3.1 Introduction

The two experiments in this chapter examine L1 Mandarin speakers' production and perception patterns of their native stop aspiration contrasts. Data from a production experiment and a perception experiment were collected from the same group of participants in order to get matched data sets used to analyze group differences across modalities (i.e. production and perception) in terms of the usage of acoustics cues.

The production experiment investigates the role of VOT, tone and the intrinsic F0 of the vowels in Mandarin aspiration contrasts by L1 Mandarin speakers. Previous studies (Xu & Xu, 2003; Luo, 2018) have observed short F0 perturbation in Mandarin although they find different F0 perturbation patterns in terms of direction. VOT is expected to have a significant influence on the contrast as suggested by previous studies (Xu & Xu, 2003; Luo, 2018). If F0 perturbation is primarily a phonetic effect due to physiological and/or aerodynamic factors (Xu & Xu, 2003; Shi, 1998), lexical tone in Mandarin and the intrinsic F0 of the vowels were expected to modulate the F0 perturbation pattern. T1 and T4 start from a high F0 range and T2 and T3 are starting from a low F0 range (Xu, 1997). Different intrinsic F0 values of the vowel require different coordination of the articulators (e.g., Whalen & Levitt, 1995). If the F0 perturbation is mainly a phonological effect due to perceptual enhancement of the

salience of the voicing feature by the speakers (Kingston & Diehl, 1994), it was expected that the perceptual enhancement would apply to all the tonal and vowel environments. The F0-unaspirated stops may or may not be higher than the F0-aspirated stops, as there is evidence supporting both directions (Xu & Xu, 2003; Luo, 2018). L1 Mandarin speakers were expected to produce the consonant-induced pitch differences with short durations, which was shown in both Mandarin and other tonal languages (e.g., Hombert et al., 1979; Luo, 2018).

The perception experiment explores the influence of VOT, post-stop F0 and tone on the perception of the Mandarin stop aspiration contrasts. The same group of participants were asked to complete a forced-choice identification task employing stimuli varying in their VOT and post-stop F0 in the four tonal environments. VOT was predicted to be the most important cue for the Mandarin listeners to distinguish aspiration, as VOT is a primary cue for aspiration in Mandarin (Lisker & Abramson, 1964; Duanmu, 2007). The participants were predicted to be able to hear the pitch differences associated with consonant aspiration. Francis et al. (2006) report that the Cantonese listeners are able to hear consonant-induced pitch differences in Cantonese, which is also a tonal language. Tone is predicted to influence the perceptual judgement. The tones in Mandarin starting from different F0 ranges, with T1 and T4 starting from a high F0 range and T2 and T3 starting from a low F0 range (Xu, 1997). The F0 variability permitted by each tone could be different. Luo (2018) report that T1 and T4 are significantly different from each other throughout the vowel except for the 50% to the 70% portion, while T2 and T3 are significantly different from each other from 60% mark

to the end of the vowel. The high initial tone group (T1 and T4) and the low initial tone group (T2 and T3) may have different influences on the participants' perception.

To sum up, the goal of the two experiments is to investigate the role of VOT, pitch and tone in distinguishing the aspiration contrast in Mandarin by L1 Mandarin speakers in both production and perception tasks. In addition, the results of the production and perception experiments also have implications of the categorical VOT boundaries in Mandarin.

3.2 Experiment 1: native Mandarin speakers' L1 production

Experiment 1 examined Mandarin speakers' productions of the stop aspiration contrast in Mandarin, focusing on the role of VOT and F0 at the vowel onset.

3.2.1 Method

3.2.1.1 Stimuli

The stimuli were CV syllables: /ta/, /t^{h}a/, /tu/, /t^{h}u/, /wa/ and /wu/. Each of the six syllables carried the four tones in Mandarin. See Figure 2 for a visualization of the four tonal contours. Because /t^{h}a/with the rising tone is lexically missing in Mandarin, it was excluded along with the unaspirated counterpart /ta2[5]/. Ohde (1984) suggests that the F0 patterns of /t^{h}a2/ and /p^{h}a2/ are consistent, so /p^{h}a2/ was used as a substitute (see Xu & Xu, 2003). In order to avoid the potential confounding effect of the place of articulation, /pa2/ was selected to form a minimal pair with /p^{h}a2/. In selecting the onset consonant of the stimuli, the number of lexically missing items was considered. Alveolar stops (/t/ and /t$^{h/}$) and the glide (/w/) yielded the least number of lexical gaps in Mandarin when they were combined with cardinal vowels /a/ and /u/, and thus were selected.

[5] Numbers are attached to the syllables to refer to its respective tone.

Written Mandarin words corresponding to each stimulus were selected based on word frequency data from the Modern Chinese Balanced Corpus (Xiao, 2010), maintained by the National Language Institute (corpus size=100 million words). Only the words labeled as most common were selected. None of the selected words act as a bound morpheme in Mandarin. The stimuli with the stop onsets served as the target words to examine the F0 perturbation patterns in Mandarin, while those with the approximant onsets served as control words to provide a baseline of the tonal contours without F0 perturbation. The set of the stimuli consisted of 24 unique words (3 onsets * 4 tones * 2 vowels). The full set of stimuli is given in the Appendix A2.

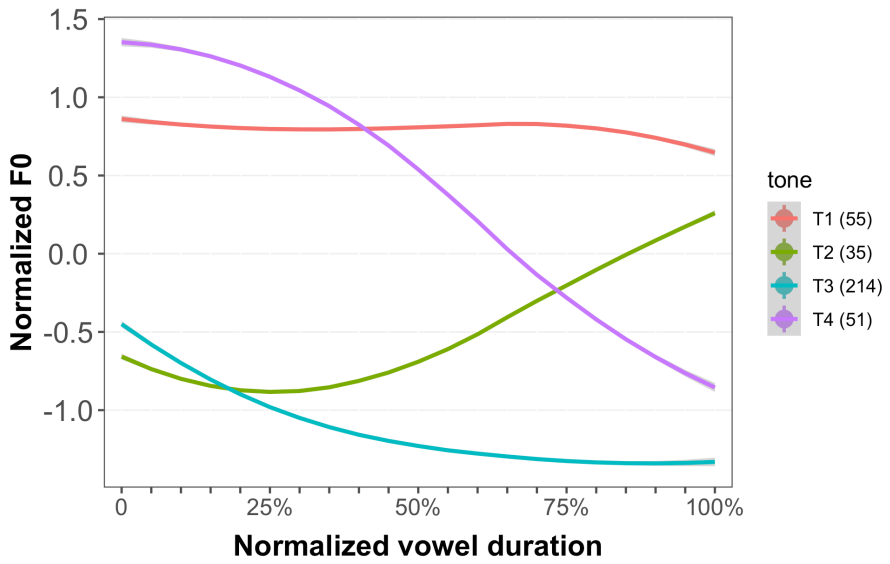


Figure 2[6]. Tonal contours of the four Mandarin tones with normalized vowel duration with data from the Mandarin speakers' L1 production experiment

[6] F0 values were extracted from twenty equidistant points of the post-stop vowel.

Each experimental word, both in Chinese characters and Pinyin, was incorporated into a carrier sentence written in Chinese characters, 请说___一次 (/tʰɕiŋ3 ʂwɔ1 ____ ji2[7] t͡sʰi4/), meaning '*Please say* ____ *one time.*' Native Mandarin speakers would not need Pinyin to read common words in Chinese. Pinyin was included in addition to the characters because this experiment was also designed for participants speaking Mandarin as an L2 for a separate study. The experimental words are fairly common for L1 speakers, but may not be accessible to beginner L2 learners.

3.2.1.2 Participants

Twenty-five native Mandarin speakers (15 female and 10 male) were recruited at George Mason University (GMU). None of them reported hearing or speaking disorders. Table 5 shows the participants' demographic information. Twenty-four of the participants were students at GMU or INTO Mason, and one of them was a Mason employee at the time of the experiment. All of them were residents of Fairfax, Virginia. Ages of the 25 participants ranged from 19 to 46, with a mean age of 26 years (s.d. = 8). They were all Mandarin-English bilinguals, dominant in Mandarin. All participants were born and grew up in China, and they self-identified as native speakers of Mandarin. At the mean age of 22 (s.d. = 3), they moved to the United States, who had lived in the States for 1 year on average (s.d. = 1), except for one participant, who had been in the U.S. for 20 years. This participant's production and perception patterns were not significantly different from the other participants, so she was included in the analysis.

[7] The tone of the syllable /ji/ is underlyingly T1 but becomes T2, because it is followed by a T4 (Duanmu, 2007, p. 245).

Table 5. Demographic information of the participants

N	Age (years)			Age of arrival (years)			Length of residence (months)			
	mean	s.d	range	mean	s.d.	range	mean	s.d	range	median
25	26	8	19-46	22	3	19-35	13	13	1-48[8]	12

3.2.1.3 Procedure

The experiment took place in a sound attenuated booth at GMU. Participants were seated comfortably in front of a Macbook and their productions were digitally recorded onto a separate Macbook Pro, using a Røde smartLav+ microphone and an external Focusrite Scarlette Solo 2nd Generation preamplifier with a sampling rate of 44.1 kHz via the Praat program (Boersma & Weenink, 2020). The microphone was attached to the participants' shirt on the upper chest, approximately 6 inches away from the speakers' mouth. Stimuli sentences were presented automatically to the participants in the middle of the computer screen in randomized orders using PsychoPy (Peirce, 2007).

In order to elicit a comparatively stable speaking rate across participants, the sentences were presented with a 3.5-second inter-stimulus interval. All instructions of the experiment were given in Mandarin. Participants were instructed to read aloud each sentence, as naturally as possible. The experiment consisted of a 2-trial practice block

[8] The participant who has been in the U.S. for 20 years is not included in the length of residence column of the demographic summary table. With her data added in, the descriptive statistics of length of residence is highly skewed towards the higher end, mispresenting the trend in the variable.

and two experimental blocks. The experiment session included altogether 144 trials (24 words * 3 repetitions/block * 2 blocks). There was a self-paced break between the blocks, and the experiment took approximately 10 minutes.

3.2.2 Acoustic measurements

All measurements were performed with Praat (Boerma & Weenink, 2020) by the author. Positive VOT was measured manually from the starting point of the target stop burst in the waveform to the first zero crossing in the waveform following the onset of the periodicity of the following vowel. The end of the VOT also marked the onset of the vowel. The end of the vowel was labeled at the offsets of the second formants. The segmentation between the vowel and the approximant was mainly determined by visual inspection of the spectral patterns. The boundary was located at the point where the second formant (F2) moved up from the steady-state position (Peterson & Lehiste, 1960). The third formant (F3) was used when F2 was not obvious (Xu & Liu, 2007). Examples of token segmentation are given in Figure 3 and 4.

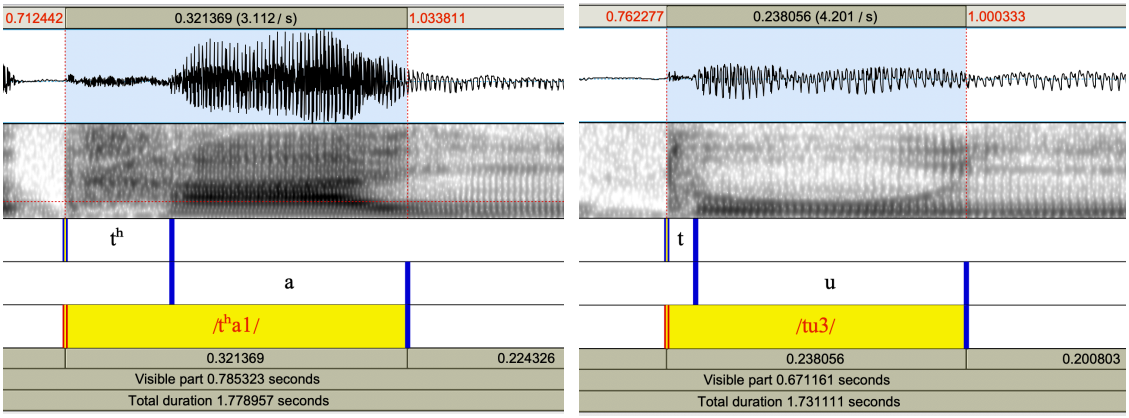


Figure 3. VOT, vowel, and word segmentation for Mandarin target words

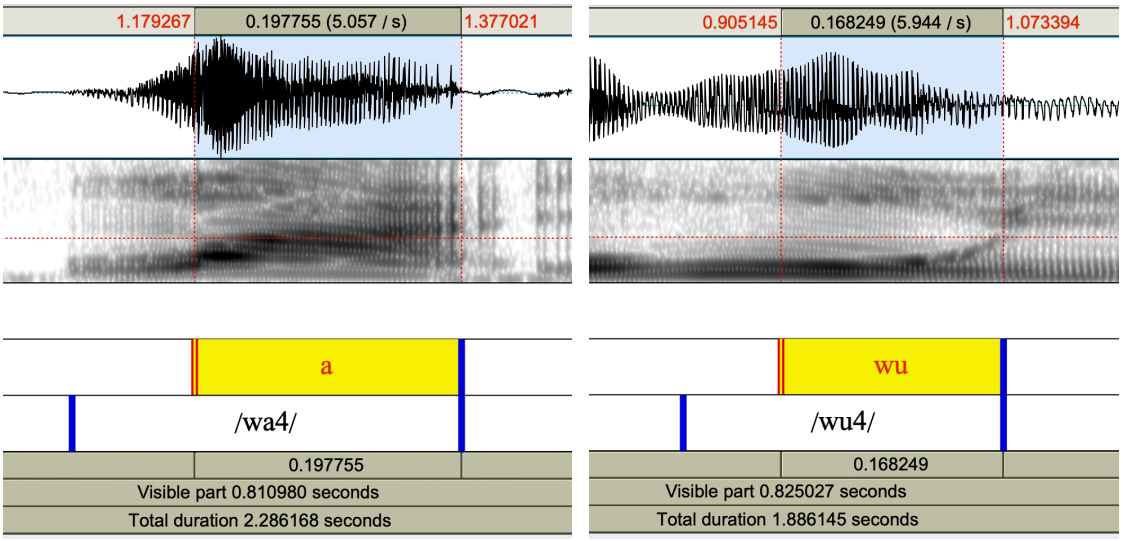


Figure 4. vowel and word segmentation for Mandarin control words

VOT durations, vowel durations, and F0 values were extracted using Praat scripts. F0 values were extracted in two different ways to account for the relation between the vowel duration and the duration of the perturbation effect. Two different methods were used because if the vowel duration and the perturbation duration are positively correlated, longer vowels were expected to show a longer effect of F0 perturbation and vice versa. The duration of Mandarin tones varies intrinsically (Howie, 1976; Whalen & Xu, 1992; Fu & Zeng, 2000; Yang, Zhang, Li & Xu, 2017). Among the four lexical tones, studies (Ho, 1976; Whalen & Xu, 1992) have suggested T3 has the longest duration, T4 has the shortest and T1 and T2 fall in between in isolated monosyllable conditions. The duration pattern is less consistent in connected speech than in isolated words. Deng, Feng and Lu

(2006) report that the duration of the four lexical tones in sentence medial position is: T2>T1>T3>T4. The vowel duration results in the current study are consistent with Deng et al.'s (2006) finding.

Table 6 presents the detailed mean vowel duration in the four tonal environments. If vowel duration and the duration of the perturbation effect were positively correlated, the duration of F0 perturbation was expected to follow the vowel duration pattern. On the other hand, if there was no correlation between the two durations, the perturbation effect was expected to be limited to a certain period of time regardless of the vowel duration. Luo (2018) reports that the perturbation duration is 35 ms for T1, 5 ms for T3, and 30 ms for T4. No significant F0 difference is observed in T2 between the F0-aspirated stop and the F0-unaspirated stop in her study. In the present study, F0 was measured in two different ways. First, F0 values were extracted from twenty equidistant points of the post-stop vowel (the normalized method). The selection of the twenty equidistant points was based on the vowel duration of each word. Therefore, the extraction of the F0 values was normalized according to the vowel duration. Figure 5 shows the normalized F0 values through the entire vowel of different consonant groups and vowel environments in each Mandarin tone. Second, F0 values were extracted every 8 ms within the first 64 ms along the F0 trajectory of the vowels (absolute method).

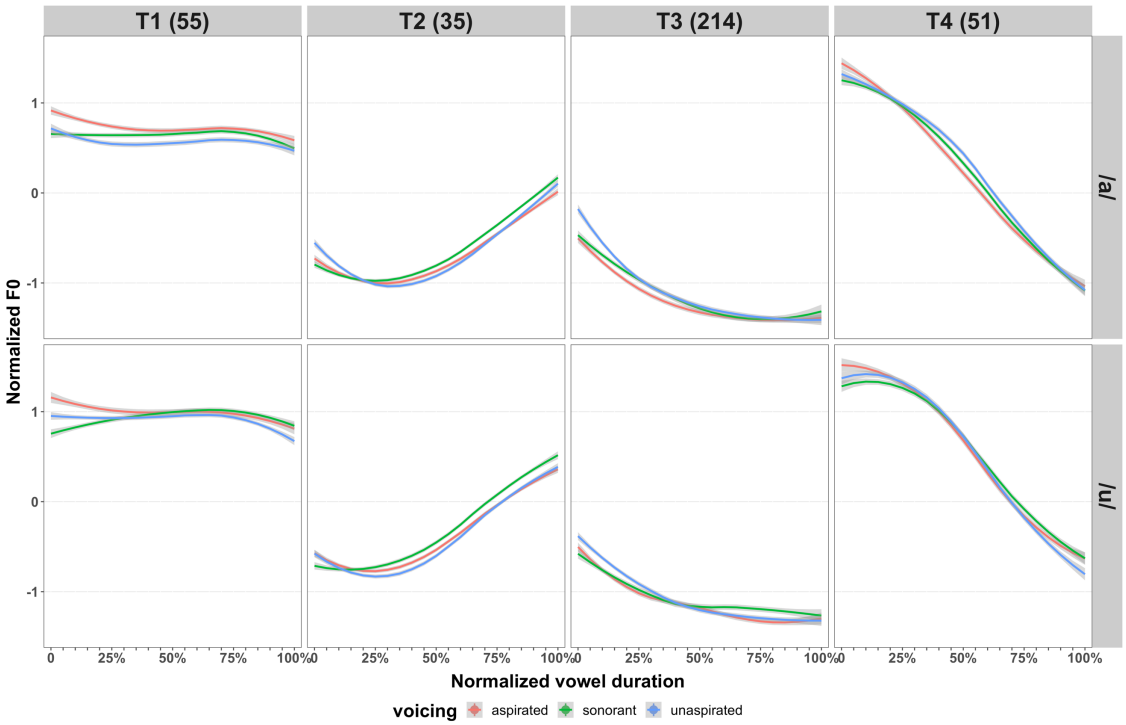


Figure 5. Normalized F0 contours by voicing, vowel and tonal environments

3.2.3 Data preparation

Data were excluded in data preparation due to low quality (0.6%), failure to extract F0 values by Praat (5%), and outliers above 2.5 standard deviations after by participant F0 normalization (0.3%).

23 of 3600 (144 experimental words * 25 participants) recorded words (0.6%) were excluded from all analyses. The excluded words included mispronunciations of the intended syllable or the tone (14), hesitation (1), self-correction (2), and excessive background noise (6).

3.2.3.1 Normalized method - F0 values extracted proportionally

The first 8 of the 20 time points were included for further statistical analysis to model the F0 perturbation effect. These 8 points took about 35% of the vowel, which was about 75ms. This decision was made based on the duration of the vowels in the current data (see Table 6). The selected duration was longer than the maximum perturbation duration (35ms) reported by Luo (2018) and shorter than the commonly agreed perturbation duration (50% of the vowel) reported in non-tonal languages (Hombert et al. 1979).

Table 6. The duration of the entire vowels and the examined portions of each tonal environment

Tone	**Duration (ms)**	**s.d.**	**Duration (35%)**
T1	215.4	40.5	75.4
T2	227.2	47.6	79.5
T3	214.8	63.6[9]	75.2
T4	201.4	40.5	70.5

[9] T3 words had a relatively larger standard deviation than the words with other tones. The target word was embedded in a carrier sentence, and the participants differed how they read the sentence. Some of them paused briefly after the target word and some of them did not. The pausing behavior had a stronger influence of the T3 words than words with other tones in terms of the vowel duration. For those who briefly paused, they tended to produce the entire contour of the T3 while for those who connected the target word and the following word, they tended to produce only half of the third tone.

