Learning to speak a minority language at school: Mandarin speech development of children in a Chinese-English bilingual program in Canada

by

Youran Lin

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science in Speech-Language Pathology
and
Doctor of Philosophy in Rehabilitation Science

Department of Communication Sciences and Disorders
Faculty of Rehabilitation Medicine
University of Alberta

© Youran Lin, 2023

Abstract

A key factor of bilingual speech development is speech input. Different from an immigrant setting where children who are learning the majority language as a second language (L2) can receive rich and authentic input, the speech input is usually limited when learning a minority language. For L2 learners of a minority language, the input is limited to the school settings and lacks a range of native-like models. For heritage speakers of a minority language, the input is often limited to their households and immediate communities. Two-way bilingual education programs provide a unique context where native speakers of a majority language and native speakers of a minority language are in class together and are expected to provide authentic peer input for each other in both languages. Investigating the speech development in a minority language of children who are enrolled in such bilingual programs can help understand the roles of home input, school input, and limited community input.

This dissertation examines the case of a Chinese-English bilingual education program in Western Canada and investigates bilingual students' speech development in Mandarin, a societal minority language whose phonological system is uniquely distinctive from English. This paper-based thesis consists of three journal manuscripts and chapters of introduction, educational context, methodology, and conclusion. The first manuscript proposes a conceptual model to understand L2 speech learning, focusing on the interactions between L2 learners and L1 listeners. It indicates that more research in non-English L2s is warranted, especially with child learners. The second manuscript qualitatively explores teachers' perspectives on teaching Mandarin pronunciation. Semi-structured interviews were conducted with twelve Chinese teachers, suggesting that Mandarin pronunciation, especially lexical tones which do not exist in English, is challenging to learn. The third manuscript analyzed bilingual students' tone productions using

transcription and acoustic measurements. Factors influencing bilingual students' speech learning were investigated, including their home input and school input, cross-linguistic influences from English, and the universal difficulties of tone targets.

This work provides qualitative and quantitative evidence on bilingual students' speech development in Mandarin as a minority language in Canada. It emphasizes that students' speech learning in a minority language is impacted by not only the limited input but also the language's phonology, and therefore, students need support at multiple levels to continually improve their Mandarin pronunciation. This work adds to the literature on bilingual education, bilingual speech development, and Mandarin speech acquisition. It also raises awareness of and encourages more evidence on (1) pronunciation teaching and learning in bilingual contexts, (2) the ongoing L2 speech development in child learners, (3) bilingual speech theories in the suprasegmental domain (e.g., tones), and (4) the learning of Mandarin and other world languages as a heritage language or L2 in this globalized world.

Keywords: children, bilingual education, pronunciation, Mandarin, minority language

Preface

This thesis is an original work by Youran Lin. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Speech Production in Children Enrolled in Mandarin-English Bilingual Education Programs," No. Pro00075638, August 30, 2017.

Chapter 2 of this thesis has been published as **Y. Lin**, F. Li, A.A.N. MacLeod, and K.E. Pollock, "A conceptual model of second language pronunciation in communicative contexts: Implications for children's bilingual education," *Frontiers in Psychology, Sec. Psychology of Language*, vol. 14, 1125157. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). I proposed the conceptual model through a literature review. All authors contributed to discussion sessions of literature and model revision. I was responsible for the manuscript composition, and all authors contributed to the manuscript revision.

Chapter 4 of this thesis has been submitted for publication as **Y. Lin**, F. Li, and K.E. Pollock, "Pronunciation teaching in minority languages: Perspectives of elementary school teachers in a Chinese-English bilingual program in Canada." I was responsible for the data collection, data analysis, and manuscript composition. K.E. Pollock contributed to the data analysis (coding book development). F. Li and K.E. Pollock were supervisory authors who contributed to concept formation and manuscript revision.

Dedication

谨以此作献给中华文化、语言和人民。你们生于勤奋、团结、传承、创新的热土，在四千年不间断的合作、贡献和交流中，是你们将一段灿烂的历史讲述给全世界。

This work is dedicated to the Chinese cultures, languages, and people that were born in the land of diligence, collectiveness, inheritance, and innovation and spread their splendid history around the world through 4,000 years of unceasing collaboration, contribution, and communication.

谨以此作致敬加拿大的华人社群。你们历经风雨，却从未放弃忠于内心、勇于追求、甘于奉献。你们创立中英双语项目的故事深深地鼓舞了我。

This work is a tribute to the Chinese communities in Canada who have been through struggles but have never given up on figuring out who they are, what they want to be, and what they can be. Their stories of establishing the Chinese-English bilingual program were inspiring.

This work is also a tribute to language learners who are brave enough to step out of their comfort zones, endeavour to embrace a different means of thinking, and take the initiative to access multiple cultures through their learning.

Do you know what a foreign accent is? It's a sign of bravery.

- Amy Chua -

Acknowledgements

When conducting my doctoral studies, I was financially supported by the Faculty of Rehabilitation Medicine's recruitment award and the Social Sciences and Humanities Research Council (SSHRC) through the Vanier Canada Graduate Scholarships program and through my supervisors' SSHRC Insight Grant, alongside the other scholarships and studentships provided by the University of Alberta.

My greatest appreciation is reserved for my supervisor at the University of Alberta, Dr. Karen E. Pollock, and my co-supervisor at the University of Lethbridge, Dr. Fangfang Li. Karen is the best supervisor I could ever ask for. I could not imagine my first five years in Canada without her kindness, positivity, and exceptional academic support. I learned so much from her as not only an outstanding researcher but also an impressive person. Thanks to Fangfang for being my role-model researcher and always providing selfless care, guidance, and support. I would also like to thank my supervisory committee members. Thanks to Dr. Benjamin V. Tucker who taught me a lot about phonetics and statistics and also about being an amazing instructor. Thanks to Dr. Andrea A. N. MacLeod who helped me a lot with the concept formation of the studies and shared her care for diversity and inclusivity. I also want to thank my examiners, Dr. Yue Wang and Dr. Monique Charest, for their comments and suggestions.

I express my heartfelt thanks to my research participants, including the principals, teachers, parents, and bilingual students. Special thanks to Teacher Terence Kong and Teacher Max Liu who helped us coordinate the research activities at the schools. I also send my thoughts to Principal Dave Kowalchuk and his beloved family. Principal Kowalchuk wholeheartedly supported our research and impacted many people with kindness and warmth. I would like to thank my fellow researchers for their enthusiastic assistance with the studies: my labmates Lujia

Yang and Nan Xing, our CSD 900 group members (I will not list all your names, but your projects and your insights continued to inspire me), our international interns (especially Xiaozhu Chen, Binqing Yu, Minjia Tao, Candice Lee, and Dorcas Chen who helped me with data collection), and our research assistants and volunteers (especially Maria Pollock and Dr. Alesha Reed who helped me with data collection and processing).

I am grateful to the people who shared their knowledge with me which helped with this dissertation. Thanks to Dr. Johanne Paradis and Dr. Trelani Chapman for examining my candidacy and providing valuable suggestions. Thanks to Dr. Jalal Al-Tamimi, Dr. Harald Baayen, Scott Perry, Dr. Adriana Soto Corominas, and Berhanu Wubie for teaching me about statistics. Thanks to Yina Liu and Dr. Wei Liu who provided feedback on one of my manuscripts. Thanks to Dr. John Archibald and Dr. Robert Frender who shared their research works with me. Thanks to Dr. Yvan Rose, Dr. Barbara Dodd, Dr. Valter Ciocca, Dr. Lisa Goffman, Dr. Françoise Brosseau-Lapré, Dr. Daniel Aalto, Teri Slade, Dr. Catrine Demers, Hsin-Hsiu Yeh, Pegah Firouzeh, Angela Feehan, Lindsay McHutchion, Jessica Harasym O'Byrne, Carlee Wilson, and Chloe Korade for discussing research and researcher life with me.

I want to thank the amazing faculty and staff of the Faculty of Rehabilitation Medicine and the Department of Communication Sciences and Disorders, including our professors, administrative staff, and clinical educators. Special thanks are due to Prof. Lu-Anne McFarlane, Dr. Carol Boliek, and Dr. Melanie Campbell who helped me with my application for the program, Ms. Vicki Trombley, Ms. Angela Libutti, and Ms. Shelley Richmond who answered a thousand questions of mine and were always understanding and supportive, Dr. Trelani Chapman and Laura Brompton who led a wonderful instructor team and taught me a lot about teaching, and Dr. Mary Roberts and Dan Gillespie who supported me as a new instructor.

I want to thank my friends in Canada and around the world who gave me tremendous emotional support: Yina and Alex (my writing buddy, food buddy, lovely couple, and parents to my godson), Lujia (my labmate, gym buddy, bridesmaid, and lovely friend), Zifeng (my hometown sister and conjugate maid of honour), Weinan and Jianfei (who chat every day and tried to meet up whenever we could), Athena (the best knitting podcaster), George (who kindly proofread my candidacy exams), Shi Cheng (who helped discover music and food), Yuke, Lin Zhou, Anqi, and Junfei (wonderful gaming friends), Gloria and Crystal (my loved dormmates who travelled far to meet me in Harbin), Lili (my loved dormmate who cooked for me in Beijing), Joyce (with great waves of laughter), SLP friends (Joyce, Jessie, Nicole, Steph, Mikayla, Kelly, and plus ones), Devin and Susan (who took care of me when I was sick in Toronto), Nikki (who helped with PRship), Louis and Keely (clinical partners), and so on.

Finally, my deepest gratitude goes to my family in China and in BC. My mom is a great person who overcomes hardship and never stops being kind and curious. She quickly made friends with my friends and many travellers at the airport during her first visit to Canada. She is an extraordinary embroidery artist and is still learning English, Star Battle, Zoom, and passing customs. My parents-in-law and Dikim and Diku continuously fed me with food and supported my work. My HEARTIEST THANKS are due to Kris, who does not like seeing all-upper-case exclamations. They are not only a partner who is always there for me, but also a friend who always understands me, a teacher who shares knowledge, skills, and constructive feedback with me, a peer who can talk about research for hours with me, and an interesting soul who brought a lot of penguins and joy into my life.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations, Key Chinese Terminologies, and Equations

Abbreviations

AIC: Akaike information criterion

ALEQ: Alberta Language Environment Questionnaire

AR(1): autoregressive model

CAH: Contrastive Analysis Hypothesis

Chinese bilingual program: the Chinese-English two-way bilingual program in Edmonton

CLT: Communicative Language Teaching

COREQ: Consolidated criteria for reporting qualitative research

CPH: Critical Period Hypothesis

ECBEA: Edmonton Chinese Bilingual Education Association

f0: fundamental frequency

FFI: Form-Focused Instruction

GA(M)M: Generalized Additive (Mixed) Model

HL: heritage language

IISLE: The Institute for Innovation in Second Language Education

IPA: International Phonetic Alphabet

L1: first language (or native language)

L2: second language

L2LP: the Second Language Linguistic Perception model

LEAP-Q: Language Experience and Proficiency Questionnaire

M: mean

MD: mean difference

NLM: the Native Language Magnet theory

PAM: Perceptual Assimilation Model

PAM-S: Perceptual Assimilation Model – Suprasegmental

PCA: principal component analysis

PCC: Percentage of Consonants Correct

PD: professional development

PPC: Percentage of Phones Correct

PRIMIR: processing rich information from multidimensional, interactive representation

P(N)TC: Percentage of (Number) Tones Correct

SD: standard deviation

SLM: Speech Learning Model

SSANOVA: Smoothing Spline ANOVA (analysis of variance)

Key Chinese terminologies

拼音 *pīnyīn*: a romanized transcription system of Mandarin speech

方言 *fāngyán*: a regional variety of Chinese, dialects

普通话 *pǔtōnghuà*: Standard Chinese in Mainland China

Equations

Equation 5.1. Percentage Number Tones Correct (PNTC, Hedlund & Rose, 2020)

$$\begin{aligned} PNTC &= correct \,/\, (correct + incorrect) \times 100 \\ &= correct \,/\, (correct + deletion + substitution) \times 100 \end{aligned}$$

Equation 6.1. T values to normalize *f0* (Shi & Wang, 2006)

$$T = 5 \times (\log_{10}(f0) - \log_{10}(f0_{min})) \,/\, (\log_{10}(f0_{max}) - \log_{10}(f0_{min}))$$

Chapter 1 Introduction

Speech input is a key factor of bilingual children's speech development. The quantity and quality of speech input in the first language (L1) and second language (L2) (Flege & Bohn, 2021) may be influenced by their social status (MacLeod & Stoel-Gammon, 2010), the age of learning and length of exposure (Flege, 1995), the current frequency of input and output, and the availability of native-like models (Flege & Bohn, 2021). For example, children who immigrated at an early age are more likely to produce the local majority language with native-like speech, or at least fully intelligible speech, compared to adult learners of an L2. These children usually have high-quantity and -quality L2 input from a variety of sources, including schools and community (Derwing, 2020). On the contrary, children who are learning a minority language as their L2 or maintaining it as a heritage language (HL) are challenged by limited speech input. For an L2 learner of a minority language, speech input may be limited in the school environment and lack a range of native speakers' input. For an HL child, speech input may be limited to the home environment and immediate communities. Therefore, research on speech development in a minority language can further our understanding of the roles played by speech input.

Two-way bilingual education programs provide a unique opportunity to examine the roles of school input in a minority language. L1 speakers of a majority language and L1 speakers of a minority language are together in class to provide peer input for each other in both languages, and students are expected to develop functional proficiency in the minority language at no cost to the majority language proficiency (Dicks & Genesee, 2017). Evidence from Spanish-English and Gaelic-English bilingual programs suggested that L2 learners' speech productions in the minority language are similar to their L1 peers, indicating the sufficiency of input in two-way bilingual schools (Menke, 2017; Nance, 2020). However, during our

interactions with educators in a Chinese-English bilingual program in Edmonton, Alberta, an Associate Principal asked: “Why after many years in the program do the children still have accents in their Mandarin?”

On one hand, this mismatch between existing research evidence and educator observations may still be due to speech input. Although L2 students are expected to receive high-quality speech input in the minority language in a two-way bilingual classroom, Western Canada is an English-dominant region where ambient input in Mandarin is limited in the general society, so most L2 students only receive Mandarin input from school. On the other hand, speech development in L2 is influenced by the similarity between L1 and L2 (Baker & Trofimovich, 2005). Different from the L1-L2 pairs investigated in previous studies (Menke, 2017; Nance, 2020), Mandarin and English are typologically distant, which poses more challenges for speech development in Mandarin.

This dissertation investigates the speech development in Mandarin of bilingual students enrolled in a Chinese bilingual program in Western Canada and the factors that impact such learning. It is of theoretical importance because it (1) observes the ongoing learning process among child learners (Tsukada et al., 2003), (2) examines the roles of speech input in bilingual speech development, including home input, school input, and the limited community input related to the minority status of Mandarin, and (3) examines the roles of L1 and L2 similarity by investigating a distinctive pair of majority and minority languages, English and Mandarin.

