Perception-production relationship of lexical tones

by

Keith K. W. Leung

M.A., University of Hong Kong, 2010
PGDE, University of Hong Kong, 2005
B.SocSc., University of Hong Kong, 2004

Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

in the
Department of Linguistics
Faculty of Arts and Social Sciences

SIMON FRASER UNIVERSITY
Summer 2022

Declaration of Committee

Name:	**Keith K. W. Leung**
Degree:	**Doctor of Philosophy (Linguistics)**
Title:	**Perception-production relationship of lexical tones**
Committee:	**Chair:** Nancy Hedberg Professor, Linguistics

Yue Wang
Supervisor
Professor, Linguistics

Murray J. Munro
Committee Member
Professor, Linguistics

H. Henny Yeung
Committee Member
Associate Professor, Linguistics

Christian Guilbault
Examiner
Associate Professor, French

Chao-Yang Lee
External Examiner
Associate Professor, College of Health Sciences and Professions
Ohio University

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

The link between perception and production is predicted to be close, but empirical findings on this relationship are mixed. While a perception-production relationship has been found for various speech sounds, some research has failed to support such a link. To explain this apparent contradiction, a proposed view is that a perception-production relationship should be established through the use of critical perceptual cues. This dissertation project aims to examine this view by using Mandarin tones as a test case, since the perceptual cues for Mandarin tones consist of a perceptually critical pitch direction cue and a non-critical pitch height cue. As there was little research on the perception-production relationship of lexical tones based on acoustic cues, the first study explored the correlation between perception and production of Mandarin Tone 2 for each of five acoustic cues which included critical pitch direction-related cues, non-critical height-related cues and a temporal cue. A perception-production correlation was only found for the critical perceptual cues. The second study investigated the proposal systematically by examining the defining features of critical and non-critical perceptual cues and the perception-production relationship of each cue for each Mandarin tone. The perceptual stimuli in the perception experiment were created by varying one critical and one non-critical perceptual cue orthogonally. The cues for tones produced by the same group of native Mandarin participants were measured. This study found that the critical status of perceptual cues primarily influenced the within-category and between-category perception for all tones. Using cross-domain bi-directional statistical modelling, a perception-production link was found for the critical perceptual cue only. A stronger link was obtained when within-category and between-category perception data were included in the modelling, as compared to using between-category perception data alone, suggesting a phonetically and phonologically driven perception-production relationship. Finally, the third study examined if forming a perception-production relationship could be clearly attributed to the use of critical perceptual cues. Using the same critical and non-critical cues as in the second study, the learning effects on the perception and production of each cue were measured for Mandarin learners whose native language was Indonesian, a non-tonal language, in a four- to six-week interval. A simultaneous improvement in perception and production was found for the critical perceptual cue only, supporting the notion that the critical perception cue was a contributing factor driving the link between perception and production.

Keywords: speech perception, speech production, perception-production relationships, Mandarin tones

Acknowledgements

First and foremost, I would like to thank my senior supervisor, Dr. Yue Wang, for guiding me in formulating this research topic, dedicating her time to reading many drafts of this dissertation, and providing insightful feedback on my writing. She gave me the opportunity to take part in a number of research projects at the Language and Brain Lab, which helped sharpen my research skills and enabled me to collaborate with many great researchers. I acknowledge that this project was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery grant given to her, titled *Communicating pitch in clear speech: articulation, acoustics, intelligibility, neuro-processing, and computational modelling*.

I am grateful to my supervisors, Dr. Murray Munro and Dr. Henny Yeung, for their support throughout my doctoral study at SFU and their critical and thought-provoking feedback on earlier versions of this work. I want to thank my examiners, Dr. Chao-Yang Lee and Dr. Christian Guilbault, for their great interest in my work and valuable comments. I am also thankful to Dr. John Alderete and Dr. Paul Tupper for their great advice.

Running this project during the COVID-19 pandemic was by no means easy. I am indebted to Dr. Yu-An Lu and Shao-Jie Jin who helped me perform the third study of this dissertation (Chapter 4) at National Yang Ming Chiao Tung University in Taiwan. Knowing that I was not able to go to Taiwan in person due to pandemic travel restrictions, Dr. Yu-An Lu kindly agreed to recruit participants and collect data for me using her research facilities. Without her and Shao-Jie's help, I would not have been able to complete this project.

Friends have also played a significant part in my Ph.D. journey. I thank the current and past members of the Language and Brain Lab and my fellow graduate students, especially Dr. Daniel Chang, Sylvia Cho, Xizi Deng, Dr. Kyeong-min Kim, Bingqing Yu and Beverly

Hannah, and our visiting scholar, Dr. Makiko Aoyagi, for their kindness and advice. I would like to mention Jason Or, Kit Ling Sze, Rev. Angus Wu and Mazy Ng for their practical and spiritual support. I want to extend my sincerest appreciation to Dr. Peggy Mok. Without her important advice, I would not have applied to SFU Linguistics. I am deeply grateful to Dr. Vincent Ooi - an excellent friend. Without his encouragement, pursuing a Ph.D. would not have become a reality for me. I thank him for the daily conversations that have made this long journey much more amusing and enjoyable.

My sincerest gratitude goes to my family members. I thank my parents for their unwavering support in my entire life. I also thank my uncles and aunts in Vancouver, especially my third uncle's family and third aunt (三姑姐、三叔、三嬸), for their great help with my transition from Hong Kong to Vancouver. Last but not least, I am greatly indebted to my wife, Tung Tung (冬冬), for her unceasing encouragement, trust, love, understanding and sacrifice. Our two lovely daughters, Yuk-Ching (郁澄) and On-Ching (安澄), joined our family during my Ph.D study. This project could have been completed much earlier without them, but they have brought so much joy and fun to us and they have made this long journey much more happy, exciting and interesting. To my wife and daughters, I dedicate this dissertation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The perception and production of speech sounds should undoubtedly be closely related based on several theories. The two domains are either posited to be governed by a unitary system in the gestural theories of speech perception (e.g. Fowler, 1986; Fowler et al., 2003; Galantucci et al., 2006; Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000) or highly synchronized under auditory approaches to speech perception (Blumstein and Stevens, 1979; Diehl and Kluender, 1989; Diehl et al., 2004; Kuhl et al., 2008) and through the feedforward and feedback system in speech production acqusition (Guenther, 1994, 1995, 2015). However, empirical findings have demonstrated mixed results. While some studies revealed evidence of a perception-production link (e.g. Beddor, 2015; Bell-Berti et al., 1979; Flege, 1995, 1999, 2003; Flege et al., 1997, 1999; Flege and Schmidt, 1995; Fox, 1982; Franken et al., 2017; Ghosh et al., 2010; Kato and Baese-Berk, 2020; Kirby and Giang, 2021; McAllister Byun and Tiede, 2017; Newman, 2003; Perkell et al., 2004a,b; Yang and Whalen, 2015; Yu et al., 2021; Zellou, 2017), others did not find a relationship between the two domains (Ainsworth and Paliwal, 1984; Bailey and Haggard, 1973; Beddor, 2015; Frieda et al., 2000; Paliwal et al., 1983; Schertz et al., 2015; Shultz et al., 2012; Yu et al., 2021; Zellou, 2017).

This project aims at exploring this issue by examining the underlying mechanism of perception-production relationships so as to understand the circumstances that can establish a perception-production link. A possibility explored in this project is the critical nature of perceptual cues (Newman, 2003; Shultz et al., 2012). Mandarin tones are used as an ideal test case because of the differential perceptual cue weightings given to pitch direction and

height which can be used to define the critical status of perceptual cues (Chandrasekaran et al., 2010; Gandour, 1983; Guion and Pederson, 2007; Francis et al., 2008).

Many theories and findings are segmentally-based and the prior work on lexical tones do not clearly show the acoustic cues that relate production to perception for Mandarin tones. Therefore, the first goal of this project aims to extend the study of perception-production links to lexical tones using Mandarin and explore the acoustic cues used for establishing this relationship (Chapter 2). Moreover, given that Mandarin tone perception displays differential cue weights given to pitch height and direction, the second goal of this project determines the critical status of Mandarin tone cues based on between and within tone category perception, and subsequently establish a perception-production relationship using these cues (Chapter 3). Finally, the third goal is to investigate whether the perception-production link is attributable to the critical status of perceptual cue by comparing the perception and production learning patterns of beginner-level learners of Mandarin (Chapter 4).

This introduction first reviews the theoretical perspectives on perception-production links (Section 1.1). Then, the empirical findings on the relationship between the two domains are presented (Section 1.2). Next, the cues pertaining to Mandarin tone perception and production are reviewed (Section 1.3). The final section of this chapter presents the rationales and structure of this dissertation project (Section 1.4).

1.1 Theoretical perspectives on perception-production links

1.1.1 Gestural theories of speech perception

To resolve the lack of acoustic invariance in speech perception, the gestural theories argue that perception is the process of perceiving articulatory gestures. The motor theory of speech perception, developed by Liberman and his colleagues, first proposed that human beings can decode incoming phonemes and refer to them as invariant neuromotor commands to the articulatory muscles (Liberman et al., 1967). Later on, the theory was revised to differentiate between observable articulatory movements and intended phonetic gestures. When listeners perceive incoming phonemes, they are able to recover the in-

tended gestures from the systematically varying transitions in the signals (Liberman and Mattingly, 1985; Liberman and Whalen, 2000). Similar to the motor theory, the direct-realist theory also considers speech perception as perceiving articulatory gestures. However, the direct-realist perspective suggests no distinction between observable movements and intended gestures. The detection of the observable articulatory movements itself provides the necessary information for speech perception (Fowler, 1986). Research showed that listeners can rapidly retrieve gestural information in perception, providing support for the direct-realist theory (Fowler et al., 2003). Regardless of the difference between the two gestural theories, the assumption that perceiving speech is perceiving gesture makes strong predictions for perception-production links (Galantucci et al., 2006). Because the perceptual targets and the motor commands for articulation are essentially controlled a unitary system, perception and production should be tightly linked. In particular, the motor theory of speech perception suggests that speech gestures and the perception-production link are innately specified (Liberman and Mattingly, 1985). The link between perception and production should be tight and remain unchanged from birth.

1.1.2 General auditory approaches to speech perception

The general auditory approaches to speech perception argue that the object of perception consists of acoustic signals instead of articulatory gestures because invariance does exist in the acoustic signals and it happens when certain distinctive features are extracted (Blumstein and Stevens, 1979; Diehl and Kluender, 1989; Diehl et al., 2004). For example, the measurement of gross spectral shapes sampled at onsets and offsets of consonant-vowel and vowel-consonant syllables demonstrated invariant acoustic characteristics of stop consonants independent of vowel contexts (Blumstein and Stevens, 1979). It is, therefore, possible that the perceptual system functions by extracting these spectral shape features in speech perception, involving no articulatory information in the process. Moreover, even though certain acoustic signals of speech sounds may not be invariant, it has been found that listeners can factor out contextual factors including phonetic contexts and interspeaker differences and utilize the compensated acoustic cues in speech perception (Jongman and

McMurray, 2017; McMurray and Jongman, 2011). Concerning the perception-production link, the general auditory account provides a completely different explanation of the relationship from the gestural theories. Since perception is an acoustic process without involving articulatory gestures, perception and production are separate systems coupled through learning. Infants first gain linguistic perceptual experience from their parents and these acoustic instances are stored in their memory. Then, they hear the auditory input coming from their own articulatory movements during vocal play, which enables them to learn acoustic consequences of their articulatory gestures. At the same time, their previous perceptual experience with language guides production by matching the perceptual instances in their memory and their imitations. A link between perception and production is then forged through this mapping process (e.g. Kuhl et al., 2008). As development continues, infants or listeners keep mapping regularities of speech production onto the corresponding acoustic consequences which are then used for speech sound perception, including coarticulation patterns. In order to produce auditorily distinctive targets, they also acquire the articulatory movements that can ensure maximize interphonemic distances (Diehl et al., 2004). As a result, although the general auditory approaches hold a different assumption about the object of perception, they also predict a close link between perception and production.

1.1.3 Directions Into Velocities of Articulator (DIVA) model

Similar to the general auditory approaches, the DIVA model has developed its speech production framework on a model of speech acquisition (Guenther, 1994, 1995, 2015). Speech production involves a feedforward and a feedback mechanism. The feedforward mechanism sends motor command signals to specify the articulatory configurations for producing speech sounds. The motor commands also generate efferent copies of the auditory (and somatosensory) targets. During speech acquisition (or speech production in general), the acoustic signals generated by articulatory movements are fed back and compared with the efferent copies. If there is a mismatch between acoustic signals and the

efferent copies, the motor commands are updated, leading to a coupling of the production and perception processes.

Furthermore, the model assumes the competing constraints of a listener's need for clarity and a speaker's motivation for economizing speaking effort in the planning of speech production (Guenther, 1995; McAllister Byun and Tiede, 2017; Perkell et al., 2004a). Human beings demonstrate varying degrees of perceptual acuity, and therefore demands different levels of clarity. Since an individual constantly monitors their own speech production through the feedback mechanisms, the difference in perceptual acuity will influence the between-category contrasts and within-category variability of the individual's speech production. If an individual displays a greater perceptually sensitivity to the acoustic contrasts of speech sounds, the production of these speech sounds should exhibit greater between-category contrasts and less within-category variability compared with an individual with less perceptual sensitivity.

1.1.4 Imperfect alignment between the perception and production systems

The above-mentioned theories should predict that there is a close link between perception and production. However, this close link does not necessarily mean that the two domains are perfectly aligned. In fact, Nearey (1992) proposes that perception and production are "less-than-perfect inverses of each other" (p.153) in the double-weak theory of speech perception which accepts the weak form of both a gestural theory of *speech perception* and an auditory theory of *speech production*. It treats the objects of perception as abstract elements that map onto *both* articulatory gestures and acoustic signals. More importantly, not only does it consider perception and production as two autonomous subsystems (a view in common with the general auditory approaches), it also argues that the two subsystems are not perfectly aligned due to the fact that there is always a compromise between production efficiency and rapid perceptual decoding. Two highly related ideas are perceptual errors due to hypo- and hypercorrection (Ohala, 1989) and production reductions caused by hypo- and hyperarticulation (Lindblom, 1990). In the perception of coarticulated speech, listeners factor out contextual effects, but this correction process can lead to errors, resulting in

either hypocorrection where listeners fail to completely factor out the contextual effects, or hypercorrection where overcompensation occurs (Ohala, 1989). The idea of reduction in production comes from the hypo- and hyper theory (Lindblom, 1990) which differentiates between input goal and production output. The former is always in the form of most distinctive, hyperarticulated speech. Depending on spoken context, the latter can be produced by minimizing efforts, realized in a form with phonetic reductions and thus fallen short of the input goal. A mismatch in perception and production occurs as a consequence of these factors. Therefore, although speech categories overlap in a large extent in the perception and production systems, imperfect alignment can exist between the two systems.

1.1.5 Summary

From a theoretical perspective, perception and production are closely connected although theories may posit different and sometimes conflicting object of speech perception and conceptualization of the perception and production systems. As a consequence of their different assumptions, there are differences in understanding how perception-production relationship is developed, which leads to different predictions about the strength of the link. On the one hand, gestural theories of speech perception predict a perfect and tight link between perception and production since the two processes are governed by the same motor command system (Fowler, 1986; Fowler et al., 2003; Galantucci et al., 2006; Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000). On the other hand, although auditory approaches to speech perception (Blumstein and Stevens, 1979; Diehl and Kluender, 1989; Diehl et al., 2004; Kuhl et al., 2008) and the DIVA model (Guenther, 1994, 1995, 2015) do not explicitly postulate the strength of a perception-production link, it is expected that the two domains should be closely linked through extensive exposure to a language (Diehl et al., 2004). However, since these theories view perception and production as two autonomous subsystems, it leaves the possibility that the two domains are not perfectly aligned as suggested in the double-weak theory (Nearey, 1992).

1.2 Empirical studies on perception-production links

1.2.1 Empirical evidence for perception-production links

Evidence from first language research

Previous studies found evidence for a perception-production link in plosives (Beddor, 2015; Newman, 2003), fricatives (Ghosh et al., 2010; Newman, 2003; Perkell et al., 2004b), sonorant consonants (McAllister Byun and Tiede, 2017), vowels (Bell-Berti et al., 1979; Fox, 1982; Franken et al., 2017; Zellou, 2017) and tones (Yang, 2015). In an early study, Bell-Berti et al. (1979) studied the production of English front tense and lax vowels /i, ɪ, e, ɛ/. They found that native English speakers demonstrated two different vowel orders: /i, ɪ, e, ɛ/ or /i, e, ɪ, ɛ/, due to individual differences in tongue movements. Among the two groups of participants, one group showed a greater shift of categorical boundary location in the perception of the /i-ɪ/ continuum when /i/-like stimuli were presented more often than /ɪ/-like stimuli, showing that individual variations in production had a impact on perception.

Some studies demonstrated a perception-production relationship through a correlation or regression analysis of perceptual responses and productions of the same acoustic cue or related dimensions. A statistically significant, strong positive correlation or regression result between acoustic cues used in perception and production provides the evidence for a link between perception and production. In a study on tonogenesis in Afrikaans (i.e. the voicing of stops is differentiated by fundamental frequency (F0) instead of voice onset time (VOT)) (Beddor, 2015), the amount of prevoicing in the production of voiced stops /b/ was correlated with the rate of identifying a prevoiced stop with high F0 as voiced stop /b/, and the study detected a positive, moderate correlation between the two measures (r = 0.4). It showed that the native Afrikaans individual who displayed tonogensis in their voiced /b/ and voiceless stop /p/ productions tended to identify a prevoiced stop with high F0 in the following vowel as voiceless /p/ instead of voiced /b/. In another study, Newman (2003) explored the relationship between perceptual prototypes and production norms. The VOT values that native English individuals perceived to be typical of /p/ were found to be moderately correlated with the mean VOT values of /p/ produced by the same group of participants (r = 0.522). A similar design was used to compare the perception and pro-

duction of English fricatives /s/ and /ʃ/ in the same study by varying peak frequency, and a significant moderate perception-production correlation was found (r = 0.5). For vowels, a perception and production link was found between perceptual cue weightings and formant values. Based on stepwise multiple regressions, a perception-production relationship was shown by the finding that, for native English individuals, the perceptual dimensions (vowel height, backness and roundedness) were predicted by the formant(s) of the vowels /i, a, u/ that are acoustic correlates of those dimensions (Fox, 1982). In addition, the production of nasal coarticulation in English vowels and patterns of perceptual compensation were found to be correlated in a mixed-effects regression model. Specifically, native English individuals who produced more vowel nasality demonstrated more perceptual compensation in the discrimination of nasalized and oral vowels. For instance, individuals who nasalized their vowel in a vowel-nasal consonant context (i.e. nasal coarticulation) would be more likely to perceive [e] and [ẽ] as the same vowel than individuals who did not produce nasal coarticulation (Zellou, 2017).

Other studies explored the relationship between perceptual acuity and the distinctiveness of production categories, which supported the prediction of the DIVA model. For instance, native English individuals showed a positive correlation between perceptual acuity in the discrimination of the intermediate steps in a continuum between /s/ and /ʃ/ and the acoustic distance between the two fricatives. The perceptual acuity was either measured by discrimination scores (r = 0.63) (Perkell et al., 2004b) or just noticeable difference (JND) ($r = -0.41$) (Ghosh et al., 2010), both indicating that individuals who were more sensitive to small acoustic differences between the two fricatives could produce the two categories with greater acoustic distance. Similarly, in the native perception and production of English vowels /ɑ vs. ʌ/ and /u vs. ʊ/, the "high" discriminators who performed with 100% accuracy in a vowel discrimination task consistently produced greater contrast distances than "low" discriminators who did not perform in the same discrimination task with 100% accuracy (Perkell et al., 2004a). In addition, the perceptual acuity in the discrimination of vowel continua of the Dutch vowel pairs /ɛ-ɪ/ and /ɑ-ɔ/ were predicted by both between-category vowel distance and within-category vowel dispersion in production in a linear regression

analysis. Specifically, better discrimination between intermediate vowel stimuli was associated with more distinctive (between-category) and more precise (within-category) vowel productions (Franken et al., 2017). This study further revealed that perceptual acuity was not only related to between-category acoustic distance, but also within-category variability. Finally, in a study of the development of American English rhotics, the categorical boundary width between /ɹ/ and /w/ was a predictor of the production distance between the two consonants for female native English individuals between the age of 9 and 14, indicating that the perception-production relationship was found among children as well as adults (McAllister Byun and Tiede, 2017). These studies, therefore, evince the prediction of the DIVA model that individual variations in perceptual acuity, or the need for clarity in perception, are related to the between-category distinctiveness and within-category precision of speech productions, presumably driven by the feedback mechanism in speech production during language acquisition.

Lastly, some studies qualitatively compared the distribution of speech categories on perceptual and production space to determine the extent of perception-production relationships. For example, the native Mandarin participants of Yang (2015) categorized resynthesized linear tone contours with varying F0 onset and offset in one of the four Mandarin tones (level, rising, dipping and falling). Their tone productions obtained in connected speech were also measured for F0 onset and offset. It was found that tone boundaries on the perceptual and production tone space defined by the same acoustic variables (F0 onset, offset) were closely aligned, suggesting a close relationship between Mandarin tone perception and production. However, the alignment of category boundaries was not perfect since the boundaries were not found on the exact locations across the perceptual and production tone space. In fact, Nearey (1992) showed that the native English sibilant perception and production space also presented imperfect alignment of category boundary location in a study of four syllables, /si/, /su/, /ʃi/ and /ʃu/. Based on this finding, the double-weak theory of speech perception predicts a less-than-perfect link between perception and production as the two domains semi-autonomous systems (Nearey, 1992). Similarly, for native Englsih vowels, the F1 $\times$ F2 perceptual vowel space formed by the participant-defined best

vowel exemplars had a size and shape comparable to the production vowel space of participants' hyperarticulated vowel productions. Furthermore, the production vowel space size was reduced when participants produced the vowels naturally, and therefore showed less perception-production alignment (Johnson et al., 1993). This phenomenon was described as the hyperspace effect of perception – the perceptual targets are stored in hyperarticulated forms and production demonstrates variability depending on speaker's production effort, corresponding to hypo- vs. hyperarticulation (Lindblom, 1990; Johnson et al., 1993). In addition, a more recent finding showed gender difference. Female perception and hyperarticulated production vowel spaces displayed a good match, but male participants showed a more obvious hyperspace effect than female speakers (Yang and Whalen, 2015).

Evidence from second language research

Apart from the above-mentioned studies which examined the perception and production of participants' native language, the link between perception and production can be demonstrated in second language (L2) acquisition research, including the Speech Learning Model (SLM) (Flege, 1995, 1999, 2003). As in some of the first language research reviewed above (Beddor, 2015; Fox, 1982; Newman, 2003), previous studies also made used of a correlation or regression analysis to investigate a perception-production relationship for consonants (Flege and Schmidt, 1995; Kato and Baese-Berk, 2020), vowels (Flege, 1993; Flege et al., 1997, 1999) and tones (Kirby and Giang, 2021; Yu et al., 2021). To demonstrate a correlation between the perception and production measurements, the perception measurement was based on the discrimination or category identification of speech items. Some studies correlated the perceptual measure with the native listeners' judgement in terms of degree of foreign accent (r = 0.424 – 0.509) (Flege, 1988, 1993) or intelligibility scores (r = 0.4 – 0.64) (Flege et al., 1999), indicating that L2 individuals who could better identify and discriminate of L2 sounds were also able to produce those sounds with more native-like accent and higher intelligibility. Other studies obtained the acoustic measurement of the target speech sound. For example, Spanish learners of English showed a positive correlation between the goodness rating difference between English plosives

produced in fast and slow speech rate and the VOT of their English aspirated plosives (r = 0.361 – 0.392), suggesting that a stronger speech rate effect on the perceived quality of English plosives was related to a longer VOT of produced English plosives by these learners (Flege and Schmidt, 1995; Schmidt and Flege, 1995). For sonorant consonants, it was found that Japanese learners of English who were more accurate in the identification of English /ɹ/ and /l/ also displayed a greater difference in F3 between the productions of these sounds (r = 0.54 – 0.77) (Kato and Baese-Berk, 2020). As for the perception and production of duration, English learners of Japanese who showed a better identification of Japanese singleton and germinate consonants also produced these consonants with a greater duration contrast (r = 0.44 – 0.49) (Kato and Baese-Berk, 2020). For the duration of English vowel preceding a voiced and voiceless consonant, Mandarin learners of English who produced a greater vowel duration difference between the two segmental contexts also showed greater duration difference between the best perceived exemplars of these vowels in each of the two contexts (r = 0.445) (Flege, 1993). The research on lexical tone perception-production relationships showed that tone discrimination and acoustic distance between tone contours are related, suggesting that learners of a tone language who could better discriminate two tone categories also produced the tones with greater acoustic contrast, as demonstrated by a positive correlation for L2 Southern Vietnamese (ρ = 0.3) (Kirby and Giang, 2021) and a regression analysis for L2 Cantonese (Yu et al., 2021). For regression analysis, one approach was to compare the variance explained by the predictor variables. The criterion variables consisted of production data from different vowel pairs whereas the predictor variables involved perception data from one vowel pair only. A perception-production relationship was demonstrated since more variance in the perception in the production data was accounted for by the perception data related to the same English vowel pair than a different English vowel pair for German, Spanish, Korean and Mandarin learners of English (Flege et al., 1997).

It should be noted that, although the perception and production of native languages are predicted to be closely linked according to theories (Blumstein and Stevens, 1979; Diehl and Kluender, 1989; Diehl et al., 2004; Fowler, 1986; Fowler et al., 2003; Galantucci et al.,

2006; Guenther, 1994, 1995, 2015; Kuhl et al., 2008; Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000), the L2 perception-production relationship can exhibit a weaker correlation than that of native languages. The SLM specifically predicted the correlation to be modest (Flege, 1995, 1999, 2003). Since some studies have found evidence that L2 production learning was constrained by L2 perception learning (Flege and Schmidt, 1995; Flege et al., 1999), the SLM postulates that L2 production accuracy is always limited by perception accuracy, and it is possible that not all aspects of perceptual learning is incorporated in production (Flege, 1999). In contrast, previous research showed that production learning could be faster than perception learning (Yang, 2012). In addition, the strength of the L2 perception-production correlation can also vary depending on the elicitation method of L2 productions. For instance, Japanese learners of English showed a weaker correlation between the perception and production of the English /ɹ/ and /l/ when the consonants were produced in a reading task prompted by orthography than in a repetition task after hearing an auditory prompt. It was due to the greater variability in production accuracy and the general better production accuracy in the orthography prompt than the auditory prompt condition (Kato and Baese-Berk, 2020). Taken together, the prediction that the perception-production relationship for L2 is not as close as that for native languages shows the dissociation between the two domains and indicates that the two domains may be semi-autonomous.

Transfer of training effect across perception and production

The aforementioned findings of the perception-production relationship lead to the implication that perception and production are linked when a learning effect is transferred across the two domains for individuals learning a new language. Indeed, perceptual training studies conducted in a laboratory setting inform researchers about this link between perception and production as the effect of perception improvement resulting from perceptual training transfers to production. Previous studies showed that non-native speakers receiving training in the perception domain only on English consonants (Bradlow et al., 1997; Hardison, 2003; Hazan et al., 2005; Herd et al., 2013) or Mandarin tones (Wang et al., 1999, 2003)

displayed improvement in both non-native perception and production. The perceptual training on English vowels produced mixed results. Some studies yielded both perception and production improvement (Iverson et al., 2012; Lambacher et al., 2005). Brosseau-Lapré et al. (2013) found increased acoustic distance between vowel productions after perceptual training but production improvement as determined by native listener's judgement was not found. In a meta-analysis of the effect of perceptual training on production, perception-only training in general improved production and vowels had a smaller effect size of production improvement than consonants (Sakai and Moorman, 2018). The transfer suggests that the modification of perceptual representations is sufficient for the modification of motor commands, and therefore proves that the two domains are linked (although this link for vowels is less clear). Furthermore, training results indicate that perception and production are semi-autonomous domains. Research showed a lack of correlation between the degrees of perceptual and production improvements, suggesting an incomplete perception-to-production transfer (Bradlow et al., 1997; Sakai and Moorman, 2018). Sakai and Moorman (2018) only demonstrated a small and non significant relationship between perception and production gains. Bradlow et al. (1997) found that participants had high individual variability in the rate of improvement in the two domains, with perception demonstrating more rapid improvement than production for most individuals. As a result, the perception-to-production transfer does not always occur effortlessly, showing a semi-autonomous nature of perception-production relationships.

Various studies have examined the possibility of an opposite direction of transfer, i.e. from production training to perception (e.g. Adank et al., 2010; Herd et al., 2013; Hirata, 2004; Kartushina et al., 2015; Leather, 1997; Wang, 2013) and found a production-to-perception transfer for some trainees (Hirata, 2004; Kartushina et al., 2015). A criticism of these studies is that implicit perception component in the production training often occurs when participants hear their own speech, so that perception improvement may not be due to production training only. Therefore, some studies have added noise during the production learning task (Adank et al., 2010) or compared the results of a perception-only training with a perception-plus-production training (Baese-Berk, 2019; Baese-Berk and Samuel,

2016, 2022; Herd et al., 2013; Lu et al., 2015; Wang, 2013). However, some of these studies did not find any production-to-perception transfer (Adank et al., 2010; Herd et al., 2013; Lu et al., 2015; Wang, 2013), whereas others showed a disruption of perceptual learning with the perception-plus-production training (Baese-Berk and Samuel, 2016; Baese-Berk, 2019; Baese-Berk and Samuel, 2022; Leach and Samuel, 2007). Specifically, the perceptual learning was completely disrupted when the production took place before the perceptual response (Baese-Berk and Samuel, 2016; Baese-Berk, 2019; Baese-Berk and Samuel, 2022). The disruption was alleviated when the participants had learning experience of the target language or the production items were unrelated to the perceptual learning materials (Baese-Berk and Samuel, 2016). The timing of the production task also had an impact on the perceptual learning outcome, as the immediate verbal repetition of the auditory item containing the sound that was to be learned disrupted perceptual learning, but the perceptual learning improved as the delaying time of the verbal repetition increased. Moreover, this delay effect was not found for production items unrelated to the sounds to be learned (Baese-Berk and Samuel, 2022). Therefore, the disruption of perceptual learning suggests that the addition of production learning in the perception-plus-production task can lead to increased cognitive load due to distraction and task switching. There also exists competition between the exemplars learned from production and perception during training (Baese-Berk and Samuel, 2016; Baese-Berk, 2019; Baese-Berk and Samuel, 2022). Taken together, these results show that the production-to-perception transfer also does not always occur (Adank et al., 2010; Herd et al., 2013; Lu et al., 2015; Wang, 2013). The perception and production domains may also display competitions when the exemplars of a new sound category are being stored (Baese-Berk and Samuel, 2016; Baese-Berk, 2019; Baese-Berk and Samuel, 2022). All of these confirm the semi-autonomous nature of the perception and production domains.

Summary

In summary, the studies reviewed above found a link between perception and production for native and non-native individuals. However, note that some theories reviewed in section

1.1, especially the gestural theories, should predict a tight link of perception-production relationships since speech perception and production involve a unitary system. The empirical evidence is not entirely consistent with this particular theoretical prediction, since previous findings showed a moderate correlation between perception and production measures (Beddor, 2015; Ghosh et al., 2010; Newman, 2003; Perkell et al., 2004b), close but imperfect alignment of category boundaries in perception and production (Johnson et al., 1993; Nearey, 1992; Yang, 2015; Yang and Whalen, 2015), an incomplete transfer of training effect across domains (Adank et al., 2010; Bradlow et al., 1997; Herd et al., 2013; Lu et al., 2015; Sakai and Moorman, 2018; Wang, 2013), and a disruption of perceptual learning by production training (Baese-Berk and Samuel, 2016; Baese-Berk, 2019; Baese-Berk and Samuel, 2022; Leach and Samuel, 2007). These results all indicate that a perception-production link is close but the two domains may be semi-autonomous.

1.2.2 Lack of evidence for perception-production links

Although a perception-production link is predicted by many theories, previous studies did not consistently find a relationship between the two domains. The studies revealed non-significant findings involved plosives (Bailey and Haggard, 1973; Schertz et al., 2015; Shultz et al., 2012), glides (Ainsworth and Paliwal, 1984), vowels (Beddor, 2015; Frieda et al., 2000; Paliwal et al., 1983; Zellou, 2017) and tones (Yu et al., 2021). Given that there has been evidence showing a perception-production link as reviewed in the above section, it is less likely that these "non-evidence" rule out the prediction of perception-production links by the theories and suggest a lack of a relationship between perception and production. Rather, a careful review of these "non-evidence" is needed to consider the issues that might have diminished the studies' capability of finding a perception-production relationship.

Some studies contained possible methodological issues including the step size of resynthesized stimulus continuum, the data range covered by the perceptual stimuli and the sensitivity of the perception task. For instance, although stop perception and production was found to be related on the VOT dimension in later studies (Beddor, 2015; Newman, 2003),

it was not found in an earlier study presumably due to the large step size used in the VOT continuum of the perceptual stimuli (Bailey and Haggard, 1973). Moreover, in the same study, the data range of perceptual stimuli did not cover the usual VOT values of aspirated stops. Similarly, although a perception-production relationship was not found for English vowels, the formant values of the production stimuli were found to fall outside the formant values of the perceptual space (Frieda et al., 2000). Therefore, it is possible that the perception task was not able to detect the most ideal perceptual target for the participants due to the lack of acoustic details or data coverage in the perceptual stimuli, and consequently failed to relate perception to production. On the other hand, the perception task may not be sufficiently sensitive to discern individual variation, and therefore failed to establish a perception-production relationship. For instance, individuals were highly accurate in the perception task and therefore lacked individual variation (Yu et al., 2021). Likewise, the perceptual accuracy of vowel nasality did not differ as a function of nasal coarticulation in English vowel production. However, the perceptual compensation as measured by a discrimination task was able to elicit the individual variation in nasal coarticulation perception, and subsequently revealing a perception-production relationship (Zellou, 2017).

Another possible issue is due to the statistical method used to analyze the correlation between perception and production. The widely-used Pearson's correlation coefficient for bivariate correlation is obtained by the ratio of covariance and variance of two variables (Cohen and Cohen, 2003). However, in two earlier studies of English glides and vowels, the sum of products and sum of squares of the perception and production raw data to calculate the ratio (Ainsworth and Paliwal, 1984; Paliwal et al., 1983). As a result, it becomes hard to compare their results with other studies. In addition, a different approach to calculate the coefficient suggests that the results may be interpreted differently (i.e. a coefficient approaching zero may not refer to weak correlation). Therefore, one should interpret these results with caution.

More importantly, the "non-evidence" possibly revealed conceptual issues pertaining to perception-production relationships. First, the studies correlating the perception-production cue weights so far yielded non-significant results. In the study of the perception-production

correlation of plosive acoustic cues, none of the acoustic cue weights revealed a significant correlation, including weights given to F0 and VOT for English stop voicing contrast (Shultz et al., 2012) and F0, VOT and closure duration for Korean three-way stop contrast (fortis, lenis, aspirated) (Schertz et al., 2015). In contrast, the studies showing a perception-production link reviewed in the previous section involved the acoustic measurements of sound categories or the between-category acoustic distance. Therefore, the two domains may not be related by cue weights alone.

Second, a few studies suggest that, regarding the role of different speech cues in determining perception-production relationships, such links exist in the cues that are perceptually critical rather than those that are not critical. In Newman (2003), three acoustic cues were systematically varied to create the synthetic speech continuum: frication centroid, skewness, and peak frequency. The study found a significant perception-production correlation for peak frequency but not for centroid or skewness, presumably because only peak frequency was perceptually critical for native English perception of fricatives. In this study, the relative importance of peak frequency was inferred from the goodness rating results, but the critical status of these three cues was not determined independently in terms of the extent to which fricatives could be identified based on each of these cues. The relative importance of peak frequency was also supported by the acoustic modelling results that spectral peaks contributed more to fricative classification than centroid and skewness (Jongman et al., 2000). In another study, Shultz et al. (2012) examined the English stop perception-production relationship by correlating the cue weights given to the production and perception of VOT, the primary cue of stop perception (Lisker and Abramson, 1964), and onset fundamental frequency (onset F0), the secondary cue (Whalen et al., 1993). Although the two cues did not reveal a statistically significant perception-production correlation, VOT displayed a trend of positive perception-production correlation, whereas onset F0 lacks such a correlation. Similarly, a trend of positive correlation between the cue weights given to F0 in the perception and production of Korean fortis and lenis stop contrast. Among VOT, F0 and closure duration, F0 had the highest contribution to the differentiation of this stop contrast in perception (Schertz et al., 2015).

The critical status of perceptual cues

To further test the extent to which a perception-production relationship is revealed by critical perceptual cues, a perceptual study needs to be conducted first to establish the critical status of perceptual cues by examining each cue independently. It is worth noting that, for the previous studies that examined individual perceptual cues, some research did not investigate the critical status of the cues (Beddor, 2015; Ghosh et al., 2010; Perkell et al., 2004b; Yang, 2015), while others focused on perceptual weightings that did not yield a perception-production correlation with the cue weightings or acoustic measures in production (Schertz et al., 2015; Shultz et al., 2012). Consequently, the question is what kind of perceptual study can determine the critical status of perceptual cues, and potentially examine a perception-production relationship. When Newman (2003) first proposed the idea that certain perceptual cues are more "important" (or critical) than others in English fricative perception, the relative importance of the cues was primarily determined by the acoustic modelling results that spectral peaks contributed more to fricative classification than centroid and skewness, and therefore, peak frequency, as a critical cue, served as a better acoustic feature than other cues for *category classification* (Jongman et al., 2000). The implication of acoustic modelling is that the critical status of perceptual cues is defined by their contribution to *between-category* perception. However, in Newman (2003), the relative importance of peak frequency was also inferred from goodness rating results. Newman stated that the perceptually critical cue could possibly be the only cue that "makes a particular fricative token sound better to a listener", and "appeared to be related to listeners' goodness ratings" (p.2857), implying that other non-critical cues would not contribute to this kind of *within-category* perception. Therefore, Newman's view also suggests that the critical status of the cues is determined by whether listeners use them to perceive the quality of speech sounds *within a phonemic category* based on acoustic-phonetic differences. However, this argument is not consistent with the acoustic modelling results that determine *between-category* classification.

The distinction between *within-category* and *between-category* perception for defining the critical status of perceptual cues has a broader implication for perception-production

relationships. The critical status of perceptual cues as defined by *within-category* perception suggests that individuals use the cues to detect phonetic differences that are within a speech category. If the perception-production link is pertinent to critical perceptual cues defined by *within-category* perception, it implies that *within-category* phonetic differences can influence the link between perception and production. This conceptual definition does not preclude the detection of between-category differences based on phonetic differences. In terms of goodness ratings, an item that belongs to a difference phonemic category should still be rated as a poor instance of the target phonemic category. However, the *between-category* perception definition for the critical status of perceptual cue indicates that the cues only trigger phonological contrasts in perception. If the critical status of perceptual cues are defined only by *between-category* perception, the phonetic differences that signal *between-category* phonological contrasts should influence the link between perception and production. In contrast, *within-category* phonetic differences should not contribute to the perception-production relationship.

From a theoretical perspective, the perception-production relationship has been discussed in the context of phonemes. For the motor theory of speech perception, the function of speech gestures is to make "phonetic contrasts" that are "the basis of phonological categories" (Liberman and Mattingly, 1985, pp.21–22), and gestures are invariant for a given phoneme (Galantucci et al., 2006; Liberman et al., 1967). Note that speech perception is perceiving gestures under the motor theory, and therefore, the perception-production link should also be primarily based on phonological contrasts. Similarly, auditory approaches to speech perception and the DIVA model also indicate a focus on phonemes during the acoustic-gesture mapping process, and thus the formation of a perception-production link, in language acquisition. In this process, "the need of auditory distinctiveness of phonemes shapes production", and perception of the "phonemic content of speech signals" is shaped by the acoustic signals generated in speech production (Diehl et al., 2004, pp.167–168). The DIVA model also shares this view in the relationship between perception and production, and therefore, stresses that perceptual acuity and production clarity are related (Guenther, 1995; McAllister Byun and Tiede, 2017; Perkell et al., 2004a).

Therefore, investigating whether critical perceptual cues are defined by *within-category* or *between-category* perception extends our understanding of the nature of perception-production links. If a perception-production link is pertinent to critical perceptual cues, the investigation should demonstrate whether the perception-production link is influenced by the perception of *within-category* phonetic variations, or by the perception of *between-category* phonological contrasts following theoretical predictions.

Summary

Although previous empirical studies that found a close but imperfect link between the perception and production (Beddor, 2015; Ghosh et al., 2010; Nearey, 1992; Newman, 2003; Perkell et al., 2004b; Yang, 2015), this section reviewed some studies that did not show a relationship between perception and production (Ainsworth and Paliwal, 1984; Bailey and Haggard, 1973; Beddor, 2015; Frieda et al., 2000; Paliwal et al., 1983; Schertz et al., 2015; Shultz et al., 2012; Yu et al., 2021; Zellou, 2017). Aside from methodological issues, the studies show that a perception-production link is probably established by the acoustic properties of perceptually critical cues (Newman 2003; Shultz et al. 2012, also see Schertz et al. 2015). There are two potential definitions of the critical status of perceptual cues, based on a phonetic, *within-category* perception and a phonological, *between-category* perception. While these two perception types are not mutually exclusive in terms of cue use, examining which perception type defines the critical status of perceptual cues can demonstrate whether the perception-production link is established on the phonetic or phonological level and therefore, further our understanding in the nature of perception-production links.

1.3 Using Mandarin tones to study perception-production relationships

Mandarin tones are used in this dissertation project as a test case for the critical status of perceptual cues, and subsequently the perception-production relationship, since Mandarin tone perceptual cues exhibit differential cue weightings given to pitch direction and height which can be used to define the critical status of perceptual cues (Chandrasekaran et al.,

2010; Gandour, 1983; Guion and Pederson, 2007; Francis et al., 2008). Therefore, this section first reviews the Mandarin tone perceptual cues and acoustic properties of Mandarin tone productions. Then, it discusses the potential perception-production relationship of Mandarin tones, and the critical status of Mandarin tone perceptual cues.

Mandarin is a lexical tone language with four tone categories differing in pitch height and contour. Tone 1 has a high-level contour. Tone 2 has a rising contour. Tone 3 is a low dipping tone and Tone 4 is a falling tone (Chao, 1947). Each Mandarin syllable carries a tone and the lexical meaning of the syllable changes if the tone varies. For instance, /ma/ carrying each of the four tones means “mother” (/ma1/), “hemp” (/ma2/), “horse” (/ma3/) and “to scold” (/ma4/). The tone contours are displayed in Figure 1.1.