1653 of 28616 data points (5%) where Praat failed to extract F0 value were excluded from statistical analysis. Praat tends to give an undefined value when it is requested to get a pitch value in a voiceless part of a sound (Boerma & Weenink, 2020). A large portion of these excluded data was due to creakiness. Among the four Mandarin lexical tones, T3 is usually realized with a creaky voice (Hockett, 1947; Chao, 1956; Kuang, 2013). In the present study, creakiness was not limited to T3, as it presented in T2 and T4 as well. This conformed to Kuang's (2013) observation that creakiness was related to low F0 values rather than a specific tonal category. Whenever the speakers reached the lowest pitch, they tended to creak. However, the occurrences of creak were more frequent in T3. It is a salient cue of T3 rather than the tones (Kuang, 2013). Table 7 presents the number of data points excluded due to the failure to extract F0 values.

Table 7. Number of Mandarin data points excluded due to F0 extraction failure

tone	**N (Normalized method)**	**N (Absolute method)**
T1	28	30
T2	322	367
T3	1254	1284
T4	49	69
Total	1653	1750

Raw F0 values were then transformed into z-scores for each subject to facilitate comparison of pitch across subjects. Outliers, 93 of 26963 data points (0.3%) above 2.5 standard deviations from each speakers' mean F0 were excluded from the statistical analyses of the perturbation effect. In the end, 26870 data points were retained for analysis.

3.2.3.2 Absolute method - F0 values extracted by absolute values

F0 values were also extracted every 8 ms within the first 64 ms along the F0 trajectory of the vowels. 1750 of 32193 (5%) data points were excluded due to extraction failure by Praat script (see Table 7 for a breakdown of the excluded data points in each tonal category). 93 of 30443 outliers (0.3%) were excluded when F0 was transformed to z-score. In the end, 30350 data points were retained for analysis.

Separate statistical analyses were performed on the two datasets obtained by using the two different extracting methods, and the results were consistent. It appeared that the perturbation duration did not correlate with the vowel duration. As 35% of the vowel is longer than the duration examined in the absolute method, I only report here the statistical analysis done on the data extracted in the normalized method.

3.2.4 Statistical analyses and results

Linear mixed-effects models were performed with the *lme4* package in R (Bates, Maechler, Bolker & Walker, 2014) to investigate the influence of aspiration, tone, vowel height, time points and gender on the normalized F0. In the full model, the dependent variable was the normalized F0 in z-scores. Aspiration (the aspirated stop vs. the unaspirated stop vs. the sonorant), tone (T1 vs. T2 vs. T3 vs. T4), vowel height (high vs. low), time points (8 time points), and gender (female vs. male) and the interaction among

the five variables were included as fixed effects. The random effects structure of the model was determined using a forward best path algorithm (Barr, Levy, Sheepers & Tily, 2013). Subjects were included as a random effect. All fixed factors were coded using treatment (dummy) coding, with the reference level for the intercept being set to the aspirated stop, T1, low vowel, 0 time point and female. The best fitting model was selected by comparing models using the likelihood ratio test.

The fixed effect gender did not fit the dataset significantly better than the full model (χ^2=0.23, p=0.63), indicating the contribution of gender to the model fit was not significant. In order to reduce the complexity of the model, gender was excluded as a fixed effect from the full model. The best model had aspiration, tone, vowel height, time points and the interaction among the four variables as the fixed effects and subjects as the random effect. Model results are given in Table 8. Figure 6 visually demonstrates the effects of the four predictors for the normalized F0.

Table 8. The output of linear mixed effects model of normalized F0: the reference level for the intercept being set to the aspirated stop, T1, low vowel, and 0 time point

Effect	**df**	**Chisq**	**p.value**
tone	3	65385.34	<.0001 ***
time_points	7	4074.09	<.0001 ***
aspiration	2	358.19	<.0001 ***
vowel	1	3206.6	<.0001 ***
tone: time_points	21	2151.7	<.0001 ***
tone: aspiration	6	830.93	<.0001 ***
time_points: aspiration	14	360.93	<.0001 ***
tone:vowel	3	1369.4	<.0001 ***
time_points:vowel	7	178.84	<.0001 ***
aspiration:vowel	2	88.49	<.0001 ***
tone: time_points: aspiration	42	191.42	<.0001 ***
tone: time_points:vowel	21	14.49	0.85
tone: aspiration:vowel	6	92	<.0001 ***
time_points: aspiration:vowel	14	52.82	<.0001 ***
tone:X:voicing:vowel	42	9.83	>.99

*Significance codes: 0.05. **Significance codes: 0.01. ***Significance codes: 0.001

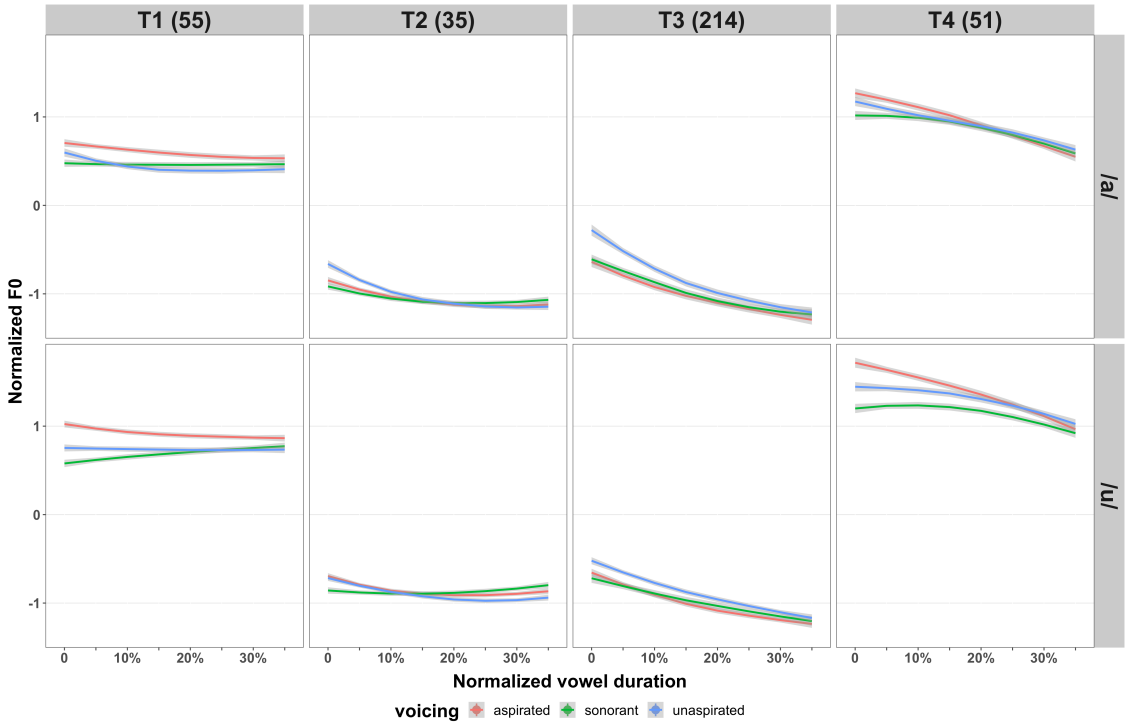


Figure 6. Normalized F0 of Mandarin words within the first 35% of the vowel

The statistical model revealed a significant interaction between tone and aspiration (χ^2=830.93, p<0.0001). The F0-sonorants tended to be lower than the F0-stops (F0 following the stops) across the four tonal environments. Moreover, in general, the F0-aspirated stops was higher than the F0-unaspirated stops in T1 and T4, while the F0-aspirated stops was lower than F0-unaspirated stops in T2 and T3. There was a significant interaction between time point, vowel and voicing (χ^2=52.82, p<0.001). The F0-aspirated stops and the F0-unaspirated stops were not significantly different from each other in T2 when the vowel was /u/, while the F0-aspirated stops was significantly lower than the F0-unaspirated stops in T2 when vowel was /a/ within the first 5% of the vowel. There was no significant difference between the F0-aspirated stops and the F0-sonorants

in T2 when the vowel was /a/, but the F0-aspirated stops was significantly higher than the F0-sonorants when the vowel was /u/ within the onset of the vowel.

Tukey's HSD tests were conducted using the emmeans package (Lenth, 2020) for R and the summary of the post-hoc pairwise comparisons are reported in Tables 9-14.

Table 9. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model in low vowel /a/ environment, F0-aspirated stops-F0-sonorants

Tone		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**	**time 6**	**time 7**
		0%	5%	10%	15%	20%	25%	30%	35%
T1	*ß*	0.228	0.203	0.168	0.139	0.112	0.090	0.071	0.068
	p	<.0001 ***	<.0001 ***	<.0001 ***	.0001 ***	.0018 **	.0170 *	.0799	.0976
T2	*ß*	0.068	0.039	0.015	0.005	-0.019	-0.038	-0.043	-0.053
	p	.1471	.5067	.9000	.9893	.8387	.5071	.4145	.2654
T3	*ß*	-0.039	-0.071	-0.059	-0.053	-0.016	-0.043	-0.042	-0.063
	p	.6005	.1576	.2709	.3307	.9076	.4910	.5278	.2497
T4	*ß*	0.248	0.194	0.111	0.072	0.036	-0.023	-0.024	-0.040
	p	<.0001 ***	<.0001 ***	.0025 **	.0764	.5327	.7731	.7604	.4592

Table 10. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model in low vowel /a/ environment, F0-sonorants-F0-unaspirated stops

Tone		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**	**time 6**	**time 7**
		0%	5%	10%	15%	20%	25%	30%	35%
T1	*ß*	-0.121	-0.039	0.025	0.056	0.067	0.062	0.074	0.054
	p	.0013 **	.4751	.7328	.2102	.1077	.1478	.0671	.2312
T2	*ß*	-0.260	-0.147	-0.067	-0.029	0.006	0.038	0.064	0.077
	p	<.0001 ***	.0001 ***	.1257	.6751	.9820	.4984	.1420	.0624
T3	*ß*	-0.316	-0.237	-0.140	-0.109	-0.093	-0.060	-0.053	-0.016
	p	<.0001 ***	<.0001 ***	.0006 ***	.0091 **	.0327 *	.2476	.3621	.9173
T4	*ß*	-0.154	-0.088	-0.020	-0.010	-0.017	-0.027	-0.034	-0.042
	p	<.0001 ***	.0226 *	.8270	.9492	.8714	.6983	.5640	.4297

Table 11. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model in low vowel /a/ environment, F0-aspirated stops-F0-unaspirated stops

Tone		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**	**time 6**	**time 7**
		0%	5%	10%	15%	20%	25%	30%	35%
T1	*ß*	0.107	0.165	0.193	0.194	0.179	0.152	0.144	0.122
	p	.0062 **	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0007 ***
T2	*ß*	-0.192	-0.108	-0.052	-0.024	-0.013	0.0002	0.021	0.024
	p	<.0001 ***	.0083 **	.2954	.7671	.9229	1.0000	.8159	.7714
T3	*ß*	-0.355	-0.308	-0.198	-0.162	-0.109	-0.103	-0.095	-0.079
	p	<.0001 ***	<.0001 ***	<.0001 ***	.0001 ***	.0122 *	.0191 *	.0402 *	.1185
T4	*ß*	0.094	0.107	0.091	0.062	0.019	-0.050	-0.058	-0.082
	p	.0147 *	.0036 **	.0161 *	.1469	.8340	.2948	.1963	.0415 *

Table 12. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model in high vowel /u/ environment, F0-aspirated stops-F0-sonorants

Tone		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**	**time 6**	**time 7**
		0%	5%	10%	15%	20%	25%	30%	35%
T1	*ß*	0.447	0.350	0.286	0.223	0.188	0.149	0.117	0.092
	p	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0011 **	.0137 *
T2	*ß*	0.167	0.077	0.032	-0.002	-0.027	-0.047	-0.060	-0.070
	p	<.0001 ***	.0508	.5874	.9973	.6858	.3339	.1668	.0854
T3	*ß*	0.071	0.005	-0.003	-0.041	-0.058	-0.039	-0.041	-0.040
	p	.0994	.9855	.9967	.4323	.2010	.5010	.4690	.5108
T4	*ß*	0.519	0.406	0.312	0.246	0.183	0.140	0.103	0.042
	p	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0001 ***	.0052 **	.4126

Table 13. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model in high vowel /u/ environment, F0-sonorants-F0-unaspirated stops

Tone		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**	**time 6**	**time 7**
		0%	5%	10%	15%	20%	25%	30%	35%
T1	*ß*	-0.176	-0.126	-0.090	-0.054	-0.023	0.001	0.022	0.038
	p	<.0001 ***	.0004 ***	.0166 *	.2315	.7556	.9989	.7748	.4854
T2	*ß*	-0.140	-0.073	-0.018	0.030	0.072	0.115	0.126	0.143
	p	.0001 ***	.0691	.8554	.6274	.0719	.0015 **	.0004 ***	<.0001 ***
T3	*ß*	-0.201	-0.151	-0.127	-0.088	-0.074	-0.063	-0.045	-0.036
	p	<.0001 ***	<.0001 ***	<.0004 ***	.0240 *	.0751	.1621	.4010	.6140
T4	*ß*	-0.250	-0.187	-0.175	-0.158	-0.127	-0.129	-0.128	-0.102
	p	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0003 ***	.0003 ***	.0003 ***	.0056 **

Table 14. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model in high vowel /u/ environment, F0-aspirated stops-F0-unaspirated stops

Tone		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**	**time 6**	**time 7**
		0%	5%	10%	15%	20%	25%	30%	35%
T1	*ß*	0.270	0.224	0.196	0.169	0.164	0.150	0.139	0.130
	p	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0001 ***	.0002 ***
T2	*ß*	0.026	0.004	0.014	0.028	0.045	0.067	0.066	0.073
	p	.7153	.9915	.8943	.6737	.3602	.0988	.1099	.0683
T3	*ß*	-0.130	-0.145	-0.129	-0.129	-0.133	-0.102	-0.086	-0.074
	p	.0004 ***	<.0001 ***	.0003 ***	.0003 ***	.0003 ***	.0093 **	.0374 *	.1011
T4	*ß*	0.269	0.219	0.138	0.088	0.055	0.010	-0.026	-0.060
	p	<.0001 ***	<.0001 ***	.0001 ***	.0218 *	.2154	.9493	.7172	.1634

The general patterns of the production results are summarized in Tables 15-17. Table 15 summarizes the relationship between F0-aspirated stops and F0-sonorants, and Table 16 summarizes the relationship between F0-unaspirated stops and F0-sonorants in terms of direction and duration. Table 17 summarizes the F0 perturbation direction and duration patterns.

Table 15. Pairwise comparisons between the F0-aspirated stops and the F0-sonorants (the duration in parenthesis was calculated based on the mean vowel duration in each tonal environment)

	/a/		/u/	
	Direction	**Duration**	**Direction**	**Duration**
T1	Asp>Son	25% (53.9 ms)	Asp>Son	35% (75.4 ms)
T2	No significant difference		Asp>Son	0% (onset)
T3	No significant difference		No significant difference	
T4	Asp>Son	10% (20.1 ms)	Asp>Son	30% (60.4 ms)

Table 16. Pairwise comparisons between the F0-unaspirated stops and the F0-sonorants (the duration in parenthesis was calculated based on the mean vowel duration in each tonal environment)

	/a/		/u/	
	Direction	**Duration**	**Direction**	**Duration**
T1	Unasp>Son	0% (onset)	Unasp>Son	10% (21.5 ms)
T2	Unasp>Son	5% (11.4 ms)	Unasp>Son	0% (onset)
T3	Unasp>Son	20% (43.0 ms)	Unasp>Son	15% (32.2 ms)
T4	Unasp>Son	5% (10.1 ms)	Unasp>Son	35% (70.5 ms)

Table 17. Pairwise comparisons between the F0-aspirated stops and the F0-unaspirated stops (the duration in parenthesis was calculated based on the mean vowel duration in each tonal environment)

	/a/		/u/	
	Direction	**Duration**	**Direction**	**Duration**
T1	Asp>Unasp	35% (75.4 ms)	Asp>Unasp	35% (75.4 ms)
T2	Asp<Unasp	5% (11.4 ms)	No significant difference	
T3	Asp<Unasp	30% (64.4 ms)	Asp<Unasp	30% (64.4 ms)
T4	Asp>Unasp	10% (20.1 ms)	Asp>Unasp	15% (30.2 ms)

The F0-unaspirated stops was consistently higher than the F0-sonorants in the four tonal environments with both vowels. The F0-aspirated stops was higher than the F0-sonorants in T1 and T4 for both vowel environments. The F0-aspirated stop was higher than the F0-sonorant in T2 with /u/ and no significant difference was found with /a/. In T3, the F0-aspirated stops was not significantly different from the F0-sonorants (see Table 15 and Table 16).

The results in Table 11, Table 14 and Table 17 suggest there was an F0 perturbation effect Mandarin. The direction of F0 perturbation was different for the high initial tones (T1, T4) and the low initial tones (T2, T3). The F0-aspirated stops was significantly higher than the F0-unaspirated stops in T1 and T4 but was significantly lower than the F0-unaspirated stops in T2 and T3. The duration of F0 perturbation was meditated by tonal and vowel environments. The longest perturbation duration was observed in T1. The F0-aspirated stops was significantly higher than the F0-unaspirated

stops throughout the selected 35% of the vowel. The duration was about 75 ms. The shortest perturbation duration was observed in T2 with the vowel /a/ and it was limited to the first 5% of the vowel. The duration was about 11 ms. No significant F0 difference was observed in T2 with the high vowel /u/ (see Table 17).

3.2.5 Interim summary

F0 perturbation direction. The direction of F0 perturbation was different for the high initial tones (T1, T4) and the low initial tones (T2, T3). The F0-aspirated stops was significantly higher than the F0-unaspirated stops in T1 and T4 but was significantly lower than the F0-unaspirated stops in T2 and T3.

F0 perturbation duration. These results confirmed the previous observation (Hombert, 1977; Gandour, 1974; Francis et al., 2006) that the F0 perturbation does not extend far into the vowel in tonal languages, ranging from 10 ms to 50 ms. The duration of F0 perturbation was meditated by tonal and vowel environments. With the low vowel /a/, the perturbation durations in T1, T2, T3 and T4 were 35% (75.4 ms), 5% (11.4 ms), 30% (64.4 ms) and 10% (20.1 ms), respectively. With the high vowel /u/, the perturbation durations in T1, T3 and T4 were 35% (75.4 ms), 30% (64.4 ms) and 15% (30.2 ms), respectively. No significant F0 difference was observed in T2 with high vowel /u/. The longest perturbation duration was about 75ms, and the shortest was about 11ms.

F0-stops vs. F0-sonorants. The F0 differences between the F0-aspirated stops and the F0-sonorants were influenced by tonal and vowel environments. The F0-aspirated stops was higher than the F0-sonorants in the high initial tone group for both vowel environments. For the low initial tone group, the F0-aspirated stops was found to be not

significantly different from the F0-sonorant except for the 0 time point in T2 with vowel /u/. By contrast, the F0-unaspirated stops was consistently higher than the F0-sonorants in the four tonal environments with both vowels.

3.2.6 Discussion of the L1 production experiment

The fact that the onset F0 of the tone and the vowel environments affected the F0 perturbation direction patterns suggests that F0 perturbation was more of an automatic effect due to the physiology of the human phonation mechanism rather than an intentional control to enhance the stop aspiration contrast. The direction of the F0 perturbation patterns observed in the present study was possibly influenced by the height of the larynx, the tension of the vocal cords, and the change of the subglottal pressure (Ps). Larynx height is positively related to the post-stop F0 (Moisik et al., 2014; Sagart et al., 1986; Hallé 1994). Overall, the larynx is higher in T1 and T4 than T2 and T3. As to the tension of the vocal cords, in general stiff vocal cords lead to high F0 and slack vocal cords lead to low F0 (Halle & Stevens, 1971; Löfqvist, Baer, McGarr & Story, 1989). The respiratory system usually generates a constant Ps during stop closures (Löfqvist, 1975; Ohala & Ohala, 1972; Slis, 1970; Ladefoged, 1967). The remaining Ps after the release of the stop can also determine the F0 of the post-stop vowel, and the remaining Ps are positively related to the onset F0 (Ladefoged, 1971; Shi, 1998; Xu & Xu, 2003). These three factors together play an important role in contrasting aspiration with F0 in Mandarin. To sum up, high larynx height, tense vocal cords and high Ps accompany high F0 at vowel onset while low larynx height, slack vocal cords and low Ps correlate with low F0 at vowel onset.

The production of high initial tones. During the process of producing high initial tones, the larynx is raised to initiate high F0 at the vowel onset (Moisik et al., 2014). Raising the larynx reduces the size of the oral cavity. When F0 is in the high range, the larynx is high, so that the horizontal movement of the hyoid bone facilitate the tilt of the thyroid cartilage, resulting in the stretching of the vocal folds (Honda, Hirai, Masaki & Shimada, 1999). When the vocal folds are stretched, they are tense, which inhibits air from flowing out of the subglottal area. The vocal fords are tenser at the onset of voicing for aspirated stops than for unaspirated stops (Shi, 1998). With tenser vocal cords, at the release of an aspirated stop, the Ps decreases minimally, which results in higher F0 after an aspirated stop than an unaspirated stop in T1 and T4.