In addition, research on speech development in bilingual education has important pedagogical implications. Pronunciation used to be the “Cinderella” in the world of L2 teaching, unfairly oppressed and marginalized (Celce-Murcia et al., 1996, p.323). With limited evidence, curriculum designs are often different across programs (e.g., Alberta Education, 2006, 2008), and

teachers report challenges in teaching pronunciation in a bilingual classroom. Qualitative and quantitative research evidence on teaching and learning pronunciation in bilingual programs can (1) document the practical evidence and share it among bilingual teachers around the world, (2) use objective data to validate teachers' observations, insights, and strategies, (3) identify challenging speech targets and challenging situations, and (4) suggest evidence-based teaching strategies for educators of bilingual programs.

Dissertation objectives and purposes

This dissertation investigates the Mandarin speech development of bilingual students enrolled in a Chinese-English bilingual education program in Canada. It reviews the theoretical and educational background of bilingual speech development and bilingual education. Subsequently, it explores Chinese teachers' perspectives of teaching and learning the pronunciation of Mandarin in a minority language context. Finally, it investigates bilingual students' production of a unique phonological dimension of Mandarin that does not exist in English, lexical tones, and the factors that influence their learning. With the understanding that tones are a small but critically important aspect of Mandarin speech development, this dissertation uses them as a probing point to showcase how bilingual children learn a unique phonological dimension in a minority language that is completely different from the majority language in a school setting.

This dissertation has three purposes that are addressed by three journal manuscripts which aim to:

- Review existing theories and evidence in pronunciation learning and obtain insights into how students learn to speak a minority language at school;

- Understand how the pronunciation of a minority language is taught at a bilingual school and present challenges perceived by teachers and strategies used to address them; and
- Investigate how students from diverse language backgrounds learn a unique phonological dimension of a minority language, i.e., Mandarin tones, in the bilingual education context in Canada.

Organization of presentation

The work starts from high-level overviews of the theoretical and educational background and dives into very specific analyses of tones.

Chapter 1, the current chapter, states the researcher's motivations for this dissertation and identifies the objectives and research questions.

Chapter 2 (Paper 1) is a published journal article (Lin et al., 2023) introducing an encompassing conceptual model of L2 pronunciation in the context of interlocutor interactions. This chapter reviews L2 speech learning theories and multidisciplinary evidence on L2 speech learning and has implications for speech development in bilingual education.

Chapter 3 introduces the Chinese bilingual program in the context of Canada's bilingual education and reviews the knowledge related to speech input and pronunciation learning in this program.

Chapter 4 (Paper 2) is a submitted manuscript examining Chinese teachers' perspectives of students' learning of Mandarin pronunciation in bilingual classrooms. In this paper, teachers reflected on the factors that influenced students' pronunciation learning and shared their strategies for teaching Mandarin pronunciation. The chapter also highlights Mandarin tones as a challenging phonological dimension for bilingual students, which motivates the next chapters.

Before presenting an empirical study on bilingual students' Mandarin tone production, Chapter 5 provides some methodological details involving transcription-based analyses in the software program Phon (Hedlund & Rose, 2020).

Chapter 6 (Paper 3) is in the form of a manuscript to be submitted for publication that examines bilingual students' production of Mandarin lexical tones using both transcription-based and acoustic analyses. Through statistical modeling, the study shows how tone productions are influenced by multifaceted factors such as intra-language factors (e.g., the universal difficulty of a tone target), inter-language factors (e.g., English's transfer effects on Mandarin tone productions), and extra-language factors (e.g., students' home language environment and learning duration at the bilingual school).

Chapter 7 summarizes the findings of the studies, their key contributions and highlights, and their limitations. It also discusses the other (current or future) research directions that are related to this dissertation.

Connecting texts between chapters are included as appropriate to elaborate on the relationships between the independent manuscripts.

横看成岭侧成峰，远近高低各不同。

不识庐山真面目，只缘身在此山中。

—— 苏轼《题西林壁》

Some see ridges yet others see a peak: What you observe depends on how you seek.

For I indulge in the mountains so deep, their true profile may forever stay in mystique.

— SŪ Shì, Written on the Wall of Xilin Temple

The speech development of a minority language in a bilingual education context is such a complex and intriguing topic. With different scopes, multiple perspectives, and a variety of methods, one can only hope to open one corner of its veil. In this dissertation, you will see me hover over the mountains, walk along the mountain trails, and closely observe the rocks and the trees deep in the mountains through three thematically related papers. With such efforts, I hope to contribute to our understanding of this topic and other relevant fields.

Chapter 2 (Paper 1) Theories and Empirical Studies on L2 Pronunciation

Lin, Y., Pollock, K. E., MacLeod, A. A. N., & Li, F. (2023). A conceptual model of second language pronunciation in communicative contexts: Implications for children's bilingual education. *Frontiers of Psychology, 14*, 1125157. doi: 10.3389/fpsyg.2023.1125157.

Chapter 2 (Paper 1) is a review article on theories and empirical evidence on L2 pronunciation. It proposes a three-layer conceptual model to depict how L2 communication is achieved through sociopsychological, acquisitional, and productive-perceptual interactions between interlocutors from diverse backgrounds (namely "L2 Learner" and "L1 Listener" in the paper). The three layers emerged from extensive literature review and weekly discussions among the authors and were not previously established in the literature. The purpose of including these three layers within one model was to promote a transdisciplinary comprehension of literature, including but not limited to sociolinguistics, L2 speech development, and phonetics, which all provide indispensable background knowledge for L2 pronunciation studies. The purpose of including both an L2 Learner and an L1 Listener in the model was to emphasize the shared responsibility of communication in an L2 context.

frontiers | Frontiers in Psychology

TYPE Review
PUBLISHED 17 April 2023
DOI 10.3389/fpsyg.2023.1125157

Check for updates

OPEN ACCESS

EDITED BY
Valéria Csépe,
University of Pannonia, Hungary

REVIEWED BY
Halil Ercan,
Atatürk Teacher Training Academy, Cyprus
Mojgan Rashtchi,
Islamic Azad University North Tehran Branch, Iran
Judit Navracsics,
University of Pannonia, Hungary

*CORRESPONDENCE
Youran Lin
youran1@ualberta.ca
Karen E. Pollock
karen.pollock@ualberta.ca

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

RECEIVED 15 December 2022
ACCEPTED 30 March 2023
PUBLISHED 17 April 2023

CITATION
Lin Y, Li F, MacLeod AAN and Pollock KE (2023) A conceptual model of second language pronunciation in communicative contexts: Implications for children's bilingual education. *Front. Psychol.* 14:1125157. doi: 10.3389/fpsyg.2023.1125157

COPYRIGHT
© 2023 Lin, Li, MacLeod and Pollock. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A conceptual model of second language pronunciation in communicative contexts: Implications for children's bilingual education

Youran Lin[1]*, Fangfang Li[2], Andrea A. N. MacLeod[1] and Karen E. Pollock[1]*

[1]Department of Communication Sciences and Disorders, University of Alberta, Edmonton, AB, Canada, [2]Department of Psychology, University of Lethbridge, Lethbridge, AB, Canada

Second language (L2) pronunciation patterns that differ from those of first language (L1) speakers can affect communication effectiveness. Research on children's L2 pronunciation in bilingual education that involves non-English languages is much needed for the field of language acquisition. Due to limited research in these specific populations and languages, researchers often need to refer to literature on L2 pronunciation in general. However, the multidisciplinary literature can be difficult to access. This paper draws on research from different disciplines to provide a brief but holistic overview of L2 pronunciation. A conceptual model of L2 pronunciation is developed to organize multidisciplinary literature, including interlocutors' interactions at three layers: the sociopsychological, acquisitional, and productive-perceptual layers. Narrative literature review method is used to identify themes and gaps in the field. It is suggested that challenges related to L2 pronunciation exist in communication. However, the interlocutors share communication responsibilities and can improve their communicative and cultural competencies. Research gaps are identified and indicate that more studies on child populations and non-English L2s are warranted to advance the field. Furthermore, we advocate for evidence-based education and training programs to improve linguistic and cultural competencies for both L1 speakers and L2 speakers to facilitate intercultural communication.

KEYWORDS

second language, pronunciation, foreign accents, communication, children, bilingual education

1. Introduction

Second language (L2) learners may acquire speech differently than first language (L1) speakers and produce speech with an accent (Munro, 1998). The concept of "foreign accents" is exonormative, as interlocutors look outward for perceived standard or prestige forms (Monfared, 2019). Thus, research on L2 pronunciation has implications for both communication efficacy and perception of identities.

Researchers from diverse disciplines have long been interested in L2 pronunciation (e.g., Lado, 1957; Giles, 1970; Munro and Derwing, 1995; Lippi-Green, 2011; Flege and Bohn, 2021). However, the research issues and approaches are often discipline-specific, which prevents a comprehensive understanding of the field and prevents researchers from studying a topic of

interest in another discipline. Therefore, a new, transdisciplinary perspective that involves psychology, education, and linguistics will advance the field of speech acquisition.

Moreover, child L2 learners have been given less attention in research and practice. In the few discussions about child L2 speech acquisition, the target language was usually English (Derwing, 2020; Levis, 2020). Applying knowledge of L2 pronunciation learning to child learners of non-English is important for bilingual education programs across the world, especially for the ones where at least one of the target language(s) is not English, for example, the international language and indigenous language programs in Canada (Dicks and Genesee, 2017) and the Russian-Hebrew bilingual program in Isreal (Schwartz et al., 2016).

The goal of this paper is to provide a brief but holistic overview of the multidisciplinary literature on L2 pronunciation through a conceptual model and present implications for child bilingual learners of non-English languages. This encompassing model disentangles the interactions between L2 learners and their interlocutors in terms of their sociopsychological characters, linguistic experiences, and speech production and perception. This paper addresses researchers who are interested in pronunciation development in child bilingual speech acquisition. However, the model can be used by researchers of L2 pronunciation in general as a tool to organize their literature and situate their studies, and its implications provide new ideas to not only researchers, but also educators, practitioners, and policymakers.

Given the long-standing history and extensive breadth of the field, a scoping review would be unrealistic. Rather than reducing the scope, a narrative literature review methodology was adopted. The model was developed through extensive reading and discussion. Multidisciplinary literature was mapped onto this model to identify themes and gaps in research. This will point out the main issues of research and raise awareness of future research venues, especially the ones that tend to be neglected at multidisciplinary intersections. This paper will first introduce the three-layer conceptual model, then briefly review L2 speech research within each layer, and finally, present implications for child L2 learners of non-English languages through themes and gaps across the layers.

2. A conceptual model of L2 pronunciation in communicative contexts

Communication involves two or more people who convey and receive information. In the context of L2 oral communication, we will refer to them as the "L2 Learner" and "L1 Listener," as if these were interlocutor roles or names. Such role assignment is oversimplified, as the interlocution is bidirectional, and communication also occurs among L2 speakers (Levis, 2020). However, such simplification allows us to discuss the speakers' speech systems and social cultures and, with cautious comparisons, has the potential to generalize to diverse interlocutor groups. Therefore, we propose a model to understand the interactions between L2 Learners and L1 Listeners at and across three layers: the sociopsychological, acquisitional, and productive-perceptual layers (Figure 1).

The sociopsychological layer focuses on communicators' attitudes toward L2 pronunciation, along with other individual and situational factors. The methods to understand attitudes include observation, interview, survey, and sociopsychological experiments (Giles, 1970; Lippi-Green, 2011). Understanding attitudes toward L2 pronunciation can help communicators become aware of biases and take mutual responsibility for communication (Clark and Wilkes-Gibbs, 1986).

The acquisitional layer addresses the roles of phonological (speech system) experience in pronunciation learning. For L2 Learners, there are a variety of theoretical models that discuss how L1 phonology impacts L2 learning (e.g., Best and Tyler, 2007; Flege and Bohn, 2021). In parallel, for L1 Listeners, the ability to listen to L2 speech is also impacted by their phonological experiences (Hau et al., 2020). Theories and studies in this layer provide frameworks for studies in speech production and perception (Flege et al., 2003) and have pedagogical implications.

The productive-perceptual layer is where L2 Learners and L1 Listeners are directly engaged in a "speech circuit (De Saussure, 1959)" and the characteristics of L2 Learners' pronunciation act on L1 Listeners' perception. Perception and production can be measured subjectively and objectively (Munro and Derwing, 1995; Flege et al., 2003), and their relationships can be identified through statistical analyzes and psychoacoustic experiments (Liu et al., 2014; Porretta et al., 2015). Such research can characterize L2 oral communication and suggest effective targets for pronunciation instruction (Trofimovich and Isaacs, 2012).

In this conceptual model, L2 speech production and perception are impacted by both sociopsychological and acquisitional factors. On the other hand, oral communication in the speech circuit can, in turn, affect interlocutors' sociopsychological characters and linguistic experiences. Therefore, the layers are interrelated and the model does not proceed in a particular order.

3. The sociopsychological layer: Attitudes toward L2 pronunciation

Researchers of child L2 pronunciation should be mindful of the ecological complexity of communication. Many sociopsychological factors at the individual and situational levels contribute to L2 oral communication, including but not limited to personality (Rivers and Ross, 2020), willingness to communicate (Baran-Łucarz, 2014), emotional state (Suzukida, 2021), and cognitive workload (Farris et al., 2008). Moreover, communication is contextualized in a bigger picture of power dynamics and cultural stereotypes. This section will focus on L1 Listeners' and L2 Learners' attitudes toward L2 pronunciation at the individual and group levels.

3.1. L1 Listeners' attitudes toward L2 pronunciation

L2 pronunciation that is accented does not necessarily cause ineffective communication, but it evokes the previously internalized attitudes toward certain groups (Derwing and Munro, 2015). L1 Listeners' negative attitudes toward L2 pronunciation are widely reported (Lippi-Green, 2011). This can be due to (mis)beliefs about their own linguistic status and/or linguistic stereotyping of the L2 pronunciation.

English exemplifies the role of language status. As English is established as an international language, English users are often

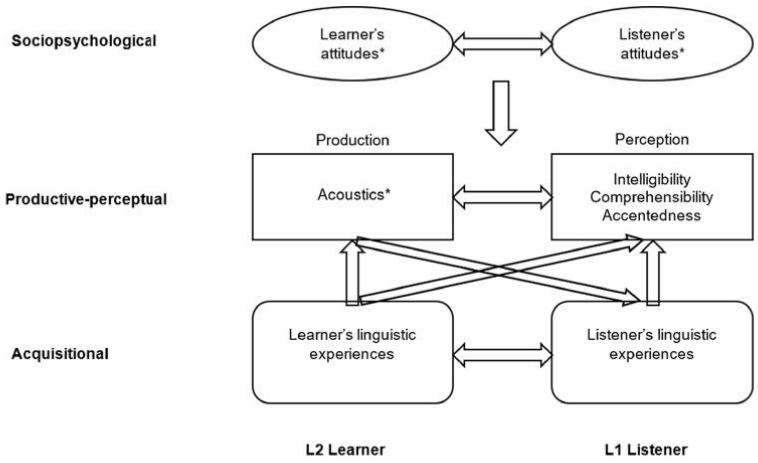


FIGURE 1
A three-layer conceptual model of L2 speech in communicative contexts. * The factor listed is only an example of the multidimensional information in the category.

dichotomized into native and non-native speakers, which gives a higher status to the former. According to a survey (Roper Public Affairs, 2006), 75% of youths in the United States believed that English was the world's most-spoken native tongue, and 38% considered speaking a foreign language "not too important." In relation, the Inner-Circle varieties of English (e.g., American and British English) often enjoy privilege (Jenkins, 2013), although varieties of English are widely used in the Outer Circle (e.g., Singapore, India, Nigeria) as an official language and constitute the countries' multilingualism (Kachru, 1990). The beliefs of language privilege impact the power dynamics between L1 Listeners and L2 Learners.