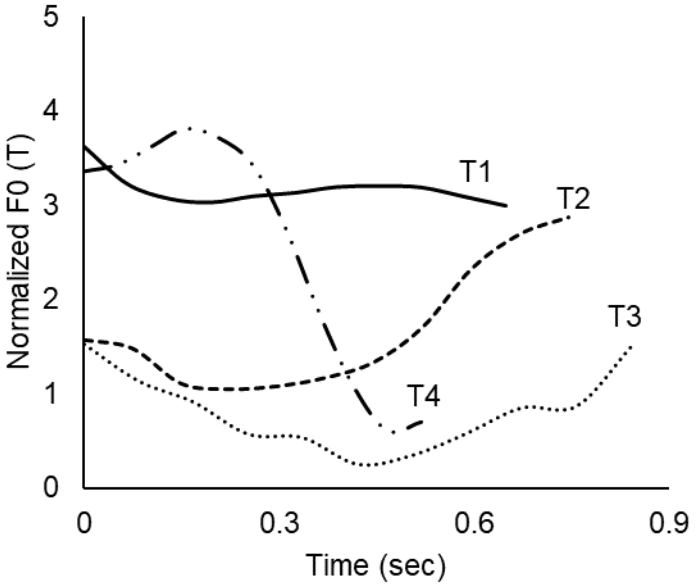


Figure 1.1: Tone contours of four Mandarin tones

Mandarin tones can be advantageous in testing whether a perception-production relationship is pertaining to critical perceptual cues, since the critical status of tone perceptual cues can be defined by both differential perceptual cue weighting in lexical tone perception and acoustic modelling of Mandarin tones. Perceptual cue weighting studies showed that pitch direction was weighted more strongly than pitch height in native Mandarin tone perception (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007), indicating that pitch direction is a more critical cue than pitch height in Mandarin tone perception from

the acoustic-phonetic perception perspective. Similarly, the acoustic modelling of tone categorization demonstrated that F0 contour-related cues provided better modelling accuracy than a combination of F0 mean and one of the F0 contour cues (Tupper et al., 2020). Note that Newman (2003) used acoustic modelling of fricative cues (Jongman et al., 2000) to define the critical status of perceptual cues. These studies showed that pitch direction (or F0 contour cues) was more critical for Mandarin tone perception than pitch height (or F0 mean cues) in both within-category and between-category perception. In addition, each Mandarin tone category may also be have its own critical cues. Both cue weighting and tone categorization showed that pitch direction could be crucial for the perception of Tone 2 and 4, but the perception of Tone 1 and 3 could require pitch height more than pitch direction (Chandrasekaran et al., 2010; Guion and Pederson, 2007; Tupper et al., 2020; Yang, 2015). The sections below review the perceptual cues and acoustic properties of Mandarin tones, and provide an in-depth discussion of the critical status of Mandarin tone perceptual cues.

1.3.1 Mandarin tone perceptual cues

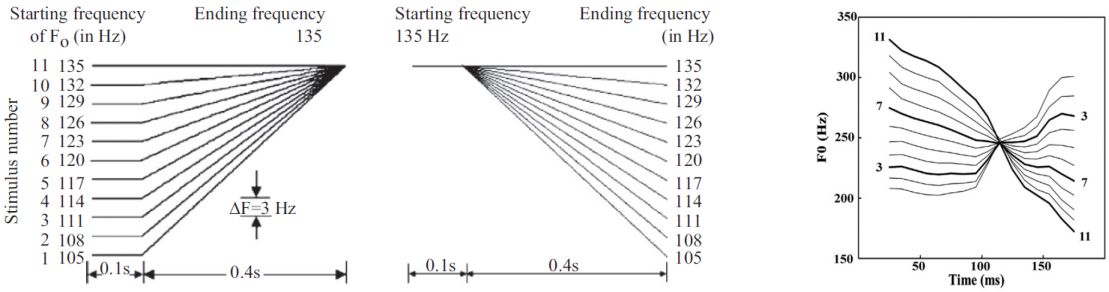
The primary cue for Mandarin tone perception is F0 (Jongman et al., 2006). Intensity and duration are secondary cues which play a role in tone perception only when F0 information is absent. For example, when native Mandarin listeners perceive Mandarin monosyllabic stimuli without F0 or formant information, amplitude contour becomes a useful cue in the perception of Mandarin Tone 2, 3 and 4 (Whalen and Xu, 1992). Duration cue is used for perceiving whispered speech (Liu and Samuel, 2004). When F0 information is present, ambiguous Mandarin tone duration do not have an effect on native listener's tone identification accuracy (Chang, 2011).

Previous research employed a weighted multidimensional scaling analysis (MDS), i.e. individual difference scaling (INDSCAL) (Carroll and Chang, 1970), to explore the number and nature of F0 dimensions required for tone perception. It is generally agreed that a two-dimension model best explains the tone perceptual data and the dimensions are interpreted as pitch height and direction (Chandrasekaran et al., 2010; Francis et al., 2008;

Gandour, 1983; Guion and Pederson, 2007). In some earlier studies (Gandour and Harshman, 1978; Gandour, 1979), lexical tone perception was found to involve five cues: average pitch, direction, length, extreme endpoint and slope. The stimuli of these studies were resynthesized tones consisted of linear tone contours (level, rising and falling). Four language groups (Cantonese, Thai, Yoruba, English) were involved in these studies as participants. A follow-up study (Gandour, 1983) was then conducted with an improvement of stimulus design (including an increase in the tone contour types by adding fall-rise and rise-fall contours, tone onset and offset levels, and controlling for duration), and an involvement of more contour language groups (Mandarin and Taiwanese, along with Cantonese, Thai, English in the previous studies). This INDSCAL analysis of tone perceptual dimensions yielded the above-mentioned two-dimension model of pitch height and direction. Among all language groups, native Mandarin results showed a higher mean weighting of pitch direction and a lower mean weighting of pitch height than non-tone language listeners (i.e. native English listeners). The same distribution of cue weightings was observed when native Mandarin listeners perceived Cantonese tones (Francis et al., 2008). Similarly, Guion and Pederson (2007) presented resynthesized linear tones to native Mandarin, English and Japanese listeners. They found that pitch direction was only used as a perceptual cue for native Mandarin listeners. In sum, native Mandarin listeners rely more on pitch direction than on pitch height to perceive tones.

Pitch height and direction are the two dimensions representing general cues that separate a group of natural or synthetic tone categories. However, native Mandarin listeners also use specific cues for perceptual differentiation of Mandarin tone pairs. Previous research manipulated the F0 onset or offset to examine the categorical perception of Mandarin tones. Lowering the F0 onset of a level tone while holding the F0 offset constant changed native Mandarin perception from Tone 1 to 2 (Chang et al., 2016; Chen and Peng, 2016; Peng et al., 2010; Wang, 1976; Xu et al., 2006). Similarly, lowering the F0 offset of a level tone while holding the F0 onset constant changed native Mandarin perception from Tone 1 to 4 (Chang et al., 2016; Peng et al., 2010) (Figure 1.2a). Previous studies also developed a Tone 2-4 continuum by co-varying F0 onset and offset in opposite direc-

tions, and proportionally changing the overall tone contour shape. As a result, Mandarin tone perception switched categorically between Tone 2 and 4 (Wang et al., 2017; Xi et al., 2010) (Figure 1.2b).



(a) Tone 1-2 (left) and Tone 1-4 (right) continua (Peng et al. 2010: 618)

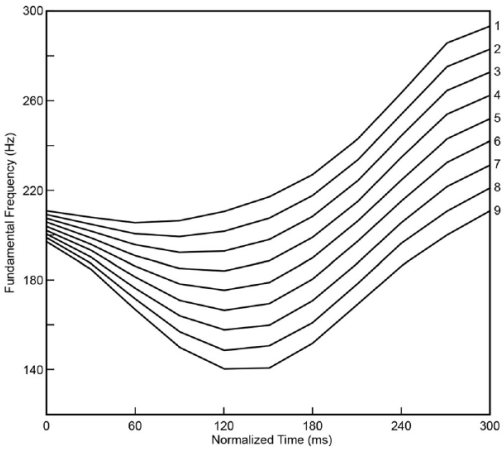
(b) Tone 2-4 continuum (Xi et al. 2010: 225)

Figure 1.2: Schematic representations of Mandarin tone continua

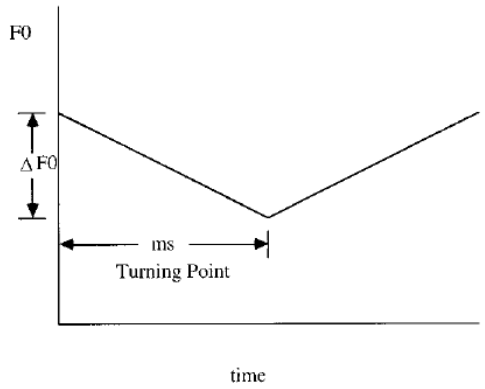
The proportional manipulation of tone contours was also applied to Tone 2 and 3 to investigate the perception of these two tones. Zhao and Kuhl (2015) investigated the change between Tone 2 and 3 perception using a Tone 2-3 continuum created by systematically changing the overall contour shape proportionally from Tone 2 to 3, which involved the changes in F0 onset and offset (Figure 1.3a).

In addition to spectral cues, this continuum also shows the change in the temporal location of F0 turning point (TP), a temporal cue, can also influence Mandarin tone perception. In fact, previous studies showed that this cue alone (Shen and Lin, 1991; Shen et al., 1993), and combined with the change in the F0 decrease from F0 onset to TP (ΔF0) (Moore and Jongman, 1997), can modulate the perception of Tone 2 and 3 (Figure 1.3b). It indicates that, although F0 (i.e., spectral cues) is the primary cue for Mandarin tone perception, TP as a temporal cue that specifies the temporal location of an F0 direction change also has an impact on Mandarin tone perception.

While F0 onset, F0 offset, TP and ΔF0 are acoustic measurements at specific temporal locations of a tone contour, the acoustic correlates of pitch height and direction are less clear. In an earlier study, Massaro et al. (1985) defined pitch height as F0 onset level and pitch direction was varied by shifting the F0 offset level, and consequently changing the slope of the tone contour. While defining pitch direction as F0 slope was generally ac-



(a) Tone 2-3 continuum (Zhao and Kuhl 2015: 1453)



(b) ΔF0 and TP measurements (Moore and Jongman 1997: 1867)

Figure 1.3: Schematic representation of Tone 2-3 continuum, ΔF0 and TP measurements

cepted (Chandrasekaran et al., 2010; Gandour, 1983; Guion and Pederson, 2007; Jongman et al., 2017; Yang, 2015), other studies regarded pitch height as F0 mean, or the average F0 of the tone contour. For instance, Yang (2015) used the level of F0 onset and offset of a linear tone contour to indicate pitch height, since a tone with high F0 onset and offset has a higher F0 mean than a tone with low F0 onset and offset. Similarly, other studies also treated F0 mean as the acoustic correlate of pitch height (Chandrasekaran et al., 2010; Gandour, 1983; Guion and Pederson, 2007; Jongman et al., 2017). It is also worth noting that, although F0 onset, F0 offset, TP and ΔF0 were found to modulate the perception of Mandarin tone categories (Chang et al., 2016; Chen and Peng, 2016; Moore and Jongman, 1997; Peng et al., 2010; Shen and Lin, 1991; Shen et al., 1993; Wang, 1976; Wang et al., 2017; Xi et al., 2010; Xu et al., 2006), relatively few studies investigated whether pitch height and pitch direction could modulate perception between tone categories since most studies relied on a discrimination task that focused on the acoustic differences between tone contours (i.e. within-category) (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). The exception was Yang (2015) which approximated F0 mean and slope by varying F0 onset and offset. Varying F0 onset and offset orthogonally had an impact on tone categorization in perception but it did

not investigate the critical status of these perceptual cues (Refer to Section 1.3.4 below for details).

1.3.2 Acoustic properties of Mandarin tone productions

For Mandarin tone productions, the F0 mean, slope and curvature were found to be the most prominent acoustic cues that separate Mandarin tone productions and four tone categories. In Tupper et al. (2020), a number of acoustic features were examined, including F0 mean, slope, curvature, onset, offset, TP and $\triangle$F0, and found that many cues were closely related. The three main groups of cues that performed within-category tone separation were F0 mean, slope and curvature. To obtain these cues, a parabola (Equation 1.1) was fitted to the tone contour and produced an intercept c_0 (F0 mean), first derivative c_1 (F0 slope) and second derivative c_2 (F0 curvature). F0 slope and curvature together provided the best between-category tone classification result. Adding F0 mean to this model did not improve the classification result, and replacing F0 slope or curvature with F0 mean lowered the classification accuracy.

$$f(t) \approx c_0 + c_1(t - \frac{1}{2}) + c_2[(t - \frac{1}{2})^2 - \frac{1}{12}] \tag{1.1}$$

The F0 mean measurement was essentially the same as obtaining the average of the F0 values obtained at all sampling points (e.g Jeng et al., 2006). Generally speaking, Tone 1 has the highest F0 mean since it maintains a high-level pitch pattern whereas Tone 3 has the lowest F0 mean. The F0 means of Tone 2 and 4 fall between those of Tone 1 and 3 (Jeng et al., 2006; Liu et al., 2007).

Measuring F0 slope using the intercept and the first derivative of a polynomial is not unprecedented. F0 slope indicates a rising and falling contour with a positive and negative value, respectively. Therefore, Tone 2 should have a positive F0 slope value, and Tone 4 should have a negative F0 slope value. The level contour of Tone 1 and the fall-rise contour of Tone 3 should result in a close to zero F0 slope value. Prom-on et al. (2009) applied the linear function (Refer to Equation 2.2) to obtain F0 slope representing the dynamic (rising or falling) nature of a pitch target. It should be noted that Prom-on et al.

(2009) approximated the pitch target at the syllable offset instead of the tone contour. It is because their model aims to capture pitch patterns in connected speech and the actual F0 contours do not always resemble the pitch target due to involuntary factors resulting from articulatory constraints. Nevertheless, the resulting pitch target corresponds to the F0 mean and F0 slope measured from the tone contour of a Mandarin monosyllable, with [high] for Tone 1 and [low] for Tone 3 indicating F0 mean, and [rising] for Tone 2 and [falling] for Tone 4 indicating F0 slope (Xu, 2001; Xu and Wang, 2001).

Another straightforward approach to measure F0 slope is deriving the ratio of F0 range (difference between maximum and minimum F0) and duration of the tone contour (Flemming and Cho, 2017; Jeng et al., 2006; Prom-on et al., 2012). However, this method produces similar values for Tone 2 and 4 since the two tones have similar F0 range (Figure 1.1). This was not an issue for previous studies that used this measurement method since rising (Flemming and Cho, 2017) and falling contours (Jeng et al., 2006; Prom-on et al., 2012) were investigated separate studies. It would not be appropriate for a study that compares the F0 slope of all four Mandarin tones.

Consistent with the pitch height $\times$ pitch direction perceptual space, a two-dimensional F0 mean $\times$ F0 slope tone space was plotted in Peng (2006) for Mandarin and Cantonese tones in order to compare the tone spaces between the two languages. The perceptual and production tone spaces of Mandarin share highly similar shapes and orientations (e.g Chandrasekaran et al., 2010) in that Tone 1 and 3 were the furthest apart among four tones along the pitch height/ F0 mean dimension indicating that they had the highest and lowest tone, respectively. Tone 2 and 4 were the furthest apart along the pitch direction/ F0 slope dimension showing that they had the most positive and negative slope, respectively. Alternatively, as in Yang (2015), F0 onset and offset were also frequently used for constructing tone space for crosslinguistic, speaking style or clinical research (Alexander, 2011; Xu Rattanasone et al., 2013; Yang, 2015; Zhou et al., 2013; Zhou and Xu, 2008). Mandarin tones can be well separated on an onset-offset plot. Four Mandarin tones fall into two groups in the F0 onset dimension, with Tone 1 and 4 falling under the high F0 onset group and Tone 2 and 3 under the low F0 onset group. Likewise, two groups emerged on

the F0 offset dimension as well. Tone 1 and 2 belong to the high offset group and Tone 3 and 4 to the low offset group (Alexander, 2011; Yang, 2015). In fact, the combination of F0 onset and offset roughly reflects F0 height and slope. If the tone space is divided by a diagonal line representing the same F0 onset and offset, level tones should appear on or near this line. The slope of these tones should be near zero. Contour tones (either rising or falling) have a substantial difference between the two points of F0 measurements and should occupy the upper left and lower right portions of the tone space. Therefore, the Mandarin F0 onset $\times$ F0 offset tone space appears like the F0 mean $\times$ F0 slope space but tilted by 45 degrees.

The F0 mean-slope or onset-offset dimensions reviewed above were effective in differentiating Mandarin tone categories. However, they failed to account for the within-contour F0 movements (Zhou and Xu, 2008), including ΔF0 and TP which are perceptually relevant for Tone 2-3 discrimination and characterize the productions of these two tones (Moore and Jongman, 1997). In fact, most previous studies examined Mandarin pitch contours with almost no dipping pattern since all tones were produced in connected speech (i.e. the dipping shape was reduced to a slight falling shape in Tone 3 tone sandhi). Jeng et al. (2006) was the only exception that included Tone 3 produced in citation form in their speech samples. However, their F0 calculation was based on F0 range, and therefore, Tone 3 had a positive slope value which could not reflect its dipping shape. Moreover, Xu (2001) acknowledged that Tone 3 produced in isolation might consist of two static pitch targets: a [low] followed by a [mid] or a [high]. Therefore, there is a need to model Mandarin tones, especially Tone 2 and 3, using a different method to take their F0 movements into account. It is possible to measure the ΔF0 and TP as separate acoustic parameters in Mandarin production studies (Wang et al., 2003). Another approach is to fit a polynomial function to these tones. In Zhao and Kuhl (2015), a sixth-order polynomial function was used to achieve the best goodness of fit without over-fitting. In Prom-on et al. (2009), the actual F0 realization of a Mandarin syllable was a combined solution of an ideal pitch target (a linear function) and carry-over effects from the preceding syllable which transferred F0 height, velocity and acceleration in transition. These were expressed in a function that contained a second-order polynomial

function. More importantly, Tupper et al. (2020) showed that F0 curvature, as the second derivative of a parabola, served as an important feature to separate Mandarin tone categories. In addition, although F0 mean and slope showed a high tone classification rate in the statistical models, F0 slope and curvature related cues yielded better performance than F0 mean and slope. Taken together, the modelling of Mandarin pitch direction probably requires a polynomial function in addition to a linear function in order to reflect the within-contour F0 movements.

1.3.3 Mandarin tone perception-production relationships

The theories reviewed in Section 1.1 are based on research on segments and previous empirical studies (Section 1.2) also mostly used segments as the target language. To date, there has been relatively little evidence supporting a perception-production link on the suprasegmental, or lexical tone, level. In Mandarin tone perceptual training, native English trainees with no prior Mandarin exposure showed improvement in both perception and production. Specifically, the trainees displayed improvement in perceptual identification of Mandarin tones after perceptual training (Wang et al., 1999). Then, their post-test Mandarin tone productions received increased identification by native Mandarin judges than pretest productions. The acoustic properties of their post-test tones were closer to native norms than pretest tones (Wang et al., 2003), which indicated a transfer from perception to production and a link between the two domains for Mandarin tones. Yang (2015) provided insights that the two domains could be acoustically linked for Mandarin tones. Synthetic linear tone contours were created by varying F0 values at F0 onset and offset of the contours and were presented in carrier sentences to native Mandarin participants for tone identification. The productions of all four tones in connected speech were recorded from the same group of participants, with the F0 onset and offset measured. It was found that tone boundaries on the perceptual and production tone space were closely aligned but with minor discrepancies, suggesting a close but imperfect relationship between Mandarin tone perception and production through the acoustic cues F0 onset and offset. Apart from Mandarin tones, there was also evidence for a tone perception-production relationship for

Cantonese. Previous research demonstrated that there was a relationship between the perception and production accuracy of native Cantonese children (Mok et al., 2019). In L2 learning, the perceptual discrimination and acoustic distance between tone contours were correlated, suggesting that learners of a tone language who could better discriminate two tone categories also produced the tones with greater acoustic contrast in Southern Vietnamese (Kirby and Giang, 2021) and Cantonese (Yu et al., 2021).

While these studies showed a perception-production relationship for tones, this relationship was based on the accuracy of perception and production (Mok et al., 2019; Wang et al., 1999) or the acoustic distance between tone categories by taking several tone cues into account (Kirby and Giang, 2021; Yu et al., 2021). It is unknown whether the relationship is related to a particular perceptual cue. Although Wang et al. (2003) compared native and trainees' F0 ranges, maximum F0, minimum F0, the temporal locations of maximum and minimum F0, as well as the F0 values at specific sampling points, the perception-production relationship was not established based on these cues. As for the tone space comparison in Yang (2015), the same cues were measured in the perception and production tasks and established a relationship based on the cues used in both domains. However, the relationship was based on a visual and qualitative comparison of boundary locations separating tone categories. Therefore, it was not able to demonstrate the strength of the perception-production relationship for native Mandarin individuals. Relating to the critical status of perceptual cues (cf. Newman, 2003), these studies did not assess the perception-production link established by critical and non-critical perceptual cues separately.

In relating Mandarin acoustic cues to perceptual cues, the majority of previous studies assumed that pitch height and direction were F0 mean and slope, respectively, as mentioned in the above section (e.g. Chandrasekaran et al., 2010; Jongman et al., 2017; Peng, 2006). However, the acoustic correlates of these perceptual cues still need to be investigated. For example, the previous section mentioned that F0 mean was not the only acoustic cue that represent pitch height, the less strongly weighted perceptual cue. For instance, F0 onset was used in at least one study (Massaro et al., 1985). For pitch direction,

F0 slope does not capture the contour movements as discussed in the previous section. In addition to the linear direction, F0 curvature should also be considered especially because it is one of the main contributing factors to the classification of Mandarin tones as shown in the acoustic analysis of Mandarin tones (Tupper et al., 2020). Apart from spectral cues, TP, as a temporal cue also plays an important role in Mandarin tone perception (Moore and Jongman, 1997; Shen and Lin, 1991; Shen et al., 1993). Since a shift in TP can change the shape of the tone contour, it is possible that this temporal cue can affect pitch direction. Therefore, given that there is little evidence on the perception-production link of Mandarin tone cues, the preliminary investigation of the impact of critical perceptual cue on the perception-production link should include these potential acoustic correlates of pitch height and direction, which can be defined as the non-critical and critical perceptual cues of Mandarin tones. The critical status of Mandarin tone perceptual cues is discussed in details in the following section.

1.3.4 The critical and non-critical Mandarin tone perceptual cues

As discussed in Section 1.2.2, the critical status of perceptual cues can potentially be defined by *within-category* and *between-category* perception (Newman, 2003). This distinction can further extend our understanding in the nature of perception-production links - whether it is phonetically or phonologically-based, if it is established by critical perceptual cues. While the cues used for *within-category* and *between-category* perception are not mutually exclusive, the perception of particular Mandarin tones appear to be different from overall perception for both types of perception, suggesting that the critical cues for overall perception and the perception of particular tones may differ.

Defining critical cues from *within-category* phonetic perception perspective

The critical status of Mandarin tone perceptual cues can be established by the well-studied cue weightings of lexical tones. Specifically, pitch height and pitch direction are both used as perceptual cues, and native Mandarin listeners give stronger weight to pitch direction than to pitch height (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). This reflects the difference in critical status of the two percep-

tual cues from a acoustic-phonetic perspective, with pitch direction being a more critical perceptual cue than pitch height.

The notion that perceptual cue weightings is acoustic-phonetically based is indicated by the research paradigm used to define cue weightings. In Gandour's phenomenal studies of lexical tone cues, cue weights were obtained by an INDSCAL model which determined the nature and dimensions of perceptual cues based on the psychological distance between tone stimuli (Gandour and Harshman, 1978; Gandour, 1979, 1983). This research paradigm has then been adopted in related studies which have depended on a discrimination task that required listeners to perceive the acoustic difference between tones in each tone pair (e.g. Chandrasekaran et al., 2010; Francis et al., 2008; Guion and Pederson, 2007; Wiener, 2017). Therefore, the perceptual dimensions and weightings obtained from these studies should reflect the acoustic-phonetic perception of tones, and therefore, using cue weightings to define the critical status of perceptual cues is consistent with Newman (2003)'s claim that the critical status of perceptual cues is related to goodness ratings within phonemic categories. Since the results of native Mandarin perception studies have consistently shown that pitch direction is weighted more strongly than pitch height by native Mandarin listeners (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007), pitch direction can thus be defined as more critical than pitch height in determining the dissimilarities between tone stimuli (i.e. psychological distance) in native Mandarin perception. Based on these results, the critical perceptual cue defined by within-category perception is pitch direction. However, it is worth noting that, although perceptual weights differ, all perceptual cue dimensions as determined by INDSCAL (e.g., pitch direction and height in tone studies) contribute to defining the psychological distances between stimuli on a perceptual space. For tone perception specifically, the perceptual space of the INDSCAL analysis showed that Tone 2 vs. 4 and Tone 1 vs. 3 were separated by the pitch direction and pitch height dimension, respectively. As a result, it is also possible that, for within-category perception, pitch direction is that critical cue for perceiving Tone 2 and 4, and pitch height for Tone 1 and 3. Based on the overall perception result, pitch direction should also determine the within-category goodness for all tones. Alternatively, pitch di-

rection only determines the within-category goodness of Tone 2 and 4, and pitch height determines that of Tone 1 and 3 (Newman, 2003). No study has yet investigated whether the within-category goodness of Mandarin tones are determined by pitch direction alone or by either pitch direction or height depending on the tone category.

Defining critical cues from *between-category* phonological perception perspective

As for tone categorization, a recent study on the acoustic modelling of Mandarin tones (Tupper et al., 2020) showed that a combination of F0 slope and curvature cues maximized Mandarin tone classification accuracy. Adding F0 mean to this model only offered little improvement to the accuracy. Combining F0 mean and F0 slope or curvature could still effectively classify Mandarin tones, but lowered the model's classification accuracy compared to the F0 slope and curvature combination. As defined in previous studies (Chandrasekaran et al., 2010; Jongman et al., 2017; Peng, 2006), F0 slope and mean are considered as the acoustic correlates of the critical pitch direction and non-critical pitch height cue. In addition, F0 curvature can be defined as an acoustic correlate of pitch direction since it reflects the shape of a tone contour. Therefore, the acoustic modelling of tone cues implies that pitch direction is more critical than pitch height for *between-category* classification, as the acoustic results should shed light on the critical status of tone cues in perception, based on Newman (2003) that depended on Jongman et al. (2000) to infer the critical status of fricative perceptual cues.

In addition to overall tone categorization, it is also possible each tone has its own critical perceptual cues. Tupper et al. (2020) examined the cues that separate individual Mandarin tone from other tones. It found that, while pitch direction-related cues, including F0 slope and curvature, are important for distinguishing Tone 2 or 4 alone from other tones, F0 mean and its correlates are important for Tone 1 and 3. The fuzzy logic model of speech perception supports these acoustic findings (Massaro et al., 1985). It views both pitch direction and height as features defining all tones, and listeners give different weights to each of them in between-category perception depending on the continuous acoustic properties. Massaro et al. (1985) varied F0 slope and F0 onset orthogonally and different acoustic

properties affected the weightings of these cues in the perception of Tone 1 and 2. Specifically, native Mandarin listeners had a higher probability to perceive a tone as Tone 1 in the high F0 onset and flat slope ranges, and as Tone 2 in the low F0 onset and rising slope ranges. As a result, both cues can influence tone category perception, and are weighted differently in determining the tone categories depending on the acoustic properties of the stimulus itself. In Yang (2015), the F0 onset and offset of linear tone contours were varied orthogonally to approximate F0 mean and slope. It showed that when F0 onset was smaller than offset (i.e., rising slope), the stimuli were identified as Tone 2. When F0 onset was greater than offset (i.e., falling slope), Tone 4 perception was dominant. Finally, when F0 onset and offset were at a comparable level, the perception was Tone 1 for high F0 onset and offset and Tone 3 for low F0 onset and offset. Note that Massaro et al. (1985) defined F0 height as F0 onset instead of F0 mean as used in other major perception studies (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour and Harshman, 1978; Gandour, 1979, 1983; Guion and Pederson, 2007). A series of F0 slope contour with a fixed F0 onset naturally led to covariations in F0 mean as well. As a result, the effects of "F0 slope" in Massaro et al.'s study actually demonstrated the influence of both F0 slope and mean on T1 and T2 perception. As for Yang (2015), the critical status of F0 onset and offset (or the approximated F0 mean and slope) were not examined. Therefore, there is still no clear evidence from perception studies to determine whether pitch direction-related cues influence Mandarin tone categorization independently and more strongly than pitch height-related cues. It is also unclear if the critical nature of the cues differ for each Mandarin tone based on perception studies. Further research is needed to examine and compare the perceptual effects of pitch direction (critical) and height (non-critical) on Mandarin tone categorization by varying these cues orthogonally, and examine the critical status of each cue for individual tone categories. So far, there has only been a study on the JND of tones that independently manipulated slope and mean (Jongman et al., 2017), but it did not investigate the effects of each cue on Mandarin tone categorization.

In sum, based on the acoustic modelling and perception results, pitch direction-related cues, including F0 slope and curvature serve as the critical cues that influence overall tone

categorization (Tupper et al., 2020). There are possible difference in the categorization of individual tones. While pitch direction-related cues remain to be critical for Tone 2 and 4 perception, Tone 1 and 3 perception possibly require pitch height-related cues, including F0 mean.

1.3.5 Summary

Taken together, there are two possible views to define the critical status of perceptual cues for Mandarin tones following *within-category* and *between-category* perception. Based on overall tone perception, the perceptual cue weighting results (the acoustic-phonetic, *within-category* perception perspective) indicate that pitch direction is the critical perceptual cue since it is weighted more strongly than pitch height for native Mandarin tone perception (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). The tone categorization perspective based on acoustic modelling data, *between-category* perception also defines pitch direction-related cues as the critical status of perceptual cues. In addition, F0 slope and curvature can be potential acoustic correlates of pitch direction, and therefore, critical for Mandarin tone perception (Tupper et al., 2020). On the other hand, both *within-category* and *between-category* perception show that F0 slope and other pitch direction cues are only critical for perception of Tone 2 and 4, whereas F0 mean and other pitch height cues are critical for the perception of Tone 1 and 3 (Tupper et al., 2020; Yang, 2015). While pitch direction-related cues constitute the critical perception for Mandarin Tone 2 and 4, the critical perceptual cues for Tone 1 and 3 could either be pitch height-related or direction-related cues. Therefore, it is conceivable that, following Newman (2003), the critical pitch direction-related cues should show a stronger perception-production relationship than pitch height-related cues following the overall perception results. When the perception-production link is examined for each Mandarin tone category, it is possible that this predicted result only applies to Tone 2 and 4 since the critical status of perceptual cues aligns with that of overall tone perception. However, Tone 1 and 3 can possibly yield a stronger perception-production link for pitch height-related than direction-related cues.

In order to examine the above predictions, this study asks whether the pitch direction-related and height-related cues have an impact on the acoustic-phonetic level of *within-category* perception or the phonemic categorization level of *between-category* perception for each tone. The outcome of this exploration furthers our understanding of the nature of perception-production relationships.

1.4 Overview of this dissertation project

1.4.1 Rationales and research questions

This chapter reviewed various theories that predict the link between perception and production (Blumstein and Stevens, 1979; Diehl and Kluender, 1989; Diehl et al., 2004; Fowler, 1986; Fowler et al., 2003; Galantucci et al., 2006; Guenther, 1994, 1995, 2015; Kuhl et al., 2008; Liberman et al., 1967; Liberman and Mattingly, 1985; Liberman and Whalen, 2000; Nearey, 1992). Empirical evidence was found in first language research (Beddor, 2015; Bell-Berti et al., 1979; Fox, 1982; Franken et al., 2017; Ghosh et al., 2010; McAllister Byun and Tiede, 2017; Newman, 2003; Perkell et al., 2004a,b; Yang, 2015; Zellou, 2017), L2 research (Flege, 1993, 1995; Flege et al., 1997; Flege, 1999; Flege et al., 1999; Flege, 2003; Flege and Schmidt, 1995; Kato and Baese-Berk, 2020; Kirby and Giang, 2021; Yu et al., 2021) and training studies (Bradlow et al., 1997; Brosseau-Lapré et al., 2013; Hardison, 2003; Hazan et al., 2005; Herd et al., 2013; Hirata, 2004; Kartushina et al., 2015; Iverson et al., 2012; Lambacher et al., 2005; Sakai and Moorman, 2018; Wang et al., 1999, 2003). However, some studies did not find a relationship between perception and production and the proposal that a perception-production link is established by critical perceptual cues emerged from the evidence and "non-evidence" (Newman 2003; Shultz et al. 2012, also see Schertz et al. 2015). To date, this proposal still needs to be tested empirically and this project argues that Mandarin tones can serve an ideal test case because of the differential weightings given to Mandarin tone perceptual cues (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007) and the acoustic modelling results of Mandarin tones (Tupper et al., 2020). Moreover, the defining features of critical perceptual cues need to be examined in order to understand whether a perception-production link is phonetically

based if *within-category* perception defines the critical status of perceptual cues and/or phonologically based according to theoretical predictions.

Previous studies found evidence supporting a perception-production link for Mandarin tones (Wang et al., 1999, 2003; Yang, 2015) and other tone languages (Kirby and Giang, 2021; Mok et al., 2019; Yu et al., 2021). However, as reviewed in section 1.3.3, it should be noted that these studies either did not compare a single cue used in perception and production (Kirby and Giang, 2021; Mok et al., 2019; Wang et al., 1999, 2003; Yu et al., 2021), or only provided a qualitative visual comparison of perception and production tone spaces (Yang, 2015). The notion that Mandarin tones show a relationship between the acoustic cues that characterizes tone production and the cues used in tone perception needs to be tested first. Apart from F0 onset and offset used in Yang (2015), this project needs to test the perception-production relationship with cues that can potentially be the acoustic correlates of pitch direction and height (i.e., the critical and non-critical cues for Mandarin tone perception). Next, the current project examines the proposal that a perception-production link is established by critical perceptual cues. Since it is unclear whether the critical status of perceptual cues is defined by *within-category* or *between-category* perception, which is crucial for the understanding of the nature of perception-production relationships (phonetically vs. phonologically based), this project investigates the defining features of critical perceptual cues prior to testing a perception-production relationship with critical and non-critical perceptual cues. Specifically, the project asks whether the critical status is defined by a *within-category*, phonetically-based perception and/or a *between-category*, phonologically-based perception. Subsequently, it tests the proposal of the perception-production relationship pertaining to critical perceptual cues using critical and non-critical perceptual cues of Mandarin. Native Mandarin individuals weigh pitch direction more strongly than pitch height in overall tone perception (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). Moreover, the pitch direction-related cues can yield better tone classification accuracy than pitch height-related cues (Tupper et al., 2020). Following these previous studies, this project, therefore, assumes that the critical and non-critical cues for Mandarin tone perception are pitch direction and height, respectively. As

indicated in this literature review, there is yet any study that systematically investigates the role of critical and non-critical perceptual cues in establishing a perception-production relationship.

To summarise, the following chapters address the main research questions of this project below:

1. Are the Mandarin tone perception and production related in terms of acoustic cues? (RQ1)

2. Is the production-perception relationship as defined by the critical perceptual cues phonetically driven or phonologically driven? (RQ2)

3. Is the formation of a perception-production relationship attributable to the use of critical perceptual cues? (RQ3)

1.4.2 Structure of this project

This dissertation reports three studies that address the above research questions.

Using Tone 2 as a test case, the first study (Chapter 2) aimed to test research question 1 by examining a tone perception-production correlation using tone-specific acoustic features. The acoustic features examined were F0 onset, F0 mean (pitch height related), F0 slope, curvature (pitch direction related) and TP (temporal), relevant to Tone 1-2 and Tone 2-3 discrimination (Chang et al., 2016; Moore and Jongman, 1997; Peng et al., 2010; Shen and Lin, 1991; Shen et al., 1993; Xu et al., 2006; Wang, 1976). As it included both critical and non-critical perceptual cues, it covered research question 3 as a preliminary study.

This first study revealed a perception-production relationship for Mandarin Tone 2 and demonstrated initial evidence supporting the proposal that critical perceptual cues could establish a perception-production link. Then, the next study (Chapter 3) investigated the perception-production relationship of all four Mandarin tones by a systematic manipulation of one critical (F0 slope) and one non-critical perceptual cue (F0 mean). First, this study examined the assumption that the defining features of the critical status of perceptual cues by revealing the effect of critical and non-critical perceptual cues on *within-category* (phonetically driven) and *between-category* perception (phonologically driven) (research question

2). Then, this study explored the relationship of native Mandarin perception and production for each cue (i.e., F0 slope and F0 mean) through a statistical learning approach, addressing research question 3.

Finally, the third study (Chapter 4) further investigates research question 3. It was necessary because the second study (Chapter 3) analyzed the perception-production relationship by a qualitative comparison of the performance of perception and production models in statistical learning. The follow-up study examined an attributable link between critical perceptual cues and perception-production relationships using a within-subject design with Mandarin learners of a non-tonal language background participating in the same experiment as in Chapter 3 in two visits while they were taking a Mandarin Chinese course. By tracking the learning effects on the critical and non-critical cues in perception and production for each learner, it asked whether establishing a perception-production link can be attributed to the use of critical perceptual cues, which was expected to be demonstrated by a simultaneous improvement in perception and production only for the critical perceptual cues.

Chapter 2

Relationship between the perceptual cues and acoustic features of Mandarin Tone 2

2.1 Introduction

The goal of this chapter is to explore the Mandarin tone acoustic cues that can relate production to perception. As a preliminary exploration, it also examines the extent to which a perception-production correlation can be established based on the critical status of perceptual cues using a set of critical and non-critical perceptual cues.

As reviewed in section 1.3, previous research evinced a perception-production relationship for Mandarin tones, but the studies did not compare the same cues used in perception and production (Wang et al., 1999, 2003) or only provided a qualitative comparison of perception and production tone spaces without considering individual variations (Yang, 2015). Therefore, to establish that perception and production can be related through Mandarin tone cues, this study analyzes the perception-production correlation using the acoustic cues that influence the perception and characterize the production of Mandarin tones. The selection of the acoustic cues examined in this study was determined by the cues used for Mandarin tone perception in order to pave the way for investigating the proposal that a perception-production relationship is established through critical perceptual cues. In section 1.3, the review of Mandarin tone perceptual cues presents the two general tone perceptual cues – pitch direction and height (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). Apart from these spectral cues, TP, a tempo-

ral cue, alone has been shown to influence Tone 2 and 3 perception (Moore and Jongman, 1997; Shen and Lin, 1991; Shen et al., 1993). Therefore, the present study involves these perceptual cues in the perception and production analysis.

Although many previous studies used F0 mean and slope to represent pitch height and direction, respectively (e.g. Chandrasekaran et al., 2010; Jongman et al., 2017; Peng, 2006), this study also included other cues that are potential acoustic correlates of these perceptual cues based on the literature. For pitch height, apart from F0 mean, F0 onset was used in at least one study (Massaro et al., 1985). For pitch direction, in addition to F0 slope, F0 curvature should also be investigated as it captures movements within a tone contour and is one of the main contributing factors to the classification of Mandarin tones (Tupper et al., 2020). In addition, the temporal cue TP (Moore and Jongman, 1997; Shen and Lin, 1991; Shen et al., 1993) is expected to be related to F0 curvature, and therefore pitch direction, since it specifies the temporal location of pitch direction change and changing TP shifts the shape of the tone contour.

Examining these cues can serve as a preliminary exploration of the proposal that a perception-production relationship is established through critical perceptual cues. As reviewed in section 1.2, this proposal is based on the finding that, among the three fricative acoustic cues (peak frequency, frication centroid, and skewness), only peak frequency can reveal a perception-production correlation (Newman, 2003). The relative importance of peak frequency was inferred from the goodness rating results, supported by acoustic modelling (Jongman et al., 2000). However, the critical status of these cues was not determined independently. Other studies also demonstrated a trend of a perception-production correlation through a critical perceptual cue (Schertz et al., 2015; Shultz et al., 2012).

To explore this proposal, studying the perception-production relationship of lexical tones has an advantage because the critical status of tone perception cues has been established by the fact that native Mandarin listeners consistently weight pitch direction more strongly than pitch height (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Massaro et al., 1985). The acoustic modelling of Mandarin tones also showed that F0 contour-related, or pitch direction, cues could better categorize Mandarin tone productions than a

combination of F0 mean and F0 contour-related cues (Tupper et al., 2020). As a result, pitch direction and height cues are assumed to be critical and non-critical perceptual cues, respectively (especially for Tone 2, the target tone of this study). Following Newman (2003), pitch direction should show a stronger perception-production relationship than pitch height. Little previous research has attempted to relate lexical tone production to perception in terms of the relative critical status of individual cues.

The current study uses Tone 2 as the target tone, since its contrast with Tone 1 involves a change of pitch direction (i.e. rising vs. level), and a change of F0 height in terms of the overall F0 mean and F0 onset (Chang et al., 2016; Massaro et al., 1985; Peng et al., 2010; Xu et al., 2006; Wang, 1976). In addition, the Tone 2-3 contrast involves a change in critical pitch direction cues (F0 curvature and TP), as well as a non-critical pitch height cue (F0 onset) (Moore and Jongman, 1997). F0 slope, F0 curvature, TP, F0 mean and F0 onset were obtained from production and perception data. To examine a tone perception-production relationship as a function of the critical status of perceptual cues, this study performs a perception-production correlation for each cue, and therefore, can examine the perception-production correlation using critical and non-critical cues separately. The study expects to find a strong, positive correlation for critical pitch direction cues (F0 slope and F0 curvature, and TP as a related temporal cue). In contrast, the non-critical pitch height cues (F0 mean and F0 onset) should reveal a weaker correlation.

2.2 Method

2.2.1 Participants

Twenty-five native Mandarin female speakers who were undergraduate students at Simon Fraser University served as participants in this study (mean age: 22.2). Only female participants were recruited to avoid potential influences of gender difference. For example, Yang and Whalen (2015) found that there is a gender difference in the perception-production relationship of English vowels, with female participants showing a closer match of perception and production than male participants. Since the current study did not intend to explore the issue of gender difference in perception-production relationships, we therefore only re-

cruited female participants, matching the gender of the talker of perceptual stimuli (refer to Section 2.2.3).

2.2.2 Production task

Stimuli

Participants produced four commonly used Mandarin monosyllabic words containing the syllable /ʈʂu/ (*zhu* in *pinyin*) with four tones, meaning “pig” in Tone 1, “bamboo” in Tone 2, “lord” in Tone 3 and “pillar” in Tone 4. Each tone word was repeated six times. A total of 600 productions were recorded (4 tone words x 6 repetitions x 25 participants).

Procedures

The stimuli were elicited in a word pair context. Each tone word was paired up with a different tone word forming a total of 12 different tone pairs (i.e. six repetitions per tone). Participants were instructed to speak at a normal speaking rate and to pause between the production of the two words to prevent any potential tone coarticulation or sandhi effect, or difference in stress. The tone pairs were displayed in Mandarin Phonetic Symbols (*Pinyin*). The order of different tone pair combinations was randomized. In each trial, one tone pair was presented in the centre of the screen and participants were instructed to read each tone pair in a clear style. Clear productions were elicited since it has been claimed that perceptual targets are stored in hyperarticulated forms and therefore clear productions should display a stronger relationship with perception than casual, reduced productions (Johnson et al., 1993).

The task was self-paced, and participants could repeat if they made any production mistake. The recording task was conducted in a sound-attenuated booth in the Language and Brain Lab at Simon Fraser University, using a Shure KSM microphone placed at a 45-degree angle, about 20cm away from the speaker’s mouth. The speech materials were digitized at a sampling rate of 48 kHz and a 16-bit resolution using Audacity 2.1.1.