The production of low initial tones. When producing low initial tones, the larynx is lowered to initiate low F0 at the vowel onset (Moisik et al., 2014). In the low F0 range, the jaw and the hyoid bone move downward together with the larynx so that the cricoid cartilage rotates along the cervical spine, which leads to the shortening and relaxation of the vocal cords (Honda et al., 1999). Both lowered larynx and relaxed vocal folds accompany low F0. At the oral release phase, the glottis is almost closed for unaspirated stops while it is widely open for aspirated stops (Löfqvist, 1975; Ohala & Ohala, 1972; Slis, 1970; Ladefoged, 1967). Therefore, upon the release of the stops, a higher volume of air is released from the subglottal area following aspirated stops than is released following unaspirated stops. With a long voice onset time, the aspirated stops exhibit a tremendous decrease of the Ps. Moreover, the slackened vocal folds facilitate the rapid airflow at the release of an aspirated stop, which speeds up the drop of Ps.

Consequently, by the time of vowel (voicing) onset, the remaining Ps is lower following an aspirated stop than following an unaspirated stop (Ohala, 1978; Ladefoged, 1963; Ohala & Ohala, 1972; Ladefoged, 1975). Lower Ps leads to lower F0 at the vowel onset for aspirated stops than for unaspirated stops in the low F0 range (Xu & Xu, 2003).

The high initial tone and low initial tone effect observed in the current study are in line with the Maddieson's (1996) proposal that F0 was controlled by two mechanisms. That is, F0 lowering results from the lowering of the larynx whereas F0 raising is influenced by the contraction of the intrinsic cricothyroid muscles.

The contour of the tone could also play a role in the duration of the F0 perturbation effect. According to Xu (1997), the tonal contours of the four lexical tone in Mandarin are as follows: T1 begins with a high F0 and maintains the same level though the entire vowel, T2 starts with a low F0 and then falls slightly until 20% into the vowel before rising throughout the rest of the vowel, T3 begins with a low F0, falls to the lowest F0 at the midpoint of the vowel, then rises sharply to the end of the syllable, and T4 starts with the highest F0 and then drops sharply from the 20% of the vowel until the end of the syllable. The tonal contours reported by Xu (1997) is in general consistent with the patterns found in this study, which are presented in Figure 2[10].

In the current findings, the duration of F0 perturbation in T2 was shorter than that in the three other tones. With the vowel /a/, the perturbation effect was observed within the first 5% of the vowel (about 11 ms). No significant differences between the F0-

[10] Like Luo (2018), T3 in our study did not retain the full contour due to the fact that the target word was embedded in a carrier sentence. The context triggered the "half-third" sandhi condition where the rising part was not produced by the participants (see Figure 2).

aspirated stops and the F0-unaspirated stops were found in T2 with the vowel /u/. This was consistent with Luo's (2018) finding that there is no significant perturbation effect in T2. The minimal or lack of perturbation effect in T2 may result from the tonal contour. It could be physiologically costly to coordinate all the articulators at the low F0 range to maintain the aspiration contrast and get ready for continuous pitch rising for the rest of the vowel within such a brief period of time. T1 is a level tone, and the change of T3 is gradual. Therefore, T1 and T3 have relatively longer duration of perturbation effect than T2 and T4 which has the dramatic change of the contour occurring near the onset of the vowel.

Taken together, the results of the current study suggested that the height of the larynx, the tension of the vocal cords and the change of subglottal pressure were the physiological factors influencing the direction of F0 perturbation. The tonal contours affected the duration of the F0 perturbation in Mandarin. These findings indicated that the F0 perturbation pattern in tonal languages was largely an automatic effect due to the physiology of the human phonation mechanism rather than a controlled process to enhance the phonological status of the consonants (aspirated vs. unaspirated). The F0 perturbation patterns in Mandarin were highly dependent on specific tonal and vowel contexts. If this were due to intentional enhancement, it would be hard to explain why different vowel environments and tonal environments exhibited different perturbation patterns either by direction or by duration.

3.3 Experiment 2: native Mandarin speakers' L1 perception

Experiment 2 examined the influence of VOT and F0 on Mandarin speakers' perception of the stop aspiration contrast in Mandarin.

3.3.1 Method

3.3.1.1 Stimuli

Mandarin perception stimuli were created from natural productions of the syllable /t^hu/ carrying four tones (i.e. /t^hu1/, /t^hu2/, /t^hu3/, /t^hu4/). A female native Mandarin speaker recorded the base tokens in isolation. Aspirated stops were selected as the baseline stimuli. That is, unaspirated tokens were created by removing the aspirated portions from the naturally produced aspirated stops. This is because it is more likely to get a natural sounding token by reducing the aspiration noise and shortening the VOT than by adding in aspiration noise and lengthening the VOT (Francis et al., 2006). The pitch information was not manipulated while removing the aspirated portion. The high back vowel /u/ was selected because /u/ provides a full set (all four tones) of real Mandarin words for both aspirated and unaspirated alveolar stops. Each of the four base tokens were then manipulated to create 49 syllables covarying in the VOT of the initial stops and the post-stop F0, by fully crossing 7 steps of VOT and 7 steps of post-stop F0 (see Table 18 for the details). Seven steps of VOT and post-stop F0 were selected in order to obtain a detailed picture of how VOT and F0 would influence listeners' perception of Mandarin stops (Schertz et al., 2015).

Table 18. VOT and onset F0 values for each acoustic dimension of the Mandarin stimuli

Tone	**Parameter**	**Base token**	**Step 1**	**Step 2**	**Step 3**	**Step 4**	**Step 5**	**Step 6**	**Step 7**
Tone 1	VOT (ms)	98.3	14.1	28.8	41.6	56.2	73.1	85.5	98.3
	F0 (Hz)	322.8	262.8	282.8	302.8	322.8	342.8	362.8	382.8
Tone 2	VOT (ms)	94.9	13.8	26.9	40.1	54.0	67.5	80.8	94.9
	F0 (Hz)	241.4	181.4	201.4	221.4	241.4	261.4	281.4	301.4
Tone 3	VOT (ms)	102.6	14.2	28.1	42.2	57.3	73.0	90.9	102.6
	F0 (Hz)	209.6	149.6	169.6	189.6	209.6	229.6	249.6	269.6
Tone 4	VOT (ms)	101.2	14.9	29.2	43.2	57.4	72.6	86.2	101.2
	F0 (Hz)	371.3	311.3	331.3	351.3	371.3	391.3	421.3	431.3

VOT manipulation: The mean VOT duration of the 4 base tokens was 99 ms, and the VOT step size was 14 ms. Starting at the nearest zero crossing point from the end of the stop burst, about 14 ms burst duration was manually removed step by step in Praat (Boersma & Weenink, 2020) until the VOT of the base token was around 14 ms. The *move curser to…* function in Praat was used to locate the position for excision.

F0 manipulation: F0 was manipulated using the Time-Domain Pitch-Synchronous-Overlap-and-Add-algorithm (TD-PSOLA, Moulines & Charpentier, 1990) as implemented in Praat. First, the first 35% of the vowel was located, and then the pitch of the selected vowel was stylized with frequency resolution as 2.0 Hz. The purpose of the *stylize* function is to have a simplified pitch curve (Boersma & Weenink, 2020). Then

all the pitch points between the initial pitch point of the vowel and the point where marked the 35% of the vowel were removed. To create the 7 steps of post-stop F0, the initial pitch point was either raised or lowered by using the *shift pitch frequencies* function in Praat. The step size of the pitch manipulation was 20 Hz. The maximum raising was 60 Hz and the maximum lowering was also 60 Hz. All the tokens were re-synthesized with TD-PSOLA after the pitch manipulation.

Four L1 Mandarin listeners were asked to test the naturalness of the synthesized tokens, and they were judged by them as good tokens of the original tones. Another four L1 Mandarin listeners were invited to pilot the experiment. The pilot participants did not respond differently to VOT step 6 (84 ms) and VOT step 7 (natural VOT) stimuli. The VOT step 6 (84 ms) stimuli were removed from the experiment due to the time constraint. After excluding this, the stimuli set of the experiment included 168 (4 tones * 7 steps of F0 * 6 steps of VOT) unique tokens.

3.3.1.2 Participants

The same group of participants completed the perception experiment after they finished the production experiment. There was a 5[11]-minute break between the two experiments.

3.3.1.3 Procedure

Listeners participated in a forced-choice identification task presented in PsychoPy (Peirce, 2007). Two Chinese characters constituting the aspirated and unaspirated pairs (e.g., 突 /tʰū/ vs. 督 /tū/) were displayed on a laptop screen while they were hearing the stimulus through a Sennheiser headset. They were instructed to choose the word they

[11] The participants can request for a longer break if they need.

heard by selecting one of the two characters using a Cedrus button box (model RB-740). Stimuli for different lexical tones were presented in separate blocks and the order among the blocks was counter-balanced across participants. There were breaks between the blocks and the next block started automatically when the participants hit continue.

Within each block, each of the 42 tokens was repeated three times in different random orders. 13 participants saw the screen with /t^h/- syllables on the left and /t/-syllables on the right, and 12 participants saw the opposite. All the participants reported to be right-handed. The task took about 20 minutes. 12600 responses (25 participants * 4 blocks * 42 tokens * 3 repetitions) were collected from the experiment. The reaction time (RT) of each response was also collected with the responses. The timer for the reaction time started from the onset of the audio syllable and stopped when the participants hit the button on the response box to make their selection.

3.3.2 Statistical analyses and results

RT was normalized by participant and responses with RT above 3 standard deviations (232 out of 12600 responses, 1.8%) were excluded from the statistical analysis. Perception responses were statistically analyzed using the logistic regression model with the *lme4* packages in R (Bates et al., 2014) to determine the influence of each acoustic cue on the identification of the prevocalic stops. In the full model, the dependent variable was the participant's response (aspirated vs. unaspirated stops). VOT step, F0 step, tone, and the interactions among the 3 variables were included as fixed effects. Tone was orthogonally contrast coded (T1, T4 vs. T2, T3; T1 vs. T4; T2 vs. T3) to examine whether there are significant response differences between the high initial tones (T1, T4)

and the low initial tones (T2, T3), as well as within the two tonal groups. T1 and T4 start from a high pitch range, while T2 and T3 start from a low pitch range. Table 19 presents the onset F0 of the four tones collected in the production experiment. Participants and words were included as random effects. VOT step, F0 step, and tone were added as random slope to the fixed effect participants.

Table 19. Onset F0 of Mandarin tones from the L1 production experiment

	Female		**Male**	
tone	**F0 (Hz)**	**s.d.**	**F0 (Hz)**	**s.d.**
T1	294	36	166	31
T2	221	26	123	22
T3	234	33	130	24
T4	325	40	184	36

The effects of the independent variables were investigated by comparing models using the likelihood ratio test. The full model was described above. VOT step significantly contributed to model fit (β= -1.813, χ^2= 43.516, $p < 0.0001$), showing that as VOT increased, the possibility of unaspirated responses decreased. Pitch step significantly contributed to model fit (β= -0.167, χ^2= 9.0093, $p < 0.001$), showing that as pitch increased, the possibility of unaspirated responses decreased. The first tonal contrast, which compared high initial tone with low initial tone, also contributed

significantly to model fit (β= -3.787, χ^2= 30.711, p<0.0001), showing that high initial tones elicited significantly less unaspirated responses than low initial tones. The second (β= -3.787, p=0.82) and third (β= 0.206, p=0.36) tonal contrasts were not significant predictors of the unaspirated responses, indicating that T1 stimuli did not elicit significantly more unaspirated responses than T4 stimuli and T2 stimuli did not elicit significantly more unaspirated responses than T3 stimuli. None of the interactions among the fixed effects were significant. The model summary of the best fitting model is given in Table 20. Figure 7 demonstrates the influences of VOT, pitch, and tone on consonant aspiration identification.

Table 20. β-coefficients, standard error and z- and p-values for the logistic regression model

	Estimate	**Std. Error**	**Z value**	**Pr (> | z |)**
(Intercept)	2.973	0.495	6.001	1.96e-09 ***
VOT step	-1.813	0.177	-10.266	< 2e-16 ***
Pitch step	-0.167	0.053	-3.156	0.0016 **
T1,T4 vs. T2,T3	-3.787	0.583	-6.499	8.07e-11 ***
T1 vs. T4	-0.088	0.398	-0.221	0.8249
T2 vs. T3	0.267	0.295	0.906	0.3650

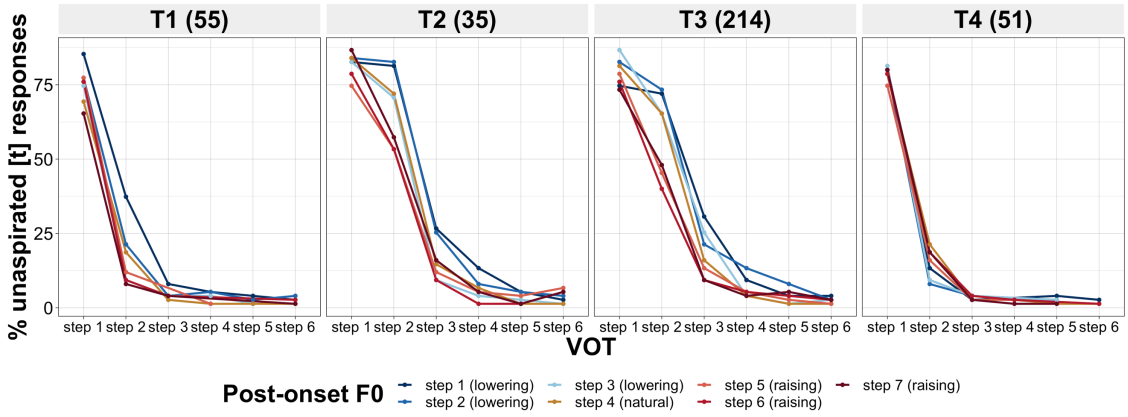


Figure 7. Percentage of unaspirated [t] responses by native Mandarin speakers

As indicated by Figure 7, VOT step 1 (14 ms) stimuli elicited the highest percentage of unaspirated responses across the four tones. In addition, stimuli with lower post-onset F0 elicited more unaspirated responses than stimuli with higher post-onset F0 across all four tones, although this pattern was not very obvious for T4 stimuli at VOT step 2. T2 and T3 stimuli elicited significantly more unaspirated responses than T1 and T4 stimuli. The percentage of the unaspirated responses dropped sharply at VOT step 2 (28 ms) for T1 and T4 stimuli, and the sharp drop occurred at VOT step 3 (42 ms) for T2 and T3 stimuli. VOT step 4, VOT step 5 and VOT step 6 elicited a low percentage of unaspirated responses across the four tones.

3.3.3 Interim summary

VOT played an important role in perceiving aspiration contrasts in Mandarin by native speakers, as predicted. As VOT became longer, the percentage of unaspirated responses decreased. The number of the unaspirated responses dropped sharply at VOT step 2 (28 ms) for the high initial tone stimuli, and the sharp drop occurred at VOT step 3

(42 ms) for the low initial tone stimuli. VOT step 1 (14 ms) elicited the highest percentage of unaspirated responses. Starting from VOT step 4 (56 ms), the number of unaspirated responses across the four tones remained low.

Pitch influenced native listeners' perception of the aspiration contrast in Mandarin. In general, as pitch became higher, the percentage of unaspirated responses decreased across the four tones. The onset pitch of the lexical tones also influenced the identification of stops in Mandarin. The low initial tone stimuli elicited significantly more unaspirated responses than the high initial tone stimuli did.

3.3.4 Discussion of the L1 perception experiment

The current findings have provided evidence that VOT is the primary cue of aspiration contrast in Mandarin. The unaspirated responses decreased as VOT became longer. At VOT step 1 (14 ms), which was the typical value of the unaspirated stop in Mandarin (Rochet & Fei, 1991), the native listeners provided the highest number of unaspirated responses, while starting from VOT 4 (56 ms), the listeners tended to give mainly aspirated responses regardless of the pitch levels. The VOT categorical boundary for the aspirated-unaspirated stops seemed to be different for the high initial tone stimuli and the low initial tone stimuli. The dramatic drop of the number of unaspirated responses indicated that the ambiguous VOT values for the high initial tone stimuli and the low initial tone stimuli were 28 ms and 42 ms, respectively. The VOT categorical boundaries occurred one step earlier for the high initial tone stimuli than for the low initial tone stimuli.

The current findings also suggest that L1 Mandarin listeners used post-stop F0 in stop identification tasks. They tended to associate high post-stop pitch with the aspirated stops and low post-stop pitch with unaspirated stops. In addition, the influence of pitch was modulated by tone. Stimuli with low initial tones (T2, T3) received significantly more unaspirated responses than stimuli with high initial tones (T1, T4). Pitch had a heavier influence on participants' responses with the low initial tone stimuli than with the high initial tone stimuli, especially when VOT was within the ambiguous region.

Taken together, these results indicated that listeners from a tonal language can extract both consonantal and tonal information from pitch.

3.4 The production-perception interface in L1

The present study offers a matched set of production-perception data from 25 L1 Mandarin speakers, focusing on the role of VOT and F0 at vowel onset in contrasting stop aspiration in Mandarin. The F0 perturbation effect was observed in the Mandarin production experiment. The results confirmed that the perturbation duration was limited to the onset of the vowel, ranging from 11 ms to 75 ms. The results were generally in line with the previous studies (Xu & Xu, 2003; Luo, 2018). However, the direction of F0 perturbation was different from those reported in the previous studies (Xu & Xu, 2003; Luo, 2018). There was a high initial tone and low initial tone effect. The F0-aspirated stops was significantly higher than the F0-unaspirated stops in T1 and T4 but was significantly lower than the F0-unaspirated stops in T2 and T3.

The results of the perception experiment indicate that L1 Mandarin listeners can decode both tonal and consonantal information from pitch. On the group level, the results

of the perception experiment partly reflected the production patterns. Pitch influenced the L1 listeners' judgement of the aspiration feature of the prevocalic consonant. The perception results also demonstrate the high initial tone and low initial tone effect. T1 and T4 paired as a group and T2 and T3 paired as a group. The onset pitch range of the tones modulated the listeners' aspiration perception. T1 and T4 elicited significant less unaspirated responses than T2 and T3. Overall, the listeners tended to associate high pitch with aspirated stops and low pitch with unaspirated stops.

The production patterns were not completely mirrored in the results from the perception experiment. In the production experiment, the F0-aspirated stops was lower than the F0-unaspirated stops in T2 and T3. However, T2 and T3 stimuli in the perception experiment received significantly more unaspirated responses than T1 and T4 stimuli. There thus appears to be a disconnection between production and perception.

Francis et al. (2006) have also reported a discrepancy between production and perception in Cantonese. They argue that the pattern in perception could be result from the participants' exposure to English rather than their L1 experience (the L2 exposure hypothesis). Their participants were recruited from Hong Kong and tended to have daily exposure to English. This could be a possible explanation for the current study, as the participants were living in the US at the time of testing and they were exposed to the American English patterns. Testing monolingual Mandarin speakers can help to verify this possibility and it can be examined in future studies.

In addition to the L2 exposure hypothesis, the discrepancy between perception and production could also be related to the robustness of the F0 perturbation patterns of

the high initial tone group and the low initial tone group. T2, one member of the low initial tone group, does not exhibit the expected F0 perturbation patterns consistently in production. This was found not just in this study but also in Luo's (2018) study. However, the high initial tones, T1 and T4, consistently demonstrate the expected patterns. In addition, the number of lexical words in each tone and the frequency of the four tones in Mandarin are different. Table 21 presents the number of words within each tone and the frequency of each tone in a Mandarin vocabulary corpus based on the Modern Chinese dictionary.

Table 21. The distribution of Mandarin tones (adapted from Liu & Ma, 1986)

Tone	**National Standard Corpus of Mandarin Words**		**Chinese Vocabulary Corpus**	
	N of words	percentage	frequency	percentage
T1	1959	25.19	24690	23.71
T2	1972	25.35	25130	24.13
T3	1300	16.71	17853	17.15
T4	2489	32.00	33560	32.23

As shown in Table 21, the most common tone for Mandarin is T4, and the least common is T3. T1 and T4 words occur 55.94% of the time in the Chinese vocabulary corpus, while T3 occurs only 17.15% of the time. Besides the amount of actual words of

each tone and tone frequency, variation of T3 such as Sandhi (Duanmu, 2007) could also play a role, where T3 becomes T2 when the following word is T3. Therefore, L1 listeners are more likely to be exposed to the F0 perturbation pattern in the high initial tone group than that in the low initial tone group. Taken together, the F0 perturbation pattern in the low initial tone group is arguably not a robust cue for consonant aspiration, since only T3 demonstrates the F0 perturbation pattern of the low initial tones, and T3 is the least frequent tone in Mandarin. Therefore, when these listeners were faced with an ambiguous token, they presumably relied on the more robust pattern: the high pitch is associated with the aspirated stop and the low pitch is associated with the unaspirated stop. In addition, the pattern is observed in many other languages, such as English. When the speakers were exposed to other languages, they would get positive feedback to reinforce the pattern. The L2 exposure could thus potentially have influenced the consonant aspiration identification process, but the current outcomes do not provide evidence for or against this possibility.