In addition, L1 Listeners' attitudes toward accents can be related to social stereotypes against certain groups. According to Lippi-Green (2011), for example, attitudes toward French accents are positive for the majority of Americans, while many have negative reactions to Asian accents. Simply due to the stereotypes of how a member of a perceived group should sound, L1 Listeners' speech perception may change, which is referred to as reversed linguistic stereotyping (Kang and Rubin, 2009). For example, Babel and Russell (2015) found that L1 Listeners had more difficulty transcribing the English speech produced by Chinese Canadian speakers when photos of their faces were presented. On the other hand, even before acquiring linguistic stereotypes, preschool-aged children already show selective trust in native-accented informants, which indicates that children are more invested in learning from members of their own cultural groups (Kinzler et al., 2011). This provides insights into how the preference for native accents is formed.

However, negative attitudes do not necessarily lead to communication failures, since their influence is mediated by communication strategies. Lindemann (2002) paired Korean English learners with two groups of English-L1 Listeners who had negative and positive attitudes to Koreans, respectively, for an interactive task. The tasks were completed successfully as long as L1 Listeners actively provided feedback. In contrast, the tasks failed when the L1 Listeners were avoidant, refusing to provide any crucial feedback and completely attributing the communication difficulties to L2 Learners. This suggests that communication can be improved through interventions on communication strategies even when attitudes are not directly addressed.

3.2. L2 Learners' attitudes toward L2 pronunciation

L2 Learners have various attitudes to foreign accents. Some learners, especially those who are in the Expanding Circle (Kachru, 1990) and learn English as a foreign language, admire native speech as the perfect example and ascribe higher status to the Inner-Circle varieties (Carrie, 2017). For example, Japanese and Korean English learners disapproved of their varieties of English and prioritized "nativeness" in their English pronunciation (Tokumoto and Shibata, 2011).

In contrast, L2 Learners accept their accents better when they perceive themselves as users of a legitimate variation of the language (Lippi-Green, 2011), for example, English speakers from the Outer Circle. In comparison to the Japanese and Korean learners aforementioned, Malaysian English learners valued message conveyance more than nativeness (Tokumoto and Shibata, 2011). In addition, L2 Learners may have positive attitudes toward their accent when it marks their identity as desired. For example, among French-English bilinguals in Québec, stronger non-native accents in English were associated with sociopolitical affiliation to the group (Gatbonton and Trofimovich, 2008).

Sociopsychological factors play a role in L2 Learners' pronunciation learning outcomes (Sardegna et al., 2014). Saito et al. (2017) found that L2 Learners who were able to improve their pronunciation over one academic semester tended to show motivations to learn English as a long-term resource. These students

produced L2 speech that was easier for listeners to understand, even though their pronunciation might not be native-like. Nonetheless, the contribution of attitudes and motivations of learning L2 pronunciation should be examined with caution. Sardegna et al. (2018) suggested that although L2 Learners' strong motivation was associated with more efforts to improve L2 pronunciation, it also predicted negative emotions with regard to L2 pronunciation, which might in return barrier their oral communication.

3.3. Summary of the sociopsychological layer

To summarize, a preference for native pronunciation occurs at young ages, and negative attitudes toward non-native pronunciation can impede L2 communication. Researchers advocate for more inclusive attitudes toward L2 pronunciation.

Despite the rich discussion on attitudes toward L2 pronunciation, we identify the languages involved in research as a gap in this layer. Literature is rich in the attitudes toward accented English, but less is known about the attitudes toward L2 pronunciation of non-English languages (e.g., Marx, 2002, an English-L1 learner of German, reflected on their German accent and identity). In some non-English languages, research focuses on accents of native varieties. For example, Chong and Tan (2013) investigated Singaporean Chinese youths' attitudes toward the Beijing, Taiwan, and Singapore varieties of Mandarin; Lindberg and Trofimovich (2020) examined French-L2 Learners' attitudes toward the European and Québec varieties of French. That being said, our research was mainly on the body of Literature that was written in English, which limited our access to literature in other languages. It remains unclear whether our knowledge of attitudes toward accents in English are equally applicable to other languages, as English has the special status of an international language.

Researchers of child L2 pronunciation should understand that even for young learners, their L2 learning and communication are impacted by sociopsychological factors. This becomes especially complex and important when the children's L1 is a high-status language, for example, English-speaking children learning an international language through bilingual education. In addition to the sociopsychological layer, L2 pronunciation learning and communication are also impacted by the specific L1-L2 pair and the interlocutors' experiences in these languages.

4. The acquisitional layer: The impacts of linguistic experiences

Researchers of child L2 pronunciation should understand the mechanisms of learning new phonological systems so they know what learning outcomes to expect given a specific L1-L2 pair. There are several impactful L2 speech acquisition theories, each with its own assumptions and predictions, which can be challenging to access for researchers who are first attempting to tackle L2 pronunciation issues. This section will introduce important theories comparatively to help researchers understand how L2 pronunciation learning is impacted by L2 Learners' phonological experiences. We then argue that L1 Listeners are parallelly biased by their linguistic experiences when communicating with L2 Learners.

In the past 70 years, L2 speech acquisition models have evolved from the Contrastive Analysis Hypothesis (CAH, Lado, 1957) and the Critical Period Hypothesis (CPH, Lenneberg, 1967) to the Perceptual Assimilation Model of L2 Speech Learning (PAM-L2, Best and Tyler, 2007), the Speech Learning Model (SLM-r [revised], Flege and Bohn, 2021), and the Second Language Perception Model (L2LP-r, Van Leussen and Escudero, 2015). In this process, at least three themes have been discussed: (a) the mechanisms of L2 speech acquisition, (b) the roles of non-phonetic information, and (c) the bidirectional interactions between L1 and L2. See Table 1 for a summary of these models and two infant speech development models in comparison, i.e., Native Language Magnet theory [NLM-e (expanded), Kuhl et al., 2008] and Processing Rich Information from Multidimensional Interactive Representations (PRIMIR, Werker and Curtin, 2005).

4.1. The mechanisms of L2 speech acquisition

L1 speech is usually acquired rapidly and effortlessly in infanthood (Kuhl et al., 2008). However, L2 speech learning can be protracted and effortful for L2 Learners who started at an older age, and the speech learning outcome can be inaccurate and accented (Flege, 1995). Theorists discussed the cause of such differences. In the mid-20th century, CPH proposed that young children learn speech through mechanisms that are specialized for language learning, but older learners lose such abilities due to neural maturation (Lenneberg, 1967). On the other hand, CAH regarded language learning as habit formation. When the speech systems are different, L1 habits negatively transfer to L2 learning (Lado, 1957; Wardhaugh, 1970). CAH is powerful in predicting L2 speech learning difficulties by comparing the speech systems of L1 and L2.

However, different speech learning occurred as early as among young school-aged bilingual children (Netelenbos et al., 2016; Yang and Fox, 2017), which challenged the notion of "earlier is better." Since the 1990s, perception-based theories have developed, represented by PAM, SLM, and L2LP. According to these theories, speech learning mechanisms remain unchanged across the lifespan. Nonetheless, L2 Learners' perception is attuned by their L1. This hinders the acquisition of L2 pronunciation but does not completely block it (MacLeod and Stoel-Gammon, 2010). However, these theories have different views on the specific mechanisms of L2 speech perception.

PAM predicts how naïve listeners or new learners perceive a *contrastive pair* (a pair of sound categories that differentiate word meanings) in L2 based on how they are assimilated to L1 categories. For example, if two L2 sounds are perceived as exemplars of two different L1 categories, PAM predicts good discrimination of this pair; but if both sounds are perceived as equally good exemplars of the same L1 category, PAM predicts poor discrimination. To establish a new category in L2, learners need to detect the gestural features of the L2 sounds and contrast them in *minimal pairs* where the sounds differentiate word meanings in the L2 (Best and Tyler, 2007).

One of the PAM's advantages is its specific predictions of challenging targets, which guides research and pedagogical practice. Moreover, it is explanatorily powerful as it has been generalized to suprasegmental elements (So and Best, 2010). However, when

TABLE 1 A summary of the speech development theories reviewed.

	CPH	CAH	PAM	SLM	L2P2	NLM	PRIMIR
Perception/ production	Both	Both	Perception	Both	Perception	Perception	Perception
Learning mechanism	Implicit, language-specific	Behavioristic	Ecological	Psychoacoustic, statistical	Connectionist, statistical	Statistical	Statistical
Identical mechanisms for L1 and L2?	No	Yes	Yes	Yes	Yes	No	Not discussed
Why is L2 speech learning challenging?	Neural maturation	Negative transfer of L1	Perceptual assimilation	Perceptual assimilation; limited input	Weak L2 connections	L1 Neural commitment	Not accessing the acoustic cues
Object of perception	Not discussed	Not discussed	Articulatory gestures	Phonetic distance	Words (meaning errors)	Acoustic information	Multidimensional information
Level of analysis	Not discussed	Phonemic	Phonemic	Allophonic	Multiple levels	Prototypes	Multiple levels
Non-phonetic factors	Not the focus	Not the focus	Not the focus	Age of learning, L2 input	Word recognition	Social interaction	Linguistic and social information
L2-to-L1 effects?	No	No	Not the focus	Yes	Yes, when the layers are situated interactively	Not the focus	Not the focus

applying PAM, researchers should understand that (a) PAM analyzes gestural features that are contrastive, but not the phonetic or allophonic details of speech production (e.g., /t/ in "cat" can be released or unreleased, but such phonetic differences do not change the meaning); and (b) PAM's intent is to account for the perception of new learners instead of experienced learners (Best and Tyler, 2007).

Different from PAM, SLM is interested in the establishment of new phonetic categories in L2, which is based on the phonetic dissimilarity between the L2 sound and its closest L1 counterpart (Flege, 1995). Therefore, SLM's learning objects are sounds instead of sound pairs, and the analysis is phonetic. Flege (1995) stated that L2 speech acquisition was predicted by phonetic dissimilarity and age of onset. In SLM-r (Flege and Bohn, 2021), age of onset was respecified as a macro-variable related to the quantity and quality of L2 speech input. Moreover, SLM initially focused on experienced learners, while SLM-r embraced an unchanged mechanism of speech acquisition: statistical learning (see Kuhl et al., 2008 for statistical learning in infant speech acquisition). Therefore, it aims to account for the full process of L2 speech acquisition.

Despite the evolution of SLM-r, researchers should understand: (a) according to SLM, L2 speech acquisition is impacted by perceived phonetic dissimilarity, while a measurement of such dissimilarity remains undefined (Flege and Bohn, 2021); (b) the quantity and quality of L2 speech input have not been operationalized (although Flege, 2021, proposed a method, it was a self-reported survey that heavily focused on L2 Learners' output instead of input); and (c) SLM discusses pronunciation deviances from the native norm, which is not fully compatible with the focus on intelligibility in L2 pronunciation education (Munro and Derwing, 1995; Levis, 2020).

In contrast to PAM and SLM, L2LP is interested in the *connections* between the acoustic, phonological, and lexical levels (Van Leussen and Escudero, 2015). For new L2 Learners, the acoustic-phonological connection is inherited from L1, so the weak connection in L2 constrains the learner from choosing the appropriate path. As L2 experiences increase, the appropriate L2 connections are strengthened. Meanwhile, the L1-inherited path is weakened whenever a meaning error (misunderstanding in communication) occurs, and as a result, a more plausible path is accessed.

L2LP uses computational models to simulate learning, which allows for quantifiable and testifiable predictions. However, when applying L2LP, some caveats should be considered: (a) several parameters in the computational model are arbitrarily set up, which might not fully represent reality; and (b) the results of simulated learning are not ground truths and need to be tested empirically (Van Leussen and Escudero, 2015).

In summary, perception-based theories argue that L2 speech learning is hindered but not blocked by L1-attuned perception. Through them, researchers can understand how L2 Learners' linguistic experiences impact their L2 pronunciation acquisition and predict specific challenges in learning by examining their L1 and L2 phonological systems.

4.2. The roles of non-phonetic information

Non-phonetic information, such as lexical and social-interactive information, is important in speech acquisition (Clark and Wilkes-Gibbs, 1986; Werker and Curtin, 2005; Kuhl et al., 2008). The following paragraph will compare how perception-based theories consider non-phonetic information in L2 pronunciation acquisition.

PAM focuses on contrastive pairs and involves a lexical perspective by nature. Furthermore, PAM predicts that there is more communicative pressure to learn L2 sound pairs that involve high-frequency words, dense phonological neighborhoods, and/or importance in social communication (Best and Tyler, 2007). However, PAM does not make specific hypotheses about these factors. On the

other hand, SLM-r focuses on the distribution of phonetic information in L2 input (Flege and Bohn, 2021). Such a distributional perspective potentially involves word frequency, learner factors, and social interactions. However, these factors are not yet unpacked in SLM-r. Different from the other two models, L2LP argues that learning is driven by lexical information (Van Leussen and Escudero, 2015). Whenever L2 pronunciation causes a misunderstanding, the L2 Learner will attempt to improve their speech perception until a more plausible path is accessed. This mechanism of using multidimensional information in speech learning is similar to PRIMIR's proposals about infant speech acquisition (Werker and Curtin, 2005).

To summarize, L2 speech acquisition models take linguistic, non-phonetic (e.g., lexical) information into account to different degrees. However, none of them directly addresses the effects of the linguistic-external factors in the sociopsychological layer. Sociopsychological factors such as language status, language attitudes, and motivation play an important role in L2 speech learning and communication (Lindemann, 2002; Meziane and MacLeod, 2017; Saito et al., 2017; Sardegna et al., 2018) and should be further incorporated into the theories mentioned in the acquisitional layers.

4.3. The bidirectional interactions between L1 and L2 speech systems

For L2 Learners, the interaction between two languages is not unidirectional from L1 to L2. Instead, the L2 phonology can also influence their L1. PAM focuses on new L2 Learners and pays limited attention to L2 effects. In contrast, SLM and L2LP discuss L2-to-L1 influences.

SLM has a radical view on L2-to-L1 influence. It believes that L1 and L2 sounds occupy the same phonetic space, therefore L2 effects are immediate and inevitable. When an L2 category is not established, the neighboring L1 categories are assimilated because they are perceptually linked. When an L2 category is established, the L1 categories are dissimilated to maintain phonetic contrast (Flege and Bohn, 2021). Some evidence supports this hypothesis (Flege et al., 2003), but other work shows that L2 effects are more complicated, impacted by language dominance and communicative partners (de Leeuw et al., 2010; Yang and Fox, 2017).