Acoustic measurements and modelling

All T2 productions were analyzed acoustically in Praat (Boersma and Weenink, 2018). Two phonetically-trained native Mandarin listeners were recruited to evaluate the accuracy of the stimulus tones. Five out of 150 T2 productions were identified as error productions and therefore were removed from the acoustic measurements and modelling below. All acoustic measurements were conducted in Praat (Boersma and Weenink, 2018). The tone contour was measured from the beginning and ending of waveform periodicity of each *zhu* word in T2. The total duration of each tone contour was measured, and each tone contour was divided into 100 intervals of equal distance (Tupper et al., 2020). F0 values in Hertz were then obtained at the 101 equidistant time points along a tone contour (Tupper et al., 2020) using the autocorrelation method and a time step of 0.015s (F0 range: 50Hz – 500Hz). The F0 values were manually checked for accuracy by the author and two phonetically trained research assistants. When a missing data point or an inaccurate value was identified, manual measurement was conducted by taking the inverse of the duration of a single period. To normalize for inter-speaker pitch range differences, each frequency value was then converted from *Hz* to a logarithm-based *T* value using Eq. 2.1 was used for F0 normalization (Wang et al., 2003):

$$T = \frac{logx - logL}{logH - logL} \times 5 \tag{2.1}$$

where *x* was F0 value in Hz at any given point, *L* and *H* were the minimum and maximum F0, respectively, of all four tones produced by the speaker. T had a range of 0 to 5, corresponding to the pitch scale for lexical tones developed by Chao (1947).

For the critical cues, polynomial fits were used to estimate F0 slope using Eq. 2.2 and F0 curvature using Eq. 2.3 based on the normalized F0 values at the 101 points along the tone contour (*cf.* Tupper et al., 2020):

$$f(t) = mt + k \tag{2.2}$$

$$f(t) = at^2 + bt + c \tag{2.3}$$

where *t* represented the time elapsed from the tone onset, obtained by multiplying the total duration of each tone contour by the relative location of each time point along the tone contour. The linear coefficient of Eq. 2.2 (*m*) and the quadratic coefficient of Eq. 2.3 (*a*) represented F0 slope and F0 curvature, respectively. F0 slope indicates a rising and falling contour with a positive and negative value, respectively. Positive and negative F0 curvature values denote upward and downward opening parabolic shape, respectively. Another critical cue, TP, was obtained at the temporal location relative to the total duration.

For non-critical cues, F0 mean was obtained by averaging the F0 values obtained from all measurement points in Praat. (time step: 0.015s). F0 onset in normalized frequency *T* was obtained at the first sampling point of each tone production which was the starting time point of the tone contour.

2.2.3 Perception task

Stimuli

Four repetitions of Mandarin words containing *zhu* with four tones (4 tones x 4 repetitions), which are the same tone words elicited in the production task, were used to create the stimuli for the perception task. These words were provided by one female native Mandarin talker who did not participate in the production and perception tasks of this study. The productions were judged as correct productions of the intended tones by two phonetically-trained research assistants who were native speakers of Mandarin. As will be seen in the results, the tonal productions of this talker were comparable to those by the participants of this study. In particular, the T2 productions of this talker had a mean F0 slope of 7.52T/s (5.26 - 10.68T/s), a mean F0 curvature of 45.83T/s^2 (26.05 - 77.54T/s), a mean TP at 20% (0 - 31%), a F0 mean of 2.11T (1.80 - 2.59T), and a mean F0 onset of 1.69T (1.28 - 2.00T). Based on these natural productions, the tone contours of the perception stimuli were resynthesized by separately manipulating TP and F0 onset, using the pitch-synchronous overlap and add (PSOLA) method in Praat (Boersma and Weenink, 2018). The goal was to create a perceptual space that situated the T2-like stimuli between two

end points that simulate the F0 trajectories of a T1 (providing high F0 height bound and low F0 contour bound) and T3 (providing low F0 height bound and high F0 contour bound).

All the resynthesized tone contours were set at a duration of 410ms, the mean duration of the speaker's T1, T2 and T3 productions (Chang et al., 2016; Moore and Jongman, 1997). The F0 onset series (high to low bound) was created with F0 onset endpoints based on the talker's F0 mean of all T1 productions (295Hz) (high F0 bound) and the minimum F0 of the talker's T3 (159Hz) (low F0 bound) (Figure 2.1 (left panel)). At the high F0 onset endpoint, a level F0 contour was first created, and the F0 onset was systematically lowered by a step size of 8Hz, the JND of Mandarin tone (Jongman et al., 2017). This follows previous research that used a level-to-rising tone continuum (Chang et al., 2016). The manipulation was uni-directional (i.e., beginning from the level contour endpoint) because Chang et al. (2016) did not find an influence of the direction of manipulation (i.e. from level to rising or from rising to level) on Mandarin tone perception. A T3 value was used to determine the low F0 onset endpoint because T3 typically had a lower F0 onset than T2 (Moore and Jongman, 1997). As a result, the F0 onset range encompassed the F0 onset from T1 to T2, and to T3.

TP endpoints were based on the earliest and latest TP location of the talker's T2 and T3 productions (i.e. 0% and 60% of the total tone duration). Each F0 onset had a TP series, except for the high bound F0 onset endpoint (a level tone). The other tone contours in the TP series began with an initial flat contour followed by a rising contour (Chang et al., 2016). For each F0 onset level, TP was varied by adjusting the temporal location where the contour changed from flat to rising (Figure 2.1 (right panel)) with the interval of 10% of the total duration (410ms) (Moore and Jongman, 1997). The change in TP consequently alters tone contour shape and is thus expected to critically influence perception (Moore and Jongman, 1997).

As a result, an 18-step (F0 onset) x 7-step (TP) grid of stimuli was formed, with fixed F0 offset level and duration. F0 slope, F0 curvature and F0 mean covaried with F0 onset and TP during acoustic manipulation. Each tone contour in the grid of stimuli had a distinctive

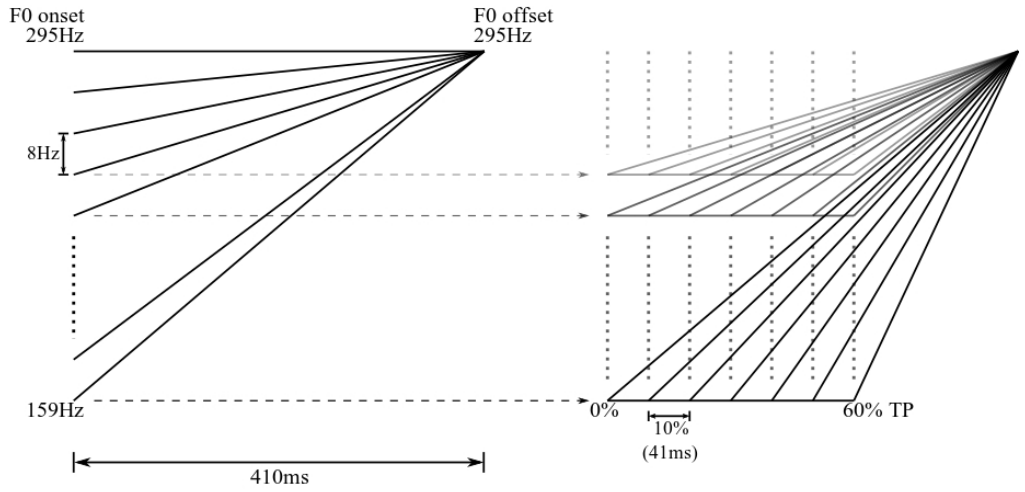


Figure 2.1: Schematic representations of perceptual stimulus series of F0 onset (left panel) and TP (right panel).

set of critical F0 slope, F0 curvature and TP values, as well as non-critical F0 mean and F0 onset values.

Procedures

The procedures of the perception task used the Method of Adjustment which required participants to select their preferred item representing a speech sound, therefore providing information about the acoustic properties of participants' typical perceptual categories (Johnson et al., 1993). This task was adopted because it would reveal the perceptual stimuli that represented participants' preferred perceptual exemplars of T2. The acoustic properties of these perceived preferred T2 exemplars could be subsequently used to correlate with the acoustic data extracted from the same participants' productions. In each trial, a stimulus grid consisting of 126 boxes (18 F0 onsets x 7 TPs) and a rating scale was displayed on a computer screen (Figure 2.2). When the participant clicked a box, one of the 126 *zhu* stimuli was played out over the headphones. Participants were instructed to first listen to each of the corner stimuli (i.e., the boundary tones) and rate each stimulus on a scale of 1 (poor exemplar of T2) to 5 (good exemplar of T2). Then, they selected the stimulus that represented the preferred exemplar of T2 and rated it on the same rating scale. If

a participant's preferred exemplar of T2 fell inside the stimulus grid, the preferred exemplar they selected should be rated higher than the boundary tones. To ensure participants determined the location of the preferred T2 exemplar auditorily (not visually), the orientations of the two axes or position of the axes were switched from trial to trial. Each participant completed 16 trials (8 orientation combinations x 2 repetitions).

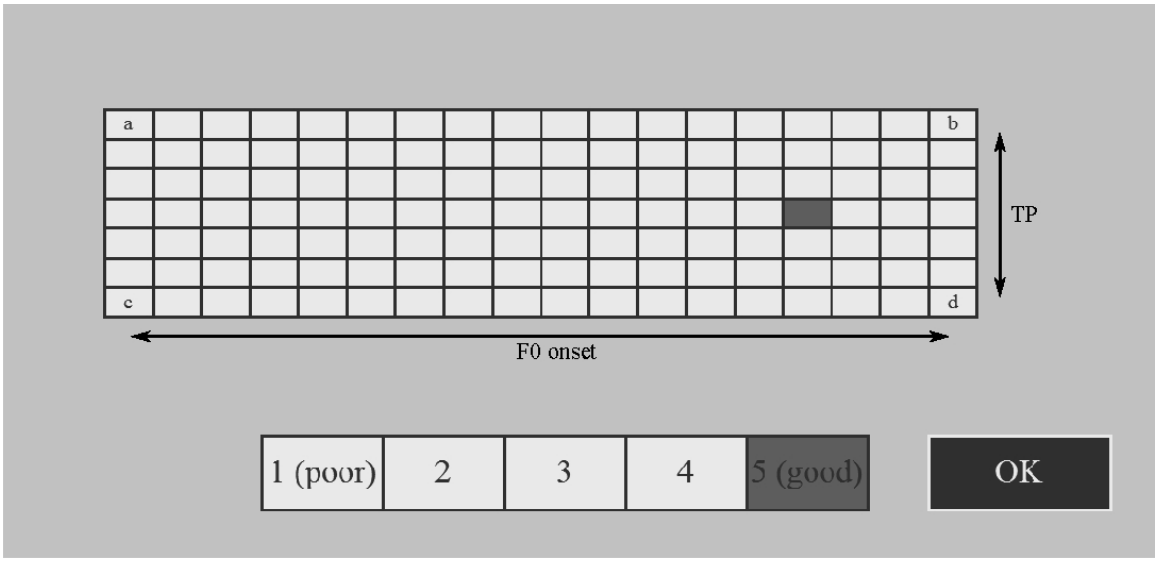


Figure 2.2: A screen capture of the method of adjustment and good rating in the perception experiment. Each box in the stimulus grid was associated to a *zhu* word with one resynthesized tone contour. The F0 onset and TP labels are for illustration and were not displayed in the experiment.

2.3 Results

2.3.1 Production results

The following acoustic measures were based on participants' T2 productions. The participants' mean F0 slope was 4.82*T*/s (range: 1.44 - 8.68*T*/s), representing a (rising contour). The F0 curvature mean was 31.88T/s^2 (range: 8.93 - 77.33T/s^2), indicating an (upward opening parabolic shape). The participants' mean TP was 25% (range 2 - 39%). The mean of F0 mean was 2.32*T* (range: 1.40 - 3.25*T*) and the mean of F0 onset was 2.14*T* (range: 1.52 - 3.16*T*. These values are similar to those obtained from the talker who provided the speech materials and reference values for creating perceptual stimuli. Figure 2.3 displays

the distribution of the production data (as well as perception data, section below) for each acoustic cue.

To determine how well the two polynomial fits explained the variance of the T2 production data, the R^2 values of the model fits were compared. The values were 67% (Standard deviation, or SD: 17%) for F0 slope and 96% (SD: 2%) for F0 curvature. The R^2 values were then submitted to a paired-sample *t*-test, which showed that F0 curvature yielded a significantly higher R^2 value than F0 slope ($t(24) = 7.96$, $p < .001$). Therefore, the variance of the T2 production data was better explained by F0 curvature than F0 slope.

2.3.2 Perception results

The same acoustic features were extracted from the preferred T2 exemplars that participants selected in the perception task. In order to establish that the stimuli were indeed the participants' preferred T2 stimuli, the mean goodness rating scores between these selected stimuli and each of the corner stimuli were compared using paired sample *t* tests. The results showed that the preferred T2 exemplars (Mean = 4.8, SD = 0.52) were significantly rated higher than the four corner stimuli (Mean < 3.05, SD < 0.98) ($ts(24) > 8.79$, $ps < .001$).

The same logarithm-based normalization procedure (Eq. 2.1) was performed to obtain T transformed scores for all F0 measurements. For each participant, the critical and non-critical cues were obtained from their preferred T2 exemplars. Using the same method as in the acoustic analysis of production data, the critical F0 slope and F0 curvature cues of the preferred T2 exemplars were derived using Eq. 2.2 and 2.3. The other critical cue TP was obtained from TP of the preferred T2 exemplars. For non-critical cues, F0 mean was determined using the same method as in the production analysis by averaging the F0 values across all sampling points along the tone contour. F0 onset was obtained from F0 onset of the preferred T2 exemplars. The mean values of the perceived preferred T2 exemplars were $10.52T$/s for F0 slope (range: 4.66 - $13.22T$/s) $18.92T/s^2$ for F0 curvature (range: -0.92 - $49.71T/s^2$), 30% for TP (range: 10 - 53%), $2.56T$ for F0 mean (range: 1.57 - $3.85T$), and $1.02T$ for F0 onset (range: 0.05 - $3.10T$ (Figure 2.3). The polynomial

fits yielded a mean R^2 value of 90% (SD: 7%) for F0 slope and 97% (SD: 2%) for F0 curvature. A significant difference was found between the two R^2 values in a *t*-test ($t(24) = 4.92$, $p < .001$), showing that F0 curvature better explained the variance of the preferred T2 contours in perception than F0 slope.

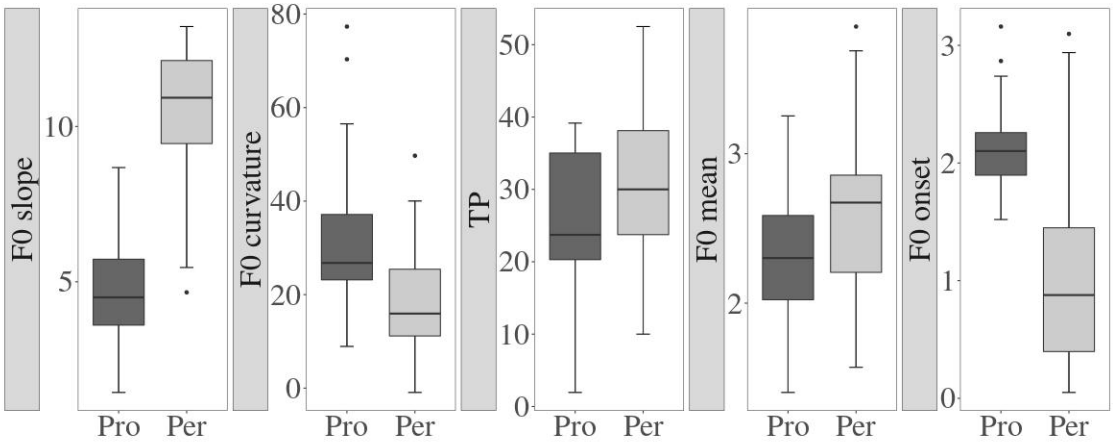


Figure 2.3: Boxplots showing the distribution of F0 slope (in *T*/s, *T* refers to normalized F0). F0 curvature (in T/s^2), TP (in % of total tone duration), F0 mean (in *T*) and F0 onset (in *T*) of Tone 2 stimuli produced by all participants (in dark grey) and the preferred Tone 2 exemplars perceived by all participants (in light grey). The outliers were defined by any point that fell more than 1.5 times of the interquartile range above 3rd quartile or below 1st quartile.

2.3.3 Perception-production relationship

The above data show that the participants produced and perceived Tone 2 with comparable ranges of acoustic data for F0 curvature, TP and F0 mean. However, the F0 slope values of the productions appear to be lower than those of the preferred Tone 2 exemplars in perception. For F0 onset, the data range of the produced Tone 2 is smaller than that of the preferred Tone 2 exemplars in perception.

In order to further quantify the relationship between production and perception as revealed by the current data, a Spearman's rank-order correlation was conducted to relate production data to perception data for each acoustic cue. Non-parametric tests were conducted because outliers were found in our data (Figure 2.3). Based on the assumption that

the perception-production correlation would be positive, one-tailed correlation analysis was conducted.

A significant positive correlation was found for F0 curvature ($\rho(23) = 0.402$; $p = 0.024$), F0 slope ($\rho(23) = 0.378$; $p = 0.032$) and TP ($\rho(23) = 0.391$; $p = 0.027$). No significant result was found for F0 onset ($\rho(23) = 0.080$; $p = 0.352$) and F0 mean ($\rho(23) = -0.022$; $p = 0.543$). Scatter plots of perception-production data with a significant result were shown in Figure 2.4. These results suggest that only critical F0 contour cues (i.e. F0 curvature, F0 slope, and TP) displayed a significant perception-production correlation, even though these cues did not always display comparable ranges of acoustic values from production and perception results (e.g. F0 slope).

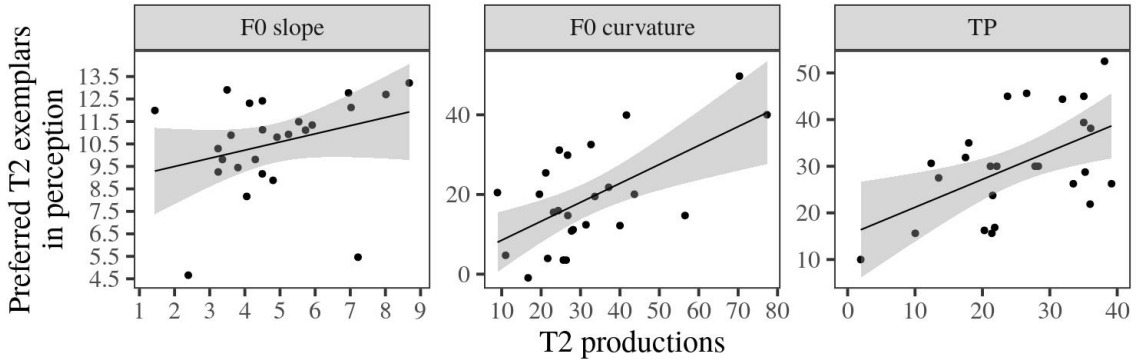


Figure 2.4: Scatterplots of the mean F0 slope (in T/s), F0 curvature in (T/s^2) and TP (in %) of all participants' Tone 2 production and preferred Tone 2 exemplars in perception.

2.4 Discussion

This study examines whether a perception-production relationship can be established using Mandarin tone cues and revealed through perceptually critical acoustic features (Newman, 2003), by comparing the Mandarin tone perception-production correlations established by pitch direction (critical) and height (non-critical) cues (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). The results of the current study revealed significant positive perception-production correlation results, indicating that the two domains can be related for lexical tones through the acoustic cues used in perception and production. Moreover, the results revealed a perception-production link for the perceptually

critical pitch direction cues (F0 curvature, F0 slope) and the temporal cue (TP), but not for the non-critical pitch height cues (F0 mean, F0 onset). This study defines the more strongly weighted pitch direction cues as perceptually more critical than the less strongly weighted F0 height cues based on perceptual cue weighting and tone modelling studies (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Massaro et al., 1985; Tupper et al., 2020). Therefore, more critical perceptual cues contribute to a stronger a perception-production link compared to less critical perceptual cues, supporting the hypothesis of this study (Newman, 2003).

The current results further show that the strength of the perception-production relationship may differ among the critical cues. Among the two possible acoustic correlates of pitch direction, the polynomial fits yielded a higher R^2 value for F0 curvature than for F0 slope, for both participants' Tone 2 productions and their preferred T2 exemplars in perception. Therefore, F0 curvature explains the variance of Tone 2 contours better than F0 slope, presumably because F0 curvature captures greater details of the Tone 2 contour shape than F0 slope (Shih and Lu, 2015; Tupper et al., 2020). Future perception studies should also consider F0 curvature to be a representative cue for F0 contour of Mandarin tones. More importantly, F0 curvature yielded a slightly stronger perception-production correlation than F0 slope (ρ = 0.402 vs. 0.378). In addition, the production data did not show a significant correlation between F0 curvature and F0 slope. Therefore, this finding further extends the previous hypothesis, in that the strength of the perception-production relationship may depend on the level of critical status of cues. That is, those acoustic cues that are more critical in perception exhibit a stronger perception-production link.

Likewise, the temporal feature TP was another cue that showed a significant correlation between perception and production. The critical status of TP can be attributed to its close relationship with other pitch direction cues. Note that TP showed a significant correlation with pitch direction cues in both perception and production, and is also linked to the magnitude of F0 curvature in tone modelling (Shih and Lu, 2015). As a result, the current study shows that critical perceptual cues are adopted from both temporal and spectral domains in establishing perception-production links.

This study also supports the hypothesis that perceptually relevant but non-critical cues exhibit a weak perception-production relationship, as demonstrated by the results for non-critical pitch height cues: F0 mean and F0 onset. Their non-critical status is supported by perceptual weighting studies. In the meantime, the non-critical cues appear to be perceptually relevant since cue weighting studies have shown that pitch height and direction are the two perceptual dimensions of a Mandarin tone perceptual space (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). Additionally, F0 onset also contributes to the perception of Tone 1 and 2 in addition to the primary pitch direction cues (Chang et al., 2016; Massaro et al., 1985). These findings thus suggest that perception-production links are not likely established through non-primary, non-critical cues, even though these cues may contribute to perception to some degree (Jongman et al., 2000; Newman, 2003; Shultz et al., 2012).

Finally, this study focused on the perception-production relationship of one target tone, Tone 2. However, the critical status of pitch direction vs. height cues may change in different Mandarin tone contexts (Massaro et al., 1985), which may then lead to different perception-production relationship patterns for different Mandarin tones. Such patterns may further demonstrate how the effects of the critical status of perceptual cues on perception-production relationships can be aligned with the intrinsic characteristics of individual tones.

2.5 Conclusion

Taken together, this study showed that Mandarin tone cues can exhibit a perception-production relationship. More importantly, perceptually critical pitch direction cues yielded a positive perception-production correlation, whereas non-critical pitch height cues did not, supporting the hypothesis of this study and the previous segmentally-based predictions (Newman, 2003). In addition, the critical status of tonal cues applies in both spectral and temporal domains. These findings extend our understanding of the cues that are pertinent to perception-production links, thus informing the relationship of speech production and perception in general.

Chapter 3

Perception-production relationships and the critical status of Mandarin perceptual cues

3.1 Introduction

The objective of this chapter is to determine the critical status of Mandarin tone perceptual cues for each Mandarin tone and examine the perception-production relationships of critical and non-critical perceptual cues. In Chapter 2, the critical status of perceptual cues was defined by the perceptual cue weightings given to Mandarin tone perception and the acoustic modelling of Mandarin tones. Pitch direction was given a stronger weighting than pitch height for native Mandarin tone perception (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). In addition, pitch direction, or F0 contour, cues showed better tone categorization performance than a combination of F0 mean (i.e., pitch height) and contour cues (Tupper et al., 2020). Therefore, pitch direction was considered to be more critical than pitch height in perceiving Mandarin tones. The results suggest that critical perceptual cues are pertinent to establishing a perception-production link (Newman, 2003). However, the critical status of perceptual cues based on cue weights and acoustic modelling is an assumption that needs to be further investigated since perceptual cue weighting and acoustic modelling imply that the critical status of perceptual cues can be defined by *within-category* and *between-category* perception, respectively, relating to two different levels of perception – phonetic and phonological (refer to section 1.2.2). After determining the critical status of pitch direction and height, the perception-production relationship of each cue

will be examined for each Mandarin tone. The contribution of this chapter is to understand what defines a critical cue for perception and how it can predict a link between perception and production. Specifically, this study examines whether a perception-production link is established by cues that have an impact on an acoustic-phonetic (i.e. within-category) or phonological, speech category (i.e. between-category) level of perception. The perception experiment employs resynthesized stimuli created by orthogonally varying one critical pitch direction (F0 slope) and one non-critical pitch height cue (F0 mean). The acoustic analysis on the tones produced by the same participants involves the same cues. The analysis of perception-production relationships involves a bi-directional modelling across the perception and production by the same participants.

3.1.1 The critical status of perceptual cues and perception-production relationships

Chapter 1 reviewed the critical status of perceptual cues that were defined by *within-category* and *between-category* perception. In Newman (2003), the critical status of perceptual cues was primarily determined by speech category classification results of acoustic modelling which suggested that it was determined by *between-category* perception (Jongman et al., 2000). At the same time, critical perceptual cues were the only cue that influenced the goodness of speech sounds within a speech category, so the critical status of cues could be determined by *within-category* perception (Newman, 2003).

One definition of perceptually critical cues in Chapter 2 is based on the differential cue weightings in Mandarin tone perception. As mentioned in Section 1.3.4, perceptual cue weight results reflect listener's ability in differentiating acoustic-phonetic differences of tones. As a result, the finding that native Mandarin individuals weighting pitch direction more strongly than pitch height suggests pitch direction is a more critical perceptual cue than pitch height in *within-category* perception (e.g. Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Wiener, 2017). On the other hand, another definition is based on the acoustic modelling of Mandarin tones in which F0 slope and curvature (i.e., pitch direction-related cues) showed the best tone category

classification result (Tupper et al., 2020), implying that pitch direction-related cues should also be critical for *between-category* perception.

Apart from overall perception, it is also possible each tone has its own critical perceptual cues. Both cue weighting (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007) and tone categorization (Massaro et al., 1985; Tupper et al., 2020; Yang, 2015) indicated that pitch direction-related cues were important for Tone 2 and 4 perception, but the perception of Tone 1 and 3 require pitch height-related cues. Therefore, it is possible that the critical perceptual cue is pitch direction for Tone 2 and 4 and pitch height for Tone 1 and 3 based on both *within-category* and *between-category* perception.

Therefore, this chapter first examines whether the critical tone perceptual cues (e.g. F0 slope) are defined by *within-category* and/or *between-category* perception. Moreover, it also investigates whether the definition is tone-general so that all tones share the same critical perceptual cue, or tone-specific so that a perceptual cue is critical for certain tone categories but not others. Then, it continues to evaluate the proposal that a perception-production relationship is pertinent to the critical status of perceptual cues (Newman, 2003). The exploration of the definition of critical perceptual cues can extend our understanding on perception-production relationships. Firstly, the theoretical predictions mainly focus on the phonological level of perception-production relationships as reviewed in section 1.2.2 (Diehl et al., 2004; Guenther, 1995; Liberman and Mattingly, 1985). The result of this chapter demonstrates whether within-category, phonetic level of perception has an impact on the relationship between the two domains. Secondly, there is still no clear evidence from perception studies to determine whether pitch direction influences Mandarin tone perception independently and more strongly than pitch height for overall and tone-specific perception. The result of this chapter indicates whether a perception-production relationship is specific to individual speech categories.

3.1.2 Outline of the present study

This chapter first reports a Mandarin tone perception study that employs tone stimuli resynthesized by orthogonally varying F0 slope (pitch direction-related) and F0 mean (pitch height-related), a critical and a non-critical cue, respectively, as defined by previous studies (e.g. Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Jongman et al., 2017; Massaro et al., 1985; Tupper et al., 2020). This orthogonal variation of critical and non-critical cues enables us to examine the impact of each cue on *between-category* tone categorization and *within-category* goodness ratings, and subsequently, the relationship between perception and production for each cue in order to examine whether such relationship is pertinent to critical perceptual cues. So far, there has only been a study on the JND of tones that orthogonally varied F0 slope and mean (Jongman et al., 2017), but it did not investigate the effects of each cue on Mandarin tone perception, goodness ratings, or the perception-production link. Although Chapter 2 used a goodness rating task, it was not designed to test whether only the critical pitch direction cues were used to determine the perceived quality of Tone 2. For this study, the predictions about the perception-production relationship for each tone are guided by the results of tone categorization and goodness ratings, depending on whether each tone has a different set of critical and non-critical perceptual cues.

For production data, the tone productions were obtained from the same participants. Acoustic modelling was performed on the four Mandarin tone productions to measure the same cues of each tone (i.e., F0 slope and F0 mean), matching the cues varied in the perceptual stimuli of the perception experiment so that the perception and production cues that entered into the subsequent perception-production analysis were comparable. As a result, both perception and production data contained a critical (F0 slope) and a non-critical perceptual cue (F0 mean).

To investigate the relationship between perception and production, this study adopts a statistical learning approach that compares the predicted outcomes of the logistic regression models trained and tested by data from each domain. A logistic regression was used to reveal participants' mapping between acoustic cues and speech categories in produc-

tion data (McMurray and Jongman, 2011; Nearey, 1990, 1997; Redmon et al., 2020). Then, these studies qualitatively compare the categorical performance of the model with the human perceptual accuracy of speech categories, which provides insights about the acoustic cues contributing to the categorization in the perceptual process. This study adopts the same technique to identify the cues contributing to the link between perception and production.

However, this study has two major differences compared to previous work. First, instead of using naturally produced speech sounds as perceptual stimuli, this study uses resynthesized tone stimuli by varying F0 slope (critical perceptual cue) and F0 mean (non-critical perceptual cue) orthogonally. As mentioned earlier, this design can explore the effect of each cue on *within-category* and *between-category* perception. In addition, to relate the perception performance to production data, the logistic regression classifiers trained by production data of this study should also model the perceptual process as in the previous studies. Therefore, the models can be used to classify the perceptual stimuli used in the perception experiment and compare the models' categorization performance with the actual perceptual categorization performed by the same group of human participants. Second, in previous studies, the speakers whose productions used to train and test the logistic regression model did not participate in the perception task as perceivers, but the participants in this study took part in both production and perception tasks. This difference enables the current study to investigate the link between the perception and production systems of the same group of individuals, achieving the goal of this study. The perception-production link should be demonstrated if the categorization performance of the logistic regression models match the perceptual categorization of same set of perception stimuli. By adjusting the cues used to train the models, the categorization performance should provide insights about whether a critical perceptual cue can establish a perception-production link (Newman, 2003). Likewise, logistic regression classifiers trained by the perception data should also inform us about the perceptual process of tone categorization. Using the perception model to categorize production data should also inform us about how perception and production are related. The same assumption about perception-production links still

applies, that is, the model should be able to correctly categorize the items according to the intended tones produced by participants using a critical perceptual cue. This method allows the use of individual perception responses and production items by the same participants to train and test the models, which complements the perception-production correlation analysis for each cue used in Chapter 2 and other previous work (e.g Beddor, 2015; Newman, 2003), which only included averaged values per participant.

Based on the past work discussed in this section and Chapter 2, the general hypotheses for this study are as follows:

1. Based on the acoustic-phonetic perception perspective, the critical status of perceptual cues is defined by *within-category* perception. The previous cue weighting studies (Chandrasekaran et al., 2010; Gandour, 1983; Guion and Pederson, 2007) suggest that the critical cue, F0 slope, should modulate the *within-category* goodness ratings of each tone category. The non-critical cue, F0 mean, should either have no effect, as predicted by Newman (2003), or a smaller effect as predicted by cue weighting studies (e.g. Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007) on this level of tone perception.

2. Alternatively, assuming that acoustic modelling results based on production have an implication on the critical status of perception cues (Jongman et al., 2000; Newman, 2003), the critical cues should influence *between-category* tone categorization since acoustic modelling explores the cues that influence the phonological categorization of speech sounds.

3. The *within-category* and *between-category* perception both indicate that the critical status of perceptual cues can be either tone-general or tone-specific (Chandrasekaran et al., 2010; Gandour, 1983; Guion and Pederson, 2007; Tupper et al., 2020). That is, a tone-general view predicts that all tones should have the same perceptually critical cue - F0 slope, so that this cue should display a perception-production relationship for all tones. Alternatively, the tone-specific view predicts that each tone should have its own set of critical and non-critical perceptual cues. F0

slope should be critical for the perception of Tone 2 and 4, and F0 mean for Tone 1 and Tone 3. Therefore, F0 slope should show a perception-production link for Tone 2 and 4, and F0 mean for Tone 1 and 3, following the tone-specific view.

3.2 Method

3.2.1 Participants

There were 33 participants in this study (Female = 26; Male = 8) who were native Mandarin speakers attending the undergraduate and graduate programs at Simon Fraser University between the age of 18 and 30 (Mean: 22.8). They were raised in China for the first 12 years of life. A small number of speakers also had knowledge of another Chinese dialect (e.g. Shanghainese), but they all learned Mandarin first and mainly spoke Mandarin in their daily life (reported over 90% of daily use among all Chinese dialects they speak). English was an additional language that the participants spoke on a daily basis. All participants took part in the production task and perception experiment described below.

3.2.2 Production task

The production task followed part of the procedures of Tupper et al. (2020).

Materials

The production items were the Mandarin read monosyllabic words /ɤ/, /i/ and /u/ (*e*, *yi* and *wu* in *pinyin*) with four Mandarin tones, carrying the meaning of “graceful” (/ɤ1/; Tone 1), “goose” (/ɤ2/; Tone 2), “nauseous” (/ɤ3/; Tone 3), “hungry” (/ɤ4/; Tone 4), "clothing" (/i1/), "move" (/i2/), "chair" (/i3/), "translate" (/i4/), "dirty" (/u1/), "none" (/u2/), "dance" (/u3/) and "error" (/u4/). Each tone word item was produced in isolation for 5 times in each speaking style - plain or clear. Therefore, each participant produced a total of 120 items (3 monosyllable $\times$ 4 tones $\times$ 5 repetitions $\times$ 2 speaking styles).

Procedures

Participants took part in two blocks of production task. In each trial, one word, displayed in both Chinese character and *pinyin*, was presented in the centre of the screen each

time. They were asked to produce the items naturally in the first block (i.e. plain speech), and clearly, as if they were talking in a noisy environment, in the second block (i.e. clear speech). As in Chapter 2, plain and clear productions were elicited since clear productions would possibly demonstrate a stronger perception-production relationship than plain productions (Johnson et al., 1993; Yang and Whalen, 2015).

The task was self-paced, and participants could repeat if they made any production mistake. The same recording setting and equipment as described in 2.2.2 were used.

Acoustic measurements and modelling

As in section 2.2.2, all productions were analyzed acoustically in Praat (Boersma and Weenink, 2018). The tone contour was measured from the beginning and ending of waveform periodicity of each item to obtain the duration of each tone. The tone contour was divided into 10 intervals of equal distance. F0 values in Hertz were then obtained at the 11 equidistant time points along a tone contour. The F0 values were manually checked for accuracy by the author and phonetically trained research assistants. Any inaccurate or missing F0 data points were manually measured by taking the inverse of the duration of a single period. The data points that were not measurable were removed and treated as missing data (due to aperiodic cycles caused by creaky voice). The number of sampling points was fewer than 101 points used in chapter 2 and Tupper et al. (2020) because this study involved a number of creaky voice productions in Tone 3. The number of sampling points were also reduced by one-tenth if any manual measurements were required in Tupper et al. (2020). Moreover, the use of 11 sampling points was consistent with previous tone acoustic studies in which 10 or fewer sampling points were used (Khouw and Ciocca, 2007; Mok et al., 2013; Rose, 1987; Wang et al., 2003; Wong et al., 2017). To normalize for inter-speaker pitch range differences, Eq. 2.1 was used to convert the F0 values to logarithm-based *T* values (Wang et al., 2003).

To model the F0 mean and slope values of each tone item produced by the participants, the normalized F0 values at the 11 sampling time points were assigned to Eq. 3.1 with normalized time *t* falling in the interval of $0 \leqslant t \leqslant 1$ (*cf.* Tupper et al., 2020):

$$f(t) \approx c_0 + c_1(t - \frac{1}{2}) \tag{3.1}$$

where c_0 and c_1 were the values of F0 mean, the non-critical cue, and slope, the critical cue, respectively. As mentioned in section 2.2.2, F0 slope of each tone contour was then divided by the duration of that contour (in s) to obtain a time-scaled slope. A positive and negative F0 slope value indicate a rising and falling contour, respectively.

3.2.3 Perception experiment

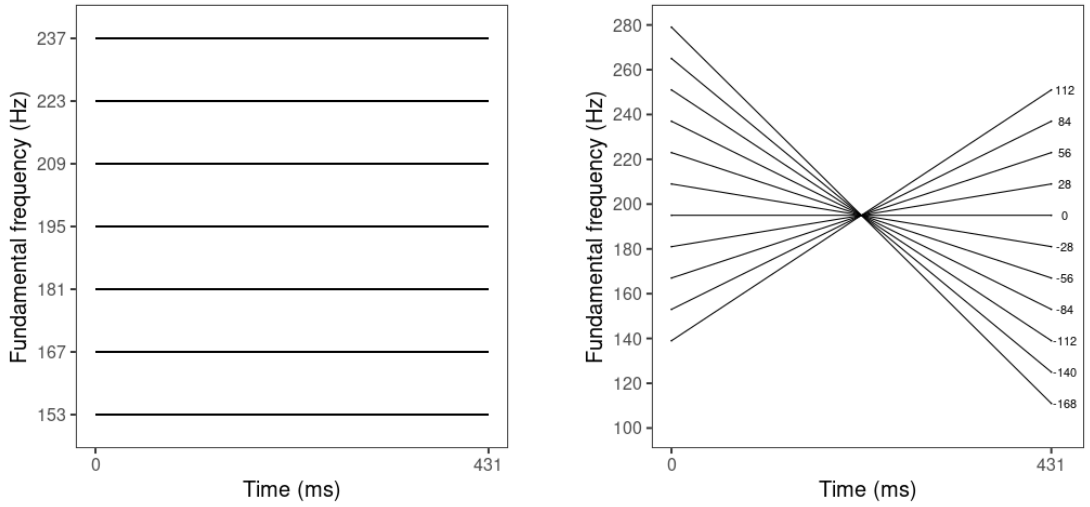
Stimuli

The stimuli were Mandarin syllable /ɤ/ with superimposed resynthesized and natural tone contours provided by 2 native Mandarin talkers who were participants of Tupper et al. (2020) (1 male, 1 female). Resynthesized tone contours were used to determine the critical status of each perceptual cue and the natural tone contours served as references. The tone contour resynthesis was carried out by varying F0 mean (7 steps) and F0 slope (11 steps) orthogonally. The parameters of the resynthesized tone contours were based on the talkers' tone productions. Each talker produced 98 instances of Mandarin words /ɤ/ with four tones, of which half were in plain speech and the other half in clear speech [(11 Tone 1 + 12 Tone 2 + 15 Tone 3 + 11 Tone 4) $\times$ 2 styles]. Two phonetically-trained native Mandarin listeners rated the goodness of each talker's productions as a member of the intended category on a scale of 1 (poor) to 5 (excellent) in the evaluation task. The tone contours that were creaky or rated below 3 in an evaluation task were excluded. Eventually, 96 productions of the male talker and 69 productions of the female talker were retained for the subsequent acoustic modelling.

The same modelling method as in the section 3.2.2 was applied to these productions to obtain the F0 mean and slope values of each tone contour. The endpoints of the resynthesized tone contours' F0 mean and slope values were based on the mean ranges of F0 mean (c_0) and slope values (c_1) across two talkers and two styles. Based on these productions, F0 mean ranged from 153 to 237Hz (Fig 3.1a) and F0 slope ranged from 112 to $-$168Hz/*normalized time* (Fig 3.1b), forming the endpoints of these parameters.

Intermediate steps between the F0 mean endpoints were determined by a step size of 14 Hz, the maximum JND value for native Mandarin listeners (Jongman et al., 2017) (Fig 3.1a, F0 mean values between 167 and 223Hz). For the F0 slope series, the onset and offset values of the F0 slope endpoints were derived from Eq. 3.1 (at $t = 0$ and 1). Then, the onset and offset were adjusted in opposite directions in 14 Hz steps to obtain the onset and offset values of the intermediate steps. Finally, their slope values were derived from Eq. 3.1 (Fig 3.1b, F0 slope values between 84 to -140Hz/*normalized time*).

The tone stimuli were resynthesized using PSOLA in Praat (Boersma and Weenink, 2018). For each talker, one excellent tone production (rated 5 by both evaluators) was selected. First, the tone duration was modified to 431ms, the mean tone duration across 2 talkers. Then, all the tone sampling points except the onset and offset were removed to create a linear tone contour, and the onset and offset values were adjusted to create flat contours corresponding to the F0 mean values of the 7 steps of F0 mean (Figure 3.1a). After that, for each F0 mean step, the onset and offset were manipulated in opposite directions to create the 11-step F0 slope series (Figure 3.1b). Finally, all tone stimuli were scaled to 70dB in intensity.



(a) F0 mean series (F0 slope = 0Hz/*t*); F0 mean values are displayed on the y-axis

(b) F0 slope series (F0 mean = 195Hz); F0 slope values are displayed on the right of each tone

Figure 3.1: Schematic representations of perceptual stimulus series

For the natural tone stimuli, one item for each tone and speaking style was selected from each talker. All items were free of creak and rated as a good production (Goodness ratings above 4) by both evaluators. The tone duration was modified to 431ms and the intensity was scaled to 70dB, matching those of the resynthesized tone stimuli. The final stimulus set contained 170 tone stimuli [(7 F0 mean steps x 11 F0 slope steps + 4 natural tones x 2 styles) x 2 talkers]

Procedures

All participants took part in the perception experiments after completing the production task. All stimuli were presented twice (Total number of stimuli was 240). The stimuli were arranged in 4 blocks. Each block contained one repetition of the stimuli of one talker.

The experiment involved a four-alternative forced-choice identification and a goodness rating task. In each trial, one stimulus was presented to the participant. Then, they had to determine the tone category of the stimulus (T1, T2, T3, or T4). This force-choice identification task aimed to examine how each F0 cue influenced tone categorization. Following the identification task, they rated the quality of the stimulus as a member of the selected tone category on a scale of 1 (poor) to 5 (excellent). Participants were encouraged to complete each task as quickly as possible, and the identification and goodness rating task each had a time limit of 4 seconds. A goodness rating task was included to test whether listeners would use a critical perceptual cue to determine which particular token “sound better” and thus the cue “appeared to be related to listeners’ goodness ratings”, according to Newman (2003, p.2857). The natural and resynthesized tone stimuli were mixed and randomized. The block orders were counterbalanced. The design of this experiment included linear tone contours only so as to orthogonally vary one critical (F0 slope) and one non-critical cue (F0 mean) for testing the hypothesis of this study. The stimuli, however, did not represent the typical dipping contour of Tone 3. The slightly falling tone contours in the mid-low F0 range could resemble the reduced form of Tone 3 that would appear in connected speech. Therefore, the number of Tone 3 responses and the rating scores of the perceived Tone 3 items were expected to be low.