The results of the present study also have implications for the VOT categorical boundary between the aspirated stops and the unaspirated stops in Mandarin. According to a study on Chinese VOT conducted by Rochet and Fei (1991), the mean VOTs for the aspirated alveolar stop /t^h/ and unaspirated alveolar stop /t/ before a high back vowel /u/ are 105 ms and 15ms, respectively. The data from the production experiment were consistent with the findings reported by Rochet and Fei (1991). The mean, maximum and minimum VOT data are given in Table 22. Figure 8 represents the distribution of VOT durations of the two groups of stops in Mandarin.

Table 22. Mean, maximum and minimum VOT durations (ms) of Mandarin stops by native Mandarin speakers from the L1 production experiment

Tone	**voicing**	**mean by voicing**	**mean by tone**	**maximum**	**minimum**
T1	aspirated	111.3	106.7	170.5	70.3
T2	aspirated		114.4	187.7	71.7
T3	aspirated		119.0	230.8	74.5
T4	aspirated		105.0	167.4	46
T1	unaspirated	17.2	17.0	38.1	9.2
T2	unaspirated		18.5	37	8.2
T3	unaspirated		17.5	38.4	7.6
T4	unaspirated		16.0	32.1	9.2

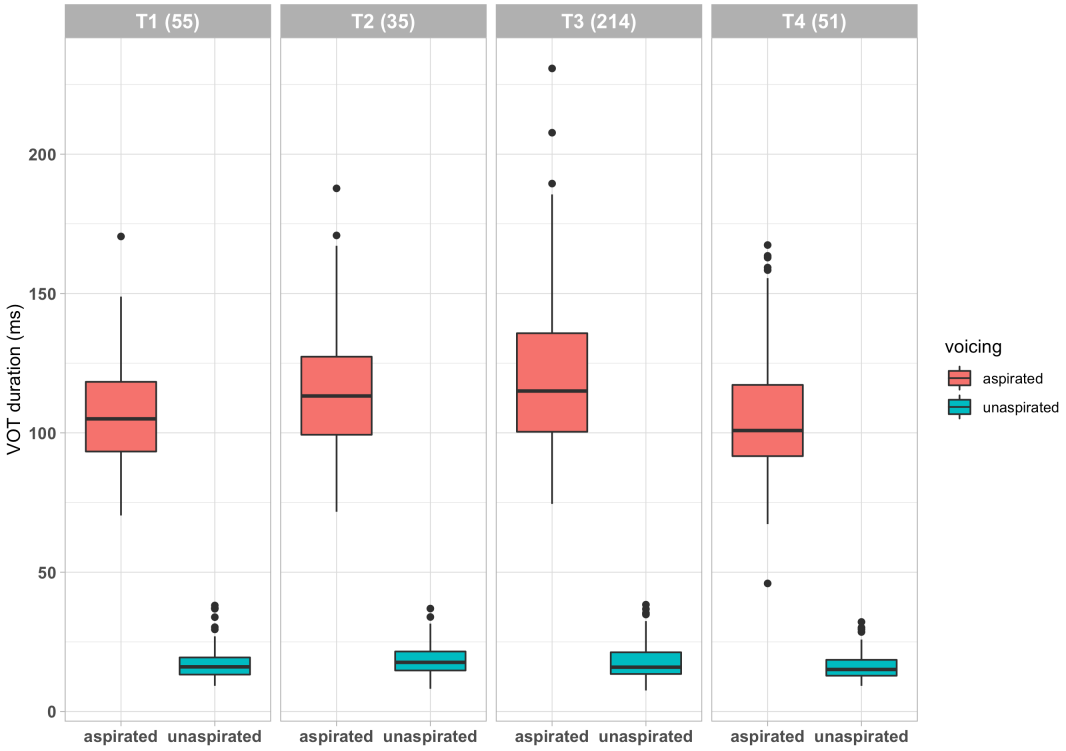


Figure 8. VOT durations of Mandarin stops by native Mandarin speakers

As indicated in Figure 8, the mean VOTs of Mandarin aspirated and unaspirated stops were around 111.3 ms and 17.2 ms, respectively, with a greater range of VOT variation of the aspirated stops than that of the unaspirated stops. The VOT of the aspirated stops does not overlap with the VOT of the unaspirated stops even the minimum VOT of the aspirated stops and the maximum VOT of the unaspirated stops. The non-overlap VOT region of the aspirated and unaspirated stops was between 40 ms to 60 ms, although one outlier of aspirated stop in T4 environment was 46 ms.

The perception experiment indicated that when VOT was located in a non-typical region of the aspirated and the unaspirated stops, the listeners were more likely to give an

aspirated response than an unaspirated response. The percentage of unaspirated responses at VOT step 3 (42 ms) in both high and low initial tonal environments was low. Even at VOT step 2 (28 ms), which was within the VOT range of unaspirated stops, participants provided mainly aspirated responses for T1 and T4 stimuli. High initial pitch of T1 and T4 played an important role on the listeners' aspiration identifications.

In summary, VOT was a primary cue and pitch was a secondary cue for native Mandarin listeners to distinguish the aspirated stops from the unaspirated ones in both production and perception. There seemed to be categorical boundaries of VOT for the unaspirated and the aspirated stops engrained as a part of the listeners' subconscious phonological knowledge. When VOT was located in the typical range of the unaspirated stops, the listeners tended to give unaspirated responses regardless of the pitch levels. When VOT was located in the typical range of the aspirated stops, the listeners tended to give aspirated responses regardless of the pitch levels. While when VOT was located in an ambiguous region, the listeners tended to give an aspirated response, indicating that the VOT of unaspirated stops permitted less variation than that of aspirated stops. Pitch had a heavier influence when VOT was ambiguous than when VOT was not ambiguous. The influence of pitch was relative to the influence of VOT. There seemed to be no categorical boundaries of pitch for the unaspirated and aspirated stops. Therefore, VOT was considered as the primary cue while pitch as the secondary cue. Tones modulated the role of pitch on consonant identification. High initial tones were more likely to be identified as following an aspirated stop than following an unaspirated stop when VOT information was ambiguous.

CHAPTER 4. F0 PERTURBATION AND STOP VOICING IDENTIFICATION IN ENGLISH BY MANDARIN SPEAKERS

4.1 Introduction

The two experiments in this chapter examine L1 Mandarin speakers' production and perception of the English stop voicing contrast. The experiments were completed by the same group of participants from chapter 3. Data collected from parallel experiments in production and perception as well as in L1 and L2 can be used to analyze group differences across modalities in L1 and L2. The experiments in this chapter focus on the role of VOT, F0 and intrinsic F0 of vowels in the English voicing contrast by Mandarin speakers of English.

The production experiment investigates whether L1 Mandarin speakers exhibit consistent F0 perturbation effects in their L2 production, and if so, how they produce the word-initial stop voicing contrast in terms of F0 perturbation direction and duration. The Mandarin participants are expected to produce F0 perturbation in their L2, as the laryngeal contrast in syllable initial position is realized as a phonetic aspiration contrast in both languages (e.g., Duanmu, 2007; Lisker & Abramson, 1964; Deterding & Nolan, 2007). If the F0 perturbation is primarily related to the phonetic aspiration of the syllable initial consonant, the Mandarin participants are expected to produce the English target-like pattern: the F0-voiceless stops is higher than the F0-voiced stops. Both English and Mandarin belong to the aspiration language group with voiceless unaspirated stops

ranging from 0 to 25 ms along the VOT continuum and 60 to 100 ms for voiceless aspirated stops (Lisker & Abramson, 1964). Deterding and Nolan (2007) found no significant VOT differences between the corresponding English and Mandarin stops in the syllable initial position. The F0 perturbation duration is expected to be shorter than 100 ms, which was the average length produced by L1 English speakers (e.g., Hombert et al., 1979). Studies that examined F0 perturbation in tonal languages suggested the perturbation duration was limited to the onset of the vowel (e.g., Hombert et al., 1979; Luo, 2018). The Mandarin speakers' L2 production may be influenced by the patterns in their L1. It is also possible that the Mandarin participants exhibit patterns that are close to neither their L1 nor the L2 patterns, which suggesting that they are in the process of learning the new pattern in their L2.

The perception experiment explores the same participants' perception of the English stop voicing contrast in a forced-choice identification task across a VOT continuum with 6 steps and a post-stop F0 continuum with 7 steps in two vowel environments. The Mandarin participants are expected to use VOT as a primary cue for the voicing contrast in English, as VOT is a primary cue for the laryngeal contrast in both languages (Lisker & Abramson, 1964). The participants are also expected to be able to use pitch as a cue for the voicing contrast in English. The F0 perturbation in English can extend 100 ms into the vowel. Francis et al. (2006) have reported that the Cantonese listeners were able to hear consonant-induced pitch differences in Cantonese when the manipulated perturbation duration was longer than 20 ms. Their study suggests that the longer the perturbation duration, the smaller pitch differences the listener can hear. The

perturbation duration in English should be long enough for the Mandarin participants in this study to hear the consonant-induced pitch differences if they are as sensitive to pitch as the L1 Cantonese listeners in Francis et al.'s (2006) study.

The intrinsic F0 of the vowel may or may not influence the F0 perturbation patterns in both production and perception. If the voicing contrast is restricted by physiological and/or aerodynamic factors (Ladefoged, 1967; Slis, 1970; Halle & Stevens, 1971; Ohala & Ohala, 1972; Löfqvist, 1975; Hombert et al., 1979; Kohler, 1984), the intrinsic F0 of the vowel is expected to influence the F0 perturbation patterns. Different vowel height requires different coordination of the articulators (e.g., Whalen & Levitt 1995). If the F0 perturbation is due to perceptual enhancement of the salience of the voicing feature by speakers (Kingston & Diehl, 1994), it is expected that the perceptual enhancement would apply to all the vowel environments. The intrinsic F0 of the vowel is also expected to influence the categorical boundary of VOT between English voiced and voiceless stops. Nakai and Scobbie (2016) argued that the vowel height affected the L1 English speakers' perceptual cutoff points for the two voicing categories in English. If the participants have attained the target-like perception of the voicing contrast, they are predicted to show the same pattern. It is also possible that the VOT categorical boundaries in Mandarin are also influenced by vowel height. The participants may transfer that knowledge into their L2.

In summary, the goal of the two experiments is to investigate the role of VOT, pitch and the intrinsic F0 of the vowels in distinguishing the voicing contrast in English by L1 Mandarin speakers in both production and perception tasks.

4.2 Experiment 1: Mandarin speakers' L2 English production

Experiment 1 examined Mandarin speakers' productions of the stop voicing contrast in English, focusing on the use of F0 within the first 50% of the vowel, which was observed to be relevant to the voicing contrast in English (Hombert et al., 1979; House & Fairbanks, 1953; Lehiste & Peterson, 1961; Lea, 1973; Hombert, 1978; Ohde, 1984; Hansen, 2009).

4.2.1 Method

4.2.1.1 Stimuli

The stimuli were minimal pairs of English monosyllabic words *two-do, tie-die, tea-D, me-knee,* and *know-mow*. The experimental stimuli were the first three pairs, which were used to examine the F0 perturbation in English, and the last two pairs were served as the control words to provide a F0 baseline without F0 perturbation. Two additional words, *tear* and *deer,* were used to create 2 practice trials. Due to limited available lexical words in English, the vowel environments were not balanced across the stop and sonorant groups. Each stimulus was embedded in an English carrier sentence '*Please say ____ again.*'

4.2.1.2 Participants

The same group of participants from the Mandarin experiments (in §3) completed the English production experiment on a separate day with at least one week between their two visits. The demographic information related to their English experience is summarized in Table 23. The participants lived in the U.S at the time of testing, so they were exposed to English every day in their normal lives. Five of them were studying in GMU degree programs, one was working at GMU, three were exchange students from

China and the rest (sixteen) of the participants were taking English language courses at INTO Mason.

Table 23. Demographic information of the participants

N	Age (years)			Age of arrival (years)			Length of residence (months)			
	mean	s.d	range	mean	s.d.	range	mean	s.d	range	median
25	26	8	19-46	22	3	19-35	13	13	1-48[12]	12

4.2.1.3 Procedure

The experiment took place in a sound attenuated booth at GMU. Participants were seated comfortably in front of a Macbook and their productions were digitally recorded onto a separate Macbook Pro, using a Røde smartLav+ microphone and an external Focusrite Scarlette Solo 2nd Generation preamplifier with a sampling rate of 44.1 kHz via the Praat program (Boersma & Weenink, 2020). The microphone was attached to the participants' shirt around their upper chest, approximately 6 inches away from the speakers' mouth. Stimuli sentences were presented to the participants in the middle of the computer screen automatically with a 2-second interval in a randomized order using PsychoPy (Peirce, 2007) in order to elicit a comparatively stable speaking rate across participants. All instructions for the experiment were given in English. Participants were instructed to read the sentence naturally. The experiment consisted of a 2-trial practice

[12] The participant who has been in the U.S. for 20 years is not included in length of residence column of the demographic information summary table. With her data added in, the descriptive statistics of length of residence is highly skewed towards the higher end, mispresenting the trend in the variable.

session and 60-trial (10 words * 3 repetitions * 2 blocks) experiment session. There was a break within the experiment session and the entire experiment took about 5 minutes.

4.2.2 Acoustic measurements

All measurements were performed with Praat (Boerma & Weenink, 2020) by the author. VOT was measured manually from the starting point of the target stop consonant burst in the waveform to the first zero crossing in the waveform following the onset of the periodicity of the following vowel. The end of the VOT also marked the onset of the vowel. The end of the vowel was marked at the offsets of the first and the second formants. The segmentation between the vowel and the consonant was mainly relied on the visual inspection of the spectral patterns and the wave forms. Examples of token segmentation are given in Figures 9 and 10.

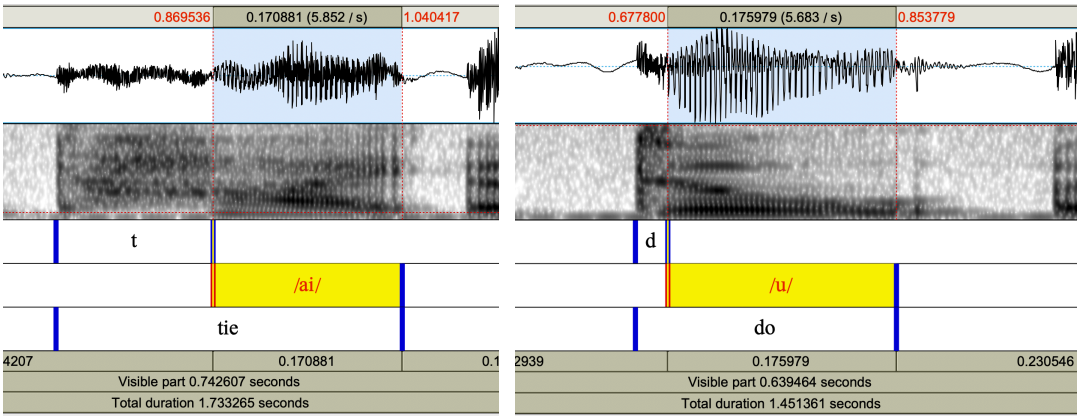


Figure 9. Segmentation of the English target words

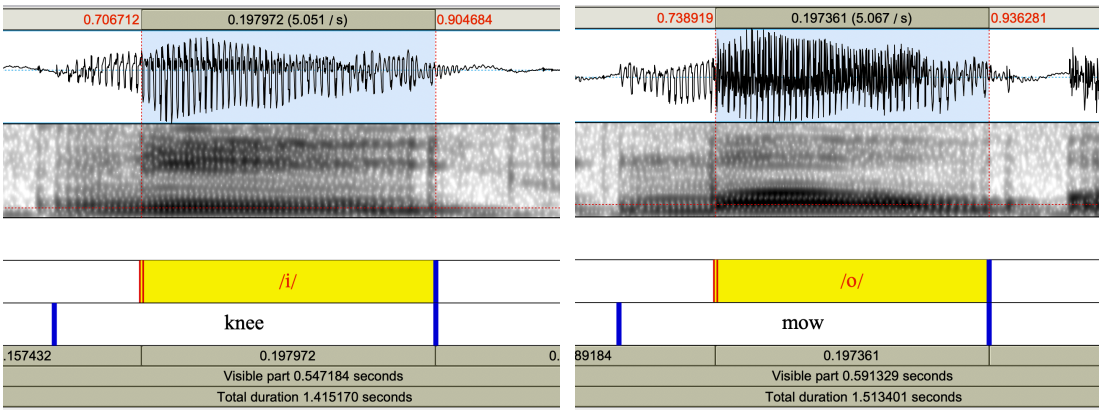


Figure 10. Segmentation of the English control words

Praat scripts were used to extract VOT durations, vowel durations and F0 values. According to previous studies (e.g., Lehiste & Peterson, 1961; Hombert et al., 1979), F0 perturbation in English can extend about 50% into the vowel. F0 was measured from ten equidistant points of the post-stop vowel. The selection of the ten equidistant points was based on the vowel duration of each word. Therefore, the extraction of the F0 values was normalized according to the vowel duration. Figure 11 shows the normalized F0 values

through the entire vowel of the different consonant groups. Vowels were not balanced in the three voicing groups, so Figure 11 represents F0 contours with the combined vowel environments (i.e., /i/, /u/ and /aɪ/).

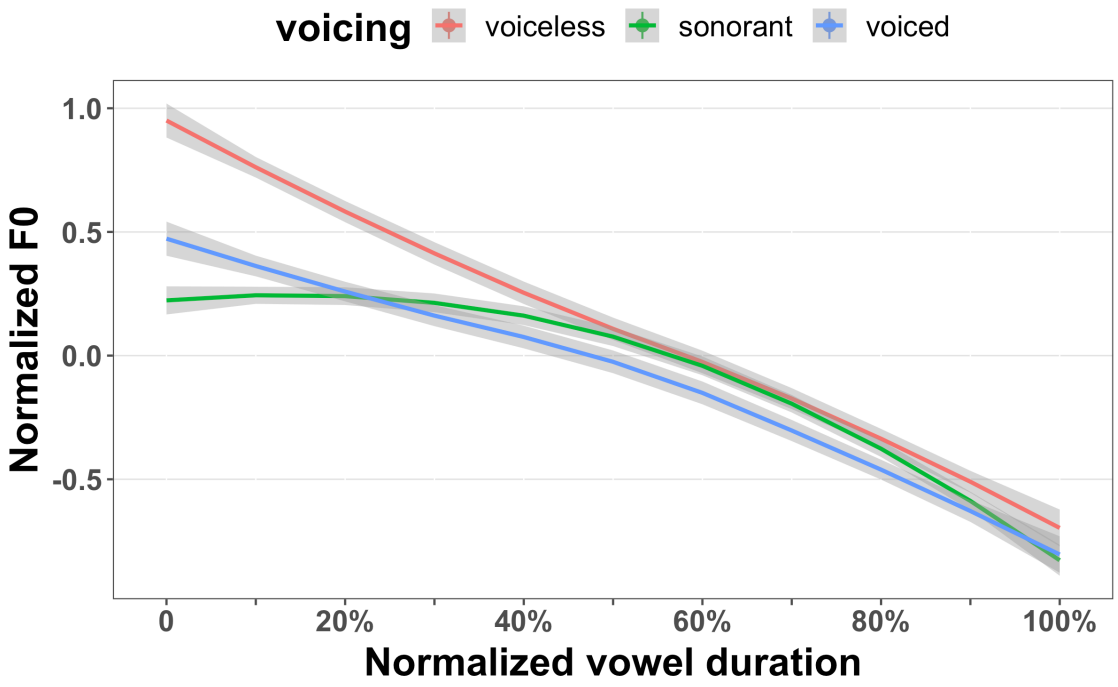


Figure 11. The normalized F0 of three consonant groups with combined vowel environments

4.2.3 Data preparation

Some data were excluded in the preparation for three reasons: mispronunciations and/or self-correction (7%), failure to extract F0 values by Praat (0.5%), and outliers above 3 standard deviations after by participant F0 normalization (0.3%).

106 of 1500 (60 experiment words * 25 participants) recorded words (7%) were excluded from all analyses. The excluded words included mispronunciations of the

intended vowel or the consonant (100) (e.g. /mo/ read as /maʊ/) and self-correction (6). The detailed number of mispronounced words are displayed in Table 24. Compared with their L1 productions, the participants mispronounced more words in the L2 task even though the stimuli were common English words. Among all the stimuli, *mow* received the highest number of mispronunciation due to the fact that many participants were not familiar with this word.