L2LP accounts for such complexity, at least in part, by assuming different models in simulated learners. In a bottom-up model, i.e., when the acoustic, phonological, and lexical strata are separated, L1 phonetic categories are retained. On the contrary, when these aspects are interactive in one stratum, learners will eventually adopt the L2 system and lose the L1 system (Van Leussen and Escudero, 2015). The authors suggest that the bottom-up model resembles adult learning that rarely reaches native-like speech. This implies that the interactive model is in line with younger learners who experience L1 attrition (e.g., Yang and Fox, 2017) and provides an insight that L2-to-L1 effects may be stronger when the L1 phonological representations are not entrenched in young children.

It is clear that L2-to-L1 effects exist and are multifaceted. Empirical evidence shows that L2 can cause both segmental and suprasegmental changes in L1 (e.g., Flege et al., 2003; Bergmann et al., 2015). Research should pay continuous attention to L2-to-L1 influence. This is a particularly relevant real-life issue for bilingual children in immigration contexts as it has implications for L1 attrition.

4.4. L1 Listeners' speech perception

Previous sections introduced how L2 Learners' perception is attuned by their linguistic experiences. Given that the speech learning mechanism, i.e., statistical learning, remains unchanged across lifespan (Flege and Bohn, 2021), we compare L1 Listeners' perception parallelly to L2 Learners'. This means L1 Listeners' perception is also attuned by their L1 phonology and experiences perceptual "learning" when encountering a new speech system, i.e., perceptual adaptation (Hau et al., 2020). L1 Listeners adapt to accented speech rapidly within 1 minute (Clarke and Garrett, 2004) and draw upon non-phonetic information to facilitate understanding (Cooper and Bradlow, 2016). Perceptual adaptation occurs in not only adults but also in school-aged children (Hu, 2021) and generalizes to novel talkers and novel accents (Baese-Berk et al., 2013). Such perceptual learning sets the foundation to train L1 Listeners to understand accented speech. Derwing et al. (2002) found that instructions about the accents of a certain language group not only facilitated a better comprehension but also improved L1 Listeners' attitudes.

4.5. Summary of the acquisitional layer

For L2 Learners, several themes were discussed by L2 pronunciation acquisition models, including the learning mechanisms, the roles of non-phonetic information, and bidirectional interactions between L1 and L2. A few research gaps are identified: First, most theories focused on speech sounds but not suprasegmental features (except for PAM, So and Best, 2010). Second, more evidence in children is needed to account for the full process of L2 speech acquisition indicated by SLM-r (e.g., Netelenbos et al., 2016; Menke, 2017; Meziane and MacLeod, 2017; Nance, 2020). Third, theories should further account for the effects of language-external factors such as social interactions, motivations, and attitudes. For L1 Listeners, research shows that relevant linguistic experiences (i.e., exposure to accented speech) facilitate perceptual adaptation and improve cultural competence. More research is expected to facilitate effective communication on the end of L1 Listeners who have the need to better understand accented speech.

5. The productive-perceptual layer: Perceptual measurements of L2 speech and their acoustic sources

Researchers of child L2 pronunciation should be familiar with the common measurements of L2 pronunciation. This section introduces two types of measurements based on the acoustics of speech production and L1 Listeners' perception, respectively. These two measurements are important because interlocutors' interaction ultimately happens in the "speech circuit (De Saussure, 1959)" when the speech is produced and perceived. It is noteworthy that such interaction is a multimodal phenomenon, where gestures, facial expressions, and environments all play a role. Among them, auditory signals have attracted the most attention, and acoustic measurement is chosen as one method to describe speech production.

5.1. Perceptual measurements of L2 speech

L2 pronunciation used to be perceptually measured by "accuracy" as if it was unidimensional (e.g., Olson and Jay Samuels, 1973; Suter, 1976). Munro and Derwing (1995) divided L1 Listeners' perception of L2 pronunciation into related but distinctive aspects, including intelligibility, comprehensibility, and accentedness. Comprehensibility is defined as the ease of understanding L2 speech, while intelligibility is the extent to which listeners can understand the message. Therefore, comprehensibility is usually rated on a scale, and intelligibility can be calculated through the percentage of words recognized (Fayer and Krasinski, 1987; Munro and Derwing, 1995). In contrast, accentedness is defined as the perceived difference compared with a reference accent and is usually rated on a scale (Southwood and Flege, 1999). By teasing them apart, Munro and Derwing (1995) argued that the goal of L2 pronunciation learning was not reduced accentedness, but increased intelligibility and comprehensibility.

Researchers often consider L1 Listeners as a homogeneous population and measure L2 pronunciation through their perception (Munro and Derwing, 2020). The literature review in the sociopsychological and acquisitional layers suggests that L1 Listeners' perception is biased by their attitudes and linguistic experiences (Kennedy and Trofimovich, 2008; Shintani et al., 2019). Therefore, it is important to be aware of these confounding factors when using perceptual measurements (Lindemann and Subtirelu, 2013). Researchers should choose carefully what speaker information to disclose: One possible option is to conceal identifying information to avoid biases based on linguistic stereotyping. The other is, contrariwise, to incorporate as much information as possible to resemble authentic communicative situations. Moreover, perceptual judgments should be paired with language background questionnaires and attitudinal measurements to account for biases (Dewaele and McCloskey, 2015; Munro and Derwing, 2020). In addition, it is important to use acoustic measurements to validate L1 Listeners' perception and provide phonetic details (Lindemann and Subtirelu, 2013).

5.2. Acoustic cues of L1 Listeners' perception of L2 speech

The source of L1 Listeners' perception of L2 pronunciation is partly contained in the acoustic signals of L2 speech production. It is intuitive to use acoustic measurements to describe L2 pronunciation. However, researchers should be cautious of using acoustic data alone as not all dimensions of acoustic deviances are equally predictive of perceptual differences (Munro and Derwing, 2020). Nonetheless, acoustic measurements can be used in combination with L1 Listeners' perception to validate the latter. In addition, such a combination can identify the acoustic dimensions that are important for intelligibility and, in turn, specify targets for efficient L2 instruction (Schertz and Clare, 2020).

As early as Ryan (1973) called for a production-based measurement of L2 pronunciation. Flege (1984) cross-spliced speech samples of English speakers and French speakers and found that L1 Listeners could detect non-native speech accurately. The study did not measure the acoustics directly, but this was an early experimental attempt to address the relationships between acoustic deviations and listener perceptions. In a later study, Flege et al. (2003) used L1 Listener judgment and acoustic measurements to measure English [eɪ] and Italian [e] produced by Italian-English bilingual speakers. Regressions revealed that the difference between the first and second formants accounted for most of the variation in listener perception. Flege et al. (2003) was different from Flege (1984) in that it took multiple acoustic measurements and explored acoustic variables' effects on perception.

Similar studies were expanded to more speech features in a variety of languages. Some showed that the perception was mainly impacted by spectral features (e.g., Wayland, 1997), while others suggested that temporal features played a role (e.g., Porretta et al., 2015; see Derwing and Munro, 2015 for a summary). However, in such studies, researchers could not establish causal relationships between acoustic deviances and perception or guarantee that the acoustic variables included were exhaustive (Porretta et al., 2015). This pointed to two directions of studies: (a) manipulated acoustic properties to establish causality and (b) a more extensive list of acoustic measurements.

Acoustic properties can be synthetically manipulated to verify causal relationships between acoustics and perception. For example, Liu et al. (2014) observed that L2 Learners might use duration as a cue to differentiate lax and tense vowels in production. To prove this hypothesis, they equalized the duration of L2 Learners' productions to find that intelligibility was reduced. In contrast with how Liu et al. (2014) removed one dimension of acoustic variance, acoustic cues can be varied to form a continuum. Chan et al. (2017) manipulated spectral features gradually and found that the frequencies of vowel formants were a primary cue for the perception of L2 speech.

On the other hand, recent studies included larger sets of acoustic measurements. Idemaru et al. (2019) examined the impacts of vowel, consonant, rhythm, pitch, and fluency properties in Japanese-L2 Learners' productions. Pitch errors were most predictive of accentedness for both English-and Mandarin-L1 learners of Japanese. L1-specific patterns were further identified. For example, vowel properties were predictive of English-L1 Learners' accentedness perceived by Japanese L1 Listeners, while consonant properties were predictive of Mandarin-L1 Learners' perceived accentedness. The large inventory of acoustic measurements provides a foundation to compare learners from a variety of language backgrounds and to explore the crucial acoustic factors for a specific pair of L1 and L2.

5.3. Summary of the productive-perceptual layer

In summary, L2 Learners' production and L1 Listeners' perception are the two ends of the speech circuit. Researchers use them to measure L2 pronunciation and examine the relationships between these two types of measurements. Such research attempts to validate the perceptual measurements, rank the gravity of acoustic deviances, and ultimately facilitate effective L2 pronunciation learning. Therefore, productive-perceptual studies have implications for speech acquisition in L2 pedagogy.

A few future directions that already emerged can be further explored in this layer. First, research generalizability in different languages should be considered. On one hand, more productive-perceptual studies in non-English languages are needed. Such studies may provide insights into the universality and uniqueness of acoustic correlates of L1 Listeners' perception in different languages and guide pronunciation instruction in these languages (Porretta et al., 2015). On the other hand, a more thorough list of speech features can

be developed with the potential to be used in any given L1-L2 pair (Idemaru et al., 2019).

Second, discourse studies are warranted to mimic more realistic communicative situations. The early productive-perceptual studies elicited single words (e.g., Flege et al., 2003), which could not provide a valid evaluation of L2 Learners' speech. In recent studies, learners were prompted to produce sentences (e.g., Idemaru et al., 2019). However, these studies are still limited to laboratory environments. Future studies can look into acoustic and perceptual measurements of conversational speech and examine the impacts of linguistic and sociopsychological information to increase ecological validity and better represent real-life communication.

6. Synthesis across layers

We have proposed a three-layer conceptual model of research on L2 pronunciation in communicative contexts between L2 Learners and L1 Listeners, which includes sociopsychological, acquisitional, and productive-perceptual layers. Through a narrative literature review, we mapped existing research onto the model and identified research themes and future directions within each layer. Here we will discuss the interconnections across layers and some forward-looking ideas for children's pronunciation acquisition of a non-English L2.

6.1. Interconnections between the layers

The layers of the model are interconnected, therefore the model does not proceed in a certain order. In the sociopsychological layer, both L1 Listeners and L2 Learners may have negative attitudes toward L2 speech. The attitudes can interact with the productive-perceptual layer. An example is reversed linguistic stereotyping (Kang and Rubin, 2009), where L1 Listeners experience perceptual difficulties solely due to the perceived group membership of the speaker.

In the acquisitional layer, L2 Learners' perception is L1-attuned (MacLeod and Stoel-Gammon, 2010). Parallelly, L1 Listeners' perception is also attuned by their L1, and they experience perceptual "learning" (adaptation) when exposed to L2 speech (Hau et al., 2020). Linguistic experiences of L2 speech can improve L1 Listeners' knowledge of L2 pronunciation and improve intelligibility in the productive-perceptual layer (Kennedy and Trofimovich, 2008). Furthermore, knowledge and experiences of L2 speech improve L1 Listeners' attitudes toward accented speech in the sociopsychological layer (Derwing et al., 2002).

In the productive-perceptual layer, perceptual and acoustic measurements of L2 pronunciation are also interconnected with the other two layers. As for perceptual measurements, L1 Listeners' attitudes and perceptual adaptation may confound their perception. As for acoustic measurements, the acoustic features were usually chosen based on language-specific comparisons, guided by the theoretical models in the acquisitional layer.

6.2. A theme across layers and the need for intervention

The common theme across layers can be summarized as follows: L2 Learners are often faced with difficulties in L2 communication, but both L1 Listeners and L2 Learners can share a mutual responsibility to improve communication effectiveness (Clark and Wilkes-Gibbs, 1986). L2 pronunciation itself is not the cause of difficulties in communication, but the difficulties related to it should not be downplayed or ignored. L2 Learners are faced with real difficulties: Their perception has been attuned by their L1, which causes difficulties learning the new phonological system. In addition, L2 oral communication is affected by negative attitudes of both L1 Listeners and L2 Learners. To address these issues, interlocutors should share the mutual responsibility of communication and be supported to improve communicative skills.

For L1 Listeners, limited listening skills and prejudicial attitudes can cause hardship in communication. This can be addressed by improving perceptual adaptation and cultural competence (Derwing et al., 2002). Proposals to mitigate L1 Listeners' attitudes and listening skills have been questioned, with a hesitation rooted in the belief that interventions aimed at L1 Listeners are too effortful and unfeasible, and that L2 oral communication is primarily a problem for L2 Learners.

However, perceptual adaptation to L2 speech can happen rapidly in both adults and children, and the learning outcomes can generalize to other accents (Clarke and Garrett, 2004; Baese-Berk et al., 2013; Hu, 2021). In addition, L1 Listeners' negative attitudes can be confronted and improved through training sessions (Kang et al., 2015), and such improvements can result in enhanced perception of L2 speech (Cooper et al., 2020). Therefore, interventions that aim to address L1 Listeners' attitudes and perception are feasible, and they are necessary at least for the groups that need to communicate with L2 Learners frequently, for example, educators, university students, healthcare providers, and public servants. Subtirelu and Lindemann (2016) proposed three aspects of L1-Listener interventions: (a) improving attitudes, (b) familiarizing with L2 pronunciations, and (c) developing communicative strategies. Future research can refer to these principles in their intervention designs.

Similarly, L2 Learners' speech proficiency and cultural competence can be improved to facilitate effective communication. In terms of speech proficiency, L2 speech acquisition is a dynamic process, and the outcomes can be improved as the L2 speech input quantity and quality increase (Flege and Bohn, 2021). Moreover, researchers investigated the acoustic cues of perceived unintelligibility (e.g., Idemaru et al., 2019), which can be translated into pedagogical targets in L2 pronunciation teaching and learning. In terms of attitudes, L2 Learners' attitudes are closely related to the language ideologies in their L2 classrooms. Unfortunately, L2 speech education often serves to ossify negative attitudes toward foreign accents (Lippi-Green, 2011). Negative feelings toward certain accents were reported among L2 teachers (Munro et al., 2006). Meanwhile, the teaching model still tends to be exonormative, i.e., British and American Englishes are often positioned as a standard (Monfared, 2019).

Fortunately, on the other hand, intervention programs have been designed in teacher education and English-L2 classes to mitigate language attitudes. For example, preservice English teachers' attitudes improved after being exposed to diverse Englishes and practicing self-reflection (Ates et al., 2015). For L2 Learners, Korean university students participated in an extracurricular project to interview diverse English users (Lee, 2019). Students reported that the lack of exposure to diverse Englishes caused their preference for American English, while the authentic communicative situations brought attitudinal changes. Different from this project-based design, pedagogies in a

university in China designed a structured program on language attitudes, including four steps: eliciting attitudes, deconstructing stereotypes, reconstructing open attitudes, and developing solutions to communication problems (Zheng and Gao, 2017). Almost half of the students embraced the concept of World Englishes after the intervention, while others remained ambivalent or conservative, indicating the necessity of continuous efforts and authentic communicative experiences to alternate the entrenched attitudes. Comparing these projects with Subtirelu and Lindemann's (2016) proposal aforementioned for L1-Listener intervention, it seems that the L2-Learner intervention should also include at least three aspects: (a) reconstructing attitudes, (b) familiarizing with a variety of pronunciations in the target L2, and (c) developing communicative strategies.