3.3 Results

This data analysis is covered in four sections. To understand the critical status of F0 dimensions in perception, the first section of the analysis examines the impact of F0 mean and slope on tone categorization (between-category perception) and goodness ratings (within-category perception) in a multinomial and ordinal logistic regression (Section 3.3.1). Then, the second and third sections report an investigation of perception-production relationships through a statistical learning method with multinomial logistic regression in two directions. The perception and production data [1] were used to train the models separately and the test data from the other domain were submitted to the model to predict tone categories. As a result, models trained by production tone category data were used to classify perceptual tone stimuli used in the perception experiment (Section 3.3.2), and models trained by perception tone response data were used to classify production tone categories obtain in the production task (Section 3.3.3). The perception data used in these two sections were based on the between-category tone categorization results. Finally, the fourth section further investigates the perception-production relationship using the perceptual tone categorization data containing good exemplars as determined by the goodness rating results. In other words, the perception data were determined by both the between-category tone categorization and within-category goodness rating results in the perception experiment. The production models reported in the second section were used to classify these perception data. Before the data were analyzed, outliers were first removed. For the resynthesized tone stimuli, participants occasionally identified a tone with an unlikely F0 dimension value (e.g. a Tone 2 with am extreme falling slope). These data were presumed to be error responses. To handle these situations, data points that were 2 standard deviations away from the mean for each F0 dimension, speaker and tone response were treated as outliers and were removed from the data set. All analyses were performed in R (R Core Team, 2021) and all figures were created using the *ggplot2* package (Wickham, 2016).

[1] Clear and plain data were included in Chapter 3 and 4 since the two production types did not show any difference in perception-production relationship a preliminary data examination

3.3.1 Examining the critical status of perception cues

In this section, the critical status of F0 mean and slope were studied in two analyses. The first analysis investigated the impact of F0 mean and slope on the tone categorization. The second analysis examined the effect of these two cues on the goodness ratings of each tone. The objective of these analyses was to understand how the critical and non-critical cues influence *between-category* tone categorization and *within-category* goodness rating tasks.

Effects of F0 mean and slope on tone categorization: A multinomial logistic regression analysis

For the first analysis, the tone categorization responses of all participants for each level of F0 mean and slope value are displayed in Figure 3.2 (Total number of responses = 10216). It shows a clear categorization of Tone 1, 2 and 4 along the F0 slope dimension. Tone 1 responses cluster around the region where F0 slope is at 0 Hz/t, which corresponds to relatively flat tone contours. Tone 2 and Tone 4 responses occupy the positive and negative slope regions, respectively, which represent rising and falling slopes. These results are all consistent with the tone contour shapes of these tones. Tone 3 responses are more scattered but are mostly found in the slow falling slope region with a small-to-mid negative slope value (-28 to -112 Hz/*t*). This is consistent with the reduced tone contour shape of Tone 3 in running speech.

In contrast, the F0 mean dimension appear to serve as a weak cue for tone categorization since every F0 mean level contains responses of all four tones. It only appears that it interacts with F0 slope to influence tone perception. In the low F0 mean region, the plot shows more Tone 2 and Tone 4 responses for near-flat slopes than in the high F0 mean region (28 and -28 Hz/t). Moreover, an increased number of falling slope items (negative F0 slope value) were categorized as Tone 3 in the low F0 mean region compared to the high F0 mean region.

To further examine the tone categorization results, a multinomial logistic regression, performed with the *nnet* package (Venables and Ripley, 2002), was used to explore the

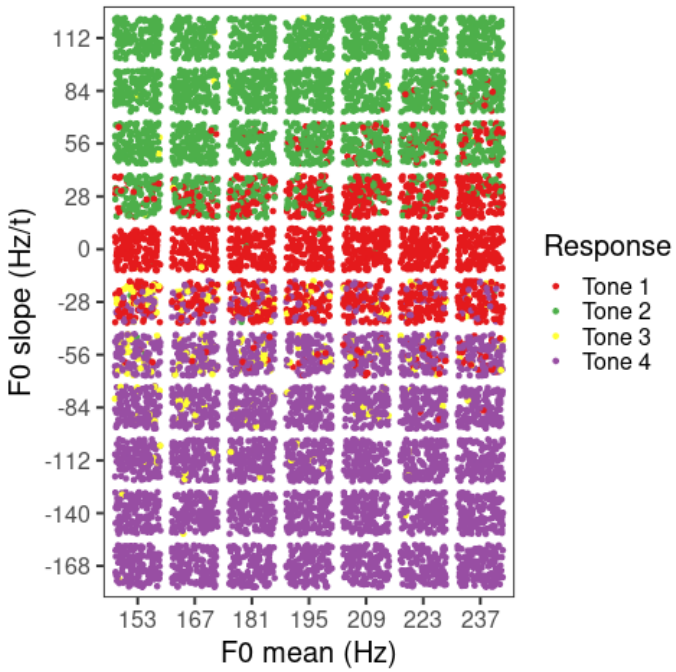


Figure 3.2: Tone responses of all participants in perception task for each F0 mean and slope level

influence of F0 mean and slope on tone responses. In the model, F0 mean (153-237Hz) and F0 slope (112 to -168Hz/t) were the predictor variables and the dependent variable was tone identification responses (Tone 1, Tone 2, Tone 3, Tone 4) using the following equations in R syntax.

Full model with interaction:

Tone Identification Responses $\sim$ F0 mean $*$ F0 slope

Full model without interaction:

Tone Identification Responses $\sim$ F0 mean $+$ F0 slope

F0 slope-only model:

Tone Identification Responses $\sim$ F0 slope

F0 mean-only model:

Tone Identification Responses $\sim$ F0 mean

Null model (with intercept only):

Tone Identification Responses ~ 1

To examine the main effects of F0 mean and F0 slope, and interaction effects between the two factors, the Deviance statistic (D) (-2 $\times$ log-likelihood of the model) was used to compare model fit qualitatively (*cf.* McMurray and Jongman, 2011; Redmon et al., 2020). Five models were created: A null model containing only the intercept (D = 23388), two single predictor models with either F0 mean (D = 23183) or F0 slope (D = 6279), a model with the linear combination of two predictors (D = 5888), and a model containing the interaction term between the two predictors (D = 5866). All models had Tone 1 as the reference level. Comparing the single predictor models with the null model, F0 mean alone only improved the model fit slightly (Difference in D or ΔD = 205), whereas F0 slope alone improved tremendously (ΔD = 17109). Moreover, the F0 mean only and the F0 mean $+$ F0 slope model differed by a huge difference (ΔD = 17295), whereas the difference between the F0 slope only and the two-predictor model was only 391 in ΔD. These show that perceptual tone categorization was explained mainly by F0 slope. The models with and without the interaction term had very similar model fit (ΔD = 22) so the contribution of the interaction term to the model is minimal. However, since a comparison of model fit still displayed a significant interaction of F0 mean and F0 slope ($\chi^2(3) = 22.0$, $p < 0.001$), this interaction effect was further investigated by follow-up tests with separate multinomial logistic regression models that examined the effect of one predictor on tone categorization for each level of the other predictor. The reference level changed from Tone 1 to Tone 2, then to Tone 3 to obtain results for each tone pair.

Table 3.1 and 3.2 display the results of the models. F0 slope consistently had an effect on tone categorization at each F0 mean level for all tone pairs and the results follow the expected pattern (Table 3.1). Specifically, the positive β estimates of the Tone 1 vs. Tone 2 comparisons showed that, at all F0 mean levels, the odds of categorizing a tone stimulus as a Tone 2 with a rising and positive slope rose as F0 slope increased, compared to Tone 1 with a flat contour and near zero slope value, as expected. Participants were more likely to perceive a tone contour as Tone 3 or 4 rather than Tone 1 as the F0 slope value lowered. These results were also expected since Tone 4 should have a falling contour and negative slope value, while the reduced form of Tone 3 in connected speech should also

have a slight falling contour and negative value. These patterns were further confirmed by the results of the Tone 3 vs. Tone 4 comparison which demonstrated that participants tended to perceive a tone stimulus as a Tone 4 rather than a Tone 3 when the F0 slope value decreased, or became more negative. These results were consistent with Figure 3.2 in that, as F0 slope varied from most positive to most negative across all F0 mean levels, the tone categorization changed from Tone 2 to Tone 1 at the near-zero slope levels. Then, the categorization changed from Tone 1 to a mix of Tone 3 and 4, and to Tone 4 only in the negative slope region.

For the effect of F0 mean on each F0 slope level (Table 3.2), F0 mean had an influence on tone categorization for a limited number of tone pairs only for most of the F0 slope levels. In the positive slope region, F0 mean was mainly used for differentiating Tone 1 from Tone 2 and Tone 3. For F0 slope levels 84, 56 and 28Hz/t, the β estimates were negative for Tone 1 vs. Tone 2 and Tone 1 vs. Tone 3 pairs, indicating that, at these rising slope levels, tone stimuli were more likely to be categorized as a Tone 2 or 3 than Tone 1 as F0 mean decreased. The models did not detect an effect of F0 mean on Tone 2 vs. Tone 3 categorization, except for the F0 slope level of 28Hz/t (i.e. slightly rising) where tone stimuli were more likely to be perceived as Tone 3 than Tone 2 as F0 mean decreased. These results aligned with the observation from the positive slope region of Figure 3.2 where more stimuli were categorized as Tone 2 as F0 mean decreased. A small number of Tone 3 responses were found in this positive slope region (4-6 responses per F0 slope level) and these were found more often at the low F0 mean region.

In the zero to negative slope region, the odds for categorizing a stimulus as Tone 1 became higher as F0 mean increased between -28 and -84Hz/t, as demonstrated by the negative β estimates for all tone pairs with Tone 1 as the reference level. The results were also consistent with the data visualized in Figure 3.2 where more Tone 1 responses were found as the F0 mean level went from low to high. At the F0 slope level of -56 and -84Hz/t, participants were more likely to identify a stimulus as a Tone 4 rather than a Tone 3 as the F0 mean level increased, with a positive β estimate for the Tone 3 vs Tone 4 pair. Figure 3.2 also illustrated that slightly more Tone 4 responses were found in the high than the low

F0 mean region. F0 mean also influenced Tone 2 perception at the slope level of -56Hz/t. It was more likely for a stimulus to be perceived as a Tone 2 rather than Tone 3 or 4 as F0 mean lowered. An effect of F0 mean was not found for the flat (0Hz/t) and the steeper falling contours (-112 $-$ -168Hz/t).

Table 3.1: Multinomial logistic regression of F0 slope on tone categorization for each level of F0 mean (The underlined tone represents the reference level in the model)

		Tone pairs:					
F0 Mean		T1 vs T2	T1 vs T3	T1 vs T4	T2 vs T3	T2 vs T4	T3 vs T4
237Hz	β	0.078***	-0.034***	-0.108***	-0.112***	-0.186***	-0.074***
	(SE)	(0.006)	(0.010)	(0.009)	(0.011)	(0.010)	(0.011)
	t	13.51	-3.51	-12.45	-9.88	-17.79	-6.47
	p	<.001	<.001	<.001	<.001	<.001	<.001
223Hz	β	0.084***	-0.049***	-0.094***	-0.133***	-0.178***	-0.044***
	(SE)	(0.006)	(0.008)	(0.007)	(0.010)	(0.009)	(0.007)
	t	13.33	-6.21	-13.44	-13.07	-18.82	-6.01
	p	<.001	<.001	<.001	<.001	<.001	<.001
209Hz	β	0.096***	-0.068***	-0.110***	-0.163***	-0.206***	-0.042***
	(SE)	(0.007)	(0.009)	(0.009)	(0.012)	(0.011)	(0.007)
	t	12.84	-7.50	-12.81	-13.89	-18.01	-6.25
	p	<.001	<.001	<.001	<.001	<.001	<.001
195Hz	β	0.097***	-0.059***	-0.105***	-0.155***	-0.201***	-0.046***
	(SE)	(0.007)	(0.008)	(0.008)	(0.011)	(0.011)	(0.007)
	t	12.60	-7.01	-12.92	-13.60	-17.94	-6.70
	p	<.001	<.001	<.001	<.001	<.001	<.001
181Hz	β	0.115***	-0.081***	-0.125***	-0.196***	-0.239***	-0.044***
	(SE)	(0.011)	(0.010)	(0.010)	(0.015)	(0.015)	(0.006)
	t	10.85	-7.77	-12.85	-13.10	-16.09	-6.74
	p	<.001	<.001	<.001	<.001	<.001	<.001
167Hz	β	0.110***	-0.063***	-0.100***	-0.173***	-0.210***	-0.037***
	(SE)	(0.010)	(0.008)	(0.008)	(0.013)	(0.013)	(0.005)
	t	11.07	-8.26	-12.88	-13.64	-16.46	-7.70
	p	<.001	<.001	<.001	<.001	<.001	<.001
153Hz	β	0.109***	-0.067***	-0.103***	-0.176***	-0.212***	-0.036***
	(SE)	(0.010)	(0.008)	(0.008)	(0.013)	(0.013)	(0.004)
	t	11.34	-8.47	-12.58	-13.91	-16.53	-8.38
	p	<.001	<.001	<.001	<.001	<.001	<.001

Note: $^{\dagger}p$<0.1; $^{**}p$<0.05; $^{***}p$<0.01

Table 3.2: Multinomial logistic regression of F0 mean on tone categorization for each level of F0 slope (The underlined tone represents the reference level in the model)

F0 Slope		*Tone pairs:* T1 vs T2	T1 vs T3	T1 vs T4	T2 vs T3	T2 vs T4	T3 vs T4
112Hz/t	β	–	–	–	0.016	–	–
	(SE)	–	–	–	(0.016)	–	–
	t	–	–	–	1.04	–	–
	p	–	–	–	.296	–	–
84Hz/t	β	-0.058***	-0.071**	–	-0.013	–	–
	(SE)	(0.015)	(0.022)	–	(0.017)	–	–
	t	-3.95	-3.18	–	-0.787	–	–
	p	<.001	.001	–	.431	–	–
56Hz/t	β	-0.041***	-0.065**	–	-0.024	–	–
	(SE)	(0.005)	(0.020)	–	(0.020)	–	–
	t	-8.66	-3.26	–	-1.24	–	–
	p	<.001	.001	–	.216	–	–
28Hz/t	β	-0.039***	-0.069***	–	-0.031***	–	–
	(SE)	(0.003)	(0.003)	–	(0.003)	–	–
	t	-12.92	-22.89	–	-10.23	–	–
	p	<.001	<.001	–	<.001	–	–
0Hz/t	β	-0.004	–	–	–	–	–
	(SE)	(0.016)	–	–	–	–	–
	t	-0.233	–	–	–	–	–
	p	.816	–	–	–	–	–
-28Hz/t	β	-0.021***	-0.025***	-0.015***	-0.003	0.007†	0.009†
	(SE)	(0.003)	(0.005)	(0.003)	(0.004)	(0.004)	(0.005)
	t	-7.61	-5.09	-5.18	-0.636	1.92	1.88
	p	<.001	<.001	<.001	0.525	0.055	0.060
-56Hz/t	β	-0.054***	-0.037***	-0.025***	0.018***	0.030***	0.012***
	(SE)	(0.003)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
	t	-16.65	-6.26	-5.55	4.057	7.98	2.96
	p	<.001	<.001	<.001	<.001	<.001	<.001
-84Hz/t	β	–	-0.062***	-0.040***	–	–	0.022***
	(SE)	–	(0.004)	(0.003)	–	–	(0.006)
	t	–	-14.68	-12.04	–	–	3.46
	p	–	<.001	<.001	–	–	<.001

Note: –: at least one tone response absent;$^{\dagger}p$<0.1; $^{**}p$<0.05; $^{***}p$<0.01

Table 3.2 – continued from previous page

F0 Mean		*Tone pairs:* T1 vs T2	T1 vs T3	T1 vs T4	T2 vs T3	T2 vs T4	T3 vs T4
-112Hz/t	β	–	–	–	–	–	0.013
	(SE)	–	–	–	–	–	(0.009)
	t	–	–	–	–	–	1.51
	p	–	–	–	–	–	.131
-140Hz/t	β	–	–	–	–	–	0.034†
	(SE)	–	–	–	–	–	(0.020)
	t	–	–	–	–	–	1.66
	p	–	–	–	–	–	.097
-168Hz/t	β	–	–	–	–	–	–
	(SE)	–	–	–	–	–	–
	t	–	–	–	–	–	–
	p	–	–	–	–	–	–

Note: –: at least one tone response absent; $^{\dagger}p$<0.1; $^{**}p$<0.05; $^{***}p$<0.01

Effects of F0 mean and slope on goodness ratings: An ordinal logistic regression analysis

The second analysis examined the effects of F0 mean and slope on goodness rating scores for each tone separately. Figure 3.3 displays the mean rating scores for each tone at each F0 mean and slope level. The score of 5 in red and 1 in yellow represent good and poor exemplars of a tone, respectively. The plots showed that the stimuli identified as Tone 2 were given a higher rating score as the slope increased. Tone 1 also appeared to show this trend, but was less obvious than Tone 2. In contrast, participants gave a higher score to stimuli with a more negative slope if they were categorized as a Tone 4. The scores given to Tone 3 items were more dispersed. The more positive or negative slope levels seemed to receive higher rating scores than slope level closer to zero.

The data were further analyzed in a ordinal logistic regression for each tone separately with the *MASS* package (Venables and Ripley, 2002). The model fit was compared using D and the same models as in the previous multinomial logistic regression analysis were included, except for the model with the interaction between F0 height and F0 slope due to

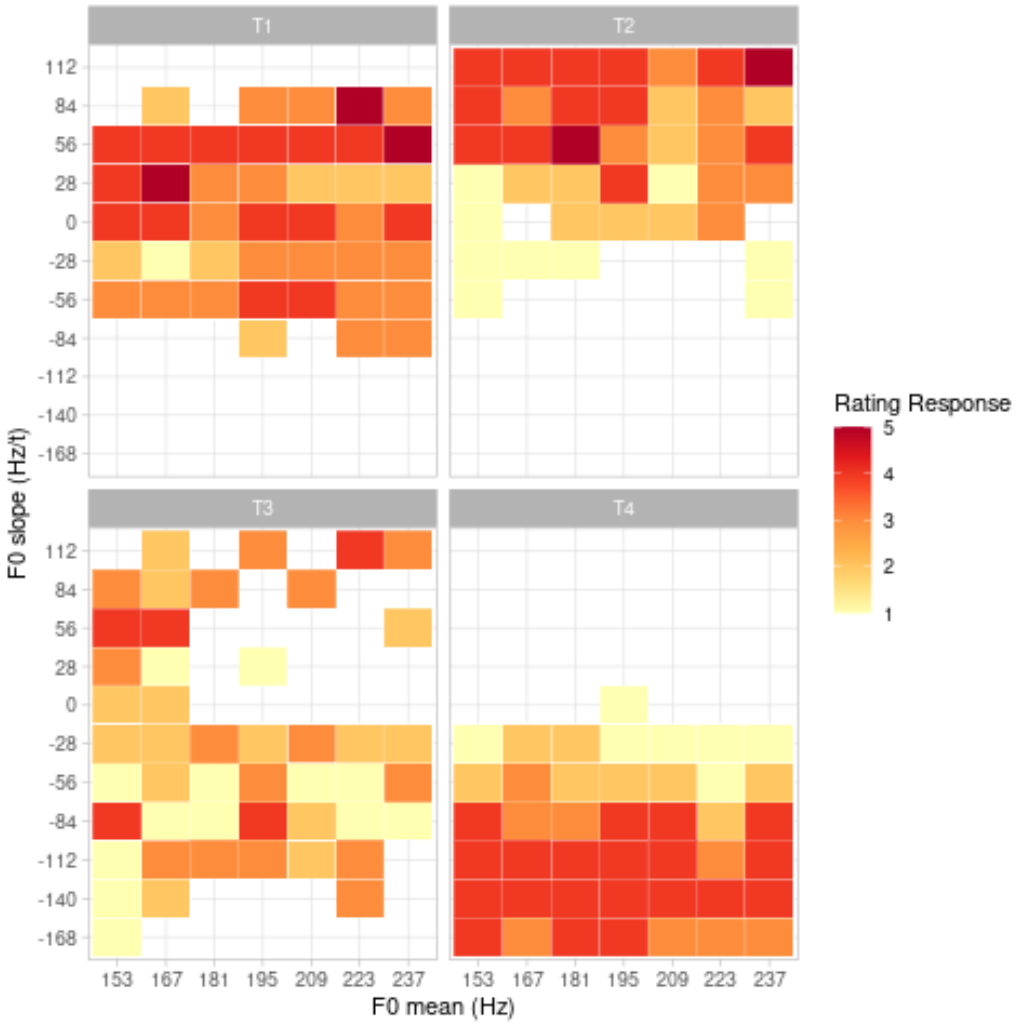


Figure 3.3: Heat plot of mean rating scores for each tone

convergence issues. The dependent variable was the rating scores (1,2,3,4,5, in which 1 was a poor exemplar and 5 was a good exemplar, respectively) (see the following equations for models in R syntax).

Full model:

Rating Scores $\sim$ F0 mean $+$ F0 slope

F0 slope-only model:

Rating Scores $\sim$ F0 slope

F0 mean-only model:

Rating Scores $\sim$ F0 mean

Null model:

Rating Scores ~ 1

The model results are shown in Table 3.3. Comparing the single predictor models with the null model, F0 mean showed a smaller ΔD than F0 slope for all tones. Comparing the one-predictor with two-predictor models, F0 mean had a larger ΔD than F0 slope for all tones. These results show that F0 slope was a stronger contributing factor for Mandarin tone goodness ratings than F0 mean. Note that the ΔD between null and F0 mean-only model for Tone 4 was exceptionally small, suggesting that the contribution of F0 mean to Tone 4 goodness ratings was minimal. The ΔD for all Tone 3 model comparisons were also very small, indicating that these two factors did not have a clear contribution to Tone 3 goodness ratings. Regarding Tone 1 and 2, the general pattern of ΔD holds. The F0 mean-only model had a much smaller ΔD than the F0 slope-only model compared to the null model. Unlike the visual inspection in which F0 slope appeared to have a smaller contribution to goodness ratings for Tone 1 than for Tone 2, F0 slope had a strong influence on goodness ratings for both tones, and the impact was comparatively much higher than F0 mean.

To further investigate the influence of F0 mean and slope on goodness ratings, the results of the two-predictor regression model presented in Table 3.4 were examined in *z*-tests. Consistent with the model fit comparisons, the β estimates of F0 mean and slope showed a significant result for all tones except for Tone 3, indicating that they both influ-

Table 3.3: Deviance statistic D of ordinal logistic regression models

Model	Tone 1	Tone 2	Tone 3	Tone 4
F0 mean + F0 slope	7158	8529	709	12807
F0 slope	7171	8683	710	12818
F0 mean	7216	9014	712	13691
Null	7234	9092	713	13693

enced the goodness ratings of these tones. Tone 3 only displayed a marginally significant result for F0 mean (p = .079). For Tone 1, both F0 mean and slope had a positive β estimate (0.011 and 0.004), showing that the goodness rating score increased with higher F0 mean or F0 slope. The results confirmed the observation from Figure 3.3 that more stimuli had a higher rating score were found on the zero to positive slope region than the negative slope region. The goodness rating pattern was less obvious on the F0 mean dimension, reflected by lower β estimate value compared to that of F0 slope. For Tone 2, the rating increased as F0 mean lowered, as shown by the negative β estimate (-0.015), or as F0 slope increased, as shown by the positive β estimate (0.026). As displayed in Figure 3.3, Tone 2 responses received higher rating scores in the low F0 mean and more positive F0 slope region. Finally, for Tone 4, the regression model results showed that rating scores went up as F0 mean and slope decreased since the β estimates were negative (-0.003 and -0.020). Figure 3.3 displayed a higher rating scores for stimuli at -84Hz/t or below, suggesting that the more falling tone contours were rated higher than the less falling ones. The effect F0 mean appeared less clear but there were slightly more stimuli got higher rating scores in the low compared to high F0 mean region. It is worth noting that the β estimate was comparatively low (-0.003), consistently with the small ΔD between the null and F0 mean-only models.

Since the Tone 3 data appeared to show higher rating scores in the more positive and negative slope regions compared to the near zero slope region, the data were divided into two sets by F0 slope. The positive slope set contained F0 slope level at 0Hz/t or above. The other set contained only negative F0 slope levels. An ordinal logistic regression was performed on each set of data. For the positive slope data set, F0 slope yielded a significant result as displayed in the lower section of Table 3.4. The positive β estimate (0.034)

revealed that Tone 3 responses received higher rating scores as the slope increased, or became more positive or rising. The negative slope data set showed marginally significant findings for F0 mean and slope. Both β estimates were negative (-0.008), suggesting that participants might have a tendency of giving higher rating scores to T3 responses that were lower in F0 mean and more negative and falling slope.

Table 3.4: Ordinal logistic regression of F0 mean and slope on tone goodness ratings for each tone

		F0 dimensions:	
		F0 mean	F0 slope
Tone 1	β	0.004***	0.011***
	(SE)	(0.001)	(0.001)
	t	3.47	7.52
	p	<.001	<.001
Tone 2	β	-0.015***	0.026***
	(SE)	(0.001)	(0.001)
	t	-12.26	21.50
	p	<.001	<.001
Tone 3	β	-0.008†	0.003
	(SE)	(0.004)	(0.002)
	t	-1.76	1.09
	p	.079	.274
Tone 4	β	-0.003***	-0.020***
	(SE)	(0.001)	(0.001)
	t	-3.30	-28.72
	p	<.001	<.001
Tone 3 - positive slope	β	-0.012	0.034**
	(SE)	(0.017)	(0.015)
	t	-0.72	2.23
	p	.473	.025
Tone 3 - negative slope	β	-0.008†	-0.008†
	(SE)	(0.005)	(0.004)
	t	-1.92	-1.87
	p	.055	.061
Note:		†p<0.1; **p<0.05; ***p<0.01	

Summary

The two analyses above showed that F0 slope strongly influenced *between-category* tone categorization at all levels of F0 mean as shown by the multinomial logistic regression models. As for *within-category* goodness ratings, the ordinal logistic regression model showed that F0 slope had an influence on the perception of Tone 1, 2 and 4. This effect was also found for Tone 3 responses in the positive, flat to rising slope region. On the other hand, F0 mean showed only a limited effect on tone categorization at some levels of F0 slope in the multinomial logistic regression models. An effect on goodness ratings was also found for Tone 1, 2 and 4, but with a lower model fit and smaller β estimate compared to F0 slope. Tone 3 only yielded a marginally significant effect on goodness ratings for negative, or falling, slope stimuli. Therefore, the two analyses indicate that F0 slope, as a critical perceptual cue, influences both *between-category* tone categorization and *within-category* goodness ratings more strongly than F0 mean, the non-critical cue, for all tones.

3.3.2 Predicting perceptual tone classification using multinomial logistic regression models trained by production data

The second set of analysis investigates the perception-production relationship of tones using a statistical learning approach. The production data used to train a multinomial logistic regression model were F0 mean and slope obtained in the acoustic modelling of the participants' tone productions (Refer to section 3.2.2). The multinomial logistic regression model had F0 mean and/or slope of the tone productions as predictors (Refer to section 3.3.2 below for details) and four Mandarin tone categories as the outcome variable, which were the tone labels presented to the participants during the production task. After cross-validating the model, it was then used to categorize the perceptual stimuli of this study. In this prediction stage, the perceptual stimuli were the resynthesized tone stimuli used in the perception experiment (Refer to section 3.2.3). Each stimulus item contained its F0 mean and slope values obtained by applying the same acoustic modelling method on the perceptual stimuli and the values were converted to logarithm-based *T* values (section 3.2.2). Each tone stimulus also contained the tone perceived by one participant obtained

in the perception experiment (Note that all participants took part in both the perception and production tasks). The predicted tone from model prediction was compared with the perceived tone from the human identification results in the perception experiment. In other words, the model either correctly or incorrectly predict the percevied tone for each stimulus item. Consequently, the mean accuracy for the model was obtained. Since the logistic regression was intended to model the perceptual process (McMurray and Jongman, 2011; Nearey, 1990, 1997; Redmon et al., 2020), it was expected that the model should be able to predict the identified tones in human perception correctly if the critical F0 slope cue was used (Newman, 2003).

The following results are presented in the next sections:

1. An overview of the production data
2. Logistic regression model training using production data and cross-validation (Input data: production)
3. Model prediction with perception data (Input data: perception)

Production data

The tone contours averaged across all productions are displayed in Figure 3.4. As shown in the figure, Mandarin tones were produced with their typical heights and contours (Tone 1 - high level; Tone 2 - rising; Tone 3 - dipping; Tone 4 - falling). As for tone duration, Tone 3 and 4 display the longest and shortest duration, respectively. Tone 1 and 2, which share similar duration, have duration between Tone 3 and 4. As shown in Table 3.5, Tone 1 had the highest F0 mean, followed by Tone 2 and 4. Tone 3 had the lowest F0 mean. For F0 slope, Tone 2 had the most positive, rising slope, whereas Tone 4 had the most negative, falling slope. Tone 1 and 3 had a slope value closer to zero than the other two tones.

Participants could possibly contribute to the variance of each cue. Therefore, a linear regression was first conducted with each cue as the dependent variable and Participants as the predictor, and the variance explained by Participants was the R^2 of the linear regression (*cf.* Redmon et al., 2020). Participants' contribution to the variance of F0 slope

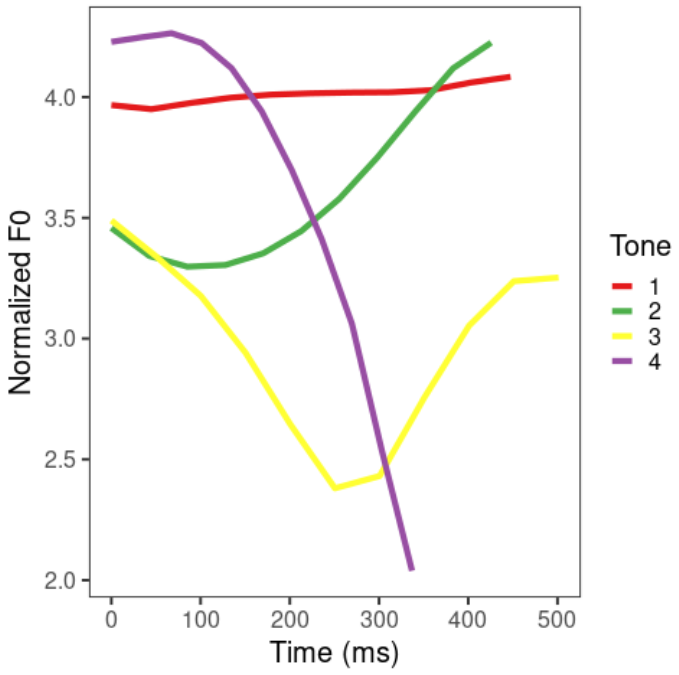


Figure 3.4: Mandarin tone contours averaged across all production items

Table 3.5: F0 mean and slope averaged across all productions (Standard deviation in parentheses)

Tone	F0 Mean (*T*)	F0 Slope (*T/s*)
1	4.01 (0.400)	0.275 (0.473)
2	3.62 (0.418)	2.27 (1.38)
3	2.99 (0.509)	-0.724 (2.22)
4	3.62 (0.413)	-7.02 (3.71)

was only 3.41% , but was 39.1% for F0 mean. It shows that there were some individual variations in overall F0 mean even when F0 range was set in the same scale after performing F0 normalization (i.e., the logarithm-based *T*-transformation, refer to section 3.2.2). As a result, in order to partial out the influence of Participants, residual values (the difference between fitted values of the linear regression model and the observed F0 mean or slope values) were used as the predictor cues for training the logistic regression models.

The distribution of tone production items in residual data in the F0 mean and F0 slope tone space are displayed in Figure 3.5. Tone 2 and 4 items are dispersed along the F0 slope dimension, whereas Tone 1 and 3 are separated on the F0 mean dimension.

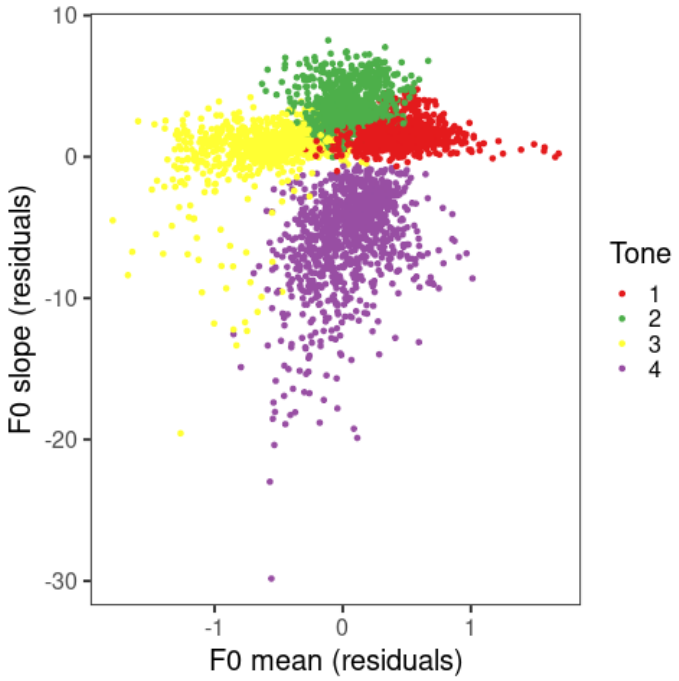


Figure 3.5: Tone productions of all participants in the residual values of F0 mean and slope

Training the production models

As mentioned above, the residual values of F0 mean and F0 slope cues obtained from the production items were used to train the logistic regression models. As in the previous section (3.3.1), five models were developed: a null model contained the intercept only; the two single predictor models had either F0 mean or F0 slope as the predictor; the two

full models included the linear combination of F0 mean and F0 slope as the independent variables, with one also containing the interaction term. The outcome variable of all models was the tone category labels presented to the participants during the production task.

Production model cross-validation

The production models were cross validated to examine the model fit and accuracy rate, demonstrating the contribution of each cue to production tone classification and potentially shed light on the use of these cues in perception. A five-fold cross validation method was used in order to avoid overfitting. The production data were equally divided into five subsets of data, not blocked by participants (n = 801 or 802). Five models were then trained with four subsets of data for each model and one subset was reserved for testing. The tone categories were determined by the tone labels presented to the participants during the production task. After testing, the Deviance statistic D and classification accuracy of each model were obtained and the mean D and accuracy were used to represent the model fit and classification accuracy of each production model. To understanding the contribution of each tone to the overall classification accuracy, the F1 score of each tone was also calculated using the following formula to represent each tone's accuracy, with a range between 0 (lowest accuracy) to 1 (highest accuracy):

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where $Precision$ was the proportion of cases correctly identified by the model as a particular tone among all cases of which the model predicted either correctly or incorrectly to be that same tone. $Recall$ was the proportion of cases correctly identified by the model as a particular tone among all cases that truly belonged to that tone. The contingency tables, accuracy rates and F1 score results for individual training models are presented in Appendix A. The mean accuracy and F1 scores are displayed in Table 3.6.

Comparing the single predictor models with the null model, the F0 slope-only models showed a greater ΔD than the F0 mean-only model. The F0 slope-only models were also more accurate than the F0 mean-only model by about 11%. Comparing with the full models

Table 3.6: Cross validation results of production models

Model	D	Accuracy (%)	*F1 Scores:* Tone 1	Tone 2	Tone 3	Tone 4
F0 mean $\times$ F0 slope	1060	92.3	0.905	0.882	0.922	0.985
F0 mean $+$ F0 slope	1205	92.3	0.909	0.888	0.922	0.975
F0 slope-only	4163	71.8	0.540	0.809	0.572	0.942
F0 mean-only	5263	60.5	0.769	0.466	0.838	0.228
Null	8889	25.3	0.385	0.353	$n.a.$	$n.a.$

without the interaction term (i.e., F0 mean + F0 slope), the F0 slope models had a smaller ΔD and overall accuracy drop than the F0 mean model. All of these suggest that F0 slope was more critical than F0 mean in classifying tones in production. For the F1 scores of individual tones, the F0 slope model yielded high F1 scores for Tone 2 and 4, with Tone 2 showing a greater drop from the F0 mean $+$ F0 slope model than Tone 4. The F1 scores plummeted in the F0 mean-only model for these two tones, with Tone 4 displaying a more drastic decrease than Tone 2. In contrast, Tone 1 and 3 only had a smaller drop of F1 scores from the F0 mean $+$ F0 slope models to the F0 mean-only model than from the F0 mean $+$ F0 slope to F0 slope-only models. The F0 slope-only models yielded the lowest F1 scores for these two tones. These results suggest that F0 slope had a more important role for tone classification than F0 mean for Tone 2 and 4. F0 mean was nevertheless more important than F0 slope for tone classifying Tone 1 and 3 than F0 slope. These are consistent with the distribution of tone productions in Figure 3.5 where Tone 1 and 3 are separated on the F0 mean dimension, whereas the F0 slope dimension determines the distribution of Tone 2 and 4. However, comparing the F0 mean and slope only models, Tone 1 and 3 had a smaller F1 score difference than Tone 2 and 4, suggesting that F0 slope was still more critical for categorizing Tone 1 and 3 than F0 mean for Tone 2 and 4.

The full models had the highest accuracy rates and the lowest D, as expected. The overall accuracy rates of the full models were higher than the F0 slope-only model by slightly more than 20% and there was a substantial lowering of D. These indicate that the classification of Mandarin tone productions requires both cues. This is not consistent with the perception results above which show that the full models did not differ from the F0 slope

model in terms of model fit (section 3.3.1). Comparing the full models with and without the interaction term, although there was a decrease in D indicating a better model fit, the interaction term models did not show any substantial improvement in overall accuracy compared to ΔD between the full and the F0 slope-only models Moreover, the changes in F1 scores was not consistent ((a change of -0.04 to 0.1).

Predicting perception responses using production models

To perform prediction, the data entered into the logistic regression models trained by production data were the perception stimuli used in the perception experiment. Note that the production models were trained by residual values of F0 mean and/or slope obtained by partialling out Participant in a linear regression. Although Participant only minimally explained the variance of F0 mean and slope of the perceptual stimuli (0.1% for both cues), the residual values of F0 mean and slope of the perception stimuli, obtained by linear regression with Participant as the predictor and either F0 mean or F0 slope, were used. The production models predicted the tone categories of the perceptual stimuli based on the residuals of their F0 mean and slope values in order to obtain the classification accuracy and F1 scores for each model. To obtain comparable results in model prediction, five unique sets of the residual perception data were randomly selected without replacement from the complete perception data set, matching the number of observations in the production data set used for cross-validating the production models.

Three out of five production models were used for prediction. Since the full model with and without the interaction term (i.e., F0 mean $\times$ F0 slope and F0 mean $+$ F0 slope) showed similar level of model accuracy, only the F0 mean $+$ F0 slope model was retained to compare the accuracy of the full model with that of the single predictor models. The prediction also involved the two single predictor models. Since the null model contained only the intercept, the prediction result would only reflect the distribution of tones in the production data, and therefore, was excluded in prediction (Refer to section 3.3.2 for the details of the five models). The mean results are shown in Table 3.7 below, and contingency tables, accuracy rate and F1 scores of individual perception models are presented in Appendix B.

Table 3.7: Mean results of predicting production tones by perception data

Model	Accuracy (%)	*F1 Scores:* Tone 1	Tone 2	Tone 3	Tone 4
F0 mean + F0 slope	64.4	0.566	0.769	0.113	0.785
F0 slope-only	76.4	0.691	0.925	0.136	0.853
F0 mean-only	23.1	0.366	0.243	0.083	0.134

Note that this model prediction compared the predicted tone of the perceptual stimuli with the actual tone responses of the participants in the perception experiment. In general, the model trained by F0 slope production data only yielded the highest accuracy among the three model (76.4%). It was similar to the F0 slope-only model accuracy obtained in cross-validation (71.8%). Including F0 mean in the model no longer improved the accuracy rate, reflecting the lack of contribution of this cue to the classification of perceptual tone responses in the perception experiment. Recalling from Section 3.3.1, the tone responses were mainly separated by the F0 slope dimension. It turned out that the accuracy rate was about 28% lower than model cross-validation for this full model (92.3% vs. 64.4%), and was lower than the F0 slope-only model accuracy by 12%. In addition, the F0 mean-only model yielded a poor accuracy rate of 23.1%, slightly lower than the 25% chance level, and was much worse than that of model cross-validation (60.5%). It should be noted that, while F0 slope continued to contribute to the classification of Tone 2 and 4 substantially as demonstrated by the high F1 scores of these tones in the F0 slope-only model (0.925 and 0.853), adding F0 mean to the model (i.e., F0 mean + F0 slope) no longer improve the classification of Tone 1 and 3 that was found in production model cross-validation. There was even a drop of F1 score for these two tones from the F0 slope-only to the full model. These two tones remained to yield a comparatively lower F1 scores than Tone 2 and 4 in the full model, rather than having all four tones showing comparable F1 scores in model cross-validation.

Taken together, the results indicate that only F0 slope, the critical cue, played a role in relating production to perception based on production data modelling. As in model cross-validation, F0 slope maintained a similar level of contribution to tone categorization ac-

curacy. Both prediction and cross-validation achieved similar model accuracy rate in the F0 slope-only model. In contrast, F0 mean showed poor performance in model prediction. This cue alone did not contribute to the prediction of perceptual tone responses and even lowered the performance of model prediction as seen in the full and F0 slope-only model comparison. As a result, the prediction based on the production logistic regression model indicated that only F0 slope was the critical cue that link production to perception.

3.3.3 Predicting tone categories in production using multinomial logistic regression models trained by perception data

The perception-production relationship was also examined using a logistic regression trained by perception data obtained in the perception experiment. This statistical learning using perception data shared the same procedures with the production model training in the previous section (3.3.2), but the training and prediction data were swapped. The training data were now the perceptual stimuli which contained residuals of F0 mean and slope in logarithm-based *T* values. The models had these residual values of F0 mean and/or slope as the predictors and the tones perceived by the participants as the outcome variable. The same five-fold cross validation was performed. The data used for prediction were the production data containing the residual values of F0 mean and slope obtained in the acoustic modelling of the participants' tone productions (Section 3.2.2) (Note that residual values were used as Participant explained 39.1% of the variance of F0 mean as reported in section 3.3.2). The production data entered into the perception models to predict the tone category of each tone production. The predicted tone categories were compared with the tone labels of the tone productions to obtain overall model accuracy and F1 scores for each tone. It was expected that the perception models should model the human perceptual process and the classification of production data should reach high accuracy if the critical perceptual cue was used.

The following results are presented in the next sections:

1. Logistic regression model training using perception data and cross-validation (Input data: perception)

2. Model prediction with production data (Input data: production)

Training the perception models

As in section 3.3.1 and 3.3.2, the training models consisted of two full models containing the linear combination of the two predictors, with (i.e., F0 mean $\times$ F0 slope) and without the interaction term (i.e., F0 mean $+$ F0 slope), two single predictor model with either F0 mean or F0 slope as the predictor, and a null model with only the intercept. The predictors were the residuals of F0 mean and/or F0 slope of the perceptual stimuli used in the perception experiment and the outcome variable was the tone responses of the participants also in the perception experiment. All perception stimuli shared the same data range for F0 mean and F0 slope, and therefore the model trained using normalized and raw F0 data, reported in section 3.3.1 should yield very similar results. A linear regression with each cue as the dependent variable and Participants as the predictor was carried out to obtain residual data. In fact, the variance explained by Participant was negligible (R^2 = 0.01% for both cues), but the residual values were still used as training data because the training data in the production models which were used as test data in this section were residual values. Moreover, with the insignificant contribution of Participant to the variance of each cue, this procedure should not affect the distribution of the perception data.