Table 24. The number of excluded English words due to mispronunciations

Stimuli	mow	tie	die	knee	know	tea
Number of excluded tokens	48	16	12	12	11	1
					Total: 100	

Based on the F0 trajectories of the three voicing groups (Figure 11), the first 6 of the 10 time points were included for further statistical analysis to model the F0 perturbation effect in Mandarin speakers' L2 productions, as F0 perturbation duration in native English productions can extend 50% into the vowel (e.g., Hombert et al., 1979). The first 6 points covered the first half of the vowel. The mean vowel durations of the three voicing groups by vowel environments are listed in Table 25. 47 data points of 13940 (0.3%) where Praat failed to extract F0 values were excluded from the statistical analysis. Table 26 presents the number of data points where Praat failed to get a F0 value. Praat tends to give an undefined value when it is requested to get a pitch value in a voiceless part of a sound (Boerma & Weenink, 2020).

Table 25. Mean English vowel durations by voicing groups and vowels

voicing	vowel	Dur (ms)	s.d.
aspirated	/aɪ/	193.5	50.6
	/i/	155.2	44.6
	/u/	156.3	40.0
sonorant	/i/	178.6	42.3
	/o/	199.4	45.1
unaspirated	/aɪ/	232.1	49.5
	/i/	178.1	41.6
	/u/	177.8	39.0

Table 26. The number of English data points excluded due to Praat extraction failure

Stimuli	die	tie	do	tea	D
Number of data points excluded	28	15	2	1	1
Total: 47					

Raw F0 values were then transformed into z-scores for each subject to facilitate the comparison of pitch across subjects. Outliers, 29 of 8317 data points (0.3%), above 3 standard deviations from each speakers' mean F0 were excluded from the analyses of F0 perturbation. In the end, 8288 data points were retained for F0 analysis.

4.2.4 Statistical analyses and results

Linear mixed-effects models were performed with the *lme4* package in R (Bates

et al., 2014) to investigate the possible influence of voicing, time point and gender on F0. In the full model, the dependent variable was the normalized F0. Voicing (the voiceless stop vs. the sonorant vs. the voiced stop), time points (6 time points) and gender (female vs. male) as well as the interaction among the three variables were included as fixed effects. The random effects structure of the model was determined using a forward best path algorithm (Barr et al., 2013). Subjects were included as a random effect. All fixed factors were coded using treatment (dummy) coding, with the reference level for the intercept being set to the voiceless, 0 time point and female. The best fitting model was selected by comparing models using the likelihood ratio test. The fixed effect gender did not fit the dataset significantly better than the full model (χ^2=2.12, p=0.14), indicating the contribution of gender to the model fit was not significant. In order to reduce the complexity of the model, gender was excluded as a fixed effect from the full model. The best model had voicing and time points as the fixed effects and subjects as the random effect. Tukey's HSD tests were conducted on this best model using the emmeans package (Lenth, 2020) for R and the summary of the post-hoc pairwise comparisons are reported in Table 27. Figure 12 visually demonstrates the effects of voicing and time points on the normalized F0 over the combined vowel condition.

Table 27. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model, F0-voiceless stops-F0-sonorants, F0-sonorants-F0-voiced stops, F0-voiceless stops-F0-voiced stops

Voicing		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**
		0%	10%	20%	30%	40%	50%
Voiceless-Sonorant	*ß*	0.735	0.533	0.363	0.211	0.106	0.045
	p	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0320 *	.5279
Sonorant-Voiced	*ß*	-0.271	-0.110	-0.012	0.045	0.073	0.092
	p	<.0001 ***	.0228 *	.9577	.5349	.1902	.0684
Voiceless-Voiced	*ß*	0.464	0.423	0.351	0.256	0.178	0.138
	p	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0001 ***	.0048 **

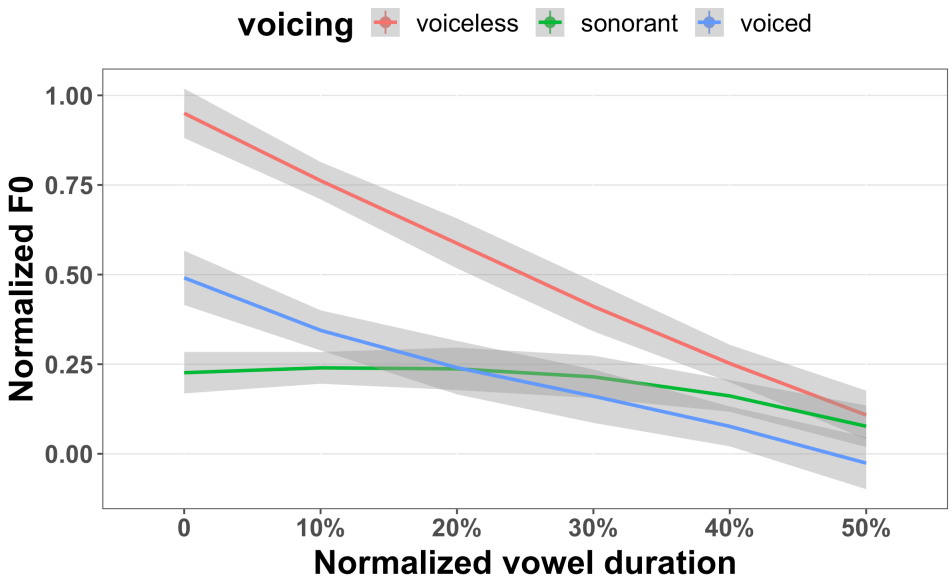


Figure 12. The first half of the normalized F0 contours of the three consonant groups with combined vowel environments

The F0-voiceless stops was significantly higher than the F0-sonorants during the first 40% of the vowel, and the F0-voiced stops was significantly higher than the F0-

sonorants during the first 10% of the vowel. The F0-voiceless stops was significantly higher than the F0-voiced stops through the entire 50% of the vowel.

The vowel environments were not balanced across the three voicing groups. In order to examine the effect of the intrinsic F0 of different vowels, a subset of data with the stops combined with 3 different vowels was selected to model the effect of voicing (voiceless vs. voiced), vowel (/aɪ/ vs. /i/ vs. /u/) and time points (6 time points). The interaction among the three fixed effects was included into the full model and participants were included as a random effect. Tukey's HSD tests were conducted using the emmeans package (Lenth, 2020) for R, and the summary of the post-hoc pairwise comparisons are reported in Table 28. Figure 13 visually demonstrates the effects of the voicing and time points on the normalized F0 over the three vowel conditions.

Table 28. Pairwise Comparisons: Results from Tukey HSD Post-hoc Analyses on the linear mixed effects model by vowel environments (F0-voiceless stops-F0-voiced stops)

Vowel		**time 0**	**time 1**	**time 2**	**time 3**	**time 4**	**time 5**
		0%	10%	20%	30%	40%	50%
/aɪ/	*ß*	0.007	0.283	0.333	0.301	0.244	0.207
	p	.9292	.0001 ***	<.0001 ***	<.0001 ***	.0006 ***	.0037 **
/i/	*ß*	0.621	0.386	0.264	0.159	0.100	0.089
	p	<.0001 ***	<.0001 ***	.0001 ***	.0175 *	.1357	.1867
/u/	*ß*	0.689	0.564	0.431	0.274	0.164	0.097
	p	<.0001 ***	<.0001 ***	<.0001 ***	<.0001 ***	.0148 *	.1483

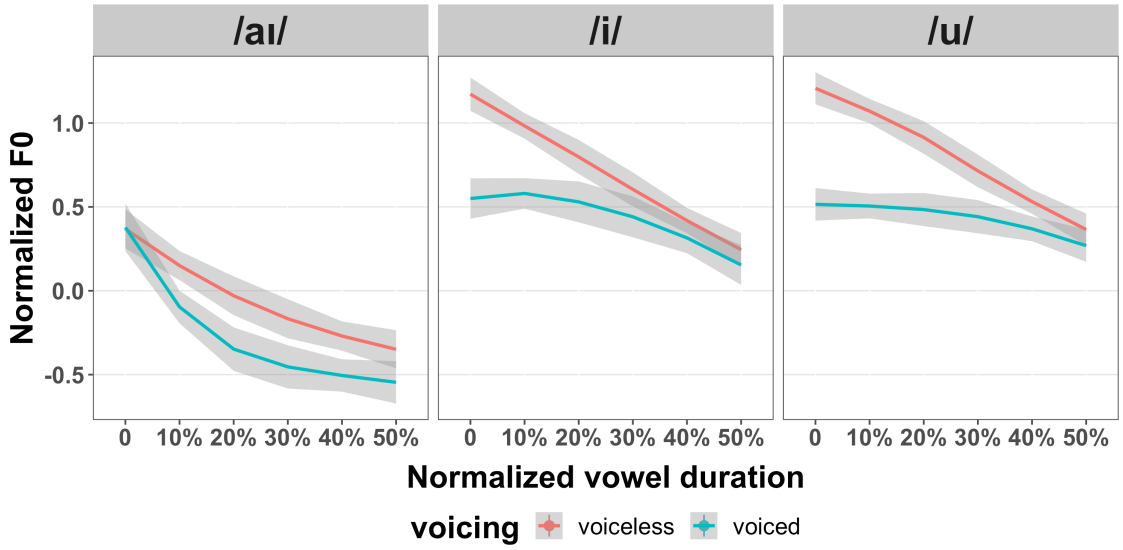


Figure 13. The normalized F0 contours of the two voicing groups by vowel environments

The statistical model revealed a significant effect of voicing (χ^2=307.98, p<0.0001). The F0-voiceless was significantly higher than the F0-voiced in all three vowel environments except for the 0 time point with vowel /aɪ/. The interaction among voicing, vowel and time points was significant (χ^2=57.59, p<0.0001). No significant F0 difference was observed between the F0 of the two voicing groups at time 0 with /aɪ/, while significant F0 difference was found between the F0 of the two voicing groups at time 0 with /i/ and /u/.

4.2.5 Interim summary

F0 perturbation direction. With the separated vowel condition (see Figure 13), the F0-voiceless stops was in general significantly higher than the F0-voiced stops in all the three vowel environments. No significant F0 difference was observed between the F0 of the two voicing groups at time 0 with /aɪ/, while significant F0 difference was found

between the F0 of the two voicing groups at time 0 with /i/ and /u/. The direction of the perturbation was primarily determined by the voicing feature of the consonant while the vowel had a modest influence.

F0 perturbation duration. With the separated vowel condition (see Figure 13), the perturbation duration was shorter than what was reported in studies by L1 English speakers (e.g., Lehiste & Peterson, 1961; Hombert et al., 1979). The durations of the perturbation with /i/ and /u/ were 50 ms (30% of the vowel) and 67 ms (40% of the vowel) respectively. There was no significant F0 difference between the two voicing groups (voiceless vs. voiced) at 0 time point with /aɪ/. The F0-voiceless stops was higher than the F0-voiced stops from time point 1 (10% of the vowel) until time point 5 (50% of the vowel). Vowel seemed to have an influence on the perturbation duration.

F0-stops vs. F0-sonorants. Both the F0-voiceless stops and the F0-voiced stops were higher than the F0-sonorants. The F0-voiceless stops was significantly higher than the F0-sonorants during the first 40% of the vowel, and the F0-voiced stops was significantly higher than the F0-sonorants during the first 10% of the vowel. With the combined vowel condition (see Figure 12), the F0-voiceless stops was significantly higher than the F0-voiced stops, which replicated the results from previous studies by L1 English speakers (e.g., Lehiste & Peterson, 1961; Hombert et al., 1979) in terms of the proportion of the duration. As to absolute duration, the F0 perturbation duration in current study was slightly less than 100 ms, especially for the voiceless group. See detailed durations in Table 29.

Table 29. Mean vowel durations by voicing with combined vowel condition

Voicing	Duration (ms)	s.d.	50% of the vowel
voiceless	167.2	48.2	83.6
sonorant	188.1	44.8	94.1
voiced	195.1	50.1	97.6

4.2.6 Discussion of the production experiment

A significant F0 perturbation effect was observed both in the combined vowel model and the separate vowel model. For the combined vowel condition, the post-stop F0 of both voicing groups was significantly higher than the F0-sonorants, suggesting that aspiration tended to raise F0 of the following vowel. Moreover, the F0-voiceless stops was significantly higher than the F0-sonorants during the first 40% of the vowel while the F0-voiced stops was significantly higher than the F0-sonorants during the first 10% of the vowel. The fact that the F0-voiceless stops deviated more from the sonorant baseline than the F0-voiced stops suggests that F0 perturbation was an effect of F0 raising of the voiceless stops rather than an effect of F0 lowering of the voiced stops.

F0 perturbation direction. For the separate vowel condition, the F0-voiceless stops was significantly higher than the F0-voiced stops except for the 0 time point with /aɪ/. The non-significant result at time point 0 with /aɪ/ could have resulted from a statistical error due to relatively sparse data points with /aɪ/. 28 (16 *tie* and 12 *die*) words were excluded due to pronunciation error, and 43 data points (28 for *die* and 15 for *tie*) were excluded due to Praat extraction failure. Altogether, 23% of the *tie-die* data points

were excluded from statistical analysis, which might lead to the non-significant results. The other possible reason for the non-significant result was the low intrinsic F0 of /aɪ/. /aɪ/ was produced in a lower F0 range than /u/ and /i/. Previous literature (e.g., Whalen & Levitt, 1994) has suggested a negative correlation between the first formant (F1) and vowel height: the lower the F1, the higher the vowel. In order to locate the transition between /a/ and /ɪ/ in /aɪ/ produced by the Mandarin speakers, F1 was measured from ten equidistant points of the post-stop vowel and F1 values for /aɪ/ are presented in Figure 14 by gender.

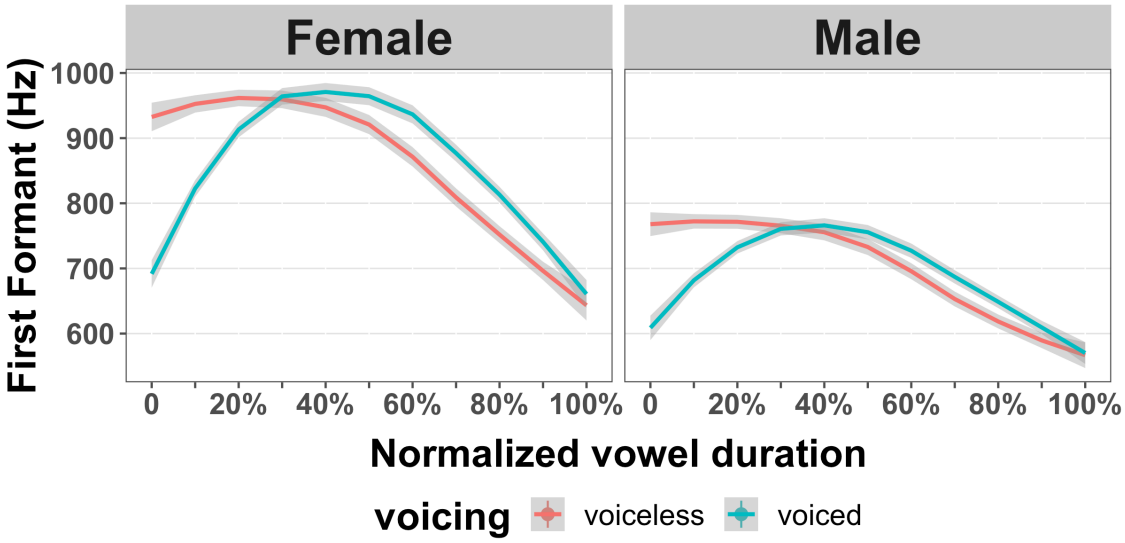


Figure 14. The first formant values of /aɪ/ throughout the entire vowel by gender

For both the male and female groups, the F1 of /aɪ/ was in the low vowel range (700-1000 Hz for female and 600-800 Hz for male). The F1 stayed at the low vowel

range until approximately the mid-vowel, where F1 started to drop sharply. The F1 of /aɪ/ reached the lowest point at the end of the vowel. However, the F1 was still not low enough to reach the typical F1 range for high vowels. The F1values of /u/ and /i/ in this study were around 500 Hz for female and 400 Hz for male. As shown in Figure 14, the F1 of the voiceless group was higher than the F1 of the voiced group, suggesting that /aɪ/ in the voiceless group was produced at a lower position than /aɪ/ produced in the voiced group. The lower vowel height of /aɪ/ in the voiceless group could potentially lower the F0-voiceless stops and the higher vowel height of /aɪ/ in the voiced group could potentially increase the F0-voiced stops. Thus, the effect of the F0 perturbation could be weakened by vowel height. In that case, the difference between the F0 of the two voicing groups with /aɪ/ may not be significant at the vowel onset. For /i/ and /u/, the F1 differences between the two voicing groups were not as obvious as in /aɪ/ (see Figure 15). It seemed that vowel height influenced Mandarin speakers' production of the voicing contrast in English.

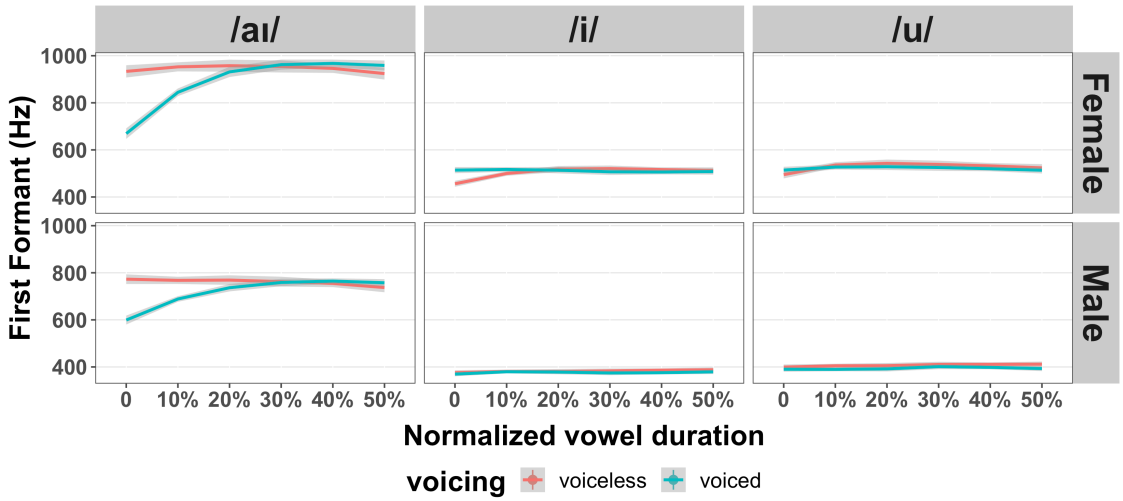


Figure 15. The first formant values by vowel environments and gender

F0 perturbation duration. For the combined vowel condition, the F0 perturbation was maintained throughout the entire 50% of the vowel, which overall matched the perturbation duration reported by L1 English speakers (e.g., Lehiste & Peterson, 1961; Hombert et al., 1979). However, for the separate vowel condition, the F0 perturbation duration was shorter both in absolute value and in percentage than what was reported in earlier studies produced by L1 English speakers, which was about 100 ms or 50% of the vowel (e.g., Lehiste & Peterson, 1961; Hombert et al., 1979). The shorter perturbation duration produced by Mandarin speakers could have resulted from the influence of the speakers' L1, in which the perturbation duration ranged from 0 to 35% of the vowel. The majority of participants in the present study were not considered fully proficient in English so far as they were enrolled at a language learning program, which prepares them to study at universities in the United States. They may not have thus far

mastered the skill to manipulate pitch to contrast the voicing feature of the prevocalic stops.

4.3 Experiment 2: Mandarin speakers' perception of English voicing contrast

Experiment 2 examined the influence of VOT, F0 and the intrinsic F0 of the vowel on Mandarin speakers' perception of the stop voicing contrast in English.

4.3.1 Method

4.3.1.1 Stimuli

English perception stimuli were created from natural productions of the syllables *too* and *tie*. The voiceless stops were selected as the baseline stimuli. That is, voiced tokens were created by removing the aspirated portions from the naturally produced voiceless stops. This is because it is more likely to get a natural sounding token by reducing the aspiration noise and shortening the VOT than by adding in aspiration noise and lengthening the VOT (Francis et al., 2006). The high back vowel /u/ was selected to create a parallel test with the Mandarin perception task. /aɪ/ was selected to form a comparison with /u/ to examine the possible influence of the vowel height on the voicing judgement task. The F1 values of the two base tokens throughout the entire vowels are presented in Figure 16. Based on the F1 range, *too* had higher intrinsic F0 than *tie*. A female native English speaker with no noticeable regional dialect accent recorded the natural productions of *too* and *tie* in a sound treated booth. Two unique tokens for each syllable with approximately the same length of VOT were selected as the bases for the acoustic manipulations. Each of the two base tokens were then manipulated to create 49 syllables co-varying in the VOT of the initial stops and the post-stop F0 by fully crossing 7 steps of post- stop F0 and 7 steps of VOT (see Table 30 for the details). Seven steps of

VOT and post-stop F0 were selected in order to obtain a detailed picture of how VOT and F0 would influence listeners' perception of English stops (Schertz et al., 2015). This yielded the parallel design of the Mandarin native perception experiment.