Synthesizing the evidence, the interactions between Educators, L2 Learners, and L1 Listeners are illustrated in Figure 2. In a vicious circle, L2 Learners form negative attitudes toward foreign accents in the classroom, feel anxious during the communication with the L1 Listener, and are frustrated by L1 Listeners' avoidant behaviors. On the other hand, when Educators foster open attitudes toward L2 pronunciation, L2 Learners feel prepared with improved pronunciation and communicative skills, and L1 Listeners are ready to adapt to L2 pronunciation, a virtuous circle can occur in L2 communication.

6.3. Research gaps and implications for child bilingual education

Research has addressed most of the layers in the model and their interconnections. However, several cross-layer gaps can be further considered to advance the field of L2 speech acquisition.

First, the vicious/virtuous circle that involves Educators, L2 Learners, and L1 Listeners (Figure 2) can and should be addressed with interventions. Recent efforts have been made to improve the cultural competence of L1 Listeners and L2 Learners, but effectiveness studies are warranted to understand what program designs are of merit. Moreover, most of the intervention programs are aimed at adolescents or adults, while stereotypes against L2 pronunciation can occur in preschool-aged children (Kinzler et al., 2011). Therefore, it remains unclear whether it is necessary and feasible to intervene in language attitudes at a younger age, especially for immigration children and their peers. Evidence is needed on whether and how bilingual education plays a role in dismissing linguistic stereotypes. Qualitative evidence shows that bilingual education in a minority language empowers students through cultural confirmation, nourishing positive self-identity, and encouraging transculturation (e.g., Wu, 2005). Little is known about how such cultural competency translates to positive attitudes toward diverse pronunciation.

Second, more research in non-English languages is needed. In this paper, we tried to include evidence from other languages (e.g., Chong and Tan, 2013; Idemaru et al., 2019; Lindberg and Trofimovich, 2020), but our access to literature in different languages was limited. However, with the dominant position of English, it is not surprising that most of the research on L2 pronunciation focused on English as the target L2 (Derwing and Munro, 2015). The issues of pronunciation in English are relevant to other languages (Levis, 2020), but learners' motivation and speech input can be different when they are English speakers learning a non-English language. Therefore, studying L2 speech in non-English languages can help understand the generalizability of research, identify different perspectives on pronunciation in different cultures, and help the learners improve their oral communication.

Third, compared with the rich literature on adult L2 pronunciation acquisition, less attention is given to child learners. Derwing (2020) pointed out that this is in part because child L2 Learners' pronunciation is usually thought to be native-like or, at least, intelligible. They discussed the L2 pronunciation difficulties in immigrant children and methods to facilitate their pronunciation learning of the societal majority language. However, little is known about how children learn the pronunciation of a minority language. For example, children who learn French as an L2 in Canada through

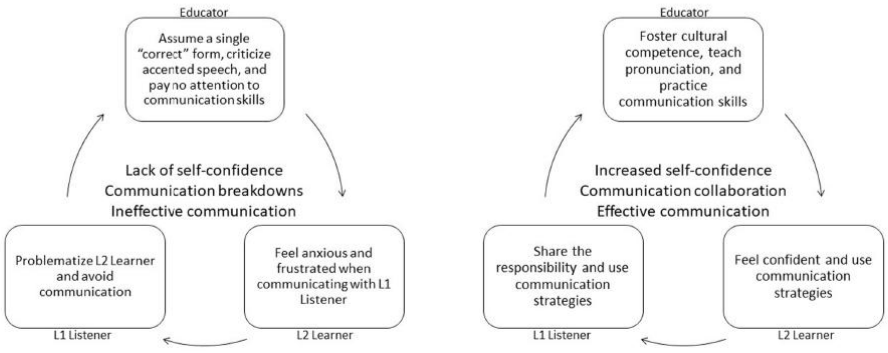


FIGURE 2
Attitudinal aspects of L2 speech in communication: vicious and virtuous circles.

immersion education showed non-native-like patterns in their consonants (Netelenbos et al., 2016), but in a Spanish-English bilingual school in the States and a Gaelic-medium school in Scotland, children's pronunciation converged despite whether they were exposed to the minority language at home or not (Menke, 2017; Nance, 2020). It seems that the high-quality interaction with native-or heritage-speaking peers played a role in the pronunciation acquisition of a minority L2. To verify this observation and understand other learning factors, we advocate for more research that focuses on the L2 pronunciation acquisition of children who are learning a minority language of the society, in addition to the immigration children who are learning the majority languages.

7. Conclusion

Despite the limitation that a review paper cannot comprehensively cover the literature across multiple disciplines and a long history, this paper provides a narrative review on L2 pronunciation that focuses on the L1 Listener and L2 Learner's interactions at the sociopsychological, acquisitional, and productive-perceptual layers. Through this review, we propose several "new ideas" for the field of language acquisition. First, we recognize that researchers in the field of L2 pronunciation acquisition often need to conduct transdisciplinary research. Therefore, a three-layer conceptual model is used to introduce the existing literature from multiple disciplines and can also be used by other researchers to organize literature during their transdisciplinary research. Moreover, we argue that it is important for future research to emphasize mutual communicative responsibility and investigate interventions for both L2 Learners and L1 Listeners to address their linguistic experiences, cultural competence, and communication strategies. Different from the unilateral effort to improve L2 Learners' pronunciation, we believe such interventions are feasible and necessary for people who need to communicate with L2 Learners frequently. Most importantly, we highlight a population which has been understudied in the field: child bilingual learners of non-English languages. Previous research, even though focused on different populations or languages, provided guidance for researchers to examine child interlocutors' attitudes to L2 pronunciation and acquisition, their phonological transfer and adaptation in a variety of L1 and L2 combinations, and their production and perception of L2 pronunciation. In the future, more studies are needed on non-English languages and the child population in the context of continued globalization and thriving bilingual education. By discussing these themes and gaps, we hope to raise awareness among not only researchers who are interested in language acquisition, but also educators, practitioners, and policymakers to better facilitate children's pronunciation learning and bilingual communication.

Author contributions

YL proposed the conceptual model through literature review and wrote the first draft of the manuscript. All authors contributed to the manuscript revision, read and approved the submitted version, and discussion sessions of the literature and model revision.

Funding

This work was supported by the SSHRC Insight Grant (435–2017-1086) and the Vanier Canada Graduate Scholarship through SSHRC (CGV–163274). The SSHRC Insight Grant funded a series of studies that focused on speech production in children enrolled in second-language education programs, which involved all the authors. The Vanier scholarship funded the YL's doctoral studies.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ates, B., Eslami, Z. R., and Wright, K. L. (2015). Incorporating world Englishes into undergraduate ESL education courses. *World Engl.* 34, 485–501. doi: 10.1111/weng.12149

Babel, M., and Russell, J. (2015). Expectations and speech intelligibility. *J. Acoust. Soc. Am.* 137, 2823–2833. doi: 10.1121/1.4919317

Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *J. Acoust. Soc. Am.* 133, EL174–EL180. doi: 10.1121/1.4789864

Baran-Łucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: the polish EFL context. *Can. Mod. Lang. Rev.* 70, 445–473. doi: 10.3138/cmlr.2666

Bergmann, C., Sprenger, S. A., and Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychol.* 161, 25–35. doi: 10.1016/j.actpsy.2015.07.015

Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities" in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. eds. M. J. Munro and O.-S. Bohn (Amsterdam: John Benjamins), 13–34.

Carrie, E. (2017). 'British is professional, American is urban': attitudes towards English reference accents in Spain. *Int. J. Appl. Linguist.* 27, 427–447. doi: 10.1111/ijal.12139

Chan, K. Y., Hall, M. D., and Assgari, A. A. (2017). The role of vowel formant frequencies and duration in the perception of foreign accent. *J. Cogn. Psychol.* 29, 23–34. doi: 10.1080/20445911.2016.1170746

Chong, R. H. H., and Tan, Y. Y. (2013). Attitudes toward accents of mandarin in Singapore. *Chin. Lang. Discourse* 4, 120–140. doi: 10.1075/cld.4.1.04cho

Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi: 10.1016/0010-0277(86)90010-7

Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131

Cooper, A., and Bradlow, A. R. (2016). Linguistically guided adaptation to foreign-accented speech. *J. Acoust. Soc. Am.* 140:EL378–EL384. doi: 10.1121/1.4966585

Cooper, B., Payne, G., Hu, X., Dixon, Q., and Kuo, L. (2020). "The impact of linguistic diversity education on L1 English speakers' ideologies, attitudes, and perceptions of international teaching assistants" in *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference*. eds. O. Kang, S. Staples, K. Yaw and K. Hirschi (Ames, IA: Iowa State University), 2380–9566.

de Leeuw, E., Schmid, M. S., and Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Biling. Lang. Congn.* 13, 33–40. doi: 10.1017/S1366728909990289

De Saussure, F. (1959). *Course in General Linguistics*. New York: Philosophical Library.

Derwing, T. M. (2020). "Issues for second language pronunciation in children" in *Child Bilingualism and Second Language Learning: Multidisciplinary Perspectives*. eds. F. Li, K. E. Pollock and R. Gibb (Amsterdam, Netherlands: John Benjamins Publishing Company), 29–44.

Derwing, T. M., and Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins Publishing.

Derwing, T. M., Rossiter, M. J., and Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *J. Multiling. Multicult. Dev.* 23, 245–259. doi: 10.1080/01434630208666468

Dewaele, J. M., and McCloskey, J. (2015). Attitudes towards foreign accents among adult multilingual language users. *J. Multiling. Multicult. Dev.* 36, 221–238. doi: 10.1080/01434632.2014.909445

Dicks, J., and Genesee, F. (2017). "Bilingual education in Canada" in *Bilingual and Multilingual Education*. eds. O. García, A. Lin and S. May (Basel: Springer), 453–467.

Farris, C., Trofimovich, P., Segalowitz, N., and Gatbonton, E. (2008). Air traffic communication in a second language: implications of cognitive factors for training and assessment. *TESOL Q.* 42, 397–410. doi: 10.1002/j.1545-7249.2008.tb00138.x

Fayer, J. M., and Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Lang. Learn.* 37, 313–326. doi: 10.1111/j.1467-1770.1987.tb00573.x

Flege, J. E. (1984). The detection of French accent by American listeners. *J. Acoust. Soc. Am.* 76, 692–707. doi: 10.1121/1.391256

Flege, J. E. (1995). "Second language speech learning: theory, findings, and problems" in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. ed. W. Strange (Timonium, MD: York Press), 233–277.

Flege, J. E. (2021). "New methods for second language (L2) speech research" in *Second Language Speech Learning: Theoretical and Empirical Progress*. ed. R. Wayland (Cambridge, UK: Cambridge University Press), 119–156.

Flege, J. E., and Bohn, O. S. (2021). "The revised speech learning model (SLM-r)" in *Second Language Speech Learning: Theoretical and Empirical Progress*. ed. R. Wayland (Cambridge, UK: Cambridge University Press), 3–83.

Flege, J. E., Schirru, C., and MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Comm.* 40, 467–491. doi: 10.1016/S0167-6393(02)00128-0

Gatbonton, E., and Trofimovich, P. (2008). The ethnic group affiliation and L2 proficiency link: empirical evidence. *Lang. Aware.* 17, 229–248. doi: 10.1080/09658410802146867

Giles, H. (1970). Evaluative reactions to accents. *Educ. Rev.* 22, 211–227. doi: 10.1080/0013191700220301

Hau, J. A., Holt, C. M., Finch, S., and Dowell, R. C. (2020). The adaptation to mandarin-accented English by older, hearing-impaired listeners following brief exposure to the accent. *J. Speech Lang. Hear. Res.* 63, 858–871. doi: 10.1044/2019_JSLHR-19-00136

Hu, C. F. (2021). Adaptation to an unfamiliar accent by child L2 listeners. *Lang. Speech* 64, 491–514. doi: 10.1177/0023830920936321

Idemaru, K., Wei, P., and Gubbins, L. (2019). Acoustic sources of accent in second language Japanese speech. *Lang. Speech* 62, 333–357. doi: 10.1177/0023830918773118

Jenkins, J. (2013). *English as a Lingua Franca in the International University: The Politics of Academic English Language Policy*. Abingdon: Routledge.

Kachru, B. B. (1990). World Englishes and applied linguistics. *World Engl.* 9, 3–20. doi: 10.1111/j.1467-971X.1990.tb00683.x

Kang, O., and Rubin, D. L. (2009). Reverse linguistic stereotyping: measuring the effect of listener expectations on speech evaluation. *J. Lang. Soc. Psychol.* 28, 441–456. doi: 10.1177/0261927X09341950

Kang, O., Rubin, D., and Lindemann, S. (2015). Mitigating US undergraduates' attitudes toward international teaching assistants. *TESOL Q.* 49, 681–706. doi: 10.1002/tesq.192

Kennedy, S., and Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: the role of listener experience and semantic context. *Can. Mod. Lang. Rev.* 64, 459–489. doi: 10.3138/cmlr.64.3.459

Kinzler, K. D., Corriveau, K. H., and Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Dev. Sci.* 14, 106–111. doi: 10.1111/j.1467-7687.2010.00965.x

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). *Philos. Trans. R. Soc. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, Michigan: University of Michigan Press.

Lee, J. S. (2019). Teacher as change agent: attitude change toward varieties of English through teaching English as an international language. *Asian Engl.* 21, 87–102. doi: 10.1080/13488678.2018.1434396

Lenneberg, E. H. (1967). The biological foundations of language. *Hosp. Pract.* 2, 59–67. doi: 10.1080/21548331.1967.11707799

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *J. Second Lang. Pronunciation* 6, 310–328. doi: 10.1075/jslp.20050.lev

Lindberg, R., and Trofimovich, P. (2020). Second language learners' attitudes toward French varieties: The roles of learning experience and social networks. *Mod. Lang. J.* 822, 841–594. doi: 10.1111/modl.12674

Lindemann, S. (2002). Listening with an attitude: a model of native-speaker comprehension of non-native speakers in the United States. *Lang. Soc.* 31, 419–441. doi: 10.1017/S0047404502020286

Lindemann, S., and Subtirelu, N. (2013). Reliably biased: the role of listener expectation in the perception of second language speech. *Lang. Learn.* 63, 567–594. doi: 10.1111/lang.12014

Lippi-Green, R. (2011). *English with an Accent: Language, Ideology, and Discrimination in the United States. 2nd*. Abingdon: Routledge.