Perception model cross-validation

The next step was to perform cross-validation of the perception models to obtain model fit and accuracy. This procedure is very similar to the comparison of model fit using D in section 3.3.1. However, since the data used to train the models were residuals instead of raw F0 mean and slope of the perceptual stimuli, it was expected that the results should be highly similar but might not be completely identical. Moreover, in addition to model fit, this analysis also contained model accuracy results in order to determine the best model for predicting production categories. Therefore, as in section 3.3.2, a five-fold cross validation was carried out to avoid overfitting. The contingency tables, accuracy rates and F1 score results for individual training models are presented in Appendix C. The mean accuracy and F1 scores of the five training models are displayed in Table 3.8.

Table 3.8: Cross validation results of perception models

Model	D	Accuracy (%)	*F1 Scores:*			
			Tone 1	Tone 2	Tone 3	Tone 4
F0 mean $\times$ F0 slope	4753	90.3	0.828	0.928	$n.a.$	0.949
F0 mean $+$ F0 slope	4787	90.2	0.826	0.927	$n.a.$	0.949
F0 slope-only	4936	89.6	0.821	0.916	$n.a.$	0.946
F0 mean-only	18536	45.5	$n.a.$	$n.a.$	$n.a.$	0.625
Null	18703	45.5	$n.a.$	$n.a.$	$n.a.$	0.625

Comparing the null and single predictor models, the F0 slope-only model again outperformed the F0 mean-only model by a lot, with a ΔD of 13767 and 167, respectively. The null and F0 mean-only model displayed the same model accuracy and F1 scores, showing participants did not use F0 mean alone to perceive Mandarin tones. Both models predicted all the perceptual stimuli to be Tone 4 since the majority of the tone responses of the participants were Tone 4. As a result, only this tone revealed an F1 score for these two models, showing that the F0 mean-only model, just like the null model, failed to perform tone classification at all. On the other hand, the ΔD was just 149 between the F0 slope-only model and the full model with linear combination of cues only, showing that the two models had very similar model fit. The overall accuracy and F1 scores of the F0 mean $+$ F0 slope model were only very slightly higher than the F0 slope-only model. These suggest that including the F0 mean cue only slightly improved the prediction of these tones. The two full models again yielded very similar D, with a ΔD of merely 34, the smallest ΔD between two models among the five perception models. The result was consistent with the results of section 3.3.1. For the full and F0 slope models, Tone 4 had the highest F1 scores and followed by the slightly lower F1 scores of Tone 2. These values were close to 1, the highest possible F1 score value, suggesting that these the models showed highly accurate classification for these two tones. Tone 1 displayed comparatively lower F1 scores than the other two tones at above 0.82. Finally, the F1 scores for Tone 3 were not available since the perception models did not predict any Tone 3 responses (Refer to Appendix C). In general, these results are consistent with the tone response distribution as shown in 3.2 where F0 slope itself appear to determine tone responses. The addition of F0 mean only

slightly modulated the tone response results. The two full models and the F0 slope model all achieved similar classification accuracy. In order to maintain consistency with the model prediction results using production models, the F0 mean + F0 slope and F0 slope models were used in the following model prediction.

Predicting tone production using perception models

After the cross validation of the perception models on the perception data, the production data were used as a new test data set. As in section 3.3.2, the residual values of the production data were used. Five unique sets of production data were randomly selected without replacement from the complete production data set (Section 3.3.2). Each data set was entered into one of the five perception models, with the number of data points matching that of the perception data test sets used for training the perception models. The mean accuracy results are presented in Table 3.9 below, and contingency tables, accuracy rate and F1 scores of individual perception models are included in Appendix D.

Table 3.9: Mean results of predicting perception responses by production data

Model	Accuracy (%)	*F1 Scores:* Tone 1	Tone 2	Tone 3	Tone 4
F0 mean + F0 slope	68.1	0.645	0.814	0.026	0.943
F0 slope-only	67.4	0.613	0.815	$n.a.$	0.936
F0 mean-only	24.3	0.011	$n.a.$	$n.a.$	0.390

This prediction compared the predicted tone categories using the perception models with the actual tone labels of the production data. As in the prediction of perceptual responses using production models (section 3.3.2), this model prediction of production tone categories yielded higher accuracy rates for the F0 slope-only than the F0 mean-only model (67.4% vs. 24.3%). The F0 mean-only model again showed a chance-level performance (25%). Since the perception model cross-validation showed F0 slope to be the main factor that separate responses, the model was presumably unable to differentiate the tones that were separated along the F0 mean dimension in the production data. Unlike the production model prediction in which full model performed worse than the F0 slope-only

model, the perception model prediction showed that the full model had a slightly better overall accuracy (68.1%) than the F0 slope-only model (67.4%), but the improvement was lower than 1%. The F1 score results generally followed the overall accuracy pattern. The full model had the highest F1 scores for all tones except for Tone 2 which had a very slight drop of 0.001 from the F0 slope-only model. Tone 1 and 4 had a lower F1 score in the F0 slope-only model. It should be noted that, while F1 scores of Tone 1, 2 and 4 yielded F1 scores above 0.8 for full and F0 slope-only models in model cross-validation, Tone 1 showed a comparatively lower F1 score for these models in prediction. This pattern mirrors that of the cross-validation and prediction of the production F0 slope-only model. The F1 scores of F0 mean-only model reflected the fact that the highest number of tone responses was Tone 4 in the perception experiment. Therefore, the perception model classified the huge majority of responses as Tone 4 when the predictor was weak in performing tone classification (Refer to Appendix D). Therefore, these results again attest that the critical F0 slope was able to establish a perception-production link.

3.3.4 Predicting perceptual tone classification using production models with between-category and within-category perception data

The final analysis investigated the relationship between perception and production data using perceptual stimuli that were rated as good as or better than the naturally produced tone stimuli, which should presumably be the resynthesized perceptual stimuli that resembled natural tones. While the previous two sections on model prediction focused on perception data based on between-category tone categorization, the stimulus selection process used in this analysis took into account the between-category and within-category perception of Mandarin tones. Comparing the results of this section with those of the second section addresses the question of whether within-category perception, in addition to between-category perception, contributes to perception-production relationships. Since the critical perceptual cue, F0 slope, was shown to strongly influence between-category and within-category perception in Section 3.3.1, filtering the perception data with goodness rating results should further improve the perception-production relationship. In addition, F0 mean, as a non-critical perceptual cue, showed a weaker influence on within-category per-

ception, the perception data screening with goodness rating results may also improve the perception-production relationship. The improvement was expected to be demonstrated by an increase in model prediction accuracy compared to the results presented in Section 3.3.2. Consistent with model prediction using production models, the analysis of this section also involved three models: F0 mean + F0 slope, F0 slope-only and F0 mean-only, with the same procedures were used. Specifically, five sets of the perceptual data entered into the production models were randomly selected without replacement to match the number of stimuli used for cross-validation. The models predicted the tone category of the perceptual stimuli and the predicted tones were compared with the perceptual tone identification responses obtained in the perception experiment to calculate the overall prediction accuracy and F1 scores.

Prediction results

The prediction results are summarized in Table 3.10. Comparing with the production model prediction results, the prediction using good exemplar perception data yielded improved overall accuracy for both full (i.e., F0 mean + F0 slope) (64.4% and 75.8%) and F0 slope-only model (76.4% and 91.2%). However, the prediction accuracy for F0 mean-only remained very similar with two sets of perception data (23.1% and 22.8%). The F1 scores also showed improvement for the full and F0 slope-only models, with the exception of Tone 3. The greatest difference was the Tone 1 F1 score for the F0 slope-only model (all perception data: 0.691; good exemplar perception data: 0.904), achieving a score that is close to that of Tone 2 and 4. The F0 mean-only model did not show clear difference in prediction accuracy and F1 scores across two sets of perception data.

Consistent with the prediction with all perception data, the F0 slope-only model continued to display the highest overall accuracy and F1 scores for the good exemplar perception data. The addition of F0 mean to the F0 slope-only model also lowered the accuracy and F1 scores, and the F0 mean-only model showed chance-level accuracy.

Taken together, the good exemplar perception data that involved both between-category and within-category perception indicated a clearer perception-production link than all per-

ception data for F0 slope with an improved model accuracy and F1 scores for the F0 slope-only model. It is also worth noting that using this set of perception data improved F1 score for all tones, especially Tone 1 (except for Tone 3). Isolating the good exemplars did not lead to any clear benefit for F0 mean in relating perception to production since the accuracy of the F0 mean-only model remained at chance level. In addition, the adverse effect of including F0 mean in the full model prediction was still present.

Table 3.10: Mean results of predicting production tones by perception data with items rated better than naturally-produced tones

		F1 Scores:			
Model	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
F0 mean + F0 slope	75.8	0.701	0.873	0.026	0.866
F0 slope-only	91.2	0.904	0.983	$n.a.$	0.947
F0 mean-only	22.8	0.372	0.264	0.017	0.157

3.3.5 Overall Result Summary

The logistic regression analysis of perception results showed that F0 slope, the critical perceptual cue, was predominantly used for *between-category* tone categorization and *within-category* goodness ratings (Section 3.3.1). Specifically, the influence of F0 slope was found at all levels of F0 mean in tone categorization and for all tones in goodness ratings (Tone 3 was the exception). F0 mean, the non-critical perceptual cue, only displayed an effect on tone categorization at certain F0 slope levels. The goodness ratings were influenced by F0 mean for all tones except Tone 3 but the effect was weaker than that of F0 slope.

Consistent with the above tone categorization results, the perception and production models demonstrated that F0 slope was critical for tone classification in model cross-validation. In contrast, F0 mean only proved crucial for the classification of tone productions, especially for Tone 1 and 3 (Section 3.3.2), whereas this cue played a minimal role in classifying perceptual tone responses (Section 3.3.3). The difference between the classification of production categories and perceptual tone responses was demonstrated by the greater ΔD between F0 slope only and F0 mean + F0 slope model, and between null and

F0 mean-only models in production than in perception models. The F1 score of Tone 1 and 3 only showed a clear improvement from F0 slope-only to the full models in production data (Table 3.6 and 3.8).

The prediction results consistently indicated that F0 slope alone, as the critical cue, exhibited a perception-production link with comparable prediction accuracy rates across both directions of prediction - i.e. predicting perceptual responses using production models and production tone categories using perception models (Table 3.7 and 3.9). This analysis did not find any evidence for F0 mean, as a non-critical cue, to be able to establish a perception-production link, since there was no improvement of prediction accuracy from the F0 slope-only model to the full model (There was even an adverse effect for production models in this regard). The F0 mean-only model also consistently yielded chance-level accuracy. In general, the model prediction yielded accuracy pattern that was similar to the perception model performance (Table 3.8). As for individual tones, the predicted F1 scores of both perception and production model again followed the F1 scores of the perception model in cross-validation. Tone 2 and 4 achieved comparatively high F1 scores, relative to Tone 1 and 3 in the F0 slope-only model. In addition, the clear improvement of F1 score for these two tones from F0 slope-only to full models as shown in the production model cross-validation was not found in model prediction across the two domains.

The final analysis used a subset of perception data that were rated as good exemplars of the perceived tones based on the within-category goodness rating task for prediction. The full and F0 slope-only models yielded greater prediction accuracy than the model prediction with all perception data. The F1 score of Tone 1 in the F0 slope-only model improved the most and reached a score close to that of Tone 2 and 4. F0 mean continued to perform poorly as including this cue in the full model lowered the prediction accuracy, and the F0 mean-only model showed only chance-level performance.

3.4 Discussion

This study evaluated the impact of F0 slope (critical perceptual cue) and F0 mean (non-critical perceptual cue) on *between-category* tone categorization and *within-category* good-

ness ratings of each tone. Using a statistical learning approach, it examined the contribution of the critical and non-critical perceptual cues separately on the perception-production relationships of Mandarin tones through the modelling of perception and production data and the bi-directional prediction across domains. It showed that the critical F0 slope perceptual cue as defined by previous cue weighting and acoustic modelling studies (e.g. Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Tupper et al., 2020) could largely determine tone categorization in perception and have a stronger influence on goodness ratings than the non-critical F0 mean cue for all tones. The analysis of perception-production relationships unequivocally demonstrated that the critical cue, F0 slope, played a vital role in linking the two domains in the bi-directional model prediction based on *between-category* perception data. Its role in the perception-production link was further strengthened when the data based on both *between-category* and *within-category* perception were used for production model prediction. None of the analysis found F0 mean, the non-critical percpetual cue, to be able to establish a perception-production relationship although it was shown to influence the *within-category* goodness ratings to a certain extent.

3.4.1 Defining features of critical and non-critical perceptual cues of Mandarin tones

The perception results indicate that F0 slope had an overwhelmingly clear influence on tone categorization in perception for native Mandarin individuals no matter how F0 mean changed (Table 3.1). Meanwhile, F0 mean only had an effect on tone categorization at certain F0 slope levels (Table 3.2). More importantly, the F0 slope-only model achieved tone classification accuracy and F1 scores similar to the full models, but the F0 mean-only model completely failed to classify tones. In addition, Tone 1, 2 and 4 had high F1 scores in the full and F0 slope-only models, and there was only a slight drop in F1 scores when F0 mean was removed from the full model. These results indicate that native Mandarin individuals rely mainly on F0 slope for Mandarin tone categorization in perception, whereas the contribution of F0 mean is relatively small. Moreover, these suggest that native Mandarin individuals can use F0 slope only to achieve high accuracy in categorizing

these three tones in perception (Table 3.8). Although the results are consistent with the cue weighting findings that F0 slope is a more critical cue in perception than F0 mean, they are not consistent with the previous findings in which both cues are required for native Mandarin tone perception and pitch height is a dimension that mainly differentiates between Tone 1 and 3 (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Tupper et al., 2020). Therefore, the general cue weighting results that pitch direction is weighted more strongly than pitch height are informative for defining the critical status of perceptual cues. In contrast, the perceptual dimension map showing the impact of each dimension on speech categories in perceptual weighting studies (e.g. Chandrasekaran et al., 2010; Francis et al., 2008) does not provide a clear picture of how the critical status of perceptual cues influence the perception of individual speech categories.

The perception results of this study also differed from Yang (2015) which showed that F0 mean could modulate the perceptual classification of Tone 1 and 3. The difference is probably attributed to the different perceptual contexts used in these study. This study presented the perceptual stimuli in an isolated context. Although the participants should be able to detect the F0 range of the perceptual stimuli in this study, the low flat tones were still predominantly identified as Tone 1 (Figure 3.2). In contrast, the perceptual stimuli in Yang (2015) were embedded in a carrier sentence, and the low flat contours were perceived as Tone 3. The methodological difference indicates that F0 mean, as a pitch height cue, only influences Mandarin tone perception with an external, sentential context. This cue only minimally influence the perception of Mandarin tones in isolation.

Similarly, there is a discrepancy between the impact of F0 slope and mean on tone classification in perception and production. Note that the tones were produced by the same group of participants in isolation. Therefore, tone production classification requires both F0 slope and mean as indicated by substantial difference between the full models and the single predictor models, as in tone production in a sentential context (Yang, 2015). While F0 slope-only models yielded a higher overall accuracy than the F0 mean-only model, Tone 1 and 3 attained high F1 scores in the F0 mean-only model, so did Tone 2 and 4 in the F0 slope-only models (Table 3.6). These results suggest that F0 slope seems to be

more critical than F0 mean in overall production tone classification. However, both cues are still required for a more accurate production tone classification and F0 mean has a critical role for classifying Tone 1 and 3 productions. These results corroborate the findings of Tupper et al. (2020) that show F0 slope-related cues are important for differentiating Tone 2 or 4 from other tones, whereas F0 mean-related cues differentiates Tone 1 or 3 from other tones. The production results also aligned with the perceptual cue weighting results described above (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). Nevertheless, they were not consistent with the perception results in which F0 slope alone could perform well in classifying tone responses of the participants. The influence of F0 mean was limited since it only moderated the perception of tone categories at certain F0 slope levels. The discrepancy between perception and production modelling indicates that the cues used for between-category perception is more restrictive than for classifying tone productions. The cues accounting for the variations of tone categories in production only partially reflect the critical status of perceptual cues.

As for within-category perception, this study shows that both F0 slope and mean are involved in modulating goodness ratings. For Tone 1, 2 and 4, the two cues were found to be significant in the ordinal logistic regression, with F0 slope always showing a higher coefficient value than F0 mean. Therefore, this study demonstrates that critical perceptual cues have a stronger influence on the perception of within-category differences than non-critical perceptual cues. The result confirms the tone perceptual cue weighting results that both cues are used in the acoustic-phonetic differences of tone contours, and the critical cue was weighted more strongly than the non-critical cue (Gandour, 1983; Guion and Pederson, 2007), as predicted by hypothesis 1. Nevertheless, Tone 3 results showed that only F0 slope significantly modulated the goodness ratings for the positive, rising slope subset of data. Therefore, for this particular tone, the present study has found some support that only the critical F0 slope cue can modulate the goodness rating results based on Newman (2003).

Taken together, this study clearly show that native Mandarin individuals make use of the critical F0 slope alone to achieve accurate *between-category* tone perception. F0 mean

provides limited modulation effects on certain F0 slope levels for all tones. While F0 slope and F0 mean are required for *within-category* goodness rating perception, F0 slope is more strongly weighted than F0 mean also for all tones. As a result, both *within-category* and *between-category* perception define the critical status of perceptual cues similarly. For Mandarin tone perception, F0 slope, representing pitch direction cues, is critical for *within-category*, phonetic level and *between-category*, phonological level of perception. Moreover, the critical status of perceptual cues displays a near tone-general feature since, generally speaking, F0 slope strongly influences both levels of perception for all tones. However, there is one caveat here: the modelling of between-category perception did not show any conclusive results for Tone 3. It is probably due to the much smaller number of and less consistent Tone 3 responses compared with other tones since the linear tone contours used in this study that did not represent the prototypical dipping contour of Tone 3, as mentioned in Section 3.2.3. It is possible that the critical status of perceptual cues for this tone can be more clearly revealed when this tone is perceived in a sentential context rather than in isolation (Yang, 2015).

3.4.2 Establishing a perception-production link through critical perceptual cue

Compared with the analysis of critical perceptual cues, the production models indicate a different picture of cue contribution. While F0 slope dominated tone categorization in perception, both F0 slope and mean are needed for classifying tone productions. F0 slope still appeared to be a better cue for classifying tones than F0 mean, since the F0 slope-only production model yielded a higher overall classification accuracy (71.8%) than the F0 mean-only production model (60.5%). However, F0 mean was also a useful cue for classifying tone productions due to the fact that the F0 mean-only model displayed a much higher accuracy than the null model which simply reflected the distribution of tone categories in the stimuli (25.3%). Moreover, including F0 mean in the full models led to a 20% increase of accuracy compared with the F0 slope-only model (Table 3.6). In contrast, the perception models, as in the critical cue analysis, showed that only F0 slope was the dominant cue for tone perceptual categorization. There was only less than 1% difference in model accuracy

between the full and F0 slope-only models, and the F0 mean-only model showed the same performance as the null model (Table 3.8). These suggest that F0 mean played a minimal role in categorizing tones in perception.

This study followed McMurray and Jongman (2011) and Redmon et al. (2020) that used logistic regression trained by production data to model the perceptual process. The difference in cue contribution in perception and production models indicated that the two domains were not perfectly aligned. However, F0 slope, as a critical cue for Mandarin tone perception, performs well in linking perception and production. The bi-directional model prediction results yielded a common picture that F0 slope-only model performed well in tone classification. In contrast, F0 mean did not show its contribution to the perception-production relationship as the F0 mean-only model only yielded a chance-level accuracy rate (Table 3.7 and Table 3.9). There was also evidence that F0 mean could weaken the perception-production link as the model accuracy and F1 scores reduced in the F0 mean $+$ F0 slope production model compared with the F0 slope-only model in the production model prediction results using perception data (Table 3.7). These results confirm the prediction that a critical perceptual cue should be able to establish a perception-production relationship based on Newman (2003) and Chapter 2.

More importantly, the F1 scores showed that this pattern was consistent across tones. Therefore, these results indicate that the perception-production link is guided by an overall, tone-general perception process. Recall that this study hypothesizes that the critical status of perceptual cue may be tone-general or tone-specific. The tone-specific prediction is based on the finding that Tone 1 and 3 perception rely on pitch height rather than pitch direction. This study only found that F0 mean, as a pitch height cue, contributed to the classification of the production of Tone 1 and 3 (Table 3.6). However, these results were not found in the perception models or in the bi-directional model predictions. All these results consistently showed that F0 slope was the only cue that facilitated the tone classification process (Table 3.7, Table 3.8 and Table 3.9). In fact, since the critical status of perceptual cues is near tone-general, it is not surprising that the perception-production link is also near tone-general as critical perceptual cues are pertinent to establishing such link. Note that

inconclusive results were obtained for Tone 3 since the models either failed or performed very poorly in predicting the classification of this tone.

The above findings showing a perception-production relationship established by a critical perceptual cue were based on tone perceptual categorization data (i.e., *between-category* perception). Following the critical perceptual cue analysis, the critical status of F0 slope was defined by both *within-category* (goodness ratings) and *between-category* perception (tone categorization). Therefore, *within-category* perception should also contribute to a perception-production relationship, and this was evinced in the present study through a tighter perception-production relationship established by F0 slope when perceptual responses were filtered by goodness rating results. It is worth noting that the F0 slope-only model achieved a high overall accuracy (91.2%) and the F1 scores were improved, especially for Tone 1 (Table 3.10, compared with Table 3.7). Therefore, this study indicates that a perception-production relationship is not only phonologically-based as described in theories (Diehl et al., 2004; Galantucci et al., 2006; Guenther, 1995; Liberman et al., 1967; Liberman and Mattingly, 1985; McAllister Byun and Tiede, 2017; Perkell et al., 2004a).

3.4.3 The possibility of a tone-specific critical perceptual cue

As mentioned above, it should be noted that this study only found weak evidence regarding the perceptual cue use in the categorization of Tone 3. F0 slope was a significant factor in perceptually differentiating Tone 3 from other tones for all F0 mean levels, and F0 mean was significant for some F0 slope levels and tone pairs. However, the models yielded very poor or no F1 scores for Tone 3 in model cross-validation and prediction, except for the cross-validation of production models. As a result, neither F0 slope nor mean was the cue that facilitated the classification of Tone 3 in perception. Since the perception stimuli were linear tone contours that did not represent the prototypical dipping contour of Tone 3 as mentioned in section 3.2.3, it was very likely that participants gave a much smaller number of and less consistent Tone 3 responses compared with other tones. The results of this study indicated that the critical perceptual cue of Tone 3 was likely not defined by F0 slope or mean. Since pitch direction is generally regarded a more strongly weighted cue for Man-

darin tone perception (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007), this study further illustrates that pitch direction, as the critical cue for Mandarin tone perception, is likely not defined by a single acoustic parameter. As reviewed in 1.3.1, previous studies indicated that Tone 3 perception involved a shift of $\triangle$F0 and TP (Moore and Jongman, 1997; Shen and Lin, 1991; Shen et al., 1993) and a systematic change of contour shape (Zhao and Kuhl, 2015). In addition, Chapter 2 also found a tone perception-production relationship to be established by F0 curvature when the Tone 2-like stimuli were situated between Tone 1 and Tone 3 extreme values. Therefore, F0 curvature is a likely candidate for the critical perceptual cue for Tone 3. Taken together, it is possible that a broadly defined pitch direction cue that is not specific to a single acoustic parameter is the tone-general critical perceptual cue for Mandarin tone perception. However, the actual acoustic correlate of pitch direction may vary among tone categories. This possibility needs to be further investigated in future studies.

3.5 Conclusion

This study demonstrated that the critical status of perceptual cues was pertinent to establishing a perception-production relationship through a bi-directional statistical modelling. It provided clearer evidence that F0 slope, as a critical cue for Mandarin tone perception, contributed predominantly to *between-category* tone categorization and *within-category* goodness ratings in perception than F0 mean. Therefore, both phonetic and phonological perception defined the critical status of perceptual cues, and consequently, contributed to establishing a perception-production relationship. In addition, this study showed that pitch direction was a tone-general critical perceptual cue. It is possible that pitch direction may be represented by several acoustic cues, instead of F0 slope alone.

Chapter 4

Examining perception-production relationships through the perception and production learning among Indonesian learners of Mandarin

4.1 Introduction

The previous chapters revealed that F0 slope, the critical cue for tone perception, influenced tone categorization and goodness ratings of Mandarin tone perception more strongly than F0 mean, the non-critical cue, and consequently could establish a perception-production link. The results were obtained from a correlation analysis (Chapter 2) and a qualitative comparison of tone classification accuracy based on a statistical learning approach (Chapter 3). While these approaches are informative of a perception-production link, the findings are not yet able to show that the perception-production link is quantitatively attributable to critical perceptual cues only. That is, in order to prove that critical perceptual cues are pertinent to establishing a perception-production link, the design of the study should be able to demonstrate that the critical status of perceptual cues is a determining factor for linking perception and production. This study aims to achieve this goal by comparing the perception and production learning patterns of native Indonesian individuals who were beginner-level learners of Mandarin. Specifically, using a within-subject design, this study aims at comparing the performance of the critical F0 slope and the non-critical F0 mean cues in perception and production for each individual learner across two

visits, so as to determine the learning effect on perception and production for each cue. Generally speaking, a perception-production link is demonstrated by a simultaneous perception and production improvement across visits. If establishing a perception-production link can be attributed to the use of critical perceptual cues, the simultaneous improvement is expected to be found for F0 slope, the critical perceptual cue, but not for F0 mean, the non-critical perceptual cue.

4.1.1 Transfer of training effect across domains

As mentioned in Section 1.2.1, a transfer of learning effect across perception and production indicates a perception-production link following the SLM (Flege, 1995, 1999, 2003). As shown in laboratory training studies, training individuals to perceive or produce speech sounds of a non-native language improved both perception and production (Bradlow et al., 1997; Brosseau-Lapré et al., 2013; Hardison, 2003; Hazan et al., 2005; Herd et al., 2013; Hirata, 2004; Iverson et al., 2012; Kartushina et al., 2015; Lambacher et al., 2005; Wang et al., 1999, 2003). The transfer of learning effect across domains is an indication of a link between perception and production. However, the two domains display a semi-autonomous nature. While perception-only training studies unequivocally demonstrated a transfer from perception training to production (Bradlow et al., 1997; Hardison, 2003; Hazan et al., 2005; Herd et al., 2013; Iverson et al., 2012; Lambacher et al., 2005; Sakai and Moorman, 2018; Wang et al., 1999, 2003), this transfer was less-than-perfect since perception and production gains lacked a correlation (Bradlow et al., 1997; Sakai and Moorman, 2018). On the other hand, production training demonstrated a transfer from production to perception only in some studies (Hirata, 2004; Kartushina et al., 2015) but not in others (Adank et al., 2010; Herd et al., 2013; Wang, 2013). Moreover, the production training could disrupt perceptual learning if the production task involved the sounds to be learned and took place immediately after perception (Baese-Berk and Samuel, 2016; Baese-Berk, 2019; Baese-Berk and Samuel, 2022). While these studies revealed a link between perception and production as semi-autonomous systems for non-native individuals learning a new language, the perception and production improvements were measured by accuracy rates of perceiv-

ing and producing sound categories in a number of studies. Specifically, the perception of speech categories was assessed by discrimination or category identification accuracy, whereas production accuracy was determined by the judgement of non-native productions performed by native listeners of the language of interest, usually through an identification task. Only a fraction of studies examined the acoustic properties of non-native productions, such as the between-category acoustic distance of L2 productions (e.g. Baese-Berk, 2019; Brosseau-Lapré et al., 2013), and the acoustic difference between native and non-native productions (e.g. Wang et al., 2003). However, very few studies have examined both perception and production in terms of particular acoustic cues. Moreover, no study has compared the perception and production improvements of critical and non-critical perceptual cues. Therefore, these studies inform us about learner's accuracy of perceiving and producing sound categories, but they do not reveal whether a critical perceptual cue impacts both learner's perception and production. This question is important for the current project that aims at understanding the role of critical perceptual cues (e.g. F0 slope in Mandarin tone perception) on perception-production links. So far, little research has directly examined whether critical perceptual cues display a perception-production link through L2 learning.

4.1.2 Learning effect on perceptual cue weighting

Though exploring whether critical perceptual cues could establish a perception-production link was not well studied through L2 learning, previous research did show that perceptual cues could become "more critical" through laboratory training and classroom learning. Non-native listener's perceptual cue weightings could be modified for consonants (Iverson et al., 2005), vowels (Ylinen et al., 2010) and tones (Chandrasekaran et al., 2010; Francis et al., 2008; Wiener, 2017). Furthermore, these studies showed that trainees developed a more native-like cue weighting pattern. In other words, they weighted the primary and critical perceptual cues of the target language more strongly after training, and consequently, the new perceptual weightings facilitated their perception of non-native speech sounds. Relating to tone perception, Chandrasekaran et al. (2010) showed that Mandarin tone train-

ing increased native English listener's perceptual weighting on F0 slope, the critical cue for Mandarin tone perception. As well, through Cantonese tone training in Francis et al. (2008), native Mandarin listeners increased the weight given to pitch height accompanied by a decrease of the weight given to pitch direction, forming a more native Cantonese-like cue weighting pattern. The modification of cue weighting is also influenced by the interaction between the sound systems of the trainee's native and non-native language, as Francis et al. showed that native English listeners who already gave a strong weighting to F0 height further increased the weighting given to this cue after Cantonese tone training. Since Cantonese has a three-way tone height distinction, the increase in weighting given to F0 height for both native Mandarin and English listeners has been attributed to a more fine-grained height distinction in Cantonese tones than in English intonations.

Regarding the non-critical cue, non-native listeners tend to retain the use of secondary, non-critical cues during the acquisition of L2 sound categories, especially when perceptual assimilation occurs. For instance, native Japanese listeners perceptually assimilate English /ɹ/ and /l/ sounds to the Japanese /ɾ/ category and show a bias towards the use of non-critical cues (e.g. F2 frequency, consonant-vowel transition duration), rather than the critical F3 cue, for English /ɹ/ and /l/ perception. The weightings of the non-critical perceptual cues were retained in the perceptual learning of English /ɹ/ and /l/, unless when stimuli were perceived to be dissimilar to the Japanese /ɾ/ category (Iverson et al., 2005). As for Mandarin tones, previous research also showed that the cue weight given to pitch height, the non-critical cue, did not change as a result of Mandarin tone perceptual training (Chandrasekaran et al., 2010; Francis et al., 2008). This is not surprising as native English individuals demonstrated perceptual assimilation in the perception of Mandarin tones through single category assimilation, especially for Tone 1-3 and Tone 1-4 pairs to the Statements intonation category (So and Best, 2011, 2014).

Apart from laboratory training, the modification of cue weightings has also been shown in a naturalistic classroom learning setting. For instance, Mandarin learners whose native language was English were found to attain more native-like cue weighting after the first three months of Mandarin classroom learning. Specifically, in the first three months

of Mandarin learning, it was shown that learners' weighting given to pitch direction had increased while the weighting given to pitch height remained unchanged (Wiener, 2017). These results corroborate the findings of the training studies reviewed above. However, it should be noted that Wiener interpreted the perceptual dimensions based on native Mandarin listener's perception results, even though the distribution of tone categories on the tone space varied with learning duration. For instance, Tone 1 and 4, instead of Tone 2 and 4, were separated by pitch direction after two months of Mandarin learning. In the third month, learners used the pitch direction dimension to separate Tone 2, with a rising contour, from other tones only. Tone 4, with a falling contour, was not correctly displayed at the other end of this dimension. As a result, Wiener (2017) displayed a more nuanced development of critical cue weighting. After three months of learning, learners were able to discriminate Tone 2 based on pitch direction better than other tones.

In sum, L2 learning redirects non-native learners to give stronger weight to the critical acoustic cue that facilitates non-native speech perception (Chandrasekaran et al., 2010; Francis et al., 2008; Iverson et al., 2005; Wiener, 2017; Ylinen et al., 2010). In particular, non-native individuals weight pitch direction more strongly after training or three months of learning in Mandarin tone perception (Chandrasekaran et al., 2010; Wiener, 2017). It is also possible that an intermediate step for cue weighting does not fully resemble that of native listeners, suggesting possible different pace of perception development across tone categories for individual learners (Wiener, 2017).

4.1.3 The present study

The present study explores if a perception-production relationship can be attributed to the use of critical perceptual cues. To demonstrate this, it investigates the learning effects on perception and production of critical and non-critical perceptual cues. Based on the review above, this study predicts that a transfer of learning effects from perception to production should occur for critical perceptual cues if these cues play a critical role in establishing a perception-production link. Specifically, the previous training and learning studies suggest that there should be an increase in weighting given to a critical perceptual cue during L2

learning (Chandrasekaran et al., 2010; Francis et al., 2008; Iverson et al., 2005; Wiener, 2017; Ylinen et al., 2010). At the same time, it should also improve the L2 production of the same acoustic cue as the result of a transfer of learning effects (Bradlow et al., 1997; Brosseau-Lapré et al., 2013; Hardison, 2003; Hazan et al., 2005; Herd et al., 2013; Hirata, 2004; Iverson et al., 2012; Kartushina et al., 2015; Lambacher et al., 2005; Wang et al., 1999, 2003). As for non-critical perceptual cues, it should demonstrate a weak learning effect in perception and should not show a clear transfer to learning effects across domains.

This study measured the learning effects by tracking the changes in the contribution of F0 slope (critical) and mean (non-critical) to the categorization of Mandarin learners' perceptual tone identification responses and tone productions in a four-week to six-week Mandarin learning period. Following Chapter 3, the cue contributions were examined by the same statistical learning method. The perception and production models with F0 slope and/or F0 mean as predicting variables were trained and tested by the data obtained from the same domain, and each model's performance was assessed by classification accuracy rates. Due to individual variability found in the learner's data (Refer to Section 4.3 for details), the models were trained by individual learner's data. As a result, each learner had their own model accuracy for each visit and domain, and consequently, this study was able to compare the improvement of the perception and production model accuracy across visits (i.e., learning effect) for F0 slope and mean quantitatively.

Since a simultaneous learning effect in perception and production indicates a link between the two domains, tracking the improvement of perception and production for each cue (i.e. F0 slope and mean) should demonstrate whether the perception-production link is attributed to the critical status of perceptual cues, achieving the aim of this study. Moreover, this design differs from previous training and learning studies which only demonstrates a perception-production transfer predominantly in terms of category identification in perception and native speaker's judgement in production. It also extends the previous research that focuses on the change of cue weighting in perception only to the change in the use of acoustic cues in both perception and production.

It is expected that the learning effect in perception should be found for F0 slope, as a critical perceptual cue. In contrast, F0 mean, as a non-critical perceptual cue, should show a minimal learning effect. Therefore, the F0 slope model classification accuracy should increase across visits while a significantly lesser degree of increase/change compared to that for F0 slope should be found for the F0 mean model. Based on previous training study results (Bradlow et al., 1997; Hardison, 2003; Hazan et al., 2005; Herd et al., 2013; Iverson et al., 2012; Lambacher et al., 2005; Sakai and Moorman, 2018; Wang et al., 1999, 2003) and the findings in Chapter 3, it is also expected that there should be a transfer of learning effect across the perception and production domains, as demonstrated by a simultaneous increase in the contribution of F0 slope to the classification of Mandarin tone productions in the statistical learning model if a perception-production link can be attributed to critical perceptual cues. In contrast, there should be a less obvious simultaneous change in the contribution of F0 mean in perception and production.

The perceptual and production gains are also compared in this study to investigate if there is a close connection between perception and production development, but based on Bradlow et al. (1997) and Sakai and Moorman (2018), the relationship between perceptual and production gains is expected to be weak. It is worth noting that, although previous studies generally found a trend of perception leading production in L2 learning as demonstrated by more perceptual gains than production gains (Bradlow et al., 1997; Flege and Schmidt, 1995; Flege et al., 1999; Sakai and Moorman, 2018), there were conflicting findings on the L2 learning of tones. Previous research showed a production lead in Mandarin learning since learners' L2 productions had a higher accuracy than perceptual identification (Yang, 2012) , but also a perception lead since learners were generally able to accurately discriminate all Southern Vietnamese tone pairs while there were much more inter-speaker variability in their tone productions (Kirby and Giang, 2021). Therefore, it is still uncertain which domain will yield greater improvement.

4.2 Method

4.2.1 Participants

The participants were 19 learners of Mandarin studying in the Mandarin Chinese program at National Chiao Tung University, Taiwan, between the ages of 19 and 28 (Mean: 23.7). They spoke Indonesian as their native language (Female = 11, Male = 8) and did not have any prior knowledge of tone languages other than Mandarin. Mandarin was the third language of the majority of the learners. All learners also spoke English as their second language. In addition, a few learners had knowledge of an additional non-tone language (e.g. Javanese, Arabic). All participants had studied Mandarin courses for 7 months when they were recruited to participate in this study. Based on the program admission interviews, these learners were regarded as beginner-level learners who had minimal knowledge of Mandarin and were still taking beginner-level courses when they participated in this study. They spent 2 to 4 class hours per week on learning Mandarin. The Mandarin classes had a focus on daily verbal communication (i.e., listening and speaking). The participants estimated that their daily use of Mandarin was between 0% and 24% outside the classroom (Mean: 10.3%). They all reported that they used the language in school. Five out of 19 learners also used the language with friends and one used it at work.

4.2.2 Stimuli

The production and perception stimuli were identical to the previous chapter (Refer to Section 3.2.2 and 3.2.3). There were 120 items of Mandarin words in the production task (3 monosyllables (/ɤ, /i/, /u/) $\times$ 4 tones $\times$ 5 repetitions $\times$ 2 speaking styles). The perception stimuli were 170 stimuli with resynthesized (7 F0 mean steps $\times$ 11 F0 slope steps) and natural tones (4 tones $\times$ 2 styles) produced by 2 talkers.

4.2.3 Procedures

In order to examine the development of Mandarin perception and production, participants took part in two visits which were 4-6 weeks apart. For each visit, the procedures were identical to the previous chapter (Refer to Section 3.2.2 and 3.2.3). The production task

was always carried out first, followed by the perception experiment. Each participant produced two sets of Mandarin tone data and responded to two sets of perception stimuli. The production data were log-normalized using Equation 2.1 and the F0 mean and slope values were obtained using Equation 3.1.

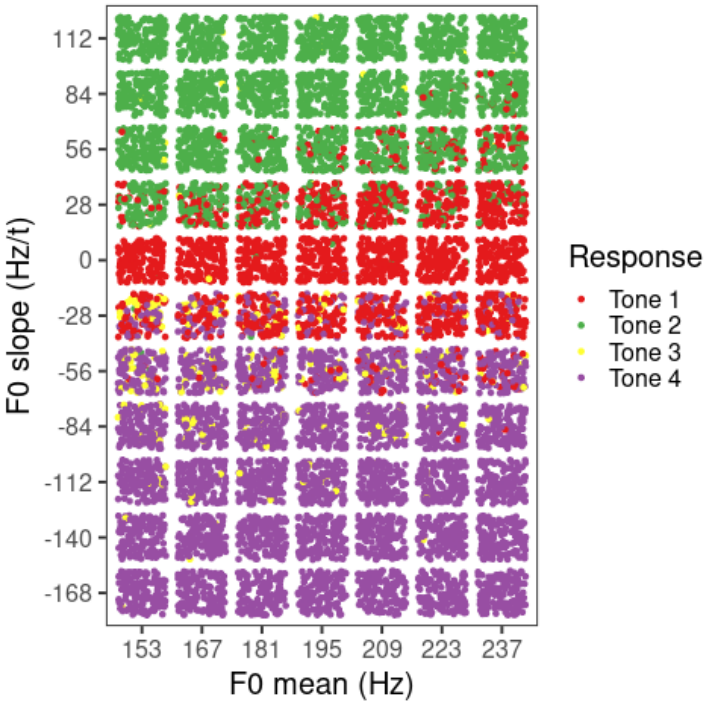
4.3 Results

The goal of this study is to compare the perception and production performance of Mandarin learners across two visits to understand the learning effect on the perception and production of F0 slope and mean. Specifically, the contributions of each acoustic cue (i.e. F0 mean or F0 slope) to tone categorization, assessed by model accuracy, were examined separately in perception and production for each visit. Since learners showed noticeable inter-speaker variability in perception (Section 4.3.1) and production (Section 4.3.2), the statistical learning was carried out separately for each learner. That is, each set of learner's perception and production data obtained in either visit 1 or 2 was used to train and test a statistical learning model that consisted of F0 slope and/or F0 mean as predictor factors. For perception data, each learner's set of data was analyzed by a multinomial logistic regression model to categorize their perceptual responses using F0 mean or slope alone. The production data were analyzed in the same manner using a linear discriminant analysis. The model classification accuracy was used to indicate the contribution of F0 mean or slope on Mandarin tone perception and production by these learners for each visit. The learning effect was shown by the improvement in model classification accuracy from visit 1 to 2. The comparison between the classification accuracy fo the F0 slope and mean models illustrated the contribution of each cue to perception and production. A simultaneous improvement across visits for perception and production should indicate a perception-production link. Finally, as a measure of the pace of learning across domains, the correlation between the gains in perception and production was also examined.

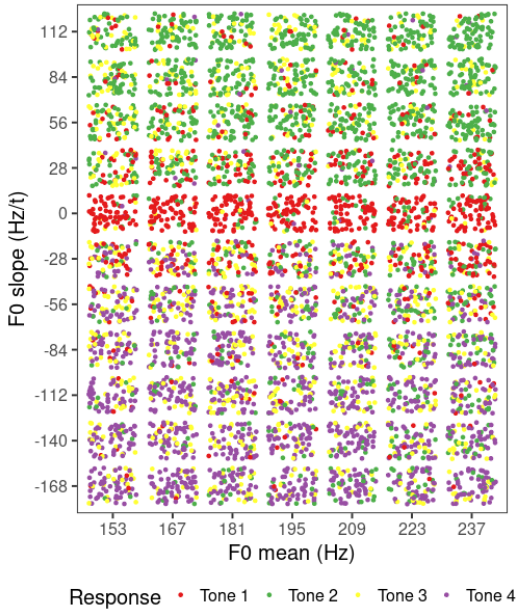
4.3.1 Perception model classification results

The tone categorization responses of all learners for each level of F0 mean and slope value are displayed in Figure 4.1b (Visit 1) and 4.1c (Visit 2). There were 5852 responses in total per visit (308 per learner = 77 stimuli per block $\times$ 4 blocks). The trials in which participants did not give a response were removed (Visit 1 = 96; Visit 2 = 30). In general, the majority of the rising, level and falling tone items were identified as Tone 2, Tone 1 and Tone 4, respectively, and were within the expected range of each of these tones. The Tone 3 responses were dispersed, but were found mostly with low F0 mean in the rising tone region, especially in Visit 2. Compared with the responses of native Mandarin participants in Figure 4.1a (or Figure 3.2), F0 slope appears to have a stronger influence on Mandarin tone perception for learners than for native Mandarin individuals. It appears that tone boundaries between Tone 1 and 2, and Tone 1 and 4 are horizontal on the scatterplots, indicating that F0 slope was the only factor that influenced the perception of Tone 1, 2 and 4. F0 mean may only have an impact on Tone 3 perception when the tone contour is rising. As for native Mandarin individuals, although F0 slope also exerted a strong influence on tone identification, F0 mean also appeared to slightly influence the perception of Tone 1, 2 and 4 in the slightly rising and falling tone regions. As the F0 mean decreased, an increased number of slightly rising and falling tones were identified as Tone 2 and 4, respectively. However, the learners' perceptual responses did not appear to show this trend.