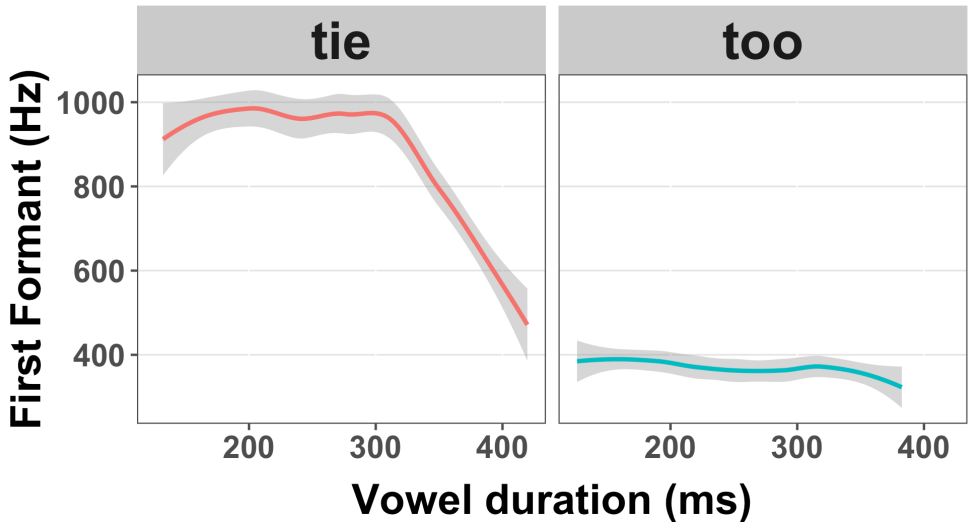


Figure 16. The F1 values of the two base stimuli of the L2 perception experiment

Table 30. VOT and onset F0 values for each acoustic dimension of the English stimuli

Word	Parameter	Base token	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7
tie	VOT (ms)	98.7	14.3	28.2	41.8	56.8	70.7	88.2	98.7
	F0 (Hz)	211.9	151.9	171.9	191.9	211.9	231.9	251.9	271.9
too	VOT (ms)	97.3	13.6	27.0	41.4	54.5	68.3	83.5	97.3
	F0 (Hz)	243.4	183.4	203.4	223.4	243.4	263.4	283.4	303.4

VOT manipulation: The mean VOT duration of the 2 base tokens was 98 ms, and the VOT step size (14 ms) was same as the native Mandarin experiment.

F0 manipulation: F0 was manipulated using Time-Domain Pitch-Synchronous-Overlap-and-Add-algorithm (TD-PSOLA, Moulines & Charpentier, 1990) as implemented in Praat. The first 50% of the vowel was selected for pitch manipulation. The methods used for VOT and pitch manipulation were exactly the same as the Mandarin perception experiment (see §3.3.1.1 for detailed information about the manipulation).

Two L1 English listeners were asked to test the naturalness of the synthesized tokens and they were judged to be as good tokens of the original target words. Four L1 Mandarin listeners were invited to pilot the experiment. The pilot participants did not respond differently to VOT step 6 (84ms) and VOT step 7 (natural VOT) stimuli. Thus, VOT step 6 (84 ms) stimuli were from the experiment to keep the duration of the experiment shorter. After excluding VOT step 6, the stimuli set of the experiment included 84 (2 words * 7 steps of F0 * 6 steps of VOT) unique tokens.

4.3.1.2 Participants

The same group of participants completed the English perception experiment after they finished the English production experiment. There was a 5-minute break between the production and perception experiment.

4.3.1.3 Procedure

Listeners participated in a forced-choice identification task presented in PsychoPy (Peirce, 2007). Two English words constituting the voiceless and voiced pairs (i.e., *too* vs. *do*, *tie* vs. *die*) were displayed on a laptop screen while they were hearing the stimulus. They were instructed to choose the word they heard by selecting one of the two words using a Cedrus button box (model RB-740). Stimuli for the two word-pairs were

presented in two separate blocks, with the order of the blocks being counter-balanced across participants. There was a break between the two blocks. Within each block, each of the 42 tokens was repeated three times in different random order. 13 participants saw the screen with the voiceless word on the left and the voiced word on the right, and 12 participants saw the opposite. All the participants reported to be right-handed. The task took about 10 minutes. A total of 6300 responses (25 participants * 2 blocks * 42 tokens * 3 repetitions) were collected from the experiment. RT was also collected from the experiment. The timer for reaction time started from the onset of the audio syllable and stopped when the participants hit the button on the response box to make their selection.

4.3.2 Statistical analyses and results

RT was normalized for each participant and the responses with RT more than 3 standard deviations away (108 out of 6300 responses, 1.7%) were excluded from the statistical analysis. The remaining responses were statistically analyzed using the logistic regression model with the *lme4* packages in R (Bates et al., 2014) to determine the influence of each acoustic cue on the identification of the prevocalic stops. In the full model, the dependent variable was the participant's response (voiceless vs. voiced stops). VOT step, F0 step, vowel, and the interactions among the 3 variables were included as fixed effects. The vowel was helmert contrast coded to examine the contribution of vowel height. Participants and words were included as random effects. VOT step, F0 step, and vowel were added as random slope to the random effect participants. The effects of the independent variables were investigated by comparing models using the likelihood ratio test. The full model was described above. Pitch step significantly contributed to model fit

(β= -0.260, χ^2= 14.883, $p < 0.001$), showing that as pitch increased, the possibility of voiced responses decreased. The interaction between VOT step and vowel significantly contributed to model fit (β= -0.465, χ^2= 12.027, $p < 0.001$), indicating that the high vowel stimuli elicited more voiced responses than the low vowel stimuli. None of the other interactions among the fixed effects were significant.

After eliminating non-contributing factors, the best model included pitch step, and the interaction between VOT step and vowel as fixed effects. Participants and words were included as random effects. VOT step, pitch step, and vowel were added as the random slope to the participants. Table 31 demonstrates the model summary of the best model. The vowel contrast was a significant predictor of voicing judgement (β= 3.862, $p<0.0001$), showing that the high vowel (*too*) elicited significantly more voiced responses than the low vowel (*tie*). VOT step was also a significant factor of voicing judgement, showing that as VOT increased, the number of voiced responses decreased. Figure 17 demonstrates the influences of VOT, pitch, and vowel on consonant voicing identification.

Table 31. β-coefficients, standard error and z- and p-values for the logistic regression model

	Estimate	**Std. Error**	**Z value**	**Pr (> | z |)**
(Intercept)	5.472	0.606	9.029	<2e-16 ***
VOT step	-2.184	0.197	-11.072	4.41e-05 ***
Pitch step	-0.260	0.064	-4.085	0.0016 **
Vowel	3.862	0.587	6.574	4.89e-11 ***
VOT step:Vowel	-0.465	0.141	-3.290	0.001 **

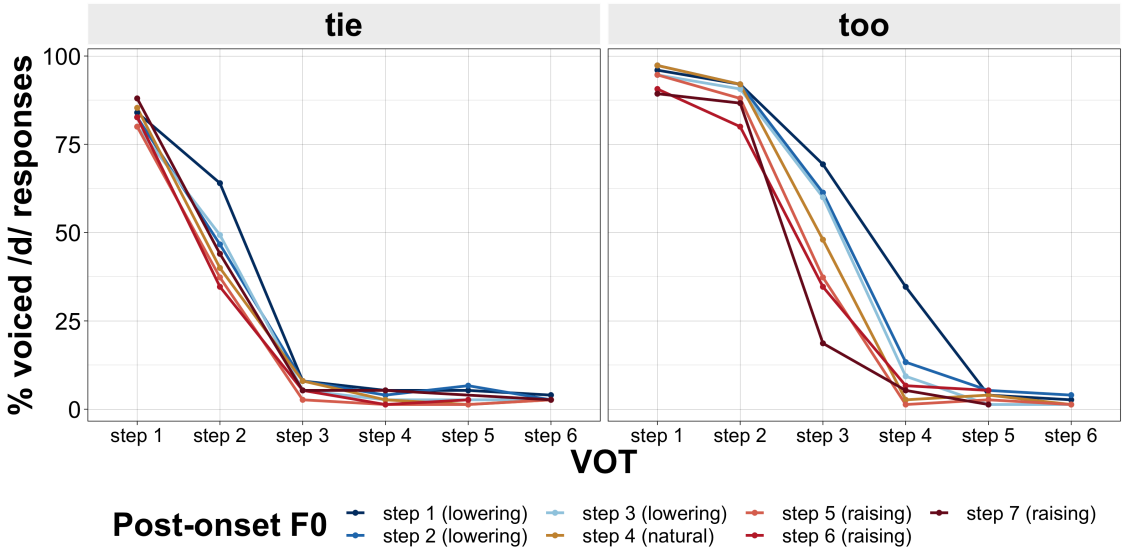


Figure 17. Percentage of voiced /d/ responses by native Mandarin speakers

As indicated by Figure 17, VOT step 1 (14ms) stimuli elicited the highest percentage of voiced responses for the two vowel environments. As VOT increased, the percentage of voiced responses decreased. The dramatic drop of the number of voiced

responses occurred at VOT step 2 (28 ms) for the *tie* stimuli set VOT step 3 (42 ms) for the *too* stimuli set. The percentage of voiced responses reached the lowest region from VOT step 3 (42 ms) and VOT step 4 (56 ms) for the *tie* and *too* stimuli sets, respectively. As pitch increased, the percentage of voiced responses tended to decrease in both vowel environments. The *too* stimuli set elicited more voiced responses than the *tie* stimuli set.

4.3.3 Interim summary

Mandarin listeners used VOT as a primary cue for the voicing judgement in their L2, as predicted. As VOT increased, the number of voiced responses decreased. Based on the perception results by native Mandarin speakers, the ambiguous VOT with vowel /aɪ/ was 28 ms(step 2) and the ambiguous VOT with /u/ was ranged from 42 ms (step 3) to 56 ms (step 4), with step 3 being more ambiguous than step 4.

Mandarin listeners used pitch as a secondary cue for the voicing judgement in their L2. As pitch increased, the number of voiced responses decreased. Pitch had a heavier influence on participants' perceptual judgement when VOT was ambiguous than when VOT was unambiguous.

The vowel environments influenced Mandarin listeners' voicing distinction in their L2. Overall, the *too* stimuli set was more likely to be identified as the voiced syllable than the *tie* stimuli set.

4.3.4 Discussion of the L1 perception experiment

VOT was the primary cue for L1 Mandarin listeners to distinguish the voicing contrast of English stops. As VOT became longer, the percentage of voiced responses decreased for both vowel environments. The VOT categorical boundary for /t/-/d/ seemed

to be different for the two vowel contexts. The dramatic drop of the number of voiced responses occurred one step earlier for the *tie* stimuli set than for the *too* stimuli set. The lowest percentage of voiced responses also occurred one step earlier for the *tie* stimuli set than for the *too* stimuli set. Therefore, the ambiguous region was one step earlier with /aɪ/ than with /u/, suggesting the Mandarin listeners allowed more VOT variations of /d/ when it was before /u/ than before /aɪ/.

Nakai and Scobbie (2016) have examined the perceptual VOT category boundary in English for L1 English listeners, and they suggest the place of articulation and vowel height influence the perceptual cutoff point for the two voicing categories while the speech rate does not play a role. They reported the /t/-/d/ boundary for /a/ was 21 ms and the boundary ranged from 36-40 ms for /u/. They do not include diphthong in their study, so it is still unknown whether the status as a diphthong may affect the boundary. As shown in Figure 16, 71% of the base perception stimuli token *tie* (/aɪ/) was produced as a /a/ and the rest of the vowel portion transited from /a/ to /i/. If the VOT categorical boundary for /aɪ/ is similar to that of /a/, then the VOT region of the voiced stop with vowel /aɪ/ is about 21 ms. The fact that the dramatic drop of the number of voiced responses occurred at VOT step 2 (28 ms) for the *tie* stimuli set VOT step 3 (42 ms) for the *too* stimuli set in the current study is thus in line with Nakai & Scobbie's (2016) findings on L1 English listeners.

In addition to VOT, post-onset F0 influenced the L1 Mandarin listeners' perception of the voicing contrast in English. The pitch lowering stimuli elicited more voiced responses than the pitch raising stimuli. It seemed that the effect of F0 on

Mandarin listeners' voicing judgement was gradient for pitch lowering and pitch raising stimuli when VOT was ambiguous: The higher the post-stop F0, the more voiceless responses; the lower the post-stop F0, the more the voiced responses. The gradient effect was not evident when VOT was unambiguous. The results suggest that the L2 learners used post-stop pitch in the identification of consonant voicing and they tended to associate higher post-stop pitch with the voiceless stop and lower post-stop pitch with the voiced stop.

The intrinsic F0 of the vowel also affected the identification of the English voicing contrast by L1 Mandarin listeners. Stimuli with higher intrinsic F0 (*too*) elicited more voiced responses than stimuli with lower intrinsic F0 (*tie*). Pitch had a heavier influence on the *too* stimuli set than on the *tie* stimuli set. There seemed to be a perceptual compensation effect in the identification of English stops by Mandarin listeners. When the listeners heard a *too* stimulus with higher F0, they tended to attribute the high F0 to the high intrinsic F0 of /u/ rather than the voicelessness of the prevocalic consonant when VOT was ambiguous. By contrast, when they heard a *tie* stimulus with higher F0, they could only attribute the high F0 to the voicelessness of the prevocalic consonant. Therefore, the L2 learners gave more voiced responses for the *too* stimuli set than the *tie* stimuli set especially at VOT step 2 (28 ms) and VOT step 3 (42 ms), which were within a range of ambiguous VOT between the English voiceless and voiced stops.

4.4 The production-perception interface in L2

The present study provided a matched set of English production-perception data from 25 L1 Mandarin speakers who were learning English as an L2. F0 perturbation was

observed in the production study (§4.2). Overall, the F0-voiceless stops was significantly higher than the F0-voiced stops. The duration of F0 perturbation matched the L1 English speakers' productions for the combined vowel condition. However, for the separate vowel condition, the duration of F0 perturbation was slightly shorter than that produced by the L1 English speakers.

For the perception experiment (§4.3), the L2 learners were able to use post-onset F0 to distinguish the voicing category of the prevocalic stop in English. They used VOT as a primary cue and post-onset F0 as a secondary cue. They tended to associate high F0 with voiceless stops and low F0 with voiced stops. There seemed to be a perceptual compensation relationship between F0 perturbation and intrinsic F0 of the vowels. The *too* stimuli set elicited significantly more voiced responses than the *tie* stimuli set. It seemed that the listeners tended to attribute the high F0 they hear to the intrinsic F0 of the vowel rather than the voicelessness of the preceding consonant when VOT was ambiguous.

In general, the production patterns were reflected in their perception task. VOT was the most salient cue for the voicing contrast in English both in production and perception, and pitch was a secondary cue. Intrinsic F0 of the vowel also influenced the learners' production and perception of English voicing contrast.

The English VOT category boundary produced by the L1 Mandarin speakers in the L2 production experiment (§4.2) was not completely mirrored in their L2 perception experiment (§4.3). For participants of the present study, /t/ had a longer mean VOT when combined with /aɪ/ (109.3 ms) than combined with /u/ (90.0 ms), while /d/ had a shorter

mean VOT when combined with /aɪ/ (15.5 ms) than combined with /u/ (20.5 ms). The non-overlap VOT region between the maximum VOT of /d/ and the minimum VOT of /t/ with /aɪ/ was from 31.3 ms to 61.6 ms, and the corresponding VOT region with /u/ was from 36.5 ms to 47.1 ms. VOT was ambiguous in a greater VOT region when the vowel was /aɪ/ than it was /u/. Table 32 summarizes the mean, maximum, minimum VOT durations produced by Mandarin speakers in the L2 production experiment (§4.2). Figure 18 visually illustrates the VOT distribution by voicing and vowel groups.

Table 32. Mean, maximum and minimum VOT durations (ms) of English stops by native Mandarin speakers from the L2 production experiment

vowel	**voicing**	**mean**	**maximum**	**minimum**
/aɪ/	voiceless	109.3	199.7	61.6
	voiced	15.5	31.3	5.3
/u/	voiceless	90.0	210.4	47.1
	voiced	20.5	36.5	7.3

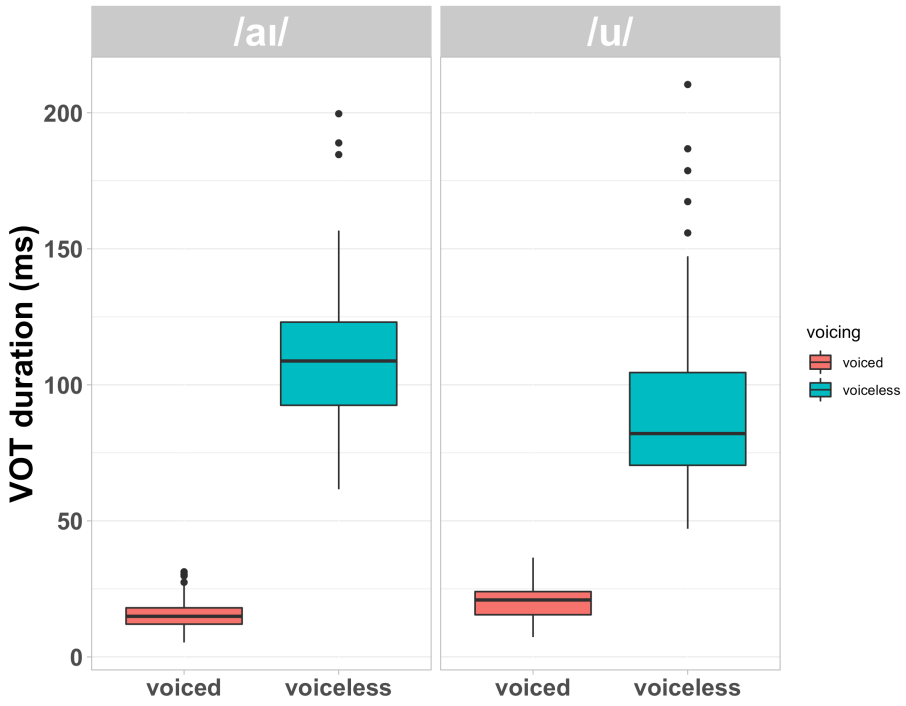


Figure 18. VOT durations of English stops by native Mandarin speakers

The production patterns were not fully represented in the perception experiment. In perception, the ambiguous VOT region occurred at the 28 ms (VOT step 2) with /aɪ/ stimuli. However, the production experiment suggested 42 ms (VOT step 3) and 56 ms (VOT step 4) could be ambiguous for the Mandarin listeners. For /u/ stimuli, both production and perception experiments suggested 42 ms (VOT step 3) was ambiguous for the Mandarin listeners. However, there was one discrepancy for the /u/ stimuli. In the production experiment, 56 ms was in the range aspirated stops, while it was a little ambiguous for the listeners in the perception experiment, especially with the pitch lowering stimuli.

One possible reason for the observed disconnection between the production and perception results is that the Mandarin listeners' perception of English voicing contrast was shaped by the inputs they were exposed to rather than their own production patterns. Nakai and Scobbie (2016) have observed that the /t/-/d/ boundary for /a/ is 21 ms and the boundary ranges from 36-40 ms for /u/, which matched the ambiguous VOT for the *tie* stimuli set and the *too* stimuli set in the current perception experiment. The Mandarin learners may be able to identify the English stop voicing contrast in a target-like manner.

In summary, VOT was a primary cue for native Mandarin speakers to distinguish English voiced and voiceless stops. Pitch was a secondary cue. The intrinsic F0 of the vowel influenced the Mandarin participants' voicing judgements. When VOT was ambiguous, vowels that have high intrinsic F0 was more likely to be identified as following the voiced stops than following the voiceless stops. The intrinsic F0 of the vowel could also influence the perceptual categorical boundary of VOT by the learners. The perceptual VOT range of the voiced stops was larger with /u/ than with /ai/

CHAPTER 5. GENERAL DISCUSSION AND CONCLUSION

5.1 Summary of the main findings

Taken together, the two sets of parallel experiments presented in the current study provide two case studies of how Mandarin speakers make use of various acoustic cues to produce and identify the laryngeal contrasts in their L1 and L2. The studies were designed to address the more general question of how speakers from a tonal language realize the stop laryngeal contrast in their L1 production, how listeners from a tonal language integrate information from multiple acoustic cues to identify the L1 laryngeal contrast, as well as how the speakers adapt the multiple acoustic cues when they acquire an L2.