Liu, C., Jin, S. H., and Chen, C. T. (2014). Durations of American English vowels by native and non-native speakers: acoustic analyses and perceptual effects. *Lang. Speech* 57, 238–253. doi: 10.1177/0023830913507692

MacLeod, A. A., and Stoel-Gammon, C. (2010). What is the impact of age of second language acquisition on the production of consonants and vowels among childhood bilinguals? *Int. J. Biling.* 14, 400–421. doi: 10.1177/1367006910370918

Marx, N. (2002). Never quite a 'native speaker': accent and identity in the L2-and the L1. *Can. Mod. Lang. Rev.* 59, 264–281. doi: 10.3138/cmlr.59.2.264

Menke, M. R. (2017). Phonological development in two-way bilingual immersion: the case of Spanish vowels. *J. Second Lang. Pronunciation* 3, 80–108. doi: 10.1075/jslp.3.1.04men

Meziane, R. S., and MacLeod, A. A. (2017). L'acquisition de la phonologie en français langue seconde: le profil phonologique d'enfants allophones en maternelle. *Can. J. Appl. Linguist.* 20, 1–17. doi: 10.7202/1042673ar

Monfared, A. (2019). Ownership of English in the outer and expanding circles: teachers' attitudes toward pronunciation in ESL/EFL teaching contexts. *Asian Engl.* 21, 207–222. doi: 10.1080/13488678.2018.1554303

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Stud. Second Lang. Learn.* 20, 139–154. doi: 10.1017/S0272263198002022

Munro, M. J., and Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Lang. Learn.* 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., and Derwing, T. M.. (2020). Collecting Data in L2 Pronunciation Research. In *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference*, pp. 2380–9566.

Munro, M. J., Derwing, T. M., and Sato, K. (2006). Salient accents, covert attitudes: consciousness-raising for pre-service second language teachers. *Prospect* 21, 67–79.

Nance, C. (2020). Bilingual language exposure and the peer group: acquiring phonetics and phonology in Gaelic medium education. *Int. J. Biling.* 24, 360–375. doi: 10.1177/1367006919826872

Netelenbos, N., Li, F., and Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: a study of voice onset time. *Int. J. Biling.* 20, 346–357. doi: 10.1177/1367006914564566

Olson, L. L., and Jay Samuels, S. (1973). The relationship between age and accuracy of foreign language pronunciation. *J. Educ. Res.* 66, 263–268. doi: 10.1080/00220671.1973.10884476

Porretta, V., Kyröläinen, A. J., and Tucker, B. V. (2015). Perceived foreign accentedness: acoustic distances and lexical properties. *Atten. Percept. Psychophys.* 77, 2438–2451. doi: 10.3758/s13414-015-0916-3

Rivers, D. J., and Ross, A. S. (2020). L1/L2 communication self-efficacy beliefs and the contribution of personality. *Lang. Learn. J.* 48, 700–714. doi: 10.1080/09571736.2018.1441895

Roper Public Affairs. (2006). *National Geographic-Roper Public Affairs 2006 Geographic Literacy Study*. National Geographic Foundation. Available at: https://media.nationalgeographic.org/assets/file/NGS-Roper-2006-Report.pdf

Ryan, E. B. (1973). "Subjective Reactions Toward Accented Speech" in *Language Attitudes: Current Trends and Prospects*. eds. R. W. Shuy and R. W. Fasold (Washington, DC: Georgetown University Press), 60–73.

Saito, K., Dewaele, J. M., and Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Lang. Speech* 60, 614–632. doi: 10.1177/0023830916687793

Sardegna, V. G., Lee, J., and Kusey, C. (2014). Development and validation of the learner attitudes and motivations for pronunciation (LAMP) inventory. *System* 47, 162–175. doi: 10.1016/j.system.2014.10.009

Sardegna, V. G., Lee, J., and Kusey, C. (2018). Self-efficacy, attitudes, and choice of strategies for English pronunciation learning. *Lang. Learn.* 68, 83–114.

Schertz, J., and Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdiscip. Rev. Cogn. Sci.* 11:e1521. doi: 10.1002/wcs.1521

Schwartz, M., Wee Koh, P., Xi Chen, B., Sinke, M., and Geva, E. (2016). Through the lens of teachers in two bilingual programmes: a look at early bilingual education. *Lang. Cult. Curric.* 29, 141–168. doi: 10.1080/07908318.2015.1103250

Shintani, N., Saito, K., and Koizumi, R. (2019). The relationship between multilingual raters' language background and their perceptions of accentedness and comprehensibility of second language speech. *Int. J. Biling. Educ. Biling.* 22, 849–869. doi: 10.1080/13670050.2017.1320967

So, C. K., and Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: effects of native phonological and phonetic influences. *Lang. Speech* 53, 273–293. doi: 10.1177/0023830909357156

Southwood, M. H., and Flege, J. E. (1999). Scaling foreign accent: direct magnitude estimation versus interval scaling. *Clin. Linguist. Phon.* 13, 335–349. doi: 10.1080/026992099299013

Subtirelu, N. C., and Lindemann, S. (2016). Teaching first language speakers to communicate across linguistic difference: addressing attitudes, comprehension, and strategies. *Appl. Linguis.* 37, 765–783. doi: 10.1093/applin/amu068

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Lang. Learn.* 26, 233–253. doi: 10.1111/j.1467-1770.1976.tb00275.x

Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: insights from research and pedagogical implications. *RELC J.* 52, 48–61. doi: 10.1177/0033688220987655

Tokumoto, M., and Shibata, M. (2011). Asian varieties of English: attitudes towards pronunciation. *World Engl.* 30, 392–408. doi: 10.1111/j.1467-971X.2011.01710.x

Trofimovich, P., and Isaacs, T. (2012). Disentangling accent from comprehensibility. *Biling. Lang. Congn.* 15, 905–916. doi: 10.1017/S1366728912000168

Van Leussen, J. W., and Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Front. Psychol.* 6:1000. doi: 10.3389/fpsyg.2015.01000

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Q.* 4, 123–130. doi: 10.2307/3586182

Wayland, R. (1997). Non-native production of Thai: acoustic measurements and accentedness ratings. *Appl. Linguis.* 18, 345–373. doi: 10.1093/applin/18.3.345

Werker, J. F., and Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Lang. Learn. Dev.* 1, 197–234. doi: 10.1080/15475441.2005.9684216

Wu, J. Z. (2005). Bilingual education in western Canada and Chinese language minority students' self perceptions of their citizenship and ethnicity. *Can. Int. Educ.* 34, 23–30. doi: 10.5206/cie-eci.v34i1.9051

Yang, J., and Fox, R. A. (2017). L1–L2 interactions of vowel systems in young bilingual mandarin-English children. *J. Phon.* 65, 60–76. doi: 10.1016/j.wocn.2017.06.002

Zheng, X., and Gao, Y. (2017). Language attitude education as ICC facilitation: an explorative class. *Chin. J. Appl. Linguist.* 40, 3–20. doi: 10.1515/cjal-2017-0001

Chapter 2 sets the stage upon which the rest of this dissertation is built. First, the model points out factors at different layers that may influence bilingual students' speech development, including the social environment, learners' motivations and attitudes, L1-L2 relationships, and the phonetic details of the speech targets in L2. It stresses the indispensable role of the education system in L2 pronunciation learning (Figure 2.2). Therefore, before presenting any qualitative or quantitative evidence, Chapter 3 explores the social and educational context of the Chinese bilingual program and discusses the expected speech input and pronunciation instruction according to the program and curriculum design.

Second, the model views L2 pronunciation learning as a dynamic process that involves not only learners' speech perception and production but also their sociopsychological aspects and their social interactions with other native and non-native speakers. In this learning process, teachers play an important role in the bilingual education system (Figure 2.2). Teachers provide explicit instructions to improve bilingual students' pronunciation learning outcomes (Dicks & Genesee, 2017), promote students' interactions with teachers and peers in the target language, and participate in shaping students' language learning motivation and language ideologies (Lippi-Green, 2011). The study in Chapter 4 (Paper 2) qualitatively presents Chinese teachers' perspectives on how they use strategies to deliver instruction in Mandarin pronunciation and promote the use of Mandarin in meaningful contexts.

Third, in the acquisitional layer of the model (Figure 2.1), several theoretical frameworks for pronunciation learning are introduced (e.g., L2LP-r, van Leussen & Escudero, 2015; PAM, Best, 1995; Best & Tyler, 2007; SLM-r, Flege & Bohn, 2021). These theories can be used to examine the relationship between L1 and L2 phonological systems and further explain and predict pronunciation learning outcomes. In the next two studies, I use these theories to

understand the factors that impact bilingual students' pronunciation learning outcomes. In Chapter 4 (Paper 2), teachers reflect on the challenging speech elements for bilingual students and the individual and social factors that influence students' pronunciation learning, and I use the theories in Chapter 2 to interpret teachers' reflections. In Chapter 6 (Paper 3), I present a quantitative study on bilingual students' production of Mandarin lexical tones, using the theories in Chapter 2 to motivate my hypotheses and explain the results.

Fourth, in the productive-perceptual layer of the model (Figure 2.1), Learners' production, especially its acoustic characteristics, are linked to listeners' perception. Chapter 2 states that it is meaningful to measure learners' speech production through native listeners' perception (Munro & Derwing, 2020), but meanwhile, it is also important to use objective measurements to validate listener perception (Lindemann & Subtirelu, 2013). Therefore, the study in Chapter 6 (Paper 3) integrates perception-based (transcription) measurements and acoustic measurements. In other words, Chapter 6 not only focuses on the categorical accuracy of students' tone productions but also investigates how students utilize the phonetic cues that can facilitate listeners' perception.

The next chapter, Chapter 3, is an overview of the Edmonton Chinese bilingual program in the context of Canada's multiculturalism and bilingual education. It especially focuses on the speech system of the minority language, Mandarin Chinese, and the type of Mandarin speech input and pronunciation instruction in the program. This chapter aims to provide important background information for readers who are unfamiliar with bilingual education, the Chinese bilingual program, or Mandarin. Although the overview in Chapter 3 is specific to the Chinese bilingual program in Edmonton, it has implications for bilingual programs in general.

Chapter 3 The Two-Way Chinese-English Bilingual Program

The studies in this dissertation took place in the Chinese-English two-way bilingual education program in Edmonton. In order to set up the social and educational background of this dissertation, Chapter 3 overviews Canada's immersive bilingual education and introduces the Chinese bilingual program of interest. Subsequently, it discusses what type of Mandarin speech input is expected according to the program and curriculum design. This provides background information for the following research studies.

Canada's bilingual education and the Chinese bilingual program

"Bilingual education" is an umbrella term that describes many different forms of education (Baker, 2007). The "bilingual education" referred to in this dissertation is different from traditional L2 education and transitional education (Baker, 2007). In an L2 education program, a minority L2 is taught as a separate subject. It provides limited input in the L2 and is not effective for the students to reach functional L2 proficiency (Genesee, 2004). In transitional education, immigrant students receive instruction in their minority L1 and transition to exclusive immersion in the majority language as soon as they have developed sufficient L2 proficiency (Cummins, 1979). The purpose of such programs is to mainstream the students with no continuous efforts on L1 maintenance (Cummins, 1979, 1981).

Stronger forms of bilingual education (Baker, 2007) include immersion in a minority language for language-minority students, language-majority students, or both, i.e., heritage language education, (one-way) immersive education, and two-way bilingual education, respectively. Heritage language education affirms language-minority students' right to access education through L1. However, its scope is limited to language-minority students, which to

some extent indicates that heritage language maintenance is an extra for the minority groups instead of a resource for the whole society (Ruíz, 1984).

A different educational design was initiated in Canada for language-majority students to be immersed in a minority language. In the 1960s, Canada established official bilingualism (Dicks & Genesee, 2017). In 1965, English-speaking parents and researchers in Québec initiated the famous St. Lambert experiment and founded the first French immersion program (Lambert & Tucker, 1972). In an early, total French immersion program, students receive 100% academic instruction in French as their L2 starting from kindergarten or grade one. The proportion of French gradually decreases to 50% by the end of middle school (Dicks & Genesee, 2017). Most French immersion students are English speakers, but an increasing number of children with diverse backgrounds are enrolling in the programs (Dagenais, 2003). Although the initial purpose of the St. Lambert experiment was to provide children with bilingual proficiency to participate locally, it has become clear that parents strategically utilize such educational opportunities to obtain multilingualism as a resource for their children to access the larger, global communities (Barrett DeWiele & Edgerton, 2020; Dagenais, 2003). Research evidence suggests that students develop French proficiency at no cost of English development (Genesee, 2004). However, they have difficulty developing native-like oral proficiency in French due to the semi-authentic French speech input in an L2 learning context and limited community input (Mougeon & Rehner, 2017; Netelenbos et al., 2016).

In two-way bilingual education, academic content is delivered in a majority language and a minority language, and native-speaking children of both languages are enrolled (Schwartz et al., 2016). After the establishment of official bilingualism, Canada embraced multiculturalism and promoted the use of non-official languages (Tavares, 2000). With this background, a Ukrainian

bilingual program was first established in Alberta in 1974. Currently, in Canada, especially in Alberta, two-way bilingual programs are available in many international languages (Cummins, 2014). As opposed to the semi-authentic input in one-way immersive education, it is believed that the two-way bilingual education design can provide authentic input from native-speaking teachers and peers of both languages (Cummins, 1979; Menke, 2017; Nance, 2020). However, more evidence is warranted for diverse language pairs, especially when the minority language contains a unique phonological dimension (e.g., Mandarin tones).

The Chinese bilingual program in this dissertation is located in Edmonton, Alberta. This program started in only two elementary schools in 1982, thrived over the past four decades, and has become one of the most highly respected Chinese bilingual programs in North America (Asia Pacific Foundation of Canada, 2013). Currently, it is offered in 14 schools, including 7 elementary, 4 junior high, and 3 high schools (ECBEA, 2007). At the elementary level, 50% of the academic content is delivered in English, including subjects such as English Language Arts, Fine Arts, Science, and Social Studies. The other 50% is delivered in Mandarin Chinese, including Chinese Language Arts, Math, and Health Science (Wu, 2005). It provides a context to investigate bilingual speech development in a phonologically unique language and a unique student population from diverse language backgrounds to investigate the roles of home speech input and school speech input. The next section discusses the expected Mandarin speech input and pronunciation instruction by reviewing the program and curriculum design.

Mandarin input and pronunciation teaching in the program

The "Chinese" speech system taught in the program: Terminology issues

In 1982, the program chose Mandarin as the language to teach instead of the more widely spoken Cantonese or Toishanese among early immigrants (Duff & Doherty, 2019), because it was the official dialect of China and was used in many post-secondary institutions (Sun, 2011). This was a choice of foresight: China experienced fast economic growth in the 1990s, waves of Mainlander immigrants arrived in Canada, and Mandarin has become the most spoken non-official language in Canada (Statistics Canada, 2022).

In the program's curriculum (Alberta Education, 2006), the terminology "Chinese" is used interchangeably with "国语 *guóyǔ*" "汉语 *hànyǔ*" "华语 *huáyǔ*" "Mandarin" "普通话 *pǔtōnghuà*" and 中文 *zhōngwén*." It is necessary to clarify these terminologies because these influenced my understanding of the "Chinese" speech system being taught at school and the coding of students' language backgrounds and speech productions in this dissertation.

- Chinese or 汉语 *hànyǔ* (literally "Han language") is a large group of languages/dialects that are historically and currently spoken by the ethnic Han Chinese people and many neighboring minority ethnic groups. Chinese linguists use the terminology "方言 *fāngyán* (literally 'regional speech')" to refer to varieties of Chinese languages/dialects (Li, 1989), but the complexity of *fāngyán* is considered comparable to the languages in the Romance language family (Han et al., 2016).
- 普通话 *pǔtōnghuà* (literally "common language") is the standardized and officialized *lingua franca* of Mainland China. It refers to the Beijing dialect as its phonology.