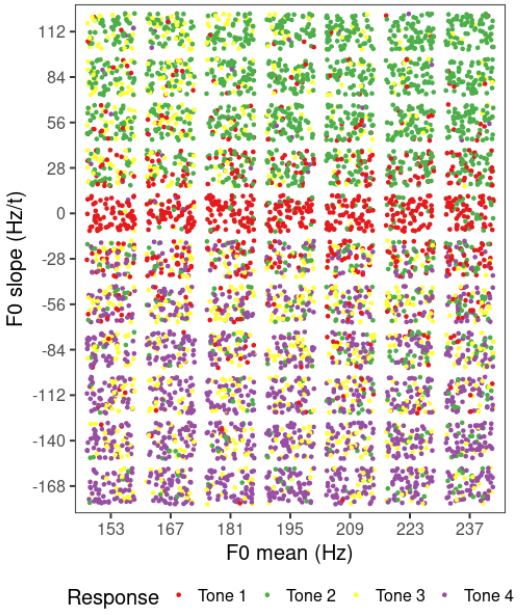
Notably, learners displayed a considerable amount of inter-subject variability, which could be demonstrated by the tone responses of two learners. In Visit 1, the first learner, LM02, who reported to use Mandarin in their workplace, showed a typical response pattern which followed the general tone categorization pattern of all learners in Figure 4.1 and native Mandarin individuals in Figure 4.1a. It showed that this learner used F0 slope only to categorize Tone 1, 2 and 4, and Tone 3 responses mainly had a falling contour in Visit 1 (Figure 4.2a. The Tone 1-2 and Tone 1-4 boundaries along the F0 slope dimension remained the same in Visit 2. The low rising tones were identified as Tone 3 instead of Tone 2, and a substantially reduced number of falling contours, especially those with steep falling trajectories, were categorized as Tone 3 compared with Visit 1. In contrast, the other



(a) Native Mandarin participants



(b) Learners: Visit 1

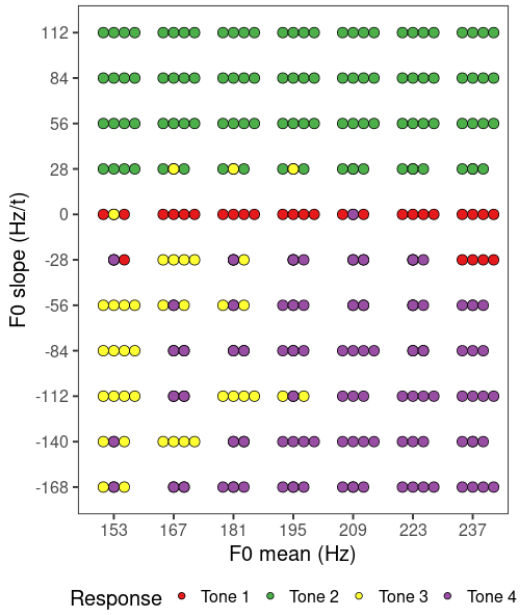


(c) Learners: Visit 2

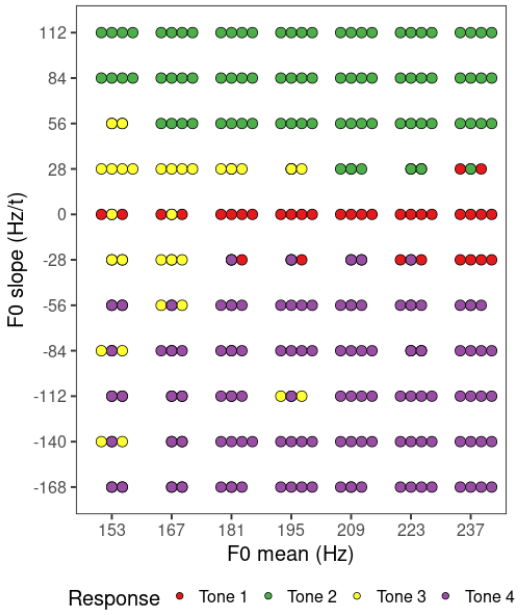
Figure 4.1: Tone responses of native Mandarin participants and Mandarin learners in perception task for each F0 mean and slope level

learner, LM05, used F0 mean to differentiate between Tone 2 (high F0 mean) and the other tones (low F0 mean). The canonical high-level Tone 1 responses were surprisingly found in the low F0 mean region. F0 slope had some impact on tone perception in this low F0 mean region in which rising tone was mostly often categorized as Tone 1 and falling tone as either Tone 3 or 4 (Figure 4.2c). In Visit 2, F0 slope became more prominent since more level or near-level tone items were categorized as Tone 1. Most Tone 2 responses were rising and most Tone 3 were falling. F0 mean remained to show influence as most of the highest F0 mean items (237Hz) were identified as Tone 2 (Figure 4.2d). Given that there was substantial variability in the learner data, the subsequent statistical learning was carried out for each learner separately since the effect of the critical F0 slope and non-critical F0 mean cues could have highly variable influence on each learner's perception. Moreover, if these learners displayed a highly uniform pattern in the learning effect of critical and non-critical cues in perception production, it should further support a perception-production link established by critical perceptual cues.

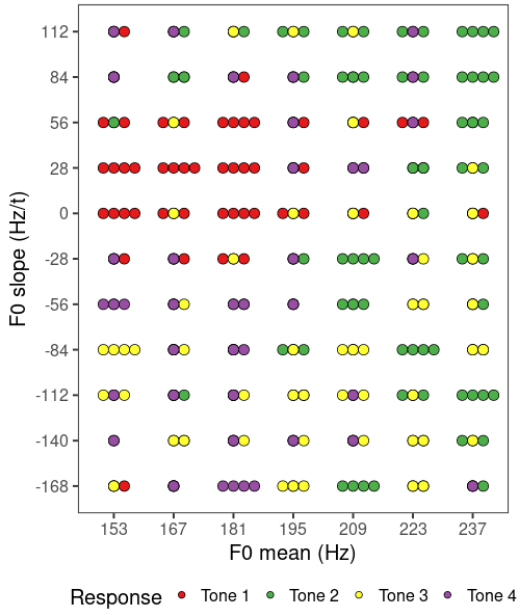
Similar to Chapter 3 (Section 3.3.1, a multinomial logistic regression (*nnet* package (Venables and Ripley, 2002)) was used to explore the influence of F0 mean and slope on tone responses. Since each data set contained a substantially fewer number of responses (at most 308 per learner), either F0 mean (153-237Hz) or F0 slope (112 to -168Hz/t) was used as a predictor variable in each model due to model convergence issues when both predictors were used for some learners. The dependent variable was tone identification responses (Tone 1, Tone 2, Tone 3, Tone 4) in each visit. All models had Tone 1 as the reference level. In order to test which variable performed better in tone categorization, the mean Deviance statistic (D) (minus 2 times log-likelihood of the model) and mean classification accuracy were obtained using the same five-fold cross validation method (Refer to Section 3.3.2). Each learner had three models - (1) Null model containing only the intercept; (2) F0 mean-only model; and (3) F0 slope-only model. The differences in *D* between the null and each one-predictor model were used to evaluate model fit qualitatively and the classification accuracy rates were used to compare between models statistically.



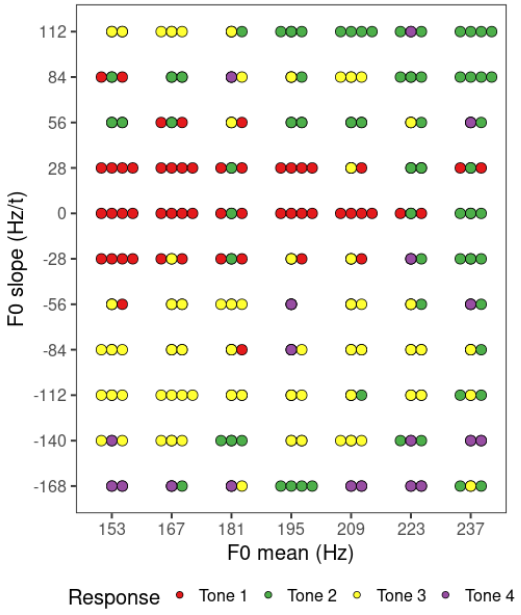
(a) LM02: Visit 1



(b) LM02: Visit 2



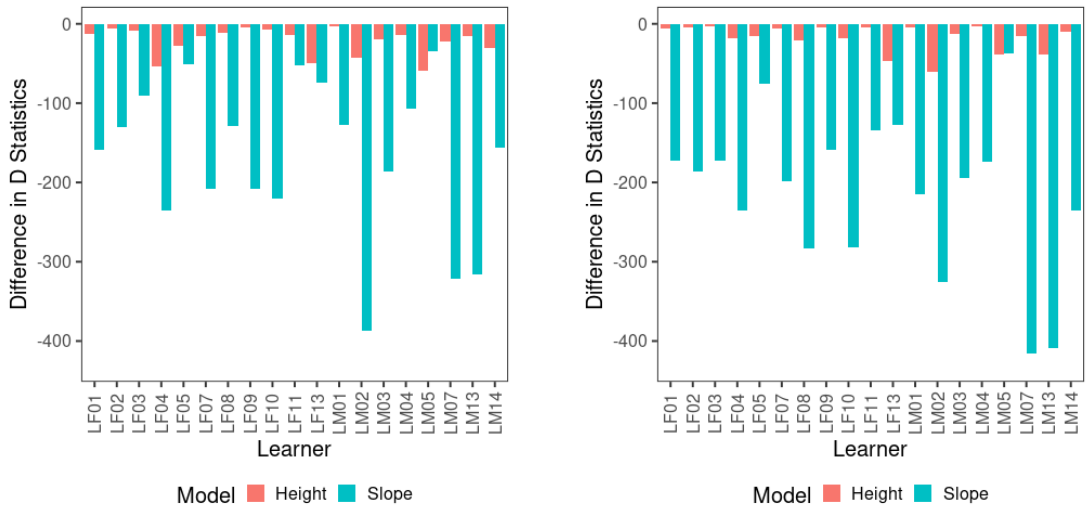
(c) LM05: Visit 1



(d) LM05: Visit 2

Figure 4.2: Tone responses of two learners (LM05 and LM02) in perception task for each F0 mean and slope level

Figure 4.3 displays the difference in D between the null model and each of the one predictor models across two visits for all learners (ΔD). Note that a smaller D represents a better model fit. Therefore, the null model containing only the intercept should have the greatest D. A greater drop in D compared with the null model indicates that the factor of the model is a better predictor than the factor of the model with a smaller drop in D. As shown in Figure 4.3, there was a huge drop in D statistics from the null model to the F0 slope-only model, and the drop in D statistics for the F0 mean-only model was comparatively limited across visits for most learners. The only exception was LM05 who yielded a greater drop in D for F0 mean than F0 slope. These results indicate that F0 slope improved model fit considerably compared to F0 mean for most learners, and therefore F0 slope had been the critical perceptual cue for tone categorization for a majority of learners since the first visit. The ΔD between the null and F0 slope-only model appears to increase in Visit 2 compared to Visit 1 for a number of learners.



(a) Visit 1

(b) Visit 2

Figure 4.3: Difference in D Statistics between the null model and one predictor models for all learners

Next, to test whether the models displayed a learning effect from Visit 1 to 2 and whether F0 slope yielded a better performance than F0 mean, the classification accuracy rates of three models (i.e., null, F0 mean-only and F0 slope-only) were compared between the two visits. As presented in Figure 4.4a, the F0 slope-only model performed much bet-

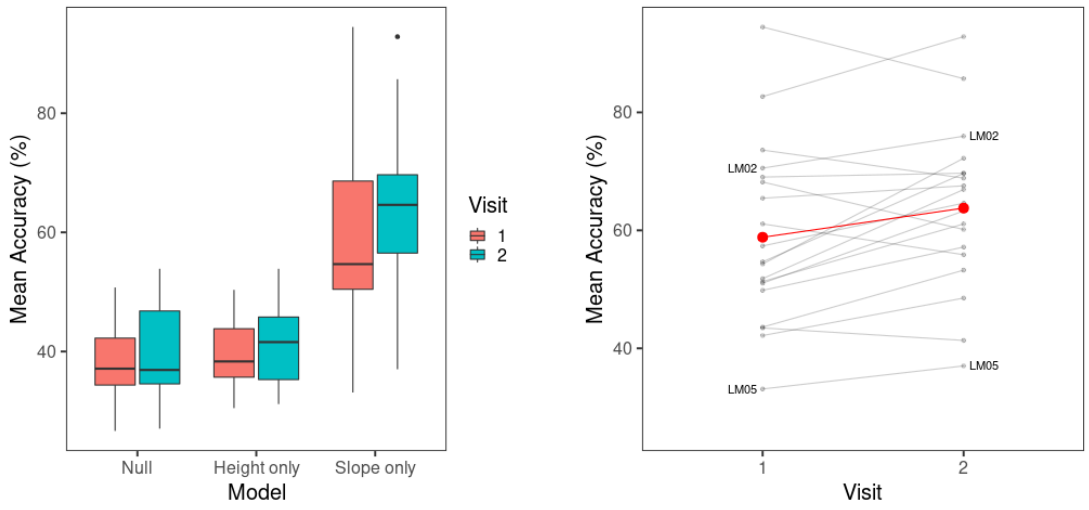
ter than the F0 mean-only model, consistent with the ΔD results. The F0 mean only and null model did not show observable difference. The accuracy difference across visits was not clear by a visual inspection of the data. The individual learner's mean classification accuracy and F1 scores of each tone are presented in Appendix E.

These results were further examined in a repeated measures ANOVA with the *afex* package (Singmann et al., 2021) in *R* (R Core Team, 2021). The independent variables were Visit (1, 2) and Model (Null, F0 Mean only, F0 Slope only). The mean classification accuracy of the multinomial logistic regression models of individual learners was the dependent variable. The tests for violation of the sphericity assumption yielded significant results for Model ($W = 0.279$, $p < 0.001$) and the Visit $\times$ Model interaction ($W = 0.279$, $p = 0.006$), so the Greenhouse-Geisser correction was used for these test results. A significant result was detected for the main effect of Visit ($F(1,18) = 6.73$, $p = 0.018$) and Model ($F(1.16,20.9) = 48.5$, $p < 0.001$). Overall, the models showed a higher mean accuracy in Visit 2 (48.3%) than in Visit 1 (45.6%). The interaction between Visit and Model was not significant ($F(1.38,24.8) = 2.02$, $p = 0.164$).

Post-hoc pairwise comparisons were conducted for Model with Tukey's adjustment using the *emmeans* package (Lenth, 2021). The F0 slope-only model had a significantly higher accuracy (Mean = 61.3%) than the F0 mean-only model (Mean = 40.6%) ($t(18) = -6.68$, $p < 0.001$) and null model (Mean = 39.1%) ($t(18) = -7.64$, $p < 0.001$). The F0 mean only and null model did not show a significant difference in model accuracy ($t(18) = -1.44$, $p = 0.340$).

The above results demonstrated that F0 slope-only model resulted in a significant improvement of model categorization accuracy compared to the Null model, but the F0 mean-only model did not perform better than the null model that contained no predictor. Therefore, the categorization performance of the F0 mean-only model simply reflected the tone response that was *a priori* more likely than others (McMurray and Jongman, 2011). Although ANOVA yielded a significant main effect of Visit, the improvement of model accuracy across visits for the F0 mean-only model did not capture the improvement of categorization performance by this acoustic cue. To confirm that F0 slope alone resulted in

a better categorization performance in Visit 2 than in Visit 1, a separate paired-sample t test was carried out with Visit as the independent variable and model accuracy as the dependent variable for F0 slope model data only. It yielded a significant difference ($t(18)$ = -2.70, p = 0.015) indicating that the F0 slope model had a higher model accuracy in Visit 2 (63.8%) than Visit 1 (58.8%), and therefore, an increase in the weighting of this cue in perception. The improvement in overall mean accuracy of this model and the mean accuracy for individual learners are presented in Figure 4.4b. Most learners showed an increase in classification accuracy from Visit 1 to 2, including LM02 whose data were presented above. As for LM05, this learner yielded the lowest mean classification accuracy among all participants, reflecting the weak influence of F0 slope on tone categorization as demonstrated in Figure 4.2c and 4.2d. As mentioned above, this learners showed some improvement in the use of F0 slope in Visit 2. Likewise, the model classification accuracy also improved from Visit 1 to 2.



(a) All models

(b) F0 slope-only model

Figure 4.4: (a) Mean classification accuracy of multinomial logistic regression models across visits. The whiskers of the boxplots represent values within 1.5 times of the interquartile range above 75% and below 25% percentile. A dot represents any value outside that range beyond either end of the box. (b) Mean classification accuracy of the F0 slope-only model, red represents overall mean accuracy; grey represents mean accuracy for individual learners (Lines for LM02 and LM05 are labelled).

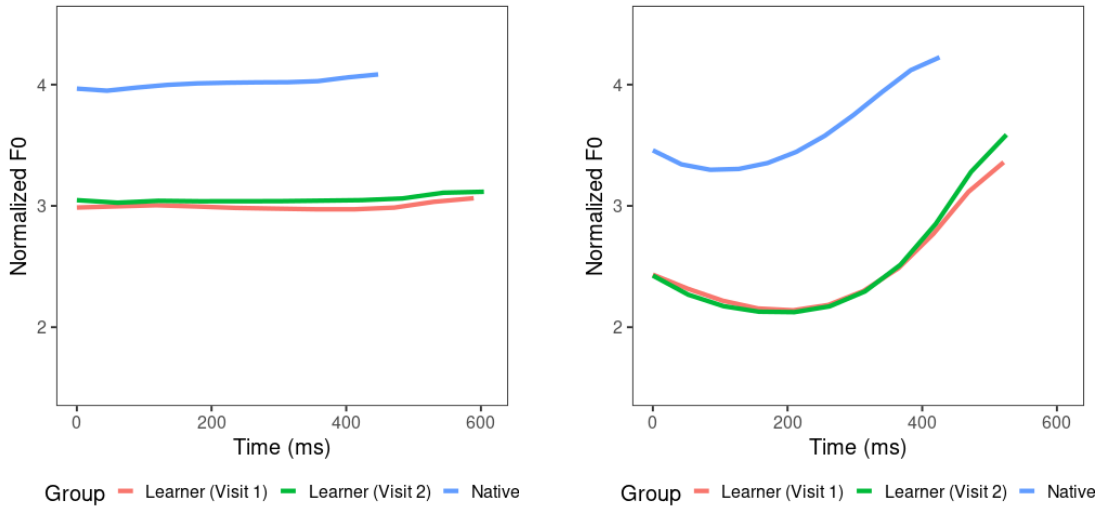
4.3.2 Production model classification results

Figure 4.5 depicts the mean tone contours for the productions of all learners compared with mean native productions. For learners' productions across visits, the tone contours indicate that learners generally produced a higher Tone 2 and 3 offset, and a slightly lower Tone 4 offset in Visit 2 than in Visit 1. These suggest that these tones became steeper as the learners went through the 4-6 weeks of Mandarin learning. The overall F0 mean remained similar across visits. Learners produced Tone 1 with the target level contour, and on average the overall F0 mean was slightly higher in Visit 2 than in Visit 1. Tone 4 was produced with a substantially shorter duration compared with other tones, leading to a steep falling slope.

Compared with native productions, learners generally produced similar contours for all tones in both visits. However, learners' Tone 3 showed a much higher F0 offset than onset, differing from the native Tone 3 that had comparable F0 onset and offset levels.

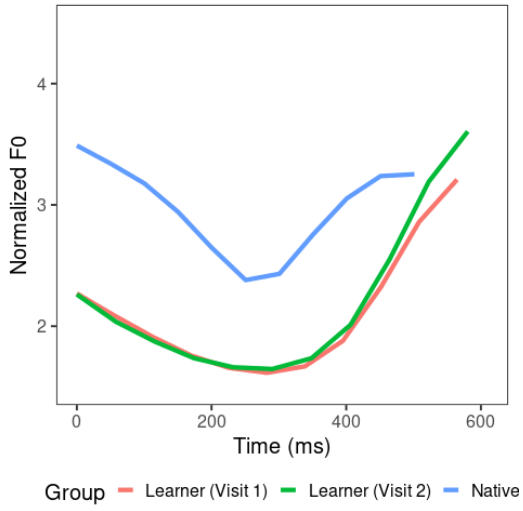
The distribution of tone production items in the F0 mean and F0 slope tone space are displayed on Figure 4.6. For both visits, Tone 1, 2 and 4 items are distributed along the F0 slope dimension. Tone 1 occupies the region close to F0 slope at zero value, indicating a flat tone contour. Tone 2 and 4 are found in the positive, rising and negative, falling slope region, respectively. A small number of Tone 2 productions are in the falling slope region. Tone 3 productions are more dispersed in Visit 1 than in Visit 2 and are mostly found in the rising slope region, overlapping with Tone 2 productions in Visit 2. Tone 2 appears to have slightly more productions in the high F0 mean region than Tone 3.

As in the perception data, the production data also contained a substantial inter-subject variability. For example, LM02, the learner who reported to use Mandarin in their workplace, produced tone contours similar to the averaged tone contours across all learners. As shown in Figure 4.7, the most noticeable change from Visit 1 to 2 is the change of Tone 3 from a falling to a dipping contour. However, this brought the F0 slope of Tone 3 closer to that of Tone 2 and two tones display a lot more overlap in the second than the first visit. The tone space plots confirm this observation (Figure 4.8). Tone 2 and 3 are less clearly separated in Visit 2 than Visit 1.

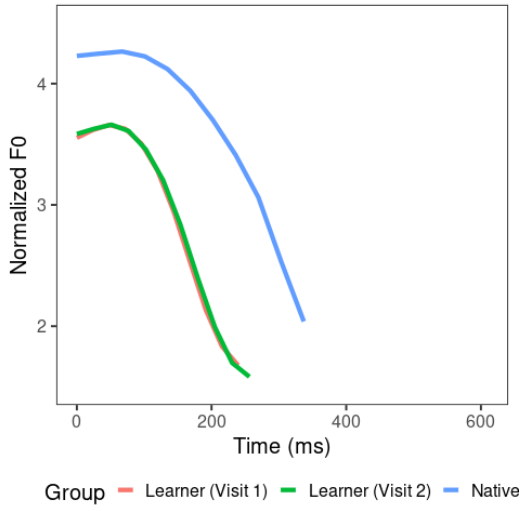


(a) Tone 1

(b) Tone 2

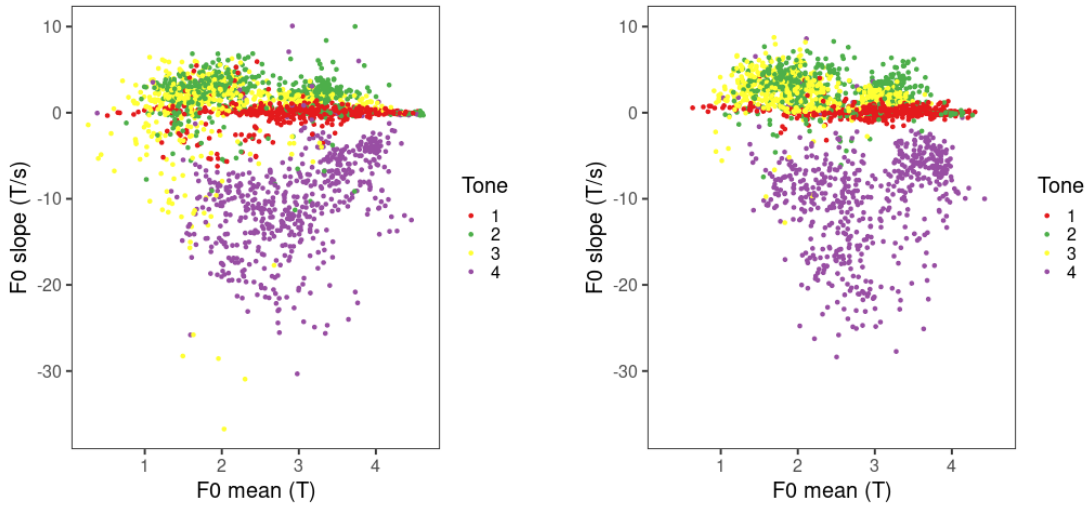


(c) Tone 3



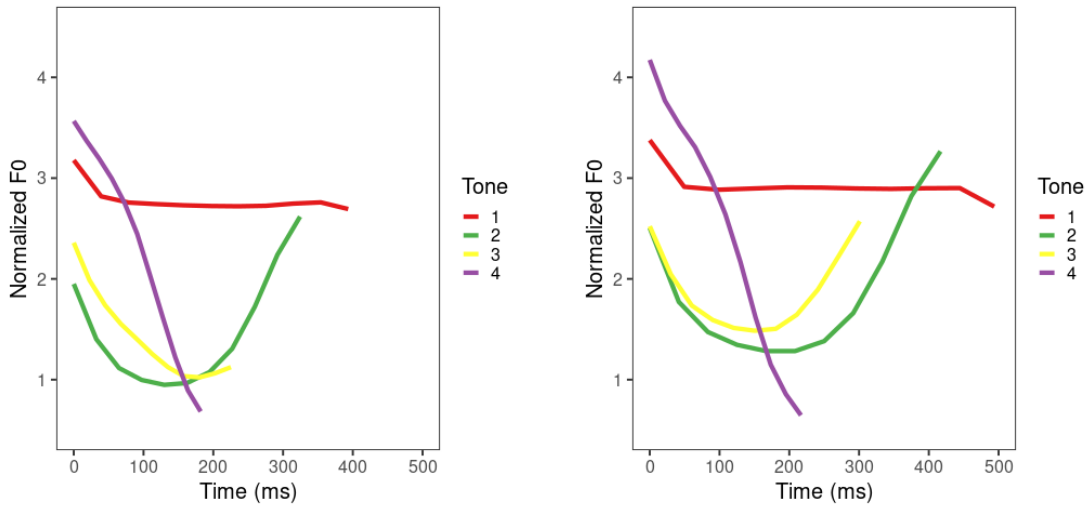
(d) Tone 4

Figure 4.5: Mandarin tone contours averaged across production items produced by all native participants and learners across visits.



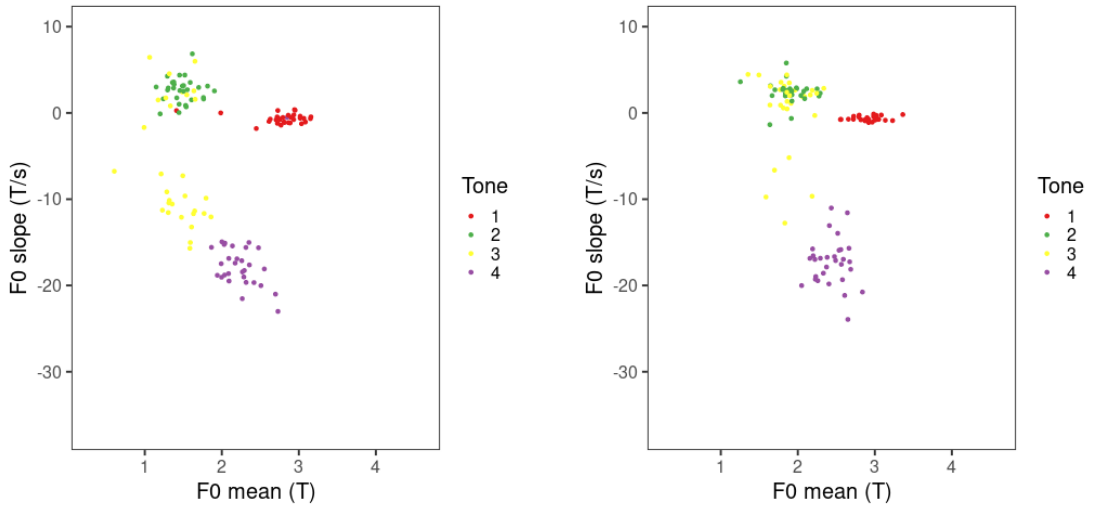
(a) Visit 1 (b) Visit 2

Figure 4.6: Mandarin tone productions of all learners in F0 mean and slope across visits. Left panel: Visit 1; Right Panel: Visit 2



(a) Visit 1 (b) Visit 2

Figure 4.7: Mandarin tone contours averaged across production items produced by LM02 across visits.

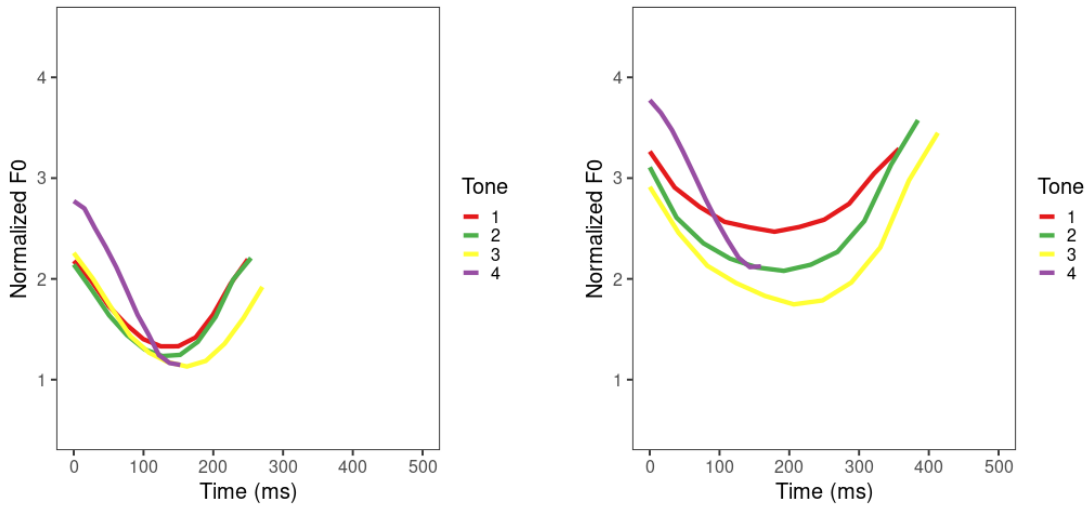


(a) Visit 1 (b) Visit 2

Figure 4.8: Mandarin tone productions of LM02 in F0 mean and slope across visits.

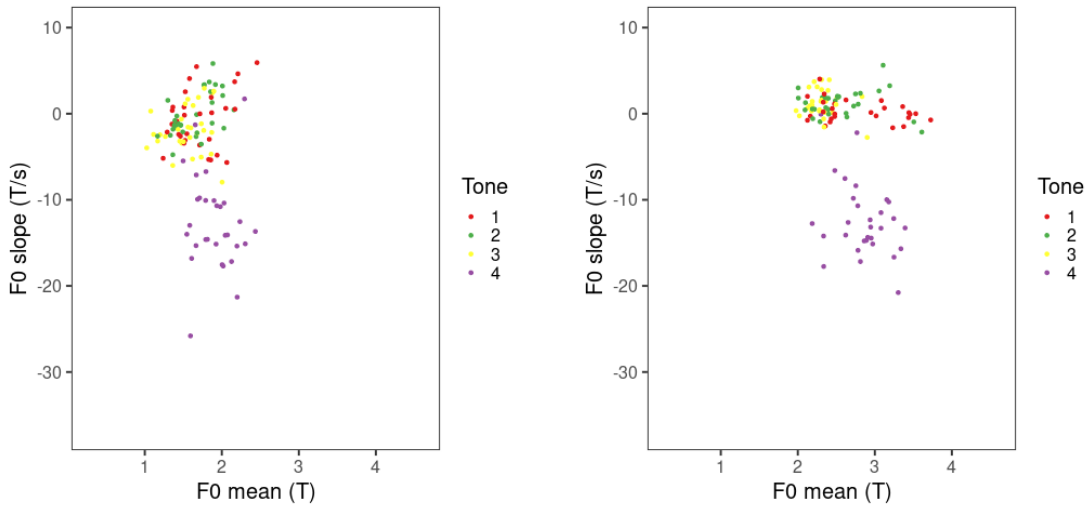
Another learners, LM04, produced Tone 4 with a distinctive falling contour, but the other three tones had a very similar dipping contour in the first visit. These three tones maintained the same contour shape in visit 2, but their averaged tone contours showed different F0 mean levels (Figure 4.9). As a result, the tone contrast improved on the F0 mean dimension as this learner had more experience in Mandarin. As shown in Figure 4.10, Tone 1, 2 and 3 show more overlap in Visit 1 than in Visit 2, and the three tones appear to be more dispersed along the F0 mean dimension in Visit 2. Given the inter-speaker variability, the following statistical modelling of production data was carried out separately for each learner.

To analyze the influence of F0 mean and slope on production tone categorization, a linear discriminant analysis (LDA) was performed for each learner and visit. An LDA was used instead of a multinomial logistic regression because there were fewer observations in each learner's production data (n = 240) than perception data (n = 680). With a limited number of observations, most learners' multinomial logistic regression model failed to converge. However, because there were equal numbers of observations in each tone category, it was possible to run LDA instead. The classification accuracy of the models with the linear combination of F0 mean and slope (i.e. F0 Mean + F0 Slope), F0 slope only and F0 mean only were compared to test which cue had a stronger influence on production tone cate-



(a) Visit 1 (b) Visit 2

Figure 4.9: Mandarin tone contours averaged across production items produced by LM04 across visits.



(a) Visit 1 (b) Visit 2

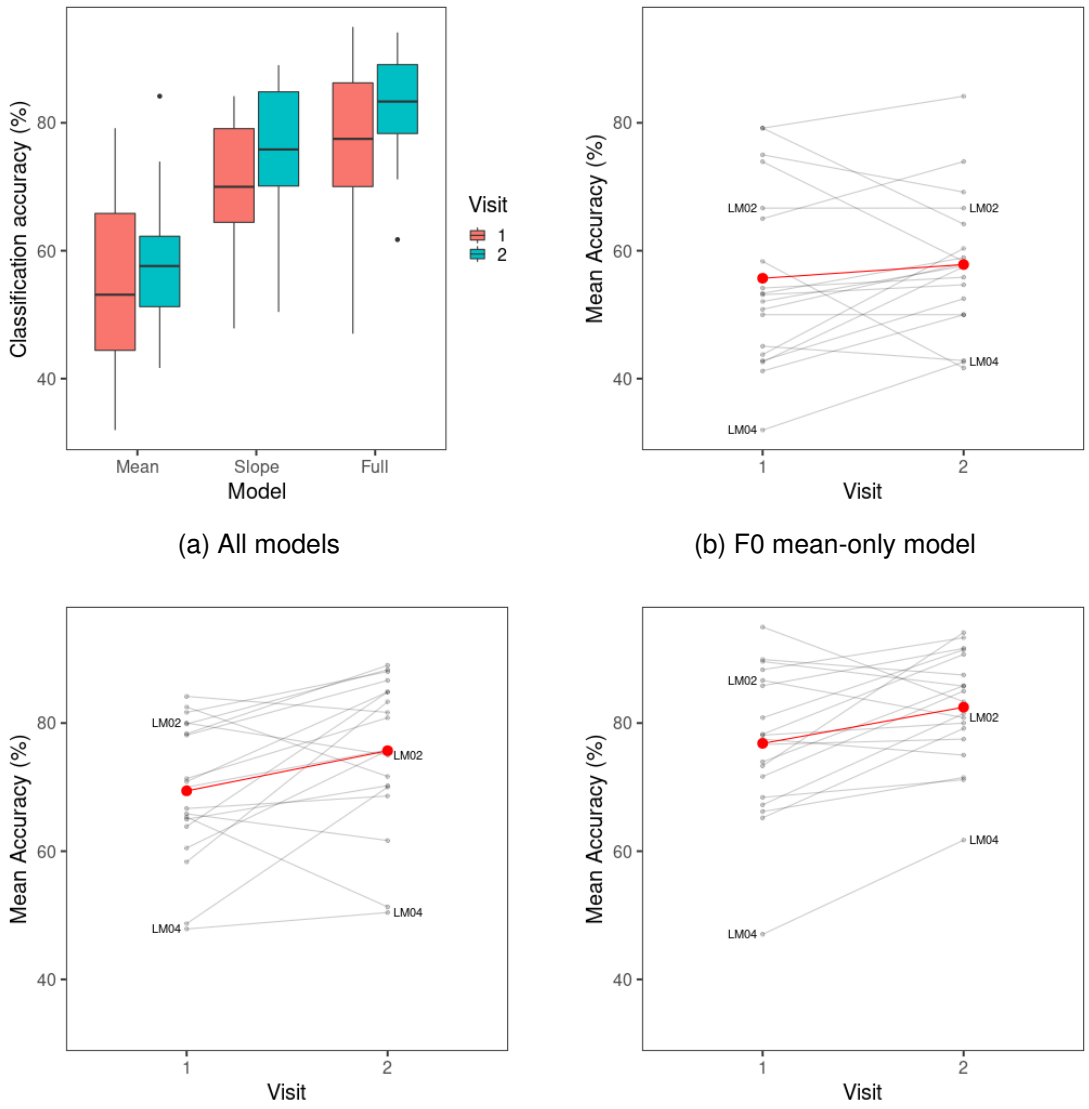
Figure 4.10: Mandarin tone productions of LM04 in F0 mean and slope across visits.

gorization across visits (*cf.* Tupper et al., 2020). The classification accuracy was obtained by the five-fold cross validation method used in Section 3.3.2). The results are presented in Figure 4.11a. The F0 mean-only model yielded the lowest model classification accuracy among all models. The classification accuracy of F0 slope-only models is noticeably higher than that of the F0 mean model but lower than the F0 mean + F0 slope models. The figure displays some increase in classification accuracy from Visit 1 to 2, especially for the F0 slope-only models. The individual learner's mean classification accuracy and F1 scores of each tone are presented in Appendix F.

A repeated measures ANOVA was carried out with the *afex* package (Singmann et al., 2021) in *R* (R Core Team, 2021). The independent variables were Visit (1, 2) and Model (F0 mean-only, F0 slope-only, F0 mean + F0 Slope). The mean classification accuracy of LDA of individual learners was the dependent variable. Model ($W = 0.602$, $p < 0.001$) showed a significant result in the test of sphericity, so the Greenhouse-Geisser correction was used for the ANOVA results of this effect. A significant main effect of Visit ($F(1,18) = 6.71$, $p = 0.018$) and Model ($F(1.21,21.7) = 88.3$, $p < 0.001$) were found. Overall, there was a higher mean classification accuracy in Visit 2 (Mean = 74.1%) than in Visit 1 (Mean = 69.5%). The Visit $\times$ Model interaction was not significant ($F(2,36) = 2.01$, $p = 0.149$).

The significant main effect of Model was followed up by post-hoc pairwise comparisons with Tukey's adjustment using the *emmeans* package (Lenth, 2021). The F0 mean + F0 slope model had a higher classification accuracy (Mean = 79.7%) than the F0 slope-only (Mean = 72.6%) ($t(18) = -6.06$, $p < 0.001$) and the F0 mean-only model (Mean = 56.8%) ($t(18) = -14.8$, $p < 0.001$). The F0 slope-only model had a higher classification accuracy than the F0 mean-only model ($t(18) = -6.69$, $p < 0.001$).

The results indicate that learners' Mandarin tone productions are best categorized using both F0 slope and mean. As in perception, F0 slope alone was more critical than F0 mean alone in tone production categorization. However, unlike perception model classification, F0 mean showed an above chance level (25%) mean classification accuracy, suggesting that this cue alone could also contribute to tone categorization in production,



(a) All models

(b) F0 mean-only model

(c) F0 slope-only model

(d) F0 mean + F0 slope model

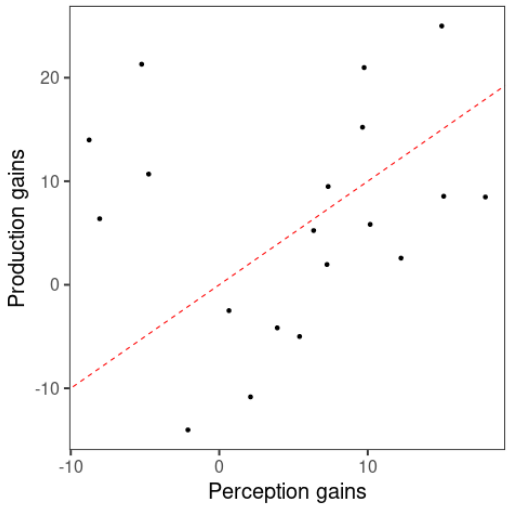
Figure 4.11: (a) Mean classification accuracy of linear discriminant analysis across visits. The whiskers of the boxplots represent values within 1.5 times of the interquartile range above 75% and below 25% percentile. A dot represents any value outside that range beyond either end of the box. (Mean: F0 mean only; Slope: F0 slope only; Full: F0 mean + F0 slope). (b–d) Mean classification accuracy of the F0 mean-only, F0 slope-only and F0 mean + F0 slope model, red represents overall mean accuracy; grey represents mean accuracy for individual learners (Lines for LM02 and LM04 are labelled).

consistent with the production model classification results of native Mandarin participants (Section 3.3.2).

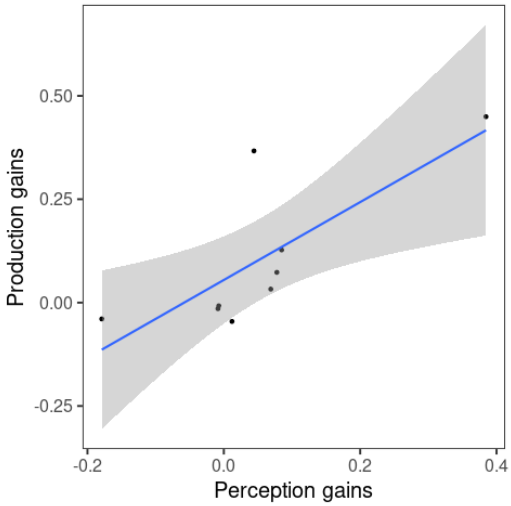
The improvement in overall mean accuracy and the mean accuracy for individual learnersof all models are presented in Figures 4.11b, 4.11c and 4.11d. Most learners showed an increase in classification accuracy from Visit 1 to 2, including LM04 who had an greater increase for the models with F0 mean as a predictor, reflecting the clearer distinction of the Tone 1, 2 and 3 contours on the F0 mean dimension in Visit 2 than in Visit 1. There was a drop of classification accuracy for a few learners, including LM02 for the models including F0 slope. The decrease reflected the higher number of change of Tone 3 productions from a slightly falling contour in Visit 1 (Figure 4.7a) to a dipping contour in Visit 2 (Figure 4.7b), which drew the F0 slope of Tone 3 closer to that of Tone 2. It potentially lowered the the classification accuracy of the models with F0 slope as a predictor.