Chapter 3 investigated how L1 Mandarin speakers contrast Mandarin stop aspiration in production and perception experiments. An F0 perturbation effect was observed in the Mandarin production experiment (§3.2). The results confirmed that the perturbation duration was limited to the onset of the vowel, ranging from 11 ms to 75 ms. The results were generally in line with the previous studies (Xu & Xu, 2003; Luo, 2018). However, the present study observed a different direction for F0 perturbation. There was a high initial tone and low initial tone effect: the F0-aspirated stops was significantly higher than the F0-unaspirated stops in T1 and T4 but significantly lower than the F0-unaspirated stops in T2 and T3. The results of the current study suggested that the height

of the larynx, the tension of the vocal cords and the change of subglottal pressure were the physiological factors influencing the direction of F0 perturbation. The change of the tonal contours affected the duration of the F0 perturbation patterns in Mandarin. The current study found that the F0 perturbation pattern in tonal languages was largely an automatic effect due to the physiology of the human phonation mechanism rather than a controlled process to enhance the phonological status of the consonants. The results of the perception experiment indicated that L1 Mandarin listeners can decode both tonal and consonantal information from pitch (§3.3). On the group level, the results of the perception experiment partly reflected the production patterns. Pitch influenced L1 listeners' judgement of the aspiration feature of the prevocalic consonant. The perception results also demonstrated the high initial tone and low initial tone effect: T1 and T4 were paired as a group, and T2 and T3 were paired as a group. T2 and T3 elicited significantly more unaspirated responses than T1 and T4. The listeners tended to associate high pitch with aspirated stops and low pitch with unaspirated stops within each tonal context and between the high initial tone and low initial tone groups.

Chapter 4 used parallel tasks to examine L1 Mandarin speakers' perception and production of the English voicing contrast. F0 perturbation was observed in the production study (§4.2). Overall, the F0-voiceless stops was significantly higher than the F0-voiced stops. The duration of F0 perturbation matched the L1 English speakers' productions for the combined vowel condition. However, for the separate vowel condition, the duration of F0 perturbation was shorter than that produced by the L1 English speakers. For the perception experiment (§4.3), the L2 learners were able to use

pitch to distinguish the voicing category of the prevocalic stop in English. They used VOT as a primary cue and post-onset pitch as a secondary cue. They tended to associate high F0 with voiceless stops and low F0 with voiced stops. There seemed to be a perceptual compensation relationship between F0 perturbation and the vowel height. The *too* stimuli set elicited significantly more voiced responses than the *tie* stimuli set. It seemed that the listeners tended to attribute the high F0 to the intrinsic F0 of the vowel rather than the voicelessness of the preceding consonant when VOT was ambiguous.

5.2 Native language influence on L2 voicing contrast

5.2.1 Stop contrasts production between L1 and L2

This study also aimed to compare how pitch was demonstrated in Mandarin and English by native Mandarin speakers, so a subset of the stimuli from the L1 and L2 production experiments were designed to share the same vowel environment (i.e., /u/). In order to examine the realization of F0 when it functions differently, the Mandarin alveolar stops combined with the high vowel /u/ in T4 and the English alveolar stops combined with the high vowel /u/ were selected for further analysis. Although the pitch contours for both groups of stimuli start from a high pitch range and drop to a low pitch range, the essential function of the falling contour was quite different. The falling pitch contour for the Mandarin words is the lexical tone, while for English the falling pitch contour is intonation. Figure 19 presents the stop contrasts in Mandarin and English produced by the same group of participants.

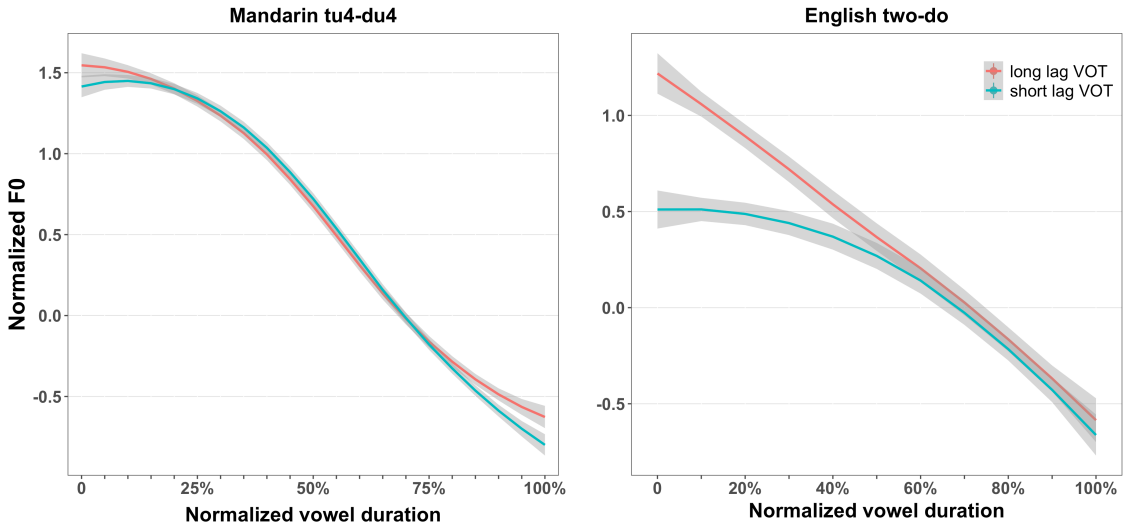


Figure 19. The F0 contours in L1 (*tu4-du4*) and L2 (*two-do*) by Mandarin speakers

As shown by Figure 19, the F0 range of Mandarin T4 was larger than the F0 range of English. For the Mandarin stimuli, the range was from below 0.5 standard deviation from the mean to around 1.5 standard deviation from the mean. While for the English stimuli, the range was from below 0.5 standard deviation from the mean to around 1.0 standard deviation from the mean. The maximum pitch of the Mandarin stimuli was higher than that of the English stimuli. The F0 difference between stops with long lag VOT and stops with short lag VOT in Mandarin was much smaller than that in English. Moreover, the F0 perturbation duration was much shorter in Mandarin than in English. Studies argued about which F0 perturbation pattern (the pattern in tonal vs. the pattern in non-tonal languages) is more fundamental (e.g., Francis et al., 2006). The essential question is, do the speakers from tonal languages actively inhibit the consonant induced F0 perturbation (inhibition hypothesis) or do the speakers from non-tonal languages actively exaggerate consonant induced F0 perturbation to enhance the voicing

feature (exaggeration hypothesis)? It seemed that the comparison between Mandarin speakers' L1 and L2 production in the current study mainly supported the inhibition hypothesis. The majority of the participants in the current study were not considered as proficient English speakers. Their L2 production was expected to be influenced by their L1 especially the primary phonetic property distinguishing the stop contrast in both L1 and L2 is aspiration. In spite of the small F0 differences produced in their L1, the participants produced big F0 differences with long perturbation durations in English. It seemed that the F0 perturbation effect was big without the interference of the tone even though it was produced by speakers from a tonal language. The effort of maintaining distinctive tonal contours may reduce the consonant induced F0 perturbation, which supports the hypothesis that the F0 perturbation is more of an automatic effect rather than a controlled effect. The outcome of this analysis, however, did not directly contradict to the exaggeration hypothesis. Given the F0 perturbation duration in the current experiment was shorter than that produced by L1 English speakers (see §4.2.5), it is still possible that the L1 English speakers take advantage of the automatic F0 perturbation effect and exaggerate the F0 perturbation effect. The automatic part could be easily achieved by the L2 learners, while the phonologically motivated enhancement may or may not be acquired by L2 learners.

To sum up, the production results from chapter 3 (Mandarin L1) and chapter 4 (Mandarin L2) suggested the F0 perturbation effect was primarily an automatic effect as the acoustic cues realized with F0 influenced F0 perturbation in the two production tasks. The comparison between Mandarin speakers' L1 and L2 production indicated that the

speakers from a tonal language inhibited the F0 perturbation effect to keep the tonal information intact.

5.2.2 Stop contrasts perception between L1 and L2

The Mandarin listeners relied on F0 information for the laryngeal contrast and tended to associate high F0 with long lag VOT and associate low F0 with short lag VOT in both L1 and L2. It seemed that the onset pitch of the tones influenced the stop identification and the perceptual VOT categorical boundaries. In the L1 perception task, the high initial tones (T1 and T4) elicited significantly less unaspirated responses than the low initial tones (T2 and T3). The dramatic drop of unaspirated response suggested that the perceptual VOT category boundaries for the high initial tones and the low initial tones were around 28 ms and 42 ms respectively. As to the L2 perception task, the intrinsic F0 of the vowels influenced the stop identification and the perceptual VOT categorical boundaries. Vowels with a high intrinsic F0 allows more VOT variations to be considered as voiced stop than vowels with a low intrinsic F0. The English VOT category boundaries for the Mandarin listeners were around 28 ms with /aɪ/ and 42 ms with /u/. It is still unknown whether such effect occurs in Mandarin speakers' L1 perception.

5.3 Implications and suggestions for future study

This study examined the production and perception of English stop contrast by L1 Mandarin speakers. The second phase of this project extends this dissertation and examines the production and perception of Mandarin stop contrast by L2 learners from a non-tonal language. L1 English speakers who are learning Mandarin as an L2 are tested in the Mandarin production and perception experiments. The F0 in the participants' L2

carries both tonal and consonantal information while it only carries consonant information in their L1. It will make a good opportunity to test how the learners incorporate the two kinds of information when they are learning their L2.

The results of the current study also inspire further investigation of the role of diphthong in F0 perturbation. As discussed in §4.2.6, the F0 perturbation pattern in English with /aɪ/ was different from that with /u/ or /i/. An interesting question to ask is whether the first part or the entire entity of the diphthong affects the F0 perturbation. Understanding of this question can help to understand the source of F0 perturbation as the long duration of diphthongs and the transition from the first member to the second member may require different articulatory gesture.

5.4 Conclusion

The parallel studies of F0 perturbation across languages and modalities in this dissertation offered insight into between-language (tonal vs. non-tonal) and within-language (production vs. perception) variations of how Mandarin speakers and listeners used different acoustic dimensions to contrast aspiration in their L1 and how they adapted the information of acoustic cues when they learned an L2. Short F0 perturbation was observed in L1 Mandarin speakers' production with a high and low initial tone effect. The findings of the production experiments suggested the F0 perturbation effect was primarily an automatic effect due to the physiology of the human phonation mechanism rather than a controlled process to enhance the phonological status of the consonants. The acoustic features such as tone and vowel height that realized with F0 influenced F0 perturbation patterns as different tones and vowels require different coordination of

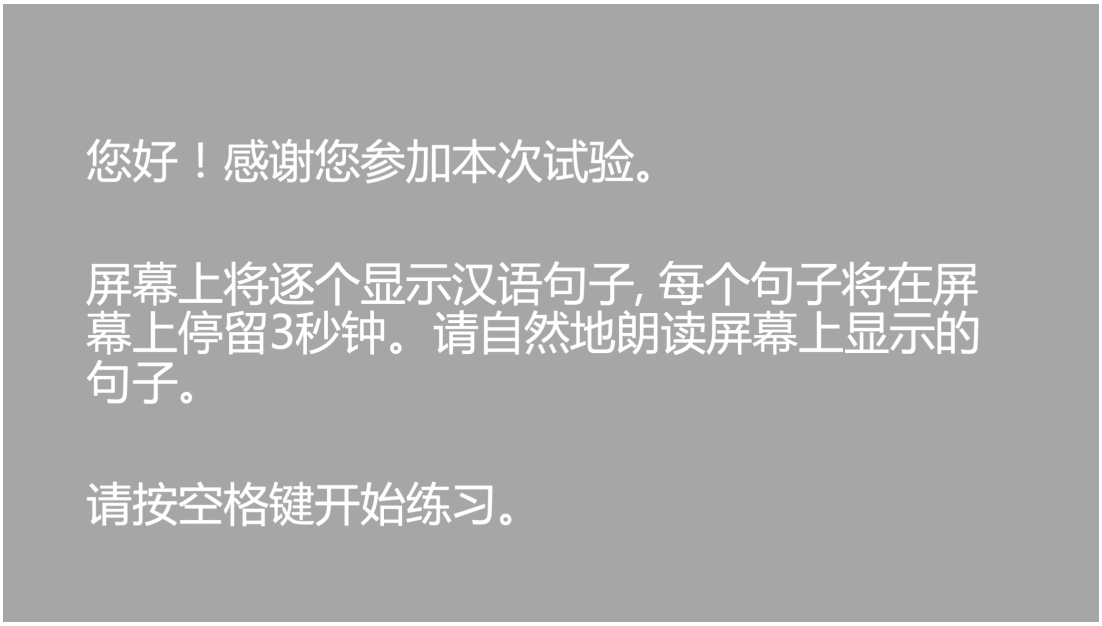
articulators. The height of larynx, the tension of vocal cords and the volume of subglottal pressure could all play a role in producing F0 perturbation. Compared with their L1 production, the Mandarin speakers produced F0 perturbation in English with larger F0 differences between the two voicing groups and longer F0 perturbation duration. It supported the inhibition hypothesis, i.e., the speakers from a tonal language inhibited the F0 perturbation effect to keep the tonal information intact, to account for the fact that the perturbation duration is in general short in non-tonal languages. In perception experiments, the Mandarin listeners used VOT and F0 for stop identification in both L1 and L2. They tended to associate high F0 with long lag VOT stops and low F0 with short lag VOT stops. Tones modulated the stop identification in Mandarin and vowel heights influenced the stop identification in English.

This dissertation attempted to account for the controversies in previous studies as to the direction of F0 perturbation in tonal languages and it is the first study that examined the perception of F0 perturbation in Mandarin. It is also the first attempt to directly compare the stop contrast in production and perception between L1 and L2 by L1 Mandarin speakers. It contributes to the understanding of F0 perturbation in both tonal and non-tonal languages. Along with the findings, this dissertation provides a balanced corpus for testing models of perception-production link as well as L1-L2 interface.

APPENDIX

Appendix A1: Mandarin production experiment instructions

The experiment instruction interface.



Translation:

Hello! Thank you for your participation.

Chinese sentences will be displayed one by one on the screen. Please read aloud the sentence on the screen naturally.

Please hit spacebar to start practice trials.

Appendix A2: Mandarin production experiment stimuli

Mandarin Words	Pinyin	IPA	Meaning
搭	dā	tɑ˥	take, build
他	tā	t^hɑ˥	he
爬	pá	p^hɑ˧˥	climb, crawl
拔	bá	pɑ˧˥	pull out
打	dǎ	tɑ˨˩˦	hit
塔	tǎ	t^hɑ˨˩˦	tower
大	dà	tɑ˥˩	big
踏	tà	t^hɑ˥˩	step on
督	dū	tu˥	superintend
突	tū	t^hu˥	suddenly
独	dú	tu˧˥	alone, sole
图	tú	t^hu˧˥	picture
堵	dǔ	tu˨˩˦	block up
土	tǔ	t^hu˨˩˦	soil
度	dù	tu˥˩	spend
兔	tù	t^hu˥˩	rabbit
屋	wū	wu˥	house
无	wú	wu˧˥	none
五	wǔ	wu˨˩˦	five
物	wù	wu˥˩	stuff
挖	wā	wɑ˥	dig
娃	wá	wɑ˧˥	baby
瓦	wǎ	wɑ˨˩˦	tile
袜	wà	wɑ˥˩	socks
爸	bà	pɑ˥˩	dad
怕	pà	p^hɑ˥˩	scared

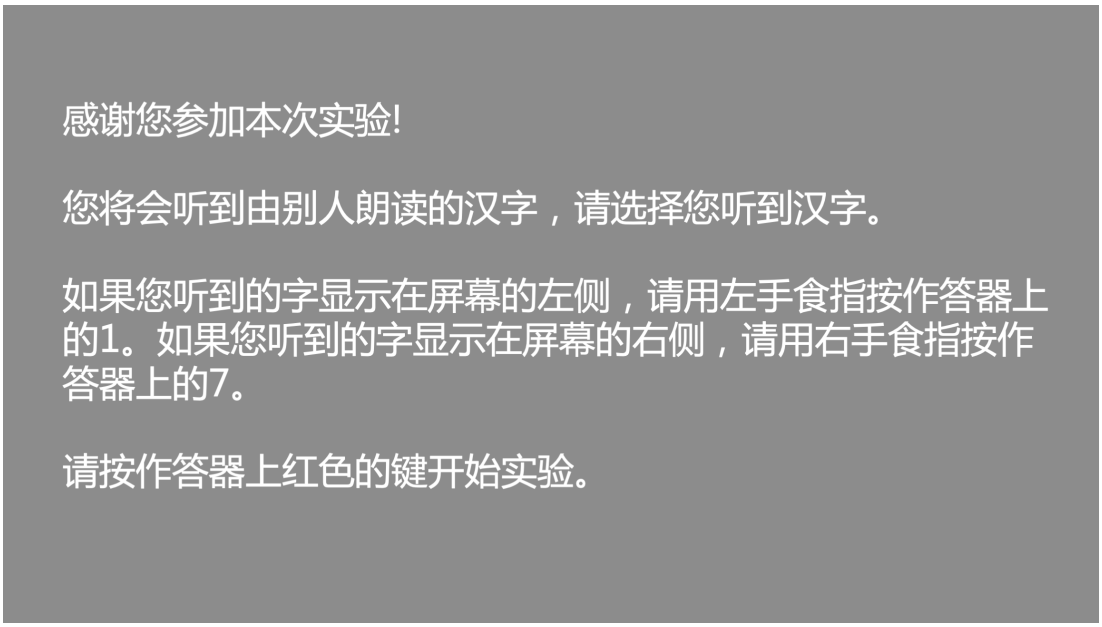
Appendix A3: Mandarin production experiment task interface

Literal translation:

Please say _____ one time.

Appendix B1: Mandarin perception experiment

The experiment instruction interface.



Translation:

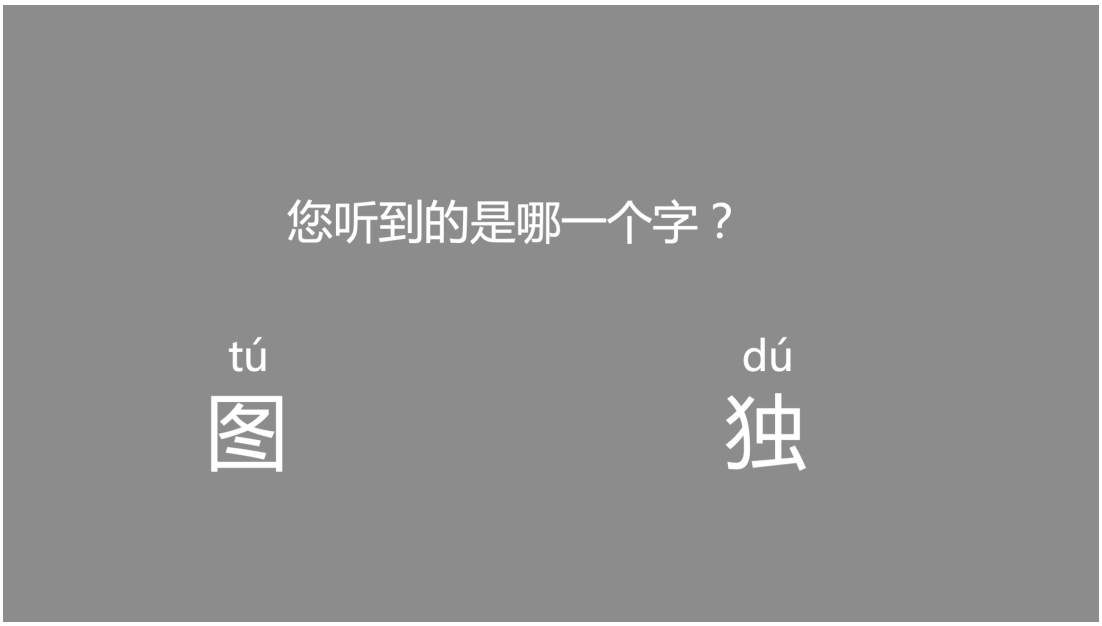
Thanks for participating!

In this experiment, you will hear someone saying Mandarin words. Please select the word you hear by pressing the associate key on the response box.

If the word you hear is displayed on the left side of the screen, please hit 1 on the response box using your left hand. If the word you hear is displayed on the right side of the screen, please hit 7 on the response box using your right hand.

Please press the red key to start the experiment.