- 国语 *guóyǔ* (literally "national language") is the terminology used in Taiwan. It is also a standardized and officialized *lingua franca* based on the modern Beijing phonology, but its phonological system is slightly different from *pǔtōnghuà* (Duanmu, 2007).
- 华语 *huáyǔ* (literally "Chinese language") is historically and currently used in contrast with any "foreign" languages. It is often used by Chinese diasporas with a stronger reference to the shared Chinese heritage as opposed to ethnicity or nationality.
- 中文 *zhōngwén* (literally "Chinese literacy") has more connection with the writing systems, but it can sometimes be used interchangeably with "Chinese" (e.g., 说中文 *shuō zhōngwén*, speak Chinese).
- Mandarin, or 官话方言 *guānhuà fāngyán* (literally "official dialects"), refers to the varieties of Chinese spoken in Northern and Southwest China with grammatical similarities (Li, 1989; Zhu, 2002). It is spoken by 70% of the Chinese population. However, meanwhile, the terminology can be used interchangeably with "普通话 *pǔtōnghuà*" or "国语 *guóyǔ*" to refer to the standardized varieties of Chinese.

Such terminology issues made the referents of "Mandarin" and "Chinese" complex in our study. For example, several teachers self-identified as native speakers of "Mandarin," but their local varieties of Mandarin lacked certain phonological contrasts in *pǔtōnghuà*. Another teacher self-identified as a native speaker of 湖北 *Húběi* dialect, stressing its differences from *pǔtōnghuà*, but some dialects spoken in *Húběi* may be categorized as varieties of Mandarin. Many parents reported their home language to be "Chinese," and we could only speculate that they were referring to Mandarin. Therefore, although language varieties are continuous and complex, arbitrary decisions were sometimes made to code or report categorical information.

In this dissertation, I refer to the bilingual program as a "Chinese" program because this is the official title, and the program promotes Chinese culture, language, and literacy (Alberta Education, 2006). But when I discuss the speech system and pronunciation, "Mandarin" is used to refer to the idealized learning target, which is the phonology of 普通话 *pǔtōnghuà* or Standard Chinese. Therefore, in the transcription-based analyses in Chapter 6 (Paper 3), the terminologies of "match" and "mismatch" are adopted to indicate the relationship between expected targets in the standardized phonology and speakers' productions perceived by transcribers, acknowledging that there are numerous legitimate variations of Mandarin. Table 3.1 is adapted from Lin et al. (2020) to provide basic background knowledge of the phonological system of Mandarin.

Table 3.1. Phonological characteristics of Mandarin (The Office of Modern Chinese, 2006).

Language: Modern Standard Mandarin Chinese, a.k.a. 普通话 *pǔtōnghuà* Language Family: Sino-Tibetan				
Words	There are more disyllabic words in Standard Mandarin. Speakers don't always reach an agreement on what is or is not a word (Li & Huang, 2010).			
Syllable shapes	Onset	Rime		
		Pre-nucleus glide (Duanmu, 2007)	Rhyme	
			Nucleus	Coda
	C_{0-1}	G_{0-1}	V_1	$V/Nasal_{0-1}$
	A syllable can be as short as 阿 *ā* /a/ or as long as 年 *nián* /njɛn/.			
Tones	Tone1: high-level tone /55/ (e.g. *mā*); Tone2: mid-rising tone /35/ (e.g. *má*); Tone3: low-dipping tone /214/ (e.g. *mǎ*); Tone4: high-falling tone /51/ (e.g. *mà*) Tone sandhi (phonological changes of tones based on contexts) can happen when there is more than one syllable.			
Syllable stress	The loss/lack of stress is more important than the presence of stress. The loss of the original tone is called 轻声 *qīngshēng* (neutral tone).			
Rimes	Monophthongs, diphthongs, and triphthongs (glide + nucleus + coda) are allowed as rimes.			
Glides	i/j, u/w, y/ɥ			
Nucleus vowels	a, o, ɤ, i, u, y, ɿ/ʅ, ʅ/ʯ, ɚ, in compound rimes (see Appendix A): ɛ, (æ), ɑ, ə			

Coda vowels Coda nasals	i/ɪ, u/ʊ n, ŋ A rime starting with a glide behaves similarly to a rime starting with this glide's counterpart vowel. In traditional Chinese Phonology, rimes were categorized into 四呼 *sìhū* (four types of rime onsets) based on either the glide or its corresponding nucleus vowel when the glide is absent (Simmons, 2016)
Onsets	p, p^h, m, f, t, t^h, n, l, k, k^h, ŋ, x, tɕ, $tɕ^h$, ɕ, ʦ, $ʦ^h$, s, tʂ, $tʂ^h$, ʂ, ʐ/ɻ, Ø/ʔ
Clusters	There are no clusters within syllables.
Examples of phonological constraints	p, p^h, m cannot be followed by y/ɥ. f has the strongest constraints on following vowels. It cannot be followed by any glides or close-front vowels. tɕ, $tɕ^h$, ɕ can only be followed by i/j and y/ɥ, while k, k^h, x, tɕ, $tɕ^h$, ɕ, ʦ, $ʦ^h$, s, tʂ, $tʂ^h$, ʂ, ʐ can only be followed by other vowels. Thus there are multiple complementary distributions among these classes.

Mandarin speech input in the bilingual program

The amount of Mandarin input in school is mainly decided by the proportion of immersion. In the elementary level of the Chinese bilingual program, 50% of the instruction time is in Mandarin (ECBEA, n.d.). Evidence from French immersion indicates that an increased proportion of immersion in the minority language is related to better learning outcomes (Genesee, 2004). Therefore, it remains a question whether the half-day input in Mandarin is enough for bilingual students from diverse backgrounds to effectively learn Mandarin pronunciation.

In school, students have a variety of opportunities to be exposed to Mandarin input: (1) subject contents, materials, and activities that are delivered in Mandarin, (2) direct linguistic instruction in Chinese Language Arts, and (3) incidental support when students struggle with Mandarin (Dicks & Genesee, 2017). This is in line with the curriculum conceptualization that language learning is intertwined with other subject-area experiences (Alberta Education, 2006). Although immersive education may seem to be an implicit language learning environment, students have opportunities to receive explicit instructions on pronunciation through (2) and (3).

The study in Chapter 4 provides evidence of the implicit and explicit strategies used in bilingual classrooms to teach Mandarin pronunciation.

The speech input in Mandarin is related to teachers' backgrounds. In the early years of the program, most Chinese teachers were Mandarin-L1 speakers recruited from Mainland China, who were recertified to teach in Alberta. A change occurred in recent decades that newly recruited teachers are required to hold bachelor's degrees in Education in Canada which involves a practicum component in local schools. According to a young, international teacher's reflection, it was not easy for international students from Mainland China to overcome the language barrier and fulfill the degree requirements at the undergraduate level. Consequently, an increasing number of bilingual teachers from Hong Kong, Taiwan, and Canada joined the program after obtaining Education degrees in Canada (Liu, 2020). Therefore, a variety of teacher input in Mandarin is expected, and the influences of such input are discussed in Chapter 4.

The speech input in Mandarin is also related to the student population. The two-way bilingual design aims to provide high-quality peer input in both Mandarin and English (Alberta Education, 2006; ECBEA, n.d.). However, many of the students are second- or third-generation Chinese immigrants whose families speak mostly English (Liu, 2020). Meanwhile, a small but increasing number of students whose L1s are languages other than Chinese and English (e.g., Korean, Hindi, Japanese) are enrolling in the program. Therefore, the program is multifunctional (Wu, 2005): It is a transitional program for Mandarin-L1 newcomers, a heritage program for Chinese-heritage students, a two-way bilingual program for Mandarin and English speakers, and an immersion program for multilingual children who are learning both English and Mandarin as L2s. The student composition does not meet Baker's (2006) ideal 1:1 ratio between the language-majority and -minority students, but it reflects the diverse population who desires to

learn Mandarin in reality. Therefore, the peer input in Mandarin may not be as sufficient as expected. This dissertation provides evidence of the influence of such reduced peer input.

Pronunciation instruction as designed by the curriculum

In the curriculum, "speaking" is an important aspect of Chinese skills: Students are expected to repeat and create oral phrases in the classroom in kindergarten and move on to more complex structures (e.g., sentences, presentations) and environments (e.g., unstructured situations) in higher grades. However, there is no explicit goal to address pronunciation (Alberta Education, 2006). Instead, pronunciation is indirectly addressed through the sound-symbol systems.

The Chinese writing system is morphosyllabic, i.e., one character corresponds to one syllable and usually maps onto one morpheme (Chen et al., 2004). Therefore, the type of grapheme-phoneme (letter-sound) correspondence in English does not exist in Chinese. To help learners access the speech of Mandarin, many romanization systems have been developed, and two were included in the curriculum: 拼音 *Pīnyīn* and 注音符号 *Zhùyīn Fúhào* (Alberta Education, 2006). In practice, *Zhùyīn Fúhào* was taught in the program until the mid-2000s, and currently, *Pīnyīn* is taught exclusively. *Pīnyīn* is widely used in Mandarin language education around the world (Commission of Written Language Reformation, 1958). Its scheme includes 声母 *shēngmǔ* (onsets), 韵母 *yùnmǔ* (rimes), and 声调 *shēngdiào* (tones) (see Appendix A).

Pīnyīn is semi-phonetic for it is highly transparent (Bassetti, 2006). Therefore, it is unsurprising that the learning of pronunciation is related to *Pīnyīn* in the program. Moreover, although the letters of *Pīnyīn* have their unique names just like English letters, these names are seldom used in practice (Lin et al., 2020). Instead, educators often use the sounds corresponding to the letters to refer to them (e.g., the letter *k* is called [k^h] or [$k^h ə$]). This also facilitates students'

learning of Mandarin pronunciation through *Pīnyīn*. However, *Pīnyīn* is *semi*-phonetic because there are mismatches between letters and sounds. Moreover, learning speech sounds through *Pīnyīn* can be especially challenging for bilingual learners because English and *Pīnyīn* use the same alphabet, but the letter-sound correspondences are sometimes different (Lin et al., 2020). These mismatches inevitably cause difficulties in learning Mandarin pronunciation through *Pīnyīn*.

Meanwhile, the focus on *Pīnyīn* only occurs in lower grades in the curriculum. In Grade 1, students are expected to listen to, identify, and produce the sounds in *Pīnyīn* and recognize Mandarin tones. In Grade 2, students are expected to practice combining sounds and combining letters (into syllables). In higher grades, students are expected to apply *Pīnyīn* in reading, use *Pīnyīn* to learn unfamiliar words, and gradually reduce the use of *Pīnyīn* to move on to Chinese characters (Alberta Education, 2006). This progress is similar to the curriculum design in Mainland China (Ministry of Education, 2012), where L1 Chinese education focuses on literacy skills. Such a curriculum design is mimicked by many private Chinese schools for heritage speakers in Canada (Duff & Doherty, 2019). However, considering the diverse language backgrounds of bilingual students, their oral language skills should not be assumed. In contrast, the curriculum of French immersion education in Alberta continues to address pronunciation in higher grades. For example, students are expected to differentiate French and English pronunciation and make use of teachers' feedback on pronunciation in Grades 4 through 7.

In summary, according to the curriculum, Mandarin pronunciation is not an explicit learning goal. Instead, it is expected to be acquired through implicit learning and through the letter-sound system. The following studies in this dissertation provide evidence of the influences of such curriculum design.

Chapter summary

Chapter 3 provided an overview of the Chinese bilingual program within the context of Canada's bilingual education. It also introduced background knowledge for readers to understand Mandarin pronunciation teaching in the Chinese bilingual program. Namely, it introduced the "Chinese" speech system that is taught in the program, the expected Mandarin speech input based on the program design, and the pronunciation instruction specified by the curriculum.

The next chapter, Chapter 4, is a submitted journal article to present Chinese teachers' perspectives on the teaching and learning of Mandarin pronunciation in bilingual classrooms.

Chapter 4 (Paper 2) Chinese Bilingual Teachers' Perspectives

Lin, Y., Pollock, K. E., & Li, F. (submitted). Pronunciation teaching in minority languages: Perspectives of elementary school teachers in a Chinese-English bilingual program in Canada.

Pronunciation teaching in minority languages: Perspectives of elementary school teachers in a Chinese-English bilingual program in Canada

Abstract

Despite an increasing interest in pronunciation instruction in English as a majority language or international language, less is known about pronunciation learning in non-English minority languages, especially among child learners. Bilingual education programs in Canada provide a unique context to address this research gap, as they involve immersive education in minority languages. Teachers in these programs thus are insightful informants. This study investigates how Chinese teachers in a Mandarin-English two-way bilingual program in Canada reflect on and address pronunciation teaching and learning in Mandarin. Semi-structured interviews were conducted with twelve Chinese teachers with diverse language backgrounds. Themes were identified through thematic analysis: (1) Mandarin pronunciation learning is difficult but progressive; (2) Pronunciation learning is impacted by multiple individual factors; (3) The societal majority language impacts the bilingual space at school; (4) Teachers incorporate direct and indirect techniques to teach pronunciation; (5) Teachers express concerns and needs about teaching pronunciation in bilingual classrooms. This study demonstrates the complexity of teaching the pronunciation of a minority language, whose speech system is distinctly different from English, in a bilingual classroom setting. It shares practical evidence of teaching strategies among bilingual teachers and identifies future directions for research and policymaking.

Keywords: bilingual education, pronunciation, minority language, children, teacher

Introduction

Second language (L2) pronunciation used to be the "Cinderella" of language teaching, unfairly oppressed (Celce-Murcia et al., 1996, p.323; Levis & Sonsaat, 2017). Among the increasing discussions on pronunciation instruction in recent decades, many were focused on adult learners of the majority language of society and/or English as an international *lingua franca* (Derwing, 2020; Isaacs, 2009), but less is known about child learners of non-English languages. It is important to examine child learners' L2 pronunciation learning to directly observe the developmental process (Flege & Bohn, 2021). When immersed in the majority language in an immigration context, children are often assumed to be able to develop native-like (or at least fully intelligible) L2 pronunciation (Derwing, 2020). However, when the target language does not enjoy a majority status, speech input is limited, and students' motivation may vary, which can impact children's learning outcomes (Flege & Bohn, 2021). Therefore, research on children's pronunciation learning of a non-English minority language will offer theoretical implications for how factors such as the age of learning, speech input, and motivation impact pronunciation learning.

Around the world, there are bilingual education programs where children learn a societal minority language through bilingual education. A few studies on pronunciation learning outcomes indicate that two-way bilingual education can level out the home language differences, as children with diverse backgrounds provide authentic input for each other (Menke, 2017; Nance, 2020). More information is needed in various languages to verify such results. In addition to implicit learning through peer interactions, teachers often need to explicitly teach the language forms, including pronunciation, in Language Arts and through incidental support (Dicks & Genesee, 2017). However, teachers are often left with few guidelines for pronunciation instruction, and not much is known about how the pronunciation of a minority language is taught

to children in practice. When little is known, qualitative data can provide insights into the situation and guide question formations to address realistic needs (Austin & Sutton, 2014).