4.3.3 Correlation analysis of perception and production gains

The relationship between perception and production gains from Visit 1 to Visit 2 was examined by a correlation analysis of the gains in model classification accuracy of perception and production. Since F0 mean did not show better classification performance than the null model in perception, the gains of the F0 slope-only models alone were examined using a Spearman's rank-order correlation due to violation of normal distribution with outliers. As expected, no significant correlation was found for F0 slope ($\rho(17) = 0.168$, $p = 0.244$). As displayed in Fig 4.12a, the data points are found on both sides of the red dashed line representing the same amounts of perception and production gains. Therefore, there was not a clear indication of greater gains in perception than in production. As for individual tones, a significant result was obtained for the gains in F1 scores of Tone 3 only ($\rho(7) = 0.800$, $p = 0.014$) (Figure 4.12b). However, this result should be interpreted with caution since only 9 learners yielded F1 scores in both perception and production for this tone.



(a) Mean accuracy of all tones; Red dashed line represents equal amounts of perception and production gains (i.e. y=x).



(b) F1 scores of Tone 3; The blue line represents the best fit line.

Figure 4.12: Scatterplot of perception and production gains of the F0 slope model. Each point represents one participant.

4.4 Discussion

This study examined the change in the contribution of F0 mean and slope to perception and production for beginner level Mandarin learners. The results confirmed the prediction that the F0 slope-only models showed improved classification accuracy across visits for both perception and production. On the other hand, only production data demonstrated an improvement in classification accuracy for the F0 mean-only model. As in previous studies (Bradlow et al., 1997; Sakai and Moorman, 2018), this study did not find a correlation between perception and production gains for overall classification accuracy.

4.4.1 Perception-production links established through F0 slope, the critical perceptual cue

This study showed that only F0 slope, as a critical perceptual cue, displayed categorization performance that was driven by the cue itself in the multinomial logistic regression analysis in both visits, as the model accuracy was higher than that of the null model. The result indicates that F0 slope played a predominant role in learners' Mandarin tone perception. The categorization performance is consistent with the native Mandarin perception reported

in Section 3.3.3 in which the F0 slope-only model showed much higher classification accuracy than the null model. In addition, this study also found an increase in classification accuracy of the F0 slope-only model in perception across two visits. Therefore, learners increased their weight given to F0 slope, the critical perceptual cue, to categorize Mandarin tones, corroborating the previous tone perception training and learning studies that found an increase in perceptual weighting on pitch direction (Chandrasekaran et al., 2010; Francis et al., 2008; Wiener, 2017) and other training studies which showed increased weights given to a critical perceptual cue in a few weeks of training (Iverson et al., 2012; Ylinen et al., 2010).

More importantly, the simultaneous improvement in perception and production for F0 slope extends the finding in Chapter 3 since it provides further evidence that establishing a perception-production link is attributable to the use of critical perceptual cues. In contrast, the F0 mean-only model showed an improvement in classification accuracy in production only. Therefore, a simultaneous improvement in both perception and production was not found for F0 mean, the non-critical perceptual cue. As demonstrated in previous training studies (Bradlow et al., 1997; Hardison, 2003; Hazan et al., 2005; Herd et al., 2013; Iverson et al., 2012; Lambacher et al., 2005; Sakai and Moorman, 2018; Wang et al., 1999, 2003), a transfer of learning effect across the perception and production domains was a factor driving the simultaneous improvement. In addition, learners performed the same tasks in the two visits and the same cues were measured in both perception and production. Therefore, this study directly examined the use of the same F0 slope and F0 mean cues in learners' perception and production of Mandarin tones. Consequently, the simultaneous learning effect on F0 slope-only model accuracy can clearly be attributed to the acoustic cue itself in both perception and production domains.

It should be noted that, comparing with native Mandarin perception, the classification accuracy of most learners' F0 slope-only model was lower (Figure 4.4a) than the overall native Mandarin individuals' F0 slope-only model (Table 3.8). It indicates that learners with 7 months of Mandarin learning experience did not rely on the F0 slope cue as heavily as native Mandarin individuals in Mandarin tone categorization. Although learners were

shifting closer to the native direction since the weighting on F0 slope continued to rise over a period of 4-6 weeks of Mandarin learning, the learners' F0 slope-only model of the second visit still attained lower accuracy rate than that of native Mandarin individuals. While the result is not surprising given that the participants were beginner-level learners and the time interval between the two visits was short, the learners' perception of Mandarin tones may involve other cues. Previous studies suggest that a possible cue involved in non-native Mandarin tone perception is pitch height (Chandrasekaran et al., 2010; Francis et al., 2008; Guion and Pederson, 2007; Gandour, 1983). This study does not fully support this view since F0 mean did not contribute to learners' Mandarin tone perception. The accuracy of the F0 mean-only model was not different from that of the null model, However, it is possible that learners used pitch height-related cues to perceive Mandarin tones, such as F0 onset as examined in Chapter 2. However, their patterns should continue to change as they gain further linguistic experience in Mandarin.

Lastly, although this study showed that learners' improvement in the perception and production of F0 slope, it does not necessarily mean that learners categorized the tone items like native Mandarin individuals. For instance, learners tended to perceive rising tone contours with a low F0 mean as Tone 3 (Figure 4.1), whereas native Mandarin individuals' Tone 3 responses were more likely to be found in the falling tone region with a low F0 mean (Figure 3.2). In addition, the slightly rising and falling tone contours with a high F0 mean were more likely to be categorized as Tone 1 for native Mandarin individuals, but were hardly noticeable for learners across two visits. In production, learners still showed a substantial overlap of Tone 3 with the other tones, especially Tone 2, in the F0 mean and slope tone space (Figure 4.6) compared with native Mandarin productions (Figure 3.5). Therefore, it suggests that learners still demonstrated a different tone categorization pattern from that of native Mandarin individuals even if the weighting given to the critical perceptual cue, F0 slope, and its contribution to tone production increased as the result of Mandarin learning.

4.4.2 Learning effect on F0 mean, the non-critical perceptual cue

In perception, the F0 mean-only model did not show any difference in classification accuracy from the null model in both visits, suggesting that F0 mean did not have a clear contribution to the perceptual tone categorization for learners whose native language was Indonesian - a non-tonal language. It is consistent with native Mandarin perception results (Section 3.3.3) in which no noticeable difference was found between the F0 mean-only and null model. Moreover, the improvement of classification accuracy across visits could not be attributed to the cue itself, since a difference between the classification accuracy of the F0 mean-only and null models was not detected in both visits. Therefore, such improvement did not show that F0 mean became a stronger perceptual cue weighting after 4-6 weeks of Mandarin learning. Together with the F0 slope-only model results, this study showed that Mandarin learners with a non-tonal language background already gave a low perceptual weighting to F0 mean as in native Mandarin perception (Table 3.8) after 7 months of Mandarin learning.

As for production, the model accuracy of the F0 mean-only models across visits was above chance level, showing that F0 mean had an influence on learners' tone production. Therefore, it is also consistent with native Mandarin production results in which F0 mean contributed to the classification of their tone productions (Table 3.6). Both native Mandarin and learner results showed that the F0 mean $+$ F0 slope models yielded a higher classification accuracy than the F0 slope-only model, further indicating that F0 mean played a role in Mandarin production for both native and non-native speakers. As discussed above, F0 mean-only model in perception did not show an improvement on tone classification that was due to the use of F0 mean, so only F0 mean-only models in production showed an actual learning effect. Therefore, F0 mean models did not show a simultaneous improvement in perception and production and indicated a lack of transfer of production improvement to perception. It suggests that there was no perception-production relationship through F0 mean, the non-critical perceptual cue, corroborating the findings of Chapter 3. The learning effect shown only for F0 mean in production suggests that the critical nature of perceptual

cue did not always correspond to the cue use in production during the course of Mandarin learning.

The fact that F0 mean played a minimal role in learners' Mandarin tone perception is not consistent with previous studies which found native English individuals to rely on both pitch direction and height in tone perception (Chandrasekaran et al., 2010; Wiener, 2017) and to show high perceptual weighting on pitch height or F0 mean in tone perception (Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). This study attributes it to the difference between the results of tone categorization in this study and the tone discrimination in previous studies that used MDS. As mentioned in Chapter 3, in contrast to the tone categorization task used in this study, MDS studies analyze data collected in a discrimination task (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007). In contrast, a tone categorization task focused on between-category perception was used in this study. The difference suggests that while both cues may be used in discriminating among four Mandarin tones, F0 mean, as a non-critical perceptual cue, is minimally used in Mandarin tone categorization for both native Mandarin individuals and beginner level Mandarin learners.

It should also be noted that the perceptual weight given to the non-critical cue was predicted to remain unchanged across two visits based on previous training studies (Chandrasekaran et al., 2010; Francis et al., 2008; Iverson et al., 2005), which should imply that F0 mean had very little influence on these learners' Mandarin tone perception at the beginning of their exposure to Mandarin. This seems to suggest that there is a possible crosslinguistic difference between the tone perception of non-tonal language speakers, since native English individuals with no Mandarin experience were found to weight pitch height strongly (Gandour, 1983; Guion and Pederson, 2007). However, another possibility is that, contrary to the prediction, the use of F0 mean in tone perception diminished in the first 7 months of Mandarin learning for the learners of this study. This speculation is supported by two reasons in this study. First, the learners who participated in this study had already studied Mandarin for 7 months. In contrast, previous studies only trained participants to learn a non-native language feature for a few weeks (Chandrasekaran et al., 2010;

Francis et al., 2008; Iverson et al., 2005). Therefore, it is possible that the weighting given to pitch height decreases in the prolonged learning period. Second, a few learners in this study (LM05, LF13, LF14) did yield a higher classification accuracy in the F0 mean only than null model, and for LM05, the accuracy decreased across two visits (See Appendix E). It is, therefore, possible that some other learners also used F0 mean to categorize Mandarin tones perceptually before they were exposed to Mandarin, but gradually reduced the use of this cue in the earlier stage of Mandarin learning prior to their participation in this study. These results indicate that learning shows individual variability and changes in different learning stages. However, the above-mentioned speculations need to be further examined in a longitudinal study covering a longer L2 learning period with individuals who just begin their L2 learning.

4.4.3 Perception and production gains

Finally, this study did not reveal a relationship between perception and production gains as predicted. Previous studies found evidence of a perception lead (Bradlow et al., 1997; Flege and Schmidt, 1995; Flege et al., 1999; Kirby and Giang, 2021; Sakai and Moorman, 2018) or a production lead (Yang, 2012) in L2 learning. However, not only did the overall model accuracy gain of this study show no significant correlation, but there was no clear sign of a perception or production lead as the perception gains did not appear to be greater than production gains, or vice versa (Figure 4.12a). The results indicate that (L2 tone) learning can display diverse patterns and learners can display a faster pace of learning in either domain. Nevertheless, when the F1 scores for each tone were examined separately, Tone 3 revealed a positive correlation between perception and production gains for some learners, suggesting that the pace of perception and production development may differ depending on speech sound categories. However, this result is based on a small number of learners (n = 9), and therefore, further studies are needed to examine whether this finding can be generalized to a larger learner population.

4.5 Conclusion

Through comparing the perception and production of Mandarin learners, this study demonstrated that a perception-production relationship was attributable to critical perceptual cue - F0 slope. It showed that perception and production development both occurred for this critical perceptual cue only, and it also found that F0 mean, the non-critical perceptual cue, had a minimal influence on tone perception for beginner level Mandarin learners of a non-tonal language with 7 months of Mandarin learning experience, even though this cue still contributed to their tone productions. This result prompts the question of how the perception and production of the critical and non-critical cues are developed in an earlier stage of language learning.

Chapter 5

General discussion

5.1 Project Summary

This project investigated the relationship of perception and production established by critical perceptual cues using Mandarin tones. Since the Mandarin tone perception-production relationship was less well studied, this project first examined the correlation between Mandarin tone perception and production. Using Mandarin Tone 2 as a test case, Chapter 2 revealed a perception-production correlation, and showed that this correlation was found for critical pitch direction cues (F0 slope, curvature and TP) only, but not for non-critical perceptual cues (F0 mean, onset). It showed preliminary results that a perception-production link was pertinent to critical perceptual cues, which was then systematically investigated in Chapter 3 by varying one critical (F0 slope) and one non-critical cue (F0 mean) orthogonally in the perception stimuli. To identify the defining features of critical perceptual cues for each Mandarin tone category, this study examined the use of F0 slope and mean in *within-category* and *between-category* perception of all four Mandarin tones, and showed that F0 slope was a near tone-general cue that was critical for *within-category* and *between-category* Mandarin tone perception (with the exception of Tone 3 in between-category perception). Consequently, it demonstrated that a perception-production relationship was established by this critical perceptual cue, and was both phonetically and phonologically based. In contrast, The non-critical F0 mean cue demonstrated weaker influence on *within-category* perception and minimal influence on *between-category* perception as compared to F0 slope, and did not show a perception-production relationship. Finally, Chapter 4 fur-

ther investigated whether establishing a perception-production relationship could be clearly attributable to the use of critical perceptual cues by examining the development of Mandarin tone perception and production in terms of the critical status of perceptual cues for Mandarin learners with no tone language background. This study showed that learners simultaneously improved in the perception and production of F0 slope during Mandarin learning, whereas F0 mean did not reveal a simultaneous improvement. These results indicated that critical perceptual cues constituted a contributing factor for establishing a perception-production link. This final chapter brings together the findings of the three studies and reviews the general research questions of this project (RQ1, 2 and 3).

5.2 Perception-production relationships for lexical tones

The first research question of this project asks whether Mandarin tone perception and production are related in terms of tone acoustic cues (RQ1). To date, there has been little previous research on perception-production relationships for lexical tones. Previous studies that showed a perception-production relationship for Mandarin tones were based on a transfer of perception training to production as evinced by increased accuracy rates (Wang et al., 1999, 2003) and a qualitative comparison between perception and production tone spaces (Yang, 2015). There was also evidence for a Cantonese tone perception-production relationship based on perception and production accuracy of native Cantonese children (Mok et al., 2019). This project is one of the first research studies that systematically examined this relationship based on Mandarin acoustic cues, and consistently revealed a link between the two domains in terms of a positive perception-production correlation (Chapter 2), a cross-domain statistical learning model with high prediction accuracy (Chapter 3) and a longitudinal Mandarin learning study showing simultaneous learning effects on perception and production (Chapter 4). These results not only corroborate previous findings, but they also extend our understanding of the tone perception-production relationship in that this relationship is established by acoustic properties of tones.

Compared with previous studies on the perception-production relationship of segmental acoustic cues, the results of this project are consistent with previous perception-production

correlation analyses in terms of the strength of the relationship. Chapter 2 showed the strength of the perception-production correlation at a modest level as revealed by correlation coefficients (ρ = 0.378 - 0.402), similar to the modest perception-production correlation found in previous segmental studies (e.g. Beddor, 2015; McAllister Byun and Tiede, 2017; Newman, 2003; Ghosh et al., 2010; Perkell et al., 2004b). One interpretation of this modest level of perception-production correlation is that the relationship between the two domains are partially independent, as opposed to the premise of the Motor Theory of speech perception (Liberman and Mattingly, 1985; Newman, 2003). The other finding of this project that supports this view is the lack of correlation between perception and production gains in L2 learning (Section 4.3.3). A tight link between perception and production implies that the learning effect in one domain should immediately be transferred to the other domain, which should manifest in a correlation between perception and production gains. Not only did this project show a lack of this correlation in Chapter 4, previous training studies also failed to find this correlation (Bradlow et al., 1997; Sakai and Moorman, 2018). These results, therefore, offered further evidence suggesting that the two domains are semi-autonomous systems.

However, the results of this project also showed evidence of a close link between perception and production. In Chapter 3, the statistical learning results of the F0 slope-only production models revealed a high classification accuracy in prediction with perception data (76.4% (Table 3.7)). Moreover, the model prediction using good exemplars of perception stimuli as determined by the native Mandarin participants reached an accuracy of 91.2% (Table 3.10). The statistical training by production data intended to model the perceptual process (cf. McMurray and Jongman, 2011; Redmon et al., 2020). In contrast to the modest correlation results as shown in Chapter 2 and previous studies (e.g. Beddor, 2015; McAllister Byun and Tiede, 2017; Newman, 2003; Ghosh et al., 2010; Perkell et al., 2004b), the statistical modelling with high prediction accuracy indicated that perception and production as semi-autonomous systems can be closely linked. Note that correlation analysis (e.g. Beddor, 2015; Ghosh et al., 2010; Newman, 2003, and Chapter 2) uses the mean values averaging across each participant's perception and production data, and

therefore, reduces the number of data points to two data points per participant. In contrast, statistical modelling involves all individual data points obtained in perception and production. The difference in the closeness of a perception-production link as shown by the two research methods may indicate that the data reduction method used in correlation analysis does not offer the most representative data points for the perception and production of an individual language user. Rather, it is speculated that a satisfactory evaluation of the perception-production link may require a large number of instances of perception and production.

If the above speculation is correct, it is still possible that perception and production are semi-autonomous systems since, as shown in Chapter 4, second language learning exhibited simultaneous perception and production improvement, but did not reveal a correlation between perception and production gains. Taken together, the results of native participants and learners suggest that perception and production may develop from a partially independent to a close relationship. The lack of correlation between perception and production gains suggest that learners can show either a perception or production lead depending on their pace of learning. For the perception-production relationship, it indicates that the two domains are partially independent during language acquisition. This phenomenon is not unique to second language learning (Bradlow et al., 1997; Sakai and Moorman, 2018). Previous first language studies with child data also found evidence of a partially independent perception-production relationship. A perception-production correlation was not always revealed among children (McAllister Byun and Tiede, 2017). When a perception-production correlation was found, the variance of the production data was only partially explained by perception data. One factor related to the production data distribution was the language background of the mother (Mok et al., 2019). Therefore, the findings of the current project and previous studies converge on the view that perception and production are related but remain partially independent during first and second language acquisition.

It is expected that, as language proficiency becomes more advanced, a close relationship as shown in Chapter 3 should be forged. This development of perception-production relationships is consistent with the general auditory approaches to speech perception

(Blumstein and Stevens, 1979; Diehl and Kluender, 1989; Diehl et al., 2004; Kuhl et al., 2008) and DIVA model (Guenther, 1994, 1995, 2015). As mentioned in Section 1.1, these theories posit that the motor commands generate both articulatory movements and mental copies of auditory targets. These copies are compared with the perceptual targets stored in the memory of the language users. Therefore, these theories imply that there could be a higher chance of a perception-production mismatch during language acquisition, but the two domains are coupled as the individuals gain more experience in the language. The findings of the current project echo the developmental process laid out in these theories, since the learners exhibited partially independent perception and production systems whereas a close perception-production relationship was found for native Mandarin individuals.

5.3 F0 slope as a critical perceptual cue of Mandarin tones defined by phonetic and phonological level of perception

The second research question (RQ2) asks whether the acoustic-phonetic level of *within-category* perception or the phonemic categorization level of *between-category* perception defines the critical perceptual cues of Mandarin tones. This question was addressed in Chapter 3 which showed that the critical perceptual cue, F0 slope, strongly influenced both within-category goodness ratings and between-category tone categorization of all Mandarin tones (except for Tone 3 in between-category perception). F0 mean, as a non-critical cue, was expected to influence the perception of Mandarin tones, especially Tone 1 and Tone 3 (e.g. Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Yang, 2015), but was found to show a much smaller influence on both between-category and within-category perception than F0 slope. In addition, the results showed that the critical status of perceptual cues was near tone-general.

As a critical perceptual cue, F0 slope influenced Mandarin tone perception as a near tone-general critical perceptual cue for non-native Mandarin individuals with a short exposure to Mandarin. An intriguing finding, as shown in Chapter 4, was that beginner-level learners who did not have prior knowledge of any tone language also used F0 slope as

the critical cue for Mandarin tone categorization after 7 months of Mandarin learning, as opposed to F0 mean as predicted by the initial greater perceptual cue weightings given to pitch height than direction for non-native Mandarin individuals with a non-tonal first language (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Wiener, 2017). Learners' results also demonstrated that that the critical status of F0 slope continued to rise and gradually approached the native norm for the non-native Mandarin individuals.

These results demonstrate that a specific acoustic cue - F0 slope, instead of a broadly-defined pitch direction cue, serves as a critical perceptual cue. Previous MDS studies that showed pitch direction was critical for the tone perception of native Mandarin individuals (Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion and Pederson, 2007; Wiener, 2017). However, the interpretation of pitch direction was not consistent in these studies. The perceptual dimensions obtained in the MDS analysis were interpreted by the researcher after obtaining the distribution of categories on the perceptual space. Although the dimension of pitch direction could clearly be interpreted as F0 slope in some studies (e.g. Gandour, 1983; Guion and Pederson, 2007), other training and learning studies sometimes showed a slightly different distribution of tones along this dimension. As mentioned in section 4.1.2, after 3 months of Mandarin learning, the participants in Wiener (2017) yielded a dimension that had Tone 2 (rising) on one end of the dimension and Tone 1 (level) on the other end, with Tone 4 (falling) in between. The same result was obtained when native English and Mandarin listeners perceived Cantonese tones (Francis et al., 2008). Moreover, Wiener (2017) found very different distribution pattern of Mandarin tone categories in the first two months compared to the third months of Mandarin learning. Tone 4 (falling), instead of Tone 2 (rising) appeared on one end of the dimension while the other end was Tone 1 (level). The acoustic correlate of pitch direction was changing across the learning period and none of the distribution could be fully interpreted as F0 slope. Therefore, baed on previous research, it is uncertain if F0 slope is the best characterization of pitch direction, which should show rising and falling tone contours on the two ends of this dimension. In contrast, this project overcame this issue by measuring the performance of the same cues in the perception models, and clearly demonstrated that F0 slope strongly

influenced Mandarin tone perception for both native Mandarin individuals and Mandarin learners. In fact, the results of Chapter 2 suggest that pitch direction can potentially be defined by or related to a set of acoustic cues, such as F0 slope, curvature and TP. It is possible that all these cues can be used as critical perceptual cues (all related to pitch direction) for Mandarin tone perception.

It is worth noting that, in a similar study (Yang, 2015), F0 slope and mean were both shown to influence the categorization of Mandarin tones for native Mandarin individuals and learners with over one year of learning experience in Mandarin. Specifically, the rising and falling resynthesized tone contours were mainly perceived as Tone 2 and 4, and the high and low level tones as Tone 1 and 3. However, the difference may further indicate the critical status of F0 slope for Mandarin tone perception. As mentioned in section 3.4.1, it seems that the discrepancy between the results of the current project and Yang (2015) may be attributed to a difference in stimulus presentation. That is, the perceptual stimuli were presented in isolation in this project whereas Yang presented the stimuli in a carrier sentence. This difference suggests that the perception of Mandarin tones makes use of F0 mean only if there is an external context. In contrast, the internal property of F0 slope is enough to influence the perception of Mandarin tone for both native Mandarin individuals and beginner-level Mandarin learners.

Moreover, the two studies differed in the cues varied in the perception stimuli. The perceptual stimuli were resynthesized by varying F0 onset and offset in Yang (2015) instead of F0 mean and slope that were varied in this project. Although the levels of F0 onset and offset can be straightforwardly translated to F0 mean and slope, the stimuli did not include the rising and falling contours at high or low F0 mean levels. These stimuli included in this project were categorized as Tone 2 and Tone 4 for rising and falling contours, respectively, by both native Mandarin participants (Figure 3.2) and learners (Figure 4.1). The perception result of this project, therefore, indicated that the influence of F0 slope remained predominant in these extreme circumstances, further indicating the critical status of F0 slope for Mandarin tone perception.

5.4 Establishing perception-production links through critical perceptual cues

The third research question of this project is to examine whether the formation of a perception-production relationship attributable to the use of critical perceptual cues (RQ3). This dissertation project has found clear evidence supporting that critical perceptual cues are essential for revealing a perception-production link. Chapter 2 showed a perception-production correlation for a set of critical pitch direction-related cues, but not for non-critical height-related cues for Mandarin tone perception, and provided preliminary evidence supporting RQ3. Then, by comparing a critical (F0 slope) and a non-critical perceptual cue (F0 mean), Chapter 3 revealed that the critical perceptual cue, F0 slope, that influenced both within-category and between-category perception could establish a close perception-production relationship. Furthermore, Chapter 4 demonstrated that establishing a perception-production link could be attributable to F0 slope in that the same critical cue revealed simultaneous improvement in perception and production.

Since Chapter 3 showed that the critical status of perceptual cues could be defined by both the *within-category*, phonetic level and the *between-category*, phonological level of perception, both levels of perception should contribute to the link between perception and production. A perception-production link can be established by tone categorization results alone, and is also amplified when the data were filtered by within-category goodness rating results. These results extend the theoretical understanding of a perception-production link. Recall that the theoretical predictions for a link between perception and production are based on phonemic contrasts. The motor theory of speech perception postulates that speech gesture is the object of speech perception, and the contrasts produced by the gestures signal phonological categories (Galantucci et al., 2006; Liberman et al., 1967; Liberman and Mattingly, 1985). Likewise, auditory approaches to speech perception and the DIVA model also focus on building phonemic contrasts in the development of perception and production during the acoustic-gesture mapping process that forms a perception-production link in language acquisition (Diehl et al., 2004; Guenther, 1995; McAllister Byun and Tiede, 2017; Perkell et al., 2004a). These theories do not explicitly include *within-*

category perception in the formation of a perception-production link. The results of this study, therefore, point to a new understanding for these theories. From the perspective of gestural theories, speech gestures as the object of perception may not only encode phonemic contrasts but also phonetic differences within a speech category. Alternatively, from the perspective of the auditory approach to speech perception and DIVA model, the acoustic-gesture mapping process should involve the phonetic details of speech sounds that are both within-category and between-category. That is, the mapping process is driven not only by the need for perceptual acuity and production clarity of phonemes, but also the phonetic details that distinguish the quality of speech sounds within a speech category. For all theories, these phonetic details should be related to the critical perceptual cues of speech sounds, in order to form a perception-production link through critical perceptual cues defined by both within-category and between-category perception. Note that critical perceptual cues should be the contributing factor that establishes a perception-production link, since Chapter 4 revealed that simultaneous improvement in perception and production occurred for the the same critical F0 slope cue.

As mentioned in the previous section, this dissertation project has demonstrated that F0 slope is the critical cue that influences Mandarin tone perception for native Mandarin individuals and beginner-level Mandarin learners. F0 mean played a comparatively much smaller role than F0 slope in perception. As for production, both F0 slope and mean are needed for the classification of tone productions for both groups of participants. This asymmetry in the cues used in perception and production suggests that not every cue that characterizes speech productions (e.g. F0 slope and mean) is required for perception (e.g. F0 slope). As a result, the perception-production relationship should be driven by the perceptually critical cue which is also the common cue (e.g. F0 slope) that influences both perception and production.

A related issue is why F0 mean, the non-critical perceptual cue, is able to emerge as a cue that characterizes speech production when it only weakly influences perception for both native Mandarin individual and beginner-level Mandarin learners. The theories, as discussed above, predict that either perception is perceiving speech gesture (gestural the-

ories (Fowler, 1986; Galantucci et al., 2006; Liberman et al., 1967; Liberman and Mattingly, 1985)) or perception and production shape each other to form a perception-production link in language acquisition (general auditory approaches and DIVA model (Diehl et al., 2004; Guenther, 1995; Kuhl et al., 2008; McAllister Byun and Tiede, 2017; Perkell et al., 2004a)). Therefore, the influence of F0 mean on Mandarin tone production should also be impacted by and have an impact on the perception of Mandarin tones. Based on this project's findings, the impact is only minimal for Mandarin tone perception and production in isolation. However, as shown in Yang (2015), it is still possible the mapping between perception and production may be achieved through the perception and production of individual Mandarin tones with a sentential context. In other words, this study does not preclude the possibility that a non-critical perceptual cue can establish a perception-production relationship on a more restrictive context of speech perception and production (e.g. isolated vs. sentential) as compared to the perception-production link established by a critical perceptual cue. This speculation enables us to explain why certain non-critical perceptual cues may have a greater impact on speech production than perception in the isolated context.

5.5 Limitations and future directions

While this dissertation project established a perception-production relationship of Mandarin tones using critical perceptual cues, the findings mainly covered Tone 1, 2 and 4. To achieve the goals of this project, Chapter 2 used one tone (Tone 2) for a preliminary test of a perception-production relationship for lexical tones. Then, one critical (F0 slope) and one non-critical (F0 mean) were used for Chapter 3 and 4. However, as mentioned in Chapter 3, resynthesized tone stimuli with a linear contour cannot represent the typical dipping tone contour of Tone 3, and therefore, only a small number of perceptual responses were obtained for this tone compared with other tones. It is expected that a perception-production relationship established through critical perceptual cues should also apply to Tone 3, and therefore, the remaining question is what cues are perceptually critical for this tone. As discussed in Chapter 3, the two cues that have been shown to be relevant to Tone 3 perception are $\triangle$F0 and TP (Moore and Jongman, 1997; Shen and Lin, 1991;

Shen et al., 1993). Related to these two cues, another possible candidate is F0 curvature that is pertinent to Mandarin tone classification (Tupper et al., 2020). The actual critical perceptual cues of Tone 3 need to be further investigated, which can extend our understanding of whether pitch direction cues can collectively serve as the tone-general critical perceptual cues for Mandarin tone perception, and therefore, show that speech categories that share the same broadly-defined critical perceptual cue (e.g. pitch direction) require specific acoustic cues to establish a perception-production relationship.

Another limitation of this project concerns the non-critical status of F0 mean as a perceptual cue. The non-critical status of F0 mean is demonstrated by the minimal influence on the within-category and between-category perception of Mandarin tones, which is not surprising as Mandarin tone categories differ primarily in terms of tone contours. However, some studies suggest that languages that require a stronger perceptual weighting on F0 mean than Mandarin also indicate a certain level of perceptual confusion when tones are perceived in isolation. For instance, although Cantonese tone perception requires a higher weighting on F0 mean than Mandarin tone perception (Francis et al., 2008; Gandour, 1983), native Cantonese individuals still demonstrate a higher degree of tone confusion in the perception of Cantonese than the native Mandarin perception of Mandarin tones in isolation (Khouw and Ciocca, 2007; Peng et al., 2012). The difference was attributed to the greater overlap of tone categories in Cantonese than in Mandarin tone system (Peng et al., 2012; Shao and Zhang, 2018). More importantly, the tone perceptual confusion was more obvious among the Cantonese level tones that differ primarily in pitch height (Peng et al., 2012). Therefore, it is uncertain whether F0 mean will show a substantial influence on tone perception (within-category and/or between-category) even if this cue serves as a critical perceptual cue for that particular language. It has to be further tested with a language in which tone categories differ primarily in terms of pitch height. It is worth noting that the impact of F0 mean on tone perception may increase in a sentential context regardless of the perceptual weighting on this cue. The perception accuracy for Cantonese tone perception was better when the stimuli were presented in a carrier sentence than in isolation (Shao and Zhang, 2018). In addition, F0 mean emerged to be a cue that separated Tone 1 and

3 when resynthesized linear tone contours were perceived in a carrier sentence by native Mandarin participants (Yang, 2015). As speculated in the previous section, this change in the critical nature of F0 mean may contribute to the perception-production mapping for Mandarin tone cues for this cue, since the theoretical basis of a perception-production link predict that a cue used in one domain should influence the use of the same cue in the other domain.

Finally, this project raises a further question about how the modification of cue weighting will influence the actual performance of Mandarin learners in tone perception and production accuracy, which should benefit the research in L2 learning and Mandarin pedagogy. Previous studies showed that an increase in the perceptual cue weighting of F0 slope was associated with an improvement in Mandarin tone perception accuracy (Chandrasekaran et al., 2010). Therefore, it is assumed that the learners who demonstrated that an increased performance of the F0 slope-only perception model should also show improved Mandarin tone category perception. It is also assumed that the increased weighting on F0 slope and mean in production should improve the production accuracy of Mandarin tones. On the other hand, it is speculated that the perceptual weighting given to F0 mean, the non-critical cue, may decrease in a short time frame as the learners gained little but sufficient experience in Mandarin. If F0 mean shows influence on Mandarin tone perception with a sentential context, it is uncertain whether this speculated decrease of cue weighting will still be observed. Examining how this potential decrease in perceptual weighting on a non-critical cue influences the perception accuracy of Mandarin tones should also benefit our general understanding of the interplay between perceptual cue weighting modifications and improvement in non-native perception.

5.6 Conclusion

In sum, this dissertation project demonstrated a perception-production relationship for lexical tones that was attributable to phonetically and phonologically defined critical perceptual cues. The findings extend our understanding that a perception-production link is not only based on phonemic contrasts as predicted by some theories but also based on phonetic

details of speech sounds within a speech category. The current project found evidence supporting a near tone-general critical perceptual cue (F0 slope), but did not preclude the possibility of specific acoustic cues for certain tone categories (i.e., Tone 3). Further research is also needed for understanding the asymmetry of the use of non-critical cues in perception and production which should further our understanding on the development of a perception-production relationship.

References

Adank, P., Hagoort, P., and Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, 21(12):1903–1909.

Ainsworth, W. and Paliwal, K. (1984). Correlation between the production and perception of the English glides /w, r, l, j/. *Journal of phonetics*, 12(3):237–243.

Alexander, J. A. (2011). The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems. In *Psycholinguistic Representation of Tone Conference 2011 (satellite of ICPhS 2011)*.

Baese-Berk, M. M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, perception, & psychophysics*, 81(4):981–1005.

Baese-Berk, M. M. and Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of memory and language*, 89:23–36.

Baese-Berk, M. M. and Samuel, A. G. (2022). Just give it time: Differential effects of disruption and delay on perceptual learning. *Attention, perception, & psychophysics*, 84(3):960–980.

Bailey, P. J. and Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. *Language and Speech*, 16(3):189–195.

Beddor, P. S. (2015). The relation between language users' perception and production repertoires. In The Scottish Consortium for ICPhS 2015, editor, *Proceedings of the 18th International Congress of Phonetic Sciences*, pages 1041.1–9, Glasgow, UK. the University of Glasgow.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., and Sawusch, J. R. (1979). Some relationships between speech production and perception. *Phonetica*, 36(6):373–383.

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

Boersma, P. and Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.43.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning

on speech production. *The Journal of the Acoustical Society of America*, 101(4):2299–2310.

Brosseau-Lapré, F., Rvachew, S., Claywards, M., and Dickson, D. (2013). Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics*, 34(3):419–441.

Carroll, J. D. and Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckhart-Young' decomposition. *Psychometrika*, 35(3):283–320.

Chandrasekaran, B., Sampath, P. D., and Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, 128(1):456–465.

Chang, D., Hedberg, N., and Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, 139(5):2432–2447.

Chang, Y.-h. S. (2011). Distinction between Mandarin Tones 2 and 3 for L1 and L2 Listeners. In *The 23rd North American Conference on Chinese Linguistics (NACCL-23)*, volume 1, pages 84–96.

Chao, Y. R. (1947). *Mandarin primer: an intensive course in spoken Chinese*. Harvard University Press, Cambridge, MA.

Chen, F. and Peng, G. (2016). Context effect in the categorical perception of Mandarin tones. *Journal of signal processing systems*, 82(2):253–261.

Cohen, J. and Cohen, P. (2003). *Applied multiple correlation/regression analysis for the behavioral sciences*. L. Erlbaum Associates, Hillsdale, NJ, 2nd edition.

Diehl, R. L. and Kluender, K. R. (1989). On the Objects of Speech Perception. *Ecological Psychology*, 1(2):121–144.

Diehl, R. L., Lotto, A. J., and Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, 55(1):149–179.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, 84(1):70–79.

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

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In Strange, W., editor, *Speech perception and linguistic experience: Issues in cross-language research*, pages 209–217. York Press, Timonium, MD.

Flege, J. E. (1999). The Relation between L2 Production and Perception. In Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D., and Bailey, A. C., editors, *Proceedings of the 14th International Congress of Phonetic Sciences*, pages 1273–1276, San Francisco, CA, USA. The Regents of the University of California.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In Meyer, A. and Schiller, N., editors, *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*, pages 319–355. Mouton de Gruyter, Berlin.

Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, 25(4):437–470.

Flege, J. E., MacKay, I. R. A., and Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, 106(5):2973–2987.

Flege, J. E. and Schmidt, A. M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica*, 52(2):90–111.

Flemming, E. and Cho, H. (2017). The phonetic specification of contour tones: Evidence from the Mandarin rising tone. *Phonology*, 34(1):1–40.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, 14:3–28.

Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, 49(3):396–413.

Fox, R. A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica*, 39(1):1–22.

Francis, A. L., Ciocca, V., Ma, L., and Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2):268–294.

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., and Hagoort, P. (2017). Individual variability as a window on production-perception interactions in speech motor control. *The Journal of the Acoustical Society of America*, 142(4):2007–2018.

Frieda, E. M., Walley, A. C., Flege, J. E., and Sloane, M. E. (2000). Adults' perception and production of the English vowel /i/. *Journal of speech, language, and hearing research*, 43(1):129–143.

Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, 13(3):361–377.

Gandour, J. T. (1979). Perceptual dimensions of Cantonese tones: A multidimensional scaling reanalysis of Fok's tone confusion data. *South-east Asian Linguistic Studies*, 4:415–429.

Gandour, J. T. (1983). Tone perception in far Eastern languages. *Journal of phonetics*, 11(2):149–175.

Gandour, J. T. and Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and speech*, 21(1):1–33.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., Guenther, F. H., Lane, H., and Perkell, J. S. (2010). An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America*, 128(5):3079–3087.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, 72:43–53.

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

Guenther, F. H. (2015). The neural control of speech: From computational modeling to neural prosthesis. In The Scottish Consortium for ICPhS 2015, editor, *Proceedings of the 18th International Congress of Phonetic Sciences*, pages 1042.1–5, Glasgow, UK. the University of Glasgow.

Guion, S. G. and Pederson, E. (2007). Investigating the role of attention in phonetic learning. In Bohn, O.-S. and Munro, M. J., editors, *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*, pages 57–77. John Benjamins, Amsterdam.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24(4):495–522.

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3):360–378.

Herd, W., Jongman, A., and Sereno, J. A. (2013). Perceptual and production training of intervocalic /d, ɾ, r/ in American English learners of Spanish. *The Journal of the Acoustical Society of America*, 133(6):4247–4255.

Hirata, Y. (2004). Computer Assisted Pronunciation Training for Native English Speakers Learning Japanese Pitch and Durational Contrasts. *Computer Assisted Language Learning*, 17(3-4):357–376.

Iverson, P., Hazan, V., and Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America*, 118(5):3267–3278.

Iverson, P., Pinet, M., and Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1):145–160.

Jeng, J. Y., Weismer, G., and Kent, R. D. (2006). Production and perception of Mandarin tone in adults with cerebral palsy. *Clinical Linguistics and Phonetics*, 20(1):67–87.

Johnson, K., Flemming, E., and Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Linguistic Society of America*, 69(3):505–528.

Jongman, A. and McMurray, B. (2017). On invariance: Acoustic input meets listener expectations. In Lahiri, A. and Kotzor, S., editors, *The Speech Processing Lexicon*, pages 1–24. Mouton De Gruyter, Berlin.

Jongman, A., Qin, Z., Zhang, J., and Sereno, J. A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America*, 142(2):EL163–EL169.

Jongman, A., Wang, Y., Moore, C. B., and Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In Li, P., Tan, L. H., Bates, E., and Tzeng, O. J. L., editors, *The Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)*, pages 209–217. Cambridge University Press, Cambridge.

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, 108(3):1252–1263.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, 138(2):817–832.

Kato, M. and Baese-Berk, M. M. (2020). The effect of input prompts on the relationship between perception and production of non-native sounds. *Journal of Phonetics*, 79:100964.

Khouw, E. and Ciocca, V. (2007). Perceptual correlates of cantonese tones. *Journal of phonetics*, 35(1):104–117.

Kirby, J. and Giang, D. L. (2021). Relating production and perception of l2 tone. In Wayland, R., editor, *Second Language Speech Learning: Theoretical and Empirical Progress*, page 249–272. Cambridge University Press.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493):979–1000.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., and Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(2):227–247.

Leach, L. and Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive psychology*, 55(4):306–353.

Leather, J. (1997). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. *Second language speech: Structure and process*, pages 75–101.

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.6.0.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, 74(6):431–461.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21:1–36.

Liberman, A. M. and Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5):187–196.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. J. and Marchal, A., editors, *Speech Production and Speech Modelling*, pages 403–439. Kluwer Academic, Dordrecht, The Netherlands.

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

Liu, H. M., Tsao, F. M., and Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43(4):912–917.

Liu, S. and Samuel, A. G. (2004). Perception of mandarin lexical tones when F0 information is neutralized. *Language and Speech*, 47(2):109–138.

Lu, S., Wayland, R., and Kaan, E. (2015). Effects of production training and perception training on lexical tone perception - A behavioral and ERP study. *Brain Research*, 1624:28–44.

Massaro, D. W., Cohen, M. M., and Tseng, C.-y. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, 13(2):267–289.

McAllister Byun, T. and Tiede, M. (2017). Perception-production relations in later development of American English rhotics. *PLoS ONE*, 12(2):e0172022.

McMurray, B. and Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2):219–246.

Mok, P. P. K., Fung, H. S. H., and Li, V. G. (2019). Assessing the link between perception and production in Cantonese tone acquisition. *Journal of Speech, Language, and Hearing Research*, 62(5):1243–1257.

Mok, P. P. K., Zuo, D., and Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in hong kong cantonese. *Language variation and change*, 25(3):341–370.

Moore, C. B. and Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, 102(3):1864–1877.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of phonetics*, 18(3):347–373.

Nearey, T. M. (1992). Context effects in a double-weak theory of speech perception. *Language and speech*, 35(1,2):153–171.

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, 101(6):3241–3254.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *The Journal of the Acoustical Society of America*, 113(5):2850–2860.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In Breivik, L. E. and Jahr, E. H., editors, *Language change: Contributions to the study of its causes*, pages 173–198. Mouton de Gruyter, Berlin.

Paliwal, K. K., Lindsay, D., and Ainsworth, W. A. (1983). Correlation between production and perception of English vowels. *Journal of Phonetics*, 11:77–83.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, 34(1):134–154.

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., and Wang, W. S.-Y. (2012). The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems. *Journal of speech, language, and hearing research*, 55(2):579–595.

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., and Wang, W. S. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, 38(4):616–624.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (2004a). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America*, 116(4):2338–2344.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. H. (2004b). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research*, 47(6):1259–1269.

Prom-on, S., Liu, F., and Xu, Y. (2012). Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modeling. *The Journal of the Acoustical Society of America*, 132(1):421–432.

Prom-on, S., Xu, Y., and Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America*, 125(1):405–424.

R Core Team (2021). R: A language and environment for statistical computing [computer program]. version 4.0.

Redmon, C., Leung, K., Wang, Y., McMurray, B., Jongman, A., and Sereno, J. A. (2020). Cross-linguistic perception of clearly spoken english tense and lax vowels based on auditory, visual, and auditory-visual information. *Journal of phonetics*, 81:100980.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, 6(4):343–352.

Sakai, M. and Moorman, C. (2018). Can perception training improve the production of second language phonemes? a meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1):187–224.

Schertz, J., Cho, T., Lotto, A., and 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.

Schmidt, A. M. and Flege, J. E. (1995). Effects of speaking rate changes on native and nonnative speech production. *Phonetica*, 52(1):41–54.

Shao, J. and Zhang, C. (2018). Context integration deficit in tone perception in Cantonese speakers with congenital amusia. *The Journal of the Acoustical Society of America*, 144(4):EL333–EL339.