Appendix B2: Mandarin perception experiment task interface

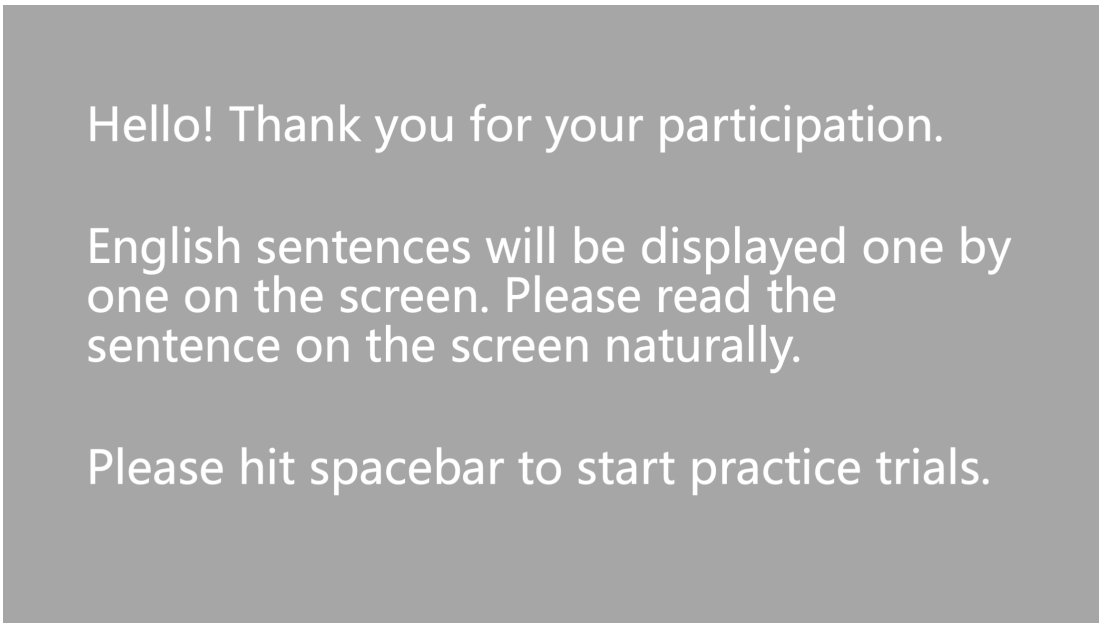


Translation of the question:

Which word did you hear?

Appendix C1: English production experiment instruction

English production instruction interface.

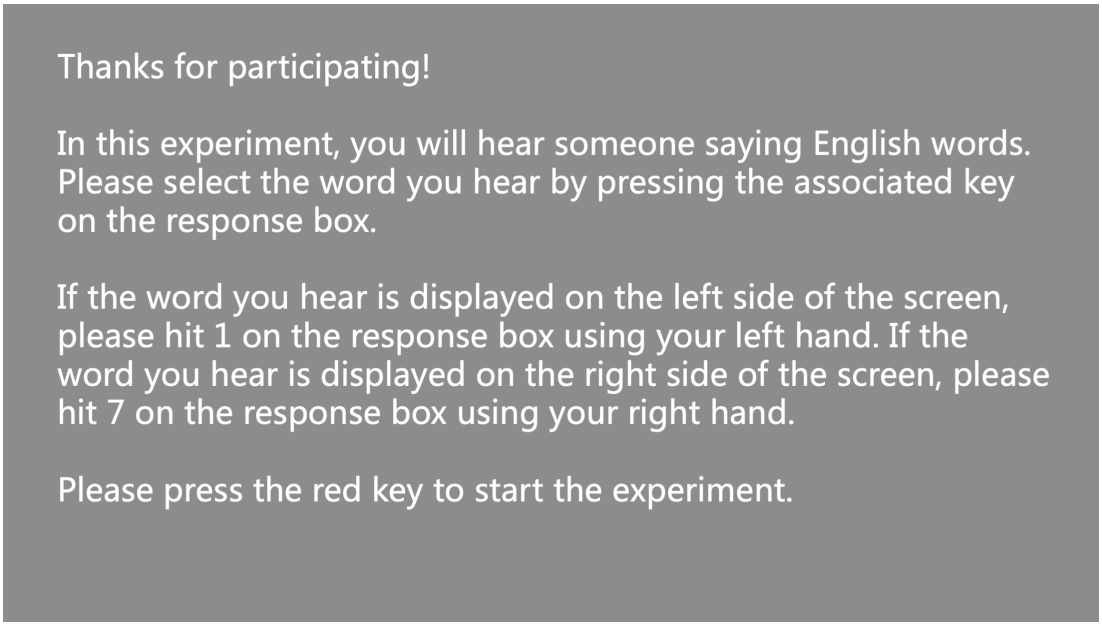


Appendix C2: English production experiment task

English production task interface.

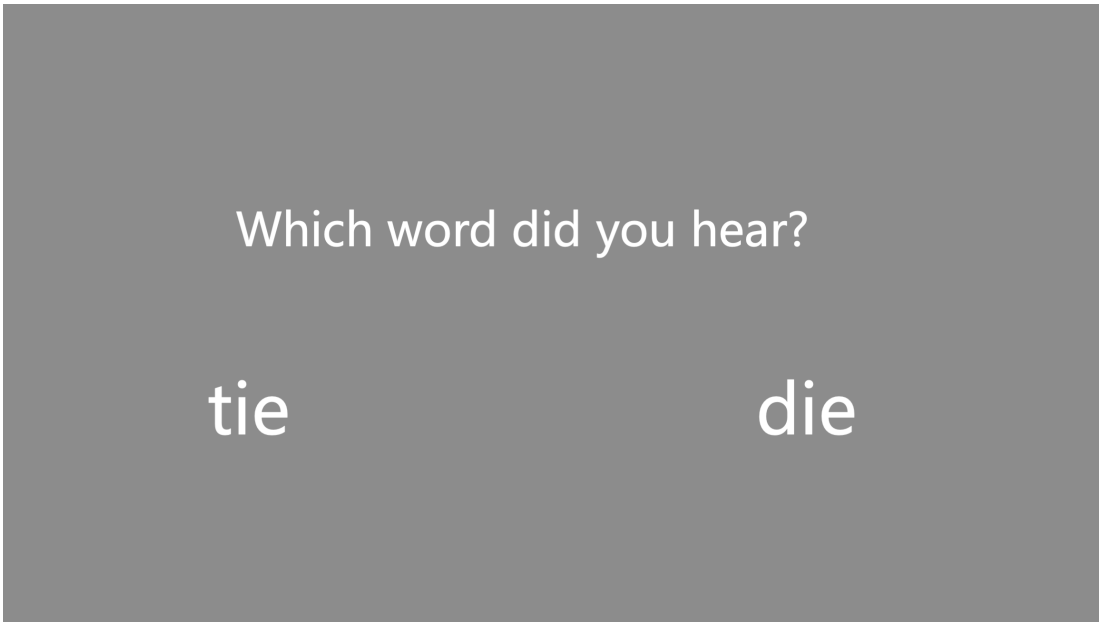
Appendix D1: English perception experiment instruction

English perception instruction interface.



Appendix D2: English perception experiment task

English perception task interface.



Appendix E: Demographical Questionnaire

1. Do you have or suffer from any hearing or speaking difficulties?

2. When and where were you born?

3. What is your gender?

4. What is your native language?

5. What are the dialects you speak?

6. What foreign languages do you speak?

7. When did you first start to learn your second language?

8. For how long have you been learning your second language?

9. How often do you expose to your second language?

10. Have you lived in a country where your second language is widely spoken? If yes, for how long?

Appendix F: Demographical Questionnaire summary

Gender	Birthplace	Age	AO	LOR (month)	English exposure
Female	Hennan	19	9	1	class
Male	Heilongjiang	19	8	1	class, communication
Female	Chongqing	21	5	1	class, communication
Male	Anhui	22	8	1	class
Male	Shandong	23	4	1	class
Female	Heilongjiang	24	10	1	class
Male	Xuzhou	22	7	2	class
Male	Shaanxi	22	7	2	class
Female	Anhui	25	10	2	class, communication
Male	Inner Mongolia	20	9	10	class, communication
Female	Shaanxi	20	4	12	class, communication
Female	Inner Mongolia	20	6	12	class, communication
Male	Tianjin	20	8	12	class
Female	Shandong	21	4	12	class
Male	Shandong	22	6	12	class
Female	Gansu	23	10	12	class, communication
Male	Shanxi	25	14	12	class
Female	Beijing	36	12	16	class
Female	Shaanxi	24	6	24	class, communication
Male	Henan	25	6	24	class
Female	Xinjiang	23	4	36	class, communication
Female	Liaoning	26	8	36	class, communication
Female	Beijing	31	10	48	class, communication
Female	Henan	25	7	24	class, communication, work
Female	Beijing	46	12	240	work, communication, family communication

AO: Age of onset

LOR: Length of residence

BIBLIOGRAPHY

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voicing timing. In V. A. Fromkin (Eds.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 25-33). New York, NY: Academic.

Alexander, J. A. (2010). The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems (Doctoral Dissertation). Northwestern University, Evanston, IL.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7*, <http://CRAN.Rproject.org/package1⁄4lme4>.

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, 255-278.

Beckman, J., Jessen, M., & Ringen, C. (2009). German fricatives: Coda devoicing or positional faithfulness? *Phonology*, 26(02), 231-268.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics, 18*(1), 37-49.

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer*, version 6.1.12, <http://www.praat.org/>.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34(3), 372-387.

Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique*, 45, 24-27.

Chao, Y. R. (1956). Tone, intonation, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. In M. Halle, H. Lunt, H. McLean, and C. V. Schooneveld (Ed.) *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday* (pp. 52-59). Mouton Press.

Chao, K. Y., & Chen, L. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing, 13*(2), 215-232.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27, 207-299.

Connell, B. (2002). Tone languages and the universality of intrinsic F0: evidence from Africa. *Journal of Phonetics*, 30, 101-129.

Deng, D., Feng, S., & Lu, S. (2006). The contrast on tone between Putonghua and Taiwan Mandarin. *Sheng Xue Xue Bao [Acta Acustica]*, 31(6), 536-541.

Deterding, D., & Nolan, F. (2007). Aspiration and voicing of Chinese and English plosives. In *Proceedings of the 16th International Congress of Phonetic Sciences, Universität des Saarlandes Saarbrücken, Germany*, 385-388.

Dmitrieva, O., Llanos F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49, 77-95.

Duanmu, S. (2007). Tone: Basic Properties. In *The phonology of standard Chinese* (pp 225). New York, NY: Oxford university press.

Ewan, W. G. (1976). Laryngeal Behavior in Speech. (Ph.D. dissertation). University of California, Berkeley.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.

Flege, J. E. & Eefting, W. (1987a). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics,*15, 67-83.

Flege, J. E. & Eefting, W. (1987b). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication,* 6, 186-202.

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

Francis, A., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant

voicing in English. *The Journal of the Acoustical Society of America*, 124, 1234.

Fu, Q. J., & Zeng, F. G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Journal of Speech, Language, and Hearing Research*, 5, 45-57.

Fujimura, O. (1971). Remarks on stop consonants: Synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jakobson, and E. Zwirner (Eds.), *Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen* (pp. 221-232). Copenhagen, Akademisk Forlag.

Gao, J., & Arai, T. (2018). F0 perturbation in a "pitch-accent" language, presented at *TAL2018, Sixth International Symposium on Tonal Aspects of Language*, Berlin, German, 2018.

Garding, E., Kratochvil, P., Svantesson, J. O., & Zhang, J. (1986). Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech*, 29(3), 281-293.

Gandour, J. T. (1974). Consonant types and tone in Siamese. *Journal of Phonetics,* 2, 337-350.

Gandour, J. T., & Harshman, R. A. (1978). Cross language differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*(1), 1-33.

Haggard, M., Ambler, S., & Callow, M. (1969). Pitch as a voicing cue. *The Journal of the Acoustical Society of America*, 47, 613-617.

Hallé, P. (1994). Evidence for tone-specific activity of the sternohyoid muscle in modern standard Chinese. *Language and Speech, 37,* 103-123.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly Progress Report*, Research Laboratory of Electronics, MIT, *101*, 198-213.

Han, M. S., and Weitzman, R. S. (1970) Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/. *Phonetica*, 22, 112-128.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, 125 (1), 425-441.

Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 28, 353-367.

Hockett, C. F. (1947). Peiping Phonology. *Journal of the American Oriental Society* 67,

253-267.
Hombert, J. M. (1975). Towards a theory of tonogenesis: an empirical, physiologically and perceptually based account of the development of tonal contrasts in languages (Doctoral Dissertation). University of California, Berkeley, CA.

Hombert, J. M. (1977). Development of tones from vowel height?. *Journal of Phonetics*, 5, 9-16.

Hombert, J. M. (1978). Consonant types, vowel quality, and tone. In V. A. Fromkin (Eds.), *Tone: A Linguistic Survey* (pp. 77-107). New York, NY: Academic.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 55, 37-58.

Honda, K., Hirai, H., Masaki, S., & Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech*, *42*(4), 401-411.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25, 105-113.

Howie, J. M. (1976). Acoustical Studies of Mandarin Vowels and Tones. Cambridge: Cambridge University Press, 1976.

Jeel, V. (1975). An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. *Technical Report No. 9*. Copenhagen, University of Copenhagen, Institute of Phonetics.

Jessen, M. (2001). Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. *Distinctive Feature Theory*, 2, 237.

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, 19(02), 189-218.

Jessen, M., & Roux, J. C. (2002). Voice quality differences associated with stops and clicks in Xhosa. *The Journal of Phonetics*, 30, 1-52.

Jia, G. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, 119, 1118.

Jun, S.A. (1996). Influence of microprosody on macroprosody: A case of phrase initial

strengthening. *Technical Report No. 92*, Los Angeles, CA: University of California at Los Angeles.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing, *Language*, 60(2), 286-319.

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language,* 70, 419-494.

Kirby, J., & Ladd, D., R. (2016). "Effects of obstruent voicing on vowel F0: evidence from "true voicing" languages, *Journal of the Acoustic Society of America*, 140 (4), pp. 2400-2411.

Kohler, K. J. (1982). F0 in the production of lenis and fortis plosives. *Phonetica*, 39, 199-218.

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica,* 41(3), 150-174.

Kuang, J. J. (2013). Phonation in tonal contrasts (Doctoral dissertation). University of California, Los Angeles.

Ladefoged, P. (1963). Some physiological parameters in speech. *Language & Speech* 6, 109-119.

Ladefoged, P. (1967). *Three Areas of Experimental Phonetics*. London: Oxford University Press.

Ladefoged, P. (1971). P*reliminaries to Linguistic Phonetics.* Chicago: Chicago University Press.

Ladefoged, P. (1975). Respiration, laryngeal activity and linguistics. In Wyke, B. (ed.), *Proceedings of the International Symposium on Ventilatory and Phonatory Control Systems*, 299-314. London: Oxford University Press.

Lai, Y., Huff, C., & 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*.

Lea, W. A. (1973). Segmental and suprasegmental influences on fundamental frequency contours. In L. M. Hyman (Eds.), *Consonant Types and Tones, Southern California Occasional Papers in Linguistics No. 1*, (pp. 15-70), Los Angeles, LA: University of Southern California Press.

Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in

Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association*, 42(2), 145–169.

Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, 41(2), 117-132.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America*, 33, 419-425.

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

Li, F. F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. *The Journal of the Acoustical Society of America*, 133 (2), 142-147.

Liang, J., & van Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes, Volume 2: Experimental studies in word and sentence prosody* (pp. 27-61). Berlin/NewYork: Mouton de Gruyter.

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

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, 29(1), 3-11.

Liu, L. Y., & Ma, Y. F. (1986). The distribution of Mandarin tones and the frequency of tonal phrases. Language Planning, 3, 21-23.

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

Löfqvist, A. (1975). Intrinsic and extrinsic F_0 variations in Swedish tonal accents. *Phonetica* 31, 228-247.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *The journal of the acoustical society of America*, *85*(3), 1314-1321.

Luo, Q. (2018). Consonantal effects on F0 in tonal languages (Doctoral dissertation).

Michigan State University, East Lansing.

Maddieson, I. (1996). Phonetic universals. *UCLA Working Papers in Phonetics*, 160-178.

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

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Attention, Perception, & Psychophysics*, 18(5), 331-340.

Moeller, J., & Fischer, J. F. (1904). Observation on the action of the cricothyroideus and thyroarytenoideus internus. *Anaals of Otology, Rhinology, and Laryngology*, 13, 42-46.

Moisik, S. R., Lin, H., & Esling, J. H. (2014). A study of laryngeal gestures in Mandarin citation tones using simultaneous laryngoscopy and laryngeal ultrasound (SLLUS). *Journal of the International Phonetic Association*, 44 (1), 21-58.

Mohr, B. (1971). Intrinsic variations in the speech signal. *Phonetica*, 23, 65-93.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9, 453-467.

Nakai, S., & Scobbie, J. M. (2016). The VOT Category Boundary in Word-Initial Stops: Counter-Evidence Against Rate Normalization in English Spontaneous Speech. Laboratory Phonology: Journal of the Association for Laboratory Phonology 7(1): 13, 1-31.

Ohala, M., & Ohala, J. J. (1972). The problem of aspiration in Hindi phonetics. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics* 6, 39-46.

Ohala, J. J. (1972). How is pitch lowered?. *The Journal of the Acoustical Society of America*, 52, 124-124.

Ohala, J. J. (1978). Production of tone. In V.A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 5-39). New York: Academic Press.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, 75, 224-230.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8-13.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, 693-703.

Reinholt Petersen, N. (1983). The effect of consonant type on fundamental frequency and larynx height in Danish. *Technical Report*, Copenhagen, University of Copenhagen, Institute of Phonetics.

Rochet, B. L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, 19(4), 105.

Sagart, L., Pierre, H., Benedicte, B. B., & Catherine, A. G. (1986). Tone production in Modern Standard Chinese: An electromygraphic investigation. *Cahiers de Linguistique Asie Orientale* 15, 205-221.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 52, 183-204.

Silverman, K. (1984). F0 perturbations as a function of voicing of prevofcalic and postvocalic stops and fricatives, and of syllable stress. In *Reproduced Sound: 1985 Autumn Conference, Windermere: Conference Handbook*. Vol. 6 (pp. 445-452). Great Britain, Cumbria, Windermere, Institute of Acoustics.

Simon, E. (2009). Acquiring a new second language contrast: An analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research,* 25, 377-408.

Shi, B. & Zhang, J. (1987) Vowel intrinsic pitch in standard Chinese. In Proceedings Xlth International Congress of Phonetic Sciences, 1 (pp. 142-145). Tallinn, Estonia: Academy of Sciences of the Estonian SSR.

Shi, F. (1998). The influence of aspiration on tones. *Journal of Chinese Linguist,* 26, 126-145.

Slis, I. H. (1970). Articulatory measurements on voiced, voiceless, and nasal consonants. *Phonetica* 21, 193-210.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of

fundamental frequency on stop consonant voicing judgments. *Phonetica*, 47, 36-49.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, 93, 2152-2159.

Whalen, D. H., & Levitt, A. G. (1995) The universality of intrinsic F0 of vowels, *Journal of Phonetics*, 23, 349-366.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49, 25-47.

Xiao, H. (2010). The construction and application of the general modern Chinese balanced corpus. *Chinese world*, 106.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61-83.

Xu, Y., & Liu, F. (2007). Determining the temporal interval of segments with the help of F0 contours. *Journal of Phonetics*, 35(3), 398-420.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *The Journal of the International Phonetic Association*, 33(2), 165-181.

Yang, Y. X., Chen, X. X., & Xiao, Q. (2020). L2 speech learning: Evidence from the acquisition of Russian stop contrasts by Mandarin speakers. *Second Language Research*, 1-27.

Yang, J., Zhang, Y., Li, A. J., & Xu, L. (2017). On the Duration of Mandarin Tones. *Interspeech,* 1407-1411.

Yu, V. Y., De Nil, L. F., & Pang, E. W. (2015). Effects of Age, Sex and Syllable Number on Voice Onset Time: Evidence from Children's Voiceless Aspirated Stops. *Language and speech*, *58*(2), 152-167.

Zee, E. (1980). The effect of aspiration on the fundamental frequency of the following vowel in Cantonese. *UCLA Working Papers in Phonetics*, 49, 90-97.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology, 27*(1), 153-201.

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time

characteristics of adults. *The journal of the acoustical society of America,* 56(3),981-994.

BIOGRAPHY

Yuting Guo graduated from Hefei University of Technology, Hefei, Anhui, China, in 2011 with a Bachelor's Degree in English. She received her first Master of Arts in foreign languages and applied linguistics from Hefei University of Technology in 2014. She earned her second Master of Arts in linguistics from Syracuse University in 2015. She was the recipient of the Full TA-ship from Syracuse University, where she taught Chinese to undergraduates for two years. She was the recipient of the 2015 GMU Presidential Scholarship and the 2020 Dissertation Completion Grant. She worked as a research assistant for four years at George Mason University and she taught English at INTO Mason for one year. She joined Amazon as a Language Data Researcher in the summer of 2020.