Shen, X. S. and Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech*, 34(2):145–156.

Shen, X. S., Lin, M., and Yan, J. (1993). F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3. *The Journal of the Acoustical Society of America*, 93(4):2241–2243.

Shih, C. and Lu, H.-y. D. (2015). Effects of talker-to-listener distance on tone. *Journal of Phonetics*, 51:6–35.

Shultz, A. A., Francis, A. L., and Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, 132(2):EL95–EL101.

Singmann, H., Bolker, B., Westfall, J., Aust, F., and Ben-Shachar, M. S. (2021). *afex: Analysis of Factorial Experiments*. R package version 0.28-1.

So, C. K. and Best, C. T. (2011). Categorizing mandarin tones into listeners' native prosodic categories: The role of phonetic properties. *Poznań Studies in Contemporary Linguistics*, 47(1):133–145.

So, C. K. and Best, C. T. (2014). Phonetic influences on english and french listeners' assimilation of mandarin tones to native prosodic categories. *Studies in second language acquisition*, 36(2):195–221.

Tupper, P., Leung, K., Wang, Y., Jongman, A., and Sereno, J. A. (2020). Characterizing the distinctive acoustic cues of mandarin tones. *The Journal of the Acoustical Society of America*, 147(4):2570–2580.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, 280(1):61–72.

Wang, X. (2013). Perception of mandarin tones: The effect of L1 background and training. *Modern Language Journal*, 97(1):144–160.

Wang, X., Wang, S., Fan, Y., Huang, D., and Zhang, Y. (2017). Speech-specific categorical perception deficit in autism: An Event-Related Potential study of lexical tone processing in Mandarin-speaking children. *Scientific Reports*, 7(February).

Wang, Y., Jongman, A., and Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2):1033–1043.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6):3649–3658.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America*, 93(4):2152–2159.

Whalen, D. H. and Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1):25–47.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wiener, S. (2017). Changes in early L2 cue-weighting of non-native speech: Evidence from learners of Mandarin Chinese. In *Proceedings of Interspeech 2017*, pages 1765–1769.

Wong, P., Fu, W. M., and Cheung, E. Y. (2017). Cantonese-speaking children do not acquire tone perception before tone production—A perceptual and acoustic study of three-year-olds' monosyllabic tones. *Frontiers in Psychology*, 8(AUG).

Xi, J., Zhang, L., Shu, H., Zhang, Y., and Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, 170(1):223–231.

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, 17(January 2001):1–31.

Xu, Y., Gandour, J. T., and Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2):1063–1074.

Xu, Y. and Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33(4):319–337.

Xu Rattanasone, N., Burnham, D., and Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics*, 41(5):332–343.

Yang, B. (2012). The gap between the perception and production of tones by American learners of Mandarin – an intralingual perspective. *Chinese as a Second Language Research*, 1(1):33–53.

Yang, B. (2015). *Perception and production of Mandarin tones by native speakers and L2 learners*. Springer, Berlin, Heidelberg.

Yang, B. and Whalen, D. H. (2015). Perception and production of English vowels by American males and females. *Australian Journal of Linguistics*, 35(2):121–141.

Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., and Näätänen, R. (2010). Training the brain to weight speech cues differently: A study of finnish second-language users of English. *Journal of Cognitive Neuroscience*, 22(6):1319–1332.

Yu, A. C. L., Lee, C. W. T., Lan, C., and Mok, P. P. K. (2021). A new system of Cantonese tones? tone perception and production in Hong Kong South Asian Cantonese. *Language and speech*, pages 1—25.

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of phonetics*, 61:13–29.

Zhao, T. C. and Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America*, 137(3):1452–1463.

Zhou, N., Huang, J., Chen, X., and Xu, L. (2013). Relationship between tone perception and production in prelingually deafened children with cochlear implants. *Otology and Neurotology*, 34(3):499–506.

Zhou, N. and Xu, L. (2008). Development and evaluation of methods for assessing tone production skills in Mandarin-speaking children with cochlear implants. *The Journal of the Acoustical Society of America*, 123(3):1653–1664.

Appendix A

Cross validation results of individual production models

Table A.1: Contingency tables of F0 mean $\times$ F0 slope production models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	178	12	13	1
	Tone 2	5	195	3	0
	Tone 3	1	1	195	0
	Tone 4	0	0	1	197

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	151	56	0	0
	Tone 2	10	199	0	0
	Tone 3	1	22	152	9
	Tone 4	0	0	7	195

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	194	8	0	0
	Tone 2	17	178	5	0
	Tone 3	0	4	191	0
	Tone 4	1	0	2	201

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	187	12	3	0
	Tone 2	16	171	16	0
	Tone 3	11	17	166	4
	Tone 4	2	0	0	197

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	201	1	0	0
	Tone 2	18	157	23	0
	Tone 3	0	1	195	3
	Tone 4	0	0	1	201

Table A.2: Contingency tables of F0 mean + F0 slope only production models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	175	14	12	3
	Tone 2	5	197	1	0
	Tone 3	0	2	194	1
	Tone 4	0	0	2	196

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	162	45	0	0
	Tone 2	10	199	0	0
	Tone 3	1	23	150	10
	Tone 4	0	0	9	193

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	194	8	0	0
	Tone 2	17	178	5	0
	Tone 3	0	4	191	0
	Tone 4	3	0	6	195

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	187	12	3	0
	Tone 2	18	170	15	0
	Tone 3	7	18	165	8
	Tone 4	6	0	0	193

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	201	1	0	0
	Tone 2	17	163	18	0
	Tone 3	0	1	196	2
	Tone 4	0	0	0	202

Table A.3: Contingency tables of F0 slope-only production models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	81	1	122	0
	Tone 2	3	199	1	0
	Tone 3	77	14	102	4
	Tone 4	0	0	2	196

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	79	102	26	0
	Tone 2	25	184	0	0
	Tone 3	63	23	74	24
	Tone 4	0	0	12	190

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	122	26	54	0
	Tone 2	29	171	0	0
	Tone 3	50	11	126	8
	Tone 4	1	0	4	199

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	148	12	42	0
	Tone 2	65	134	4	0
	Tone 3	79	6	109	4
	Tone 4	0	0	21	178

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	138	3	61	0
	Tone 2	51	131	16	0
	Tone 3	47	1	139	12
	Tone 4	0	0	24	178

Table A.4: Contingency tables of F0 mean-only production models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	146	37	1	20
	Tone 2	36	55	27	85
	Tone 3	0	0	191	6
	Tone 4	67	62	16	53

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	205	0	0	2
	Tone 2	23	96	3	87
	Tone 3	0	21	163	0
	Tone 4	27	70	44	61

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	186	12	0	4
	Tone 2	35	145	9	11
	Tone 3	0	11	184	0
	Tone 4	47	111	35	11

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	150	30	0	22
	Tone 2	31	138	12	22
	Tone 3	0	57	140	1
	Tone 4	30	119	32	18

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	172	30	0	0
	Tone 2	15	85	6	92
	Tone 3	0	0	178	21
	Tone 4	45	85	23	49

Table A.5: Contingency tables of null production models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	204	0	0	0
	Tone 2	203	0	0	0
	Tone 3	197	0	0	0
	Tone 4	198	0	0	0

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	207	0	0	0
	Tone 2	209	0	0	0
	Tone 3	184	0	0	0
	Tone 4	202	0	0	0

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	202	0	0	0
	Tone 2	200	0	0	0
	Tone 3	195	0	0	0
	Tone 4	204	0	0	0

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	202	0	0	0
	Tone 2	203	0	0	0
	Tone 3	198	0	0	0
	Tone 4	199	0	0	0

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	102	100	0	0
	Tone 2	100	98	0	0
	Tone 3	92	107	0	0
	Tone 4	95	107	0	0

Table A.6: Accuracy rates and F1 scores of individual production models

(a) F0 mean $\times$ F0 slope model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	1165	95.4	0.917	0.948	0.954	0.995
2	932	86.9	0.818	0.819	0.886	0.961
3	1151	95.4	0.937	0.913	0.972	0.993
4	899	89.9	0.895	0.849	0.867	0.985
5	1153	94.1	0.955	0.880	0.933	0.990

(b) F0 mean $+$ F0 slope model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	1310	95.0	0.911	0.947	0.904	0.985
2	1029	87.8	0.853	0.836	0.934	0.953
3	1322	94.6	0.933	0.913	0.982	0.977
4	1026	89.2	0.890	0.844	0.843	0.965
5	1337	95.1	0.957	0.898	0.945	0.995

(c) F0 slope-only model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	4367	72.1	0.444	0.954	0.481	0.985
2	3785	65.7	0.422	0.710	0.500	0.913
3	4338	77.2	0.604	0.838	0.665	0.968
4	4198	70.9	0.599	0.755	0.583	0.934
5	4128	73.2	0.630	0.787	0.633	0.908

(d) F0 mean-only model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	5058	55.5	0.645	0.308	0.884	0.293
2	5369	65.5	0.887	0.485	0.827	0.347
3	5453	65.7	0.791	0.605	0.870	0.096
4	5087	55.6	0.726	0.505	0.733	0.137
5	5349	60.4	0.793	0.427	0.877	0.269

(e) Null model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	8888	25.4	0.406	$n.a.$	$n.a.$	$n.a.$
2	8889	25.8	0.410	$n.a.$	$n.a.$	$n.a.$
3	8891	25.2	0.403	$n.a.$	$n.a.$	$n.a.$
4	8888	25.2	0.402	$n.a.$	$n.a.$	$n.a.$
5	8890	25.0	0.345	0.321	$n.a.$	$n.a.$

Appendix B

Prediction results of individual production models by perception data

Table B.1: Contingency tables of F0 mean + F0 slope production models by perception data

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	84	21	65	8
	Tone 2	50	149	29	0
	Tone 3	2	2	13	6
	Tone 4	34	0	95	244

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	93	18	46	6
	Tone 2	31	186	22	0
	Tone 3	3	0	12	6
	Tone 4	17	0	90	272

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	105	14	49	6
	Tone 2	42	169	37	0
	Tone 3	2	1	8	6
	Tone 4	28	0	104	230

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	103	19	46	8
	Tone 2	47	137	47	0
	Tone 3	0	3	14	4
	Tone 4	30	0	94	250

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	104	21	60	6
	Tone 2	44	164	28	0
	Tone 3	3	1	13	6
	Tone 4	27	0	91	233

Table B.2: Contingency tables of F0 slope-only production models by perception data

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	92	14	72	0
	Tone 2	15	213	0	0
	Tone 3	0	2	13	8
	Tone 4	0	0	84	289

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	87	25	50	1
	Tone 2	6	232	1	0
	Tone 3	0	2	8	11
	Tone 4	0	0	73	306

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	109	14	50	1
	Tone 2	27	221	0	0
	Tone 3	0	1	12	4
	Tone 4	0	0	88	274

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	113	11	52	0
	Tone 2	20	210	1	0
	Tone 3	0	5	12	4
	Tone 4	0	0	104	270

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	111	15	64	1
	Tone 2	17	219	0	0
	Tone 3	0	2	17	4
	Tone 4	0	0	95	256

Table B.3: Contingency tables of F0 mean-only production models by perception data

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	94	23	30	26
	Tone 2	82	18	84	51
	Tone 3	1	1	11	9
	Tone 4	163	32	112	65

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	84	58	34	4
	Tone 2	87	78	57	1
	Tone 3	7	5	12	1
	Tone 4	162	105	106	1

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	105	41	48	0
	Tone 2	80	68	81	0
	Tone 3	3	7	11	0
	Tone 4	148	110	99	0

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	102	59	32	5
	Tone 2	80	74	85	3
	Tone 3	6	6	6	1
	Tone 4	144	95	100	4

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	91	41	27	22
	Tone 2	91	29	74	36
	Tone 3	8	7	11	2
	Tone 4	135	64	100	63

Table B.4: Accuracy rates and F1 scores of individual production models predicted by perception data

(a) F0 mean + F0 slope model

		F1 Scores:			
Model	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	61.1	0.483	0.745	0.116	0.773
2	70.2	0.606	0.840	0.126	0.821
3	63.9	0.598	0.782	0.074	0.762
4	62.8	0.579	0.703	0.126	0.786
5	64.2	0.564	0.777	0.121	0.782

(b) F0 slope only model

		F1 Scores:			
Model	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	75.7	0.646	0.932	0.135	0.863
2	78.9	0.680	0.932	0.105	0.878
3	76.9	0.703	0.913	0.144	0.855
4	75.4	0.731	0.919	0.126	0.833
5	75.3	0.696	0.928	0.171	0.837

(c) F0 mean only model

		F1 Scores:			
Model	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	23.4	0.366	0.117	0.085	0.249
2	21.8	0.323	0.333	0.103	0.005
3	23.0	0.396	0.299	0.085	$n.a.$
4	23.2	0.385	0.311	0.050	0.022
5	24.2	0.360	0.156	0.092	0.260

Appendix C

Cross validation results of individual perception models

Table C.1: Contingency tables of F0 mean $\times$ F0 slope perception models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	377	52	0	21
	Tone 2	34	557	0	0
	Tone 3	9	0	0	4
	Tone 4	72	0	0	917

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	355	47	0	20
	Tone 2	55	563	0	0
	Tone 3	2	2	0	6
	Tone 4	81	0	0	908

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	425	46	0	20
	Tone 2	31	563	0	0
	Tone 3	24	1	0	42
	Tone 4	19	0	0	872

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	422	39	0	14
	Tone 2	40	563	0	0
	Tone 3	13	14	0	25
	Tone 4	29	0	0	885

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	399	27	0	18
	Tone 2	47	583	0	0
	Tone 3	30	2	0	76
	Tone 4	27	0	0	834

Table C.2: Contingency tables of F0 mean + F0 slope perception models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	370	59	0	21
	Tone 2	29	562	0	0
	Tone 3	9	0	0	4
	Tone 4	72	0	0	917

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	355	47	0	20
	Tone 2	55	563	0	4
	Tone 3	2	2	0	6
	Tone 4	81	0	0	908

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	422	49	0	20
	Tone 2	30	564	0	0
	Tone 3	24	1	0	42
	Tone 4	19	0	0	872

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	422	39	0	14
	Tone 2	40	563	0	0
	Tone 3	13	14	0	25
	Tone 4	29	0	0	885

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	399	27	0	18
	Tone 2	47	583	0	0
	Tone 3	30	2	0	76
	Tone 4	27	0	0	834

Table C.3: Contingency tables of F0 slope only perception models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	388	41	0	21
	Tone 2	53	538	0	0
	Tone 3	9	0	0	4
	Tone 4	72	0	0	917

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	362	37	0	23
	Tone 2	76	541	0	5
	Tone 3	2	2	0	6
	Tone 4	81	0	0	908

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	432	37	0	22
	Tone 2	48	546	0	0
	Tone 3	18	1	0	48
	Tone 4	19	0	0	872

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	426	26	0	23
	Tone 2	63	540	0	0
	Tone 3	11	14	0	27
	Tone 4	26	0	0	888

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	406	19	0	19
	Tone 2	74	556	0	0
	Tone 3	31	1	0	79
	Tone 4	27	0	0	834

Table C.4: Contingency tables of F0 mean only perception models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	450
	Tone 2	0	0	0	591
	Tone 3	0	0	0	13
	Tone 4	0	0	0	989

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	422
	Tone 2	0	0	0	622
	Tone 3	0	0	0	10
	Tone 4	0	0	0	989

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	491
	Tone 2	0	0	0	594
	Tone 3	0	0	0	67
	Tone 4	0	0	0	891

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	475
	Tone 2	0	0	0	603
	Tone 3	0	0	0	52
	Tone 4	0	0	0	914

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	444
	Tone 2	0	0	0	630
	Tone 3	0	0	0	108
	Tone 4	0	0	0	861

Table C.5: Contingency tables of null perception models

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	450
	Tone 2	0	0	0	591
	Tone 3	0	0	0	13
	Tone 4	0	0	0	989

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	422
	Tone 2	0	0	0	622
	Tone 3	0	0	0	10
	Tone 4	0	0	0	989

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	491
	Tone 2	0	0	0	594
	Tone 3	0	0	0	67
	Tone 4	0	0	0	891

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	475
	Tone 2	0	0	0	603
	Tone 3	0	0	0	52
	Tone 4	0	0	0	914

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	444
	Tone 2	0	0	0	630
	Tone 3	0	0	0	108
	Tone 4	0	0	0	861

Table C.6: Accuracy rates and F1 scores of individual perception models

(a) F0 mean $\times$ F0 slope model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	4968	90.6	0.800	0.928	$n.a.$	0.950
2	4699	89.4	0.776	0.912	$n.a.$	0.942
3	4893	91.0	0.859	0.935	$n.a.$	0.956
4	4597	91.5	0.862	0.924	$n.a.$	0.963
5	4609	88.9	0.843	0.939	$n.a.$	0.932

(b) F0 mean $+$ F0 slope model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	5016	90.5	0.796	0.927	$n.a.$	0.950
2	4728	89.4	0.776	0.912	$n.a.$	0.942
3	4925	91.0	0.856	0.934	$n.a.$	0.956
4	4610	91.5	0.862	0.924	$n.a.$	0.963
5	4654	88.9	0.843	0.939	$n.a.$	0.932

(c) F0 slope only model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	5184	90.2	0.798	0.920	$n.a.$	0.950
2	4900	88.6	0.768	0.900	$n.a.$	0.940
3	5077	90.6	0.857	0.927	$n.a.$	0.951
4	4742	90.7	0.851	0.912	$n.a.$	0.959
5	4778	87.9	0.829	0.922	$n.a.$	0.930

(d) F0 mean only model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	18762	48.4	$n.a.$	$n.a.$	$n.a.$	0.652
2	18787	48.4	$n.a.$	$n.a.$	$n.a.$	0.652
3	18403	43.6	$n.a.$	$n.a.$	$n.a.$	0.607
4	18524	44.7	$n.a.$	$n.a.$	$n.a.$	0.618
5	18204	42.1	$n.a.$	$n.a.$	$n.a.$	0.593

(e) Null model

			F1 Scores:			
Model	D	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	18943	48.4	$n.a.$	$n.a.$	$n.a.$	0.652
2	18972	48.4	$n.a.$	$n.a.$	$n.a.$	0.652
3	18572	43.6	$n.a.$	$n.a.$	$n.a.$	0.607
4	18675	44.7	$n.a.$	$n.a.$	$n.a.$	0.618
5	18351	42.1	$n.a.$	$n.a.$	$n.a.$	0.593

Appendix D

Prediction results of individual perception models by production data

Table D.1: Contingency tables of F0 mean + F0 slope perception models by production data

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	477	51	0	1
	Tone 2	108	417	0	0
	Tone 3	405	30	7	55
	Tone 4	1	0	0	491

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	465	46	0	1
	Tone 2	113	406	0	0
	Tone 3	406	27	11	51
	Tone 4	1	0	0	516

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	478	53	0	1
	Tone 2	103	424	0	0
	Tone 3	397	31	2	66
	Tone 4	1	0	0	487

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	461	50	0	2
	Tone 2	107	408	0	0
	Tone 3	410	34	0	60
	Tone 4	1	0	0	511

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	468	51	0	3
	Tone 2	114	399	0	0
	Tone 3	397	23	0	62
	Tone 4	0	0	0	526

Table D.2: Contingency tables of F0 slope-only perception models by production data

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	463	65	0	1
	Tone 2	108	417	0	0
	Tone 3	412	19	0	66
	Tone 4	1	0	0	491

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	454	57	0	1
	Tone 2	110	409	0	0
	Tone 3	416	15	0	64
	Tone 4	1	0	0	516

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	471	60	0	1
	Tone 2	99	428	0	0
	Tone 3	402	21	0	73
	Tone 4	1	0	0	487

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	452	59	0	2
	Tone 2	108	407	0	0
	Tone 3	422	21	0	61
	Tone 4	1	0	0	511

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	460	61	0	1
	Tone 2	116	397	0	0
	Tone 3	399	12	0	71
	Tone 4	0	0	0	526

Table D.3: Contingency tables of F0 mean-only perception models by production data

(a) Model 1

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	4	0	0	525
	Tone 2	0	0	0	527
	Tone 3	0	0	0	516
	Tone 4	0	0	0	471

(b) Model 2

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	4	0	0	531
	Tone 2	0	0	0	534
	Tone 3	0	1	0	481
	Tone 4	0	0	0	492

(c) Model 3

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	3	0	0	545
	Tone 2	0	0	0	485
	Tone 3	0	0	0	499
	Tone 4	0	0	0	511

(d) Model 4

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	1	0	0	528
	Tone 2	0	0	0	509
	Tone 3	0	0	0	508
	Tone 4	0	0	0	498

(e) Model 5

		Predicted			
		Tone 1	Tone 2	Tone 3	Tone 4
Actual	Tone 1	0	0	0	510
	Tone 2	0	0	0	510
	Tone 3	0	0	0	520
	Tone 4	0	0	0	503

Table D.4: Accuracy rates and F1 scores of individual perception models predicted by production data

(a) F0 mean + F0 slope model

		F1 Scores:			
Model	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	68.1	0.628	0.815	0.028	0.945
2	68.4	0.621	0.814	0.043	0.951
3	68.1	0.633	0.819	0.008	0.935
4	67.5	0.618	0.810	$n.a.$	0.942
5	68.2	0.624	0.809	$n.a.$	0.942

(b) F0 slope-only model

		F1 Scores:			
Model	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	67.1	0.612	0.813	$n.a.$	0.935
2	67.5	0.608	0.818	$n.a.$	0.940
3	67.8	0.626	0.826	$n.a.$	0.929
4	67.0	0.604	0.812	$n.a.$	0.941
5	67.7	0.615	0.808	$n.a.$	0.936

(c) F0 mean-only model

		F1 Scores:			
Model	Accuracy (%)	Tone 1	Tone 2	Tone 3	Tone 4
1	23.3	0.015	$n.a.$	$n.a.$	0.375
2	24.2	0.015	$n.a.$	$n.a.$	0.389
3	25.2	0.011	$n.a.$	$n.a.$	0.401
4	24.4	0.004	$n.a.$	$n.a.$	0.392
5	24.6	$n.a.$	$n.a.$	$n.a.$	0.395

Appendix E

Multinomial logistic regression classification accuracy of learners' perception data

Table E.1: Multinomial logistic regression classification accuracy and F1 scores of null model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	38.1	$n.a.$	$n.a.$	$n.a.$	0.55
LF01	2	47.7	$n.a.$	$n.a.$	$n.a.$	0.645
LF02	1	37.1	$n.a.$	$n.a.$	0.538	$n.a.$
LF02	2	34.5	$n.a.$	$n.a.$	0.509	$n.a.$
LF03	1	34.1	$n.a.$	0.495	$n.a.$	$n.a.$
LF03	2	34.1	$n.a.$	$n.a.$	$n.a.$	0.506
LF04	1	41	$n.a.$	$n.a.$	$n.a.$	0.581
LF04	2	42.4	$n.a.$	$n.a.$	$n.a.$	0.591
LF05	1	48.4	$n.a.$	0.652	$n.a.$	$n.a.$
LF05	2	45.9	$n.a.$	0.624	$n.a.$	$n.a.$
LF07	1	33.2	$n.a.$	0.522	$n.a.$	0.395
LF07	2	47.7	$n.a.$	0.638	$n.a.$	$n.a.$
LF08	1	32.6	$n.a.$	0.487	$n.a.$	0.329
LF08	2	36.9	$n.a.$	$n.a.$	$n.a.$	0.536
LF09	1	35.1	$n.a.$	0.519	$n.a.$	$n.a.$
LF09	2	39.2	$n.a.$	0.562	$n.a.$	$n.a.$
LF10	1	39.6	$n.a.$	$n.a.$	$n.a.$	0.564
LF10	2	35.7	$n.a.$	$n.a.$	$n.a.$	0.523
LF11	1	50.7	$n.a.$	0.672	$n.a.$	$n.a.$
LF11	2	50.3	$n.a.$	0.669	$n.a.$	$n.a.$
LF13	1	37.9	$n.a.$	0.538	$n.a.$	$n.a.$

Table E.1 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF13	2	27.1	$n.a.$	0.356	$n.a.$	0.439
LM01	1	44	$n.a.$	0.608	$n.a.$	$n.a.$
LM01	2	34.7	$n.a.$	0.513	$n.a.$	$n.a.$
LM02	1	34.6	$n.a.$	0.513	$n.a.$	$n.a.$
LM02	2	35.7	$n.a.$	$n.a.$	$n.a.$	0.524
LM03	1	32.9	$n.a.$	0.492	$n.a.$	$n.a.$
LM03	2	34	$n.a.$	0.506	$n.a.$	$n.a.$
LM04	1	35.7	$n.a.$	0.526	$n.a.$	$n.a.$
LM04	2	42.9	$n.a.$	0.598	$n.a.$	$n.a.$
LM05	1	35.4	$n.a.$	0.522	$n.a.$	$n.a.$
LM05	2	35.4	$n.a.$	0.521	$n.a.$	$n.a.$
LM07	1	43.5	$n.a.$	$n.a.$	$n.a.$	0.604
LM07	2	51.9	$n.a.$	$n.a.$	$n.a.$	0.683
LM13	1	50.3	$n.a.$	$n.a.$	$n.a.$	0.669
LM13	2	53.9	$n.a.$	$n.a.$	$n.a.$	0.7
LM14	1	24.7	$n.a.$	0.384	$n.a.$	0.334
LM14	2	29.7	$n.a.$	0.445	$n.a.$	0.34

Table E.2: Multinomial logistic regression classification accuracy and F1 scores of F0 mean-only model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	36.8	0.2	0.358	$n.a.$	0.479
LF01	2	47.1	$n.a.$	0.167	$n.a.$	0.641
LF02	1	37.1	$n.a.$	0.289	0.524	$n.a.$
LF02	2	31.2	$n.a.$	0.362	0.412	0.091
LF03	1	30.5	$n.a.$	0.464	0.243	0.133
LF03	2	31.5	$n.a.$	$n.a.$	0.182	0.477
LF04	1	41.7	$n.a.$	0.365	0.418	0.514
LF04	2	44.9	$n.a.$	0.418	$n.a.$	0.566
LF05	1	45.7	$n.a.$	0.646	$n.a.$	0.216
LF05	2	43.6	0.235	0.611	$n.a.$	$n.a.$
LF07	1	43.6	$n.a.$	0.55	$n.a.$	0.463
LF07	2	46.4	$n.a.$	0.63	$n.a.$	$n.a.$
LF08	1	31.2	$n.a.$	0.467	0.196	0.238
LF08	2	35.3	$n.a.$	0.323	$n.a.$	0.495
LF09	1	31.8	$n.a.$	0.469	0.069	0.29
LF09	2	38.9	$n.a.$	0.548	0.253	$n.a.$
LF10	1	37.9	$n.a.$	0.263	$n.a.$	0.526
LF10	2	35.4	$n.a.$	$n.a.$	0.297	0.507
LF11	1	50	0.213	0.669	$n.a.$	$n.a.$

Table E.2 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF11	2	49.4	$n.a.$	0.66	0.194	$n.a.$
LF13	1	40.5	0.325	0.608	0.22	$n.a.$
LF13	2	41.6	$n.a.$	0.573	$n.a.$	0.48
LM01	1	44	$n.a.$	0.608	$n.a.$	$n.a.$
LM01	2	34.7	$n.a.$	0.513	$n.a.$	$n.a.$
LM02	1	37.6	$n.a.$	0.335	0.445	0.437
LM02	2	42.5	$n.a.$	$n.a.$	0.549	0.556
LM03	1	34.5	$n.a.$	0.501	0.125	0.379
LM03	2	34.3	$n.a.$	0.487	0.366	0.273
LM04	1	31.2	0.303	0.429	0.214	$n.a.$
LM04	2	41.6	$n.a.$	0.587	$n.a.$	0.226
LM05	1	50.2	0.549	0.658	$n.a.$	$n.a.$
LM05	2	45.2	0.496	0.561	$n.a.$	$n.a.$
LM07	1	43.5	$n.a.$	$n.a.$	$n.a.$	0.604
LM07	2	51.9	$n.a.$	$n.a.$	$n.a.$	0.683
LM13	1	50.3	$n.a.$	$n.a.$	$n.a.$	0.669
LM13	2	53.9	$n.a.$	$n.a.$	$n.a.$	0.7
LM14	1	38.3	$n.a.$	0.474	$n.a.$	0.509
LM14	2	35.3	$n.a.$	0.456	$n.a.$	0.422

Table E.3: Multinomial logistic regression classification accuracy and F1 scores of F0 slope-only model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	57.4	0.235	0.58	$n.a.$	0.779
LF01	2	64.6	$n.a.$	0.613	$n.a.$	0.823
LF02	1	49.8	$n.a.$	0.74	0.519	0.183
LF02	2	57.2	$n.a.$	0.862	0.511	0.372
LF03	1	43.6	$n.a.$	0.505	0.118	0.619
LF03	2	53.3	0.23	$n.a.$	0.502	0.818
LF04	1	69	0.378	0.377	0.653	0.898
LF04	2	69.7	0.31	0.659	$n.a.$	0.887
LF05	1	43.4	$n.a.$	0.635	0.141	0.149
LF05	2	41.3	$n.a.$	0.618	0.21	0.174
LF07	1	65.5	$n.a.$	0.682	$n.a.$	0.864
LF07	2	67.6	$n.a.$	0.768	$n.a.$	0.79
LF08	1	54.3	$n.a.$	0.552	0.428	0.689
LF08	2	72.2	0.697	0.574	0.512	0.896
LF09	1	68.2	$n.a.$	0.762	0.63	0.736
LF09	2	60.1	$n.a.$	0.747	0.451	0.597
LF10	1	73.6	0.563	0.841	$n.a.$	0.809

Table E.3 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF10	2	68.9	0.731	0.851	$n.a.$	0.741
LF11	1	51.8	$n.a.$	0.689	$n.a.$	0.449
LF11	2	66.9	$n.a.$	0.764	0.663	$n.a.$
LF13	1	42.2	$n.a.$	0.59	0.233	0.412
LF13	2	48.5	0.1	0.538	0.245	0.732
LM01	1	51.3	$n.a.$	0.699	0.222	0.458
LM01	2	61.1	0.372	0.742	0.266	0.727
LM02	1	70.6	0.791	0.973	0.125	0.653
LM02	2	76	0.721	0.927	$n.a.$	0.827
LM03	1	61.1	$n.a.$	0.673	0.472	0.794
LM03	2	55.9	$n.a.$	0.702	0.55	0.512
LM04	1	51	0.562	0.506	$n.a.$	0.578
LM04	2	63.3	$n.a.$	0.734	$n.a.$	0.734
LM05	1	33.1	0.226	0.378	0.427	$n.a.$
LM05	2	37	0.457	0.264	0.42	$n.a.$
LM07	1	82.7	0.768	0.854	$n.a.$	0.914
LM07	2	92.8	0.891	0.92	$n.a.$	0.985
LM13	1	94.5	0.987	0.978	$n.a.$	0.959
LM13	2	85.7	0.982	0.728	0.433	0.994
LM14	1	54.7	0.412	0.614	$n.a.$	0.739
LM14	2	69.7	0.622	0.792	0.222	0.836

Appendix F

Linear discriminant analysis classification accuracy of learners' production data

Table F.1: Linear discriminant analysis classification accuracy and F1 scores of F0 mean-only model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	52.1	0.398	0.444	0.529	0.85
LF01	2	57.6	0.55	0.296	0.604	0.861
LF02	1	42.8	0.311	0.244	0.585	0.539
LF02	2	52.5	0.482	0.2	0.814	0.515
LF03	1	43.8	0.662	0.392	0.589	0.254
LF03	2	60.4	0.831	0.392	0.647	0.6
LF04	1	79.2	0.921	0.684	0.861	0.669
LF04	2	64.2	0.931	0.277	0.733	0.687
LF05	1	45.1	0.443	0.6	0.498	0.45
LF05	2	42.8	0.346	0.648	0.445	0.443
LF07	1	75	0.853	0.651	0.688	0.768
LF07	2	69.2	0.595	0.848	0.798	0.542
LF08	1	73.9	0.686	0.726	0.865	0.647
LF08	2	58.3	0.561	0.611	0.697	0.531
LF09	1	53.3	0.612	0.293	0.758	0.424
LF09	2	58.9	0.947	0.537	0.535	0.274
LF10	1	50.8	0.661	0.32	0.565	0.439
LF10	2	58	0.908	0.409	0.483	0.668
LF11	1	41.2	0.318	0.261	0.534	0.558
LF11	2	50	0.483	0.256	0.609	0.583
LF13	1	54.2	0.517	0.677	0.557	0.329

Table F.1 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF13	2	55.9	0.527	0.429	0.821	0.553
LM01	1	53.1	0.683	0.49	0.546	0.286
LM01	2	54.7	0.663	0.538	0.37	0.614
LM02	1	66.7	0.881	0.449	0.467	0.863
LM02	2	66.7	0.874	0.469	0.591	0.74
LM03	1	42.6	0.357	0.404	0.577	0.515
LM03	2	57.5	0.521	0.509	0.77	0.476
LM04	1	32	0.203	$n.a.$	0.408	0.493
LM04	2	42.6	$n.a.$	0.347	0.614	0.509
LM05	1	58.3	0.406	0.73	0.635	0.538
LM05	2	41.7	0.236	0.625	0.487	0.386
LM07	1	79.2	0.835	0.777	0.83	0.67
LM07	2	84.2	0.807	0.844	0.891	0.805
LM13	1	65	0.471	0.821	0.651	0.633
LM13	2	73.9	0.725	0.593	0.822	0.793
LM14	1	50	0.737	0.437	0.427	0.444
LM14	2	50	0.846	0.224	0.448	0.44

Table F.2: Linear discriminant analysis classification accuracy and F1 scores of F0 time-normalized slope only model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	46.2	0.222	0.317	0.665	0.574
LF01	2	67.7	0.697	0.5	0.725	0.852
LF02	1	78.1	0.789	0.346	0.874	0.971
LF02	2	82.5	0.817	0.554	0.896	1
LF03	1	68	0.801	0.661	0.336	0.913
LF03	2	76.5	0.916	0.452	0.67	0.985
LF04	1	67.5	0.879	0.372	0.418	0.96
LF04	2	69.2	0.859	0.399	0.576	0.966
LF05	1	66.2	0.724	$n.a.$	0.78	0.805
LF05	2	54.7	0.446	$n.a.$	0.892	0.732
LF07	1	73.3	0.985	0.482	0.445	1
LF07	2	80	0.985	0.623	0.551	0.982
LF08	1	76.5	0.956	0.527	0.599	0.964
LF08	2	68.3	1	0.305	0.42	1
LF09	1	79.2	0.796	0.654	0.74	0.982
LF09	2	83.8	0.949	0.699	0.678	1
LF10	1	73.3	0.825	0.56	0.568	0.964
LF10	2	82.4	0.941	0.663	0.668	0.982
LF11	1	78.9	0.908	0.698	0.678	0.929

Table F.2 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF11	2	79.2	0.985	0.575	0.55	1
LF13	1	65	0.582	0.251	0.833	0.893
LF13	2	70.2	0.678	0.333	0.806	0.982
LM01	1	62.2	0.618	0.577	0.382	1
LM01	2	75.6	0.825	0.644	0.506	1
LM02	1	75.8	0.899	0.797	0.435	0.801
LM02	2	78.3	0.889	0.763	0.414	0.941
LM03	1	53	0.929	0.549	0.249	0.836
LM03	2	72.5	0.887	0.622	0.484	0.916
LM04	1	45.3	0.515	0.323	0.405	0.783
LM04	2	54.8	0.543	0.397	0.444	0.95
LM05	1	64.2	0.582	0.354	0.711	0.865
LM05	2	58.3	0.556	0.283	0.715	0.966
LM07	1	80.8	0.901	0.748	0.543	1
LM07	2	83.3	0.969	0.703	0.653	1
LM13	1	79.3	0.848	0.985	$n.a.$	0.849
LM13	2	67.2	0.826	0.588	0.328	0.923
LM14	1	66.7	0.836	0.333	0.52	0.927
LM14	2	83.3	0.886	0.694	0.795	0.92

Table F.3: Linear discriminant analysis classification accuracy and F1 scores of F0 time-scaled slope only model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	66.7	0.685	0.28	0.666	0.931
LF01	2	68.6	0.734	0.345	0.686	0.945
LF02	1	71.3	0.684	0.416	0.886	0.971
LF02	2	80.8	0.817	0.479	0.871	1
LF03	1	60.5	0.787	0.621	0.226	0.925
LF03	2	75.7	0.879	0.365	0.676	1
LF04	1	84.2	0.936	0.626	0.773	0.982
LF04	2	81.7	0.897	0.703	0.789	0.966
LF05	1	65.3	0.724	$n.a.$	0.74	0.805
LF05	2	51.3	0.43	$n.a.$	0.773	0.701
LF07	1	82.5	0.985	0.687	0.61	1
LF07	2	71.7	0.985	0.425	0.438	1
LF08	1	79.9	0.956	0.609	0.652	0.964
LF08	2	88.3	1	0.735	0.78	1
LF09	1	81.7	0.82	0.637	0.805	0.982
LF09	2	88	0.949	0.782	0.765	1
LF10	1	78.3	0.825	0.654	0.69	0.964

Table F.3 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF10	2	89	0.956	0.776	0.815	0.982
LF11	1	78.1	0.895	0.711	0.684	0.908
LF11	2	86.7	0.969	0.671	0.772	1
LF13	1	65	0.575	0.26	0.851	0.875
LF13	2	70.2	0.668	0.444	0.806	0.982
LM01	1	63.9	0.598	0.618	0.394	1
LM01	2	84.9	0.836	0.784	0.76	1
LM02	1	80	0.887	0.777	0.55	0.881
LM02	2	75	0.841	0.723	0.395	0.954
LM03	1	48.7	0.69	0.543	0.188	0.897
LM03	2	70	0.829	0.604	0.261	0.982
LM04	1	47.9	0.667	0.298	0.48	0.891
LM04	2	50.4	0.423	0.347	0.365	0.964
LM05	1	65.8	0.64	0.41	0.7	0.865
LM05	2	61.7	0.599	0.303	0.692	0.966
LM07	1	70	0.887	0.46	0.377	1
LM07	2	75.8	0.985	0.501	0.517	1
LM13	1	70.9	0.748	0.803	$n.a.$	0.84
LM13	2	84.9	0.818	0.893	0.723	0.945
LM14	1	58.3	0.828	0.237	0.33	0.927
LM14	2	83.3	0.859	0.754	0.799	0.902

Table F.4: Linear discriminant analysis classification accuracy and F1 scores of F0 mean $+$ F0 time-normalized slope only model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	61.6	0.607	0.363	0.624	0.817
LF01	2	68.7	0.728	0.472	0.719	0.861
LF02	1	80.6	0.763	0.487	0.876	0.971
LF02	2	78.3	0.789	0.401	0.854	1
LF03	1	68	0.92	0.521	0.453	0.909
LF03	2	80.7	0.887	0.52	0.78	0.969
LF04	1	89.2	0.941	0.81	0.861	0.942
LF04	2	80.8	0.921	0.544	0.762	0.966
LF05	1	76.4	0.758	0.59	0.876	0.805
LF05	2	80.7	0.823	0.747	0.887	0.732
LF07	1	86.7	0.985	0.74	0.714	0.982
LF07	2	90.8	0.985	0.848	0.798	0.982
LF08	1	89.9	0.941	0.822	0.865	0.964
LF08	2	83.3	1	0.717	0.676	1
LF09	1	80.8	0.807	0.629	0.79	0.982

Table F.4 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF09	2	87.2	0.964	0.742	0.762	1
LF10	1	72.5	0.79	0.507	0.641	0.942
LF10	2	83.2	0.951	0.694	0.677	0.982
LF11	1	78.1	0.892	0.638	0.62	0.905
LF11	2	80.8	0.969	0.594	0.636	1
LF13	1	78.3	0.669	0.66	0.898	0.875
LF13	2	71.2	0.65	0.473	0.855	0.951
LM01	1	74.7	0.951	0.453	0.598	1
LM01	2	81.5	1	0.646	0.603	1
LM02	1	86.7	0.948	0.837	0.697	0.938
LM02	2	82.5	1	0.744	0.502	0.969
LM03	1	68.7	0.828	0.427	0.625	0.923
LM03	2	81.7	0.887	0.641	0.754	0.948
LM04	1	42.8	0.276	0.288	0.389	0.797
LM04	2	65.2	0.748	0.452	0.611	0.95
LM05	1	78.3	0.652	0.818	0.765	0.865
LM05	2	80	0.685	0.721	0.786	0.966
LM07	1	90	1	0.805	0.794	1
LM07	2	94.2	1	0.871	0.891	1
LM13	1	93.4	0.929	0.985	0.736	0.978
LM13	2	88.2	0.954	0.802	0.828	0.945
LM14	1	69.2	0.884	0.478	0.488	0.912
LM14	2	86.7	0.956	0.78	0.783	0.942

Table F.5: Linear discriminant analysis classification accuracy and F1 scores of F0 mean + F0 time-scaled slope only model

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF01	1	68.4	0.7	0.362	0.625	0.969
LF01	2	71.2	0.759	0.325	0.718	0.93
LF02	1	78.1	0.72	0.456	0.847	0.971
LF02	2	80	0.806	0.433	0.872	1
LF03	1	67.2	0.887	0.437	0.465	0.933
LF03	2	81.5	0.901	0.585	0.741	1
LF04	1	95	0.969	0.9	0.938	0.982
LF04	2	83.3	0.921	0.59	0.808	0.966
LF05	1	66.2	0.528	0.59	0.693	0.805
LF05	2	71.5	0.662	0.726	0.754	0.658
LF07	1	85.8	1	0.695	0.687	1
LF07	2	91.7	1	0.848	0.798	1
LF08	1	89.9	0.941	0.822	0.865	0.964

Table F.5 – continued from previous page

			F1 Scores:			
Learner	Visit	Mean Accuracy	Tone 1	Tone 2	Tone 3	Tone 4
LF08	2	87.5	0.966	0.707	0.789	1
LF09	1	80.8	0.779	0.636	0.807	0.982
LF09	2	91.4	0.964	0.842	0.842	1
LF10	1	73.3	0.803	0.507	0.641	0.964
LF10	2	94.1	0.951	0.897	0.929	0.982
LF11	1	73.9	0.815	0.638	0.583	0.861
LF11	2	85.8	0.969	0.671	0.753	1
LF13	1	77.5	0.648	0.658	0.898	0.875
LF13	2	75	0.673	0.533	0.896	0.982
LM01	1	78.3	0.951	0.546	0.636	1
LM01	2	90.7	1	0.784	0.833	1
LM02	1	86.7	0.948	0.837	0.702	0.938
LM02	2	80.8	1	0.713	0.463	0.969
LM03	1	65.2	0.786	0.468	0.611	0.956
LM03	2	79.2	0.912	0.521	0.681	1
LM04	1	47	0.344	0.296	0.464	0.894
LM04	2	61.7	0.637	0.412	0.597	0.964
LM05	1	76.7	0.633	0.82	0.717	0.865
LM05	2	77.5	0.605	0.721	0.76	0.966
LM07	1	88.3	0.982	0.764	0.782	1
LM07	2	93.3	1	0.858	0.869	1
LM13	1	89.6	0.971	0.87	0.68	0.944
LM13	2	85.8	0.923	0.775	0.809	0.924
LM14	1	71.7	0.884	0.479	0.511	0.927
LM14	2	85	0.905	0.809	0.784	0.902