This study takes a long-standing and well-respected Chinese-English two-way bilingual program in Alberta, Canada, as a successful case and qualitatively presents Chinese teachers' lived experiences with pronunciation teaching. Specifically, it depicts the complexity of pronunciation instruction in a minority language, presents teachers' reflections on factors of pronunciation learning, and adds practical evidence of teaching techniques for the worldwide intercultural community of bilingual teachers (Fishman, 1976). The following sections introduce our lenses to understand teachers' discussions, including the factors of pronunciation learning, the challenges of pronunciation teaching, and an empowering view of teachers' roles in bilingual education, followed by an introduction to the bilingual program of interest and its targeted language, Mandarin.

Factors of bilingual pronunciation learning

Bilingual pronunciation learning is impacted by linguistic, individual, and social factors (MacLeod & Stoel-Gammon, 2010; Netelenbos et al., 2016; Richter, 2019). This categorization of learning factors draws upon Paradis's (2011) definition of "individual differences" in L2 acquisition but analyzes native language (L1)-L2 transfer and the general language environment as separate categories, i.e., linguistic and social factors, respectively. This is because as opposed to an immigration context (Paradis, 2011), in bilingual programs, language transfer between the majority and minority language and the language environment at school and in society are largely shared among the students and therefore, can be separated from individual factors. Each category is introduced below and guides the organization of teachers' reflections.

Linguistic factors. Pronunciation learning is impacted by the phonological/phonetic similarities between L1 and L2. For example, the Speech Learning Model (Flege & Bohn, 2021) hypothesized that learners perceive an L2 category based on its most similar L1 counterpart and gradually establish its own category; The Perceptual Assimilation Model (Best & Tyler, 2007) hypothesized that learners distinguish the L2 categories using L1 categories and features. Meanwhile, the universal difficulty of the L2 sounds also plays a role (Major, 2001).

Individual factors. A learner's age of learning has been regarded as an important factor (Flege, 1995), i.e., *earlier is better*. However, recent evidence suggests that *earlier is not enough*, as pronunciation learning is impacted by the quantity and quality of input received by the individual learner, including the length of exposure, the current amount of exposure, the opportunity of output, and the authenticity of the input (Flege & Bohn, 2021). Furthermore, pronunciation learning is also impacted by individual motivation, language aptitude, and cognitive maturity (Flege et al., 1995; Paradis, 2011; Robinson, 2005).

Social factors. Input and motivation are related to language status in the wider society. When both L1 and L2 enjoy a high social status, learners have plentiful opportunities to use both languages, self-identify as dual-lingual speakers, and demonstrate balanced pronunciation competence (MacLeod & Stoel-Gammon, 2010). On the other hand, when one language is a minority language, pronunciation attrition or incomplete learning may occur (Chang et al., 2011; Flege et al., 1995).

According to evidence in teaching English as a majority second language or international language, teachers consciously reflect on and utilize these factors (Couper, 2021). The current study demonstrates how these factors are understood by teachers of young children learning a minority language and utilized in their teaching practices in a bilingual school setting.

Challenges in pronunciation teaching

With the understanding that pronunciation teaching of non-English minority languages is a distinctive issue, English teachers' perspectives can provide insights into the challenges in pronunciation teaching. Teachers of adult learners reported a reluctance to teach pronunciation due to limited resources and insufficient knowledge (MacDonald, 2002). Teachers embraced the intelligibility principle in theory but tended to set nativeness as a goal in practice (Jenkins, 2007; Levis, 2020). Their instruction techniques consisted mainly of form-focused instruction (FFI, Spada & Lightbown, 2008) such as practice and repetition (Baker, 2014; Foote et al., 2011; Murphy, 2011) and limited communicative language teaching approaches (CLT, Littlewood, 2011).

Couper (2021) surveyed teachers of child learners in Uruguay and New Zealand. Results were similar to teachers of adult learners in that they lacked confidence in pronunciation teaching due to limited knowledge of phonetics and phonology. Their pronunciation instruction was limited by time, textbooks, and curricula. In addition, non-native English teachers had concerns about their own pronunciation (see also, Çağatay, 2021).

In bilingual education, teachers reported similar needs for resources and training (Wisecup, 2017) and further expressed their particular struggles to promote the minority language (Estrada & Chacón, 2015). Therefore, it is safe to expect that the challenges in pronunciation teaching of a minority language in English-dominant environments will be equally or more complex than the aforementioned challenges reported by English teachers.

An empowering view of teachers' role

There are different approaches to understanding the complex and challenging issues in bilingual education. One approach is to analyze the problems in the system and advocate for improvements. For example, Duff and Doherty (2019) criticized the ill-designed curricula, unnatural classroom interactions, and lack of teaching resources in Chinese bilingual education.

On the other hand, Menken and García (2010) took an empowering view of teachers' roles. Teachers are active policymakers in the classroom and use their own intuition, knowledge, experience, and reflections to negotiate between language education policies and their practices. For example, Estrada and Chacón (2015) demonstrated how a teacher struggled to promote the students' use of Spanish in the United States by maintaining Spanish use in the whole class and prioritizing interactions in small groups. Schwartz et al. (2016) interviewed educators in a Mandarin-English bilingual school in Canada and a Russian-Hebrew bilingual preschool in Israel. Challenged by the lack of resources and the shifting curriculum designs, teachers adopted flexible classroom practices (e.g., code-switching), collaborated with teachers of the other language to facilitate cross-linguistic transfer, and managed the curriculum innovatively to best support student learning.

This study adopts this empowering lens and believes that teachers possess an insightful understanding of pronunciation learning and are able to address the challenges with techniques and strategies. Therefore, we aim to share these insights and strategies with other researchers and bilingual teachers through the current study.

Two-way bilingual education: The case of a Chinese-English program in Western Canada

This study addresses the research gap in child pronunciation learning in non-English minority languages through a case of a two-way bilingual education program in Alberta, Canada. In such programs, both languages are the medium for content delivery (Dicks & Genesee, 2017), and the goals are to achieve functional bilingualism (Genesee & Lindholm-Leary, 2008). In the Chinese-English bilingual program in this study, 50% of the academic content is delivered in English and the other 50% in Mandarin (a.k.a. Standard Chinese). The program has a long-standing history of 40 years and is one of the most highly respected Mandarin bilingual programs in North America (Asia Pacific Foundation of Canada, 2013). It attracts not only children with Chinese backgrounds but also children who speak English or other languages at home.

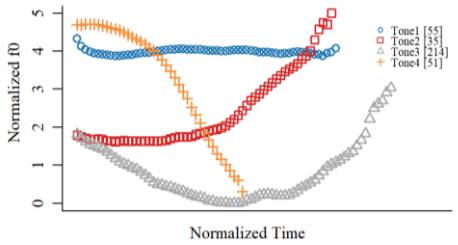
Learning Mandarin can be challenging in an English-dominant society, yet the program has thrived over the past four decades, starting in only two elementary schools in 1982 and continually expanding to be offered in 14 schools (7 elementary, 4 junior high, and 3 high schools) across the city in 2023 (ECBEA, n.d.). One important factor for such success was considered to be the availability of diverse Chinese language teachers with high professionalism and strong Mandarin competence (Liu, 2020). There is a rigorous standard of teaching qualifications and Mandarin proficiency during Chinese teacher recruitment. Thus, we believe it is especially informative to present this case and share the successful experiences of the teachers.

A brief introduction to the challenges of learning Mandarin pronunciation

In terms of pronunciation learning, it is well-recognized that Mandarin is very different from English in its phonology. Mandarin is a tonal language where the four lexical tones (Table 4.1) differentiate word meanings. Therefore, inaccurate productions of tones may lead to

miscommunication. In addition to tones, Mandarin has unique speech sounds. Its consonant inventory includes voiceless sibilant fricatives (and corresponding affricates) at three places, i.e., alveolar [s], alveopalatal [ɕ], and retroflex [ʂ], compared to only two in English, i.e., alveolar [s] and postalveolar [ʃ] (Li & Munson, 2016). The vowel inventory includes rounded front-closed vowel [y], back-mid-closed vowel [ɤ], and apical vowels [ɿ] and [ʅ], which are not present in English (Lee-Kim, 2014).

Table 4.1. Mandarin lexical tones. The four tones were produced by a female Mandarin speaker, and fundamental frequency (*f0*) was extracted using Praat.

Tone	Description	5-scale*	Example word	*f0* plot
Tone1	high-level	[55]	*dā*, 搭 build	
Tone2	mid-rising	[35]	*dá*, 答 answer	
Tone3	low-dipping	[214]	*dǎ*, 打 hit	
Tone4	high-falling	[51]	*dà*, 大 big	

* In Chinese linguistics, tones are often transcribed in the 5-scale convention, with [5] representing the highest and [1] representing the lowest pitch.

Mandarin pronunciation is also difficult due to its logographic orthography, for learners can access little pronunciation information through written materials (Chen et al., 2004). To facilitate language and literacy development, *Pīnyīn*, the official romanized transcription system of Mandarin, is widely used in teaching Mandarin in Mainland China and worldwide. The bilingual program has adopted *Pīnyīn* in its current curriculum design. The curriculum indirectly addresses it through the learning of letter-sound relationships of *Pīnyīn* and does not directly address pronunciation instruction (Alberta Education, 2006).

Beyond the linguistic and literacy complexities, the program's composition of the student population is diverse in the immigration context (Liu, 2020). Different waves of Chinese diasporas who arrived in Canada at various times may speak different Chinese *fāngyán* (regional variations of Chinese dialects) with drastically different speech systems, although the number of Mandarin speakers increased in recent decades (Duff & Doherty, 2019). Consequently, the bulk of the student population is somewhat connected to Chinese heritage but exposed to English at home (e.g., second- or third-generation children of non-Mandarin *fāngyán* speakers), followed by L1-speakers of Mandarin, other *fāngyán*, and a few other languages.

In summary, teachers are challenged by not only the generic difficulties of teaching pronunciation in a minority language but also Mandarin's unique speech system and the diverse student population. This study will provide a qualitative analysis of how teachers have approached the challenges of Mandarin pronunciation teaching in Canada and developed strategies to promote bilingual students' pronunciation learning.

Method

The study has obtained ethics approval from the Research Ethics Board at the University of Alberta (Pro00075638). The report of methods and findings is guided by COREQ (Consolidated Criteria for Reporting Qualitative Research, Tong et al., 2007). See Appendix B for the full checklist.

Participants. Voluntary and snowball sampling methods were used to recruit twelve teachers from three elementary schools. Interviewees had diverse language backgrounds, educational backgrounds, and teaching experiences. Each teacher was assigned a pseudonym. Because the Chinese bilingual program is a rather small community with a limited number of

teachers, specific information about their language, educational, or teaching experiences can be identifying. Therefore, individual teachers' information was omitted to protect participants' confidentiality. As an aggregated description, participants' L1s included not only Mandarin but Cantonese and other Chinese *fāngyán*. Seven teachers were exposed to Mandarin from birth, and the others started to learn Mandarin at school age. Among the latter, three were native speakers of another Chinese language/dialect, and two were native speakers of English in Alberta who later attended the Chinese bilingual program. Most teachers self-rated their Mandarin proficiency as high. Teachers were each teaching one of the grades from kindergarten to grade five, but most teachers had experience teaching other grades. The years of teaching ranged from one to more than twenty years. Teachers received varying amounts of training in education, Mandarin teaching, and L2 teaching from a variety of sources. Among them, five teachers received specific training in teaching Mandarin through degree programs, university courses, or professional development (PD). It should be mentioned that Chinese teachers in this program have to hold a local teaching certificate (Liu, 2020). In the earlier years of the program, most teachers were immigrants from Mainland China who recertified to teach in Alberta. In recent years, more teachers obtained their Bachelor's degree in Education in Canada. In either case, teachers were not mandated to have formal training in teaching Mandarin, teaching L2s, teaching pronunciation, or phonology and phonetics.

Interviewer. The interviewer was the first author of the paper. They had research experiences in phonetics and child language development. They were familiar with school settings through research activities but did not have work or learning experience in bilingual education. They were a native speaker of Mandarin and proficient in English.

Interviews. Semi-structured interviews were conducted in the schools. General guidelines (Appendix C) were used, but teachers were encouraged to discuss each topic in an open-ended style. Each interview was conducted in the teacher's preferred language and lasted 15 to 42 minutes. Interviews were audio-recorded with a Zoom H1n digital recorder.

Transcription. The recordings were each transcribed by one of three Mandarin-L1 transcribers who were proficient in English. Mandarin interviews were translated into English. Each transcript was reviewed by another transcriber. The first author was either the transcriber or reviewer of all transcripts.

Coding. Conventional thematic analysis was used, where codes and themes emerged from the text. An initial codebook was developed after reading the transcripts several times. The first author and a research assistant coded the transcripts in NVivo 12 independently. The first round of coding reached a weighted average agreement of 98.81% and a weighted average κ of 0.72, interpreted as "good agreement." A full consensus was reached through discussion. The initial and final codebooks are documented (Appendix D). The codes were reanalyzed into themes, and a mind map was presented to facilitate understanding (Figure 4.1). The organization is based on the three levels of pronunciation learning factors and the challenges and strategies of teaching pronunciation in minority languages discussed in the introduction.

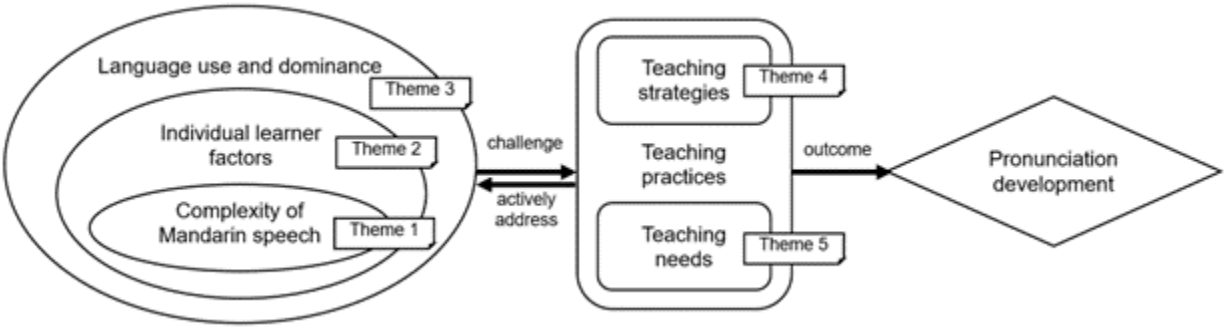


Figure 4.1. A mind map for the themes emerged from teacher interviews.

Saturation. Figure 4.2 shows when each of the 27 codes was mentioned. The first five teachers had already mentioned all the codes. Each teacher mentioned 15 to 24 codes. Four codes were mentioned by all teachers. Twenty codes were mentioned by more than six teachers. Each code was mentioned by at least two teachers. This suggests that all concepts were repeated multiple times and new concepts were unlikely to emerge (Trotter, 2012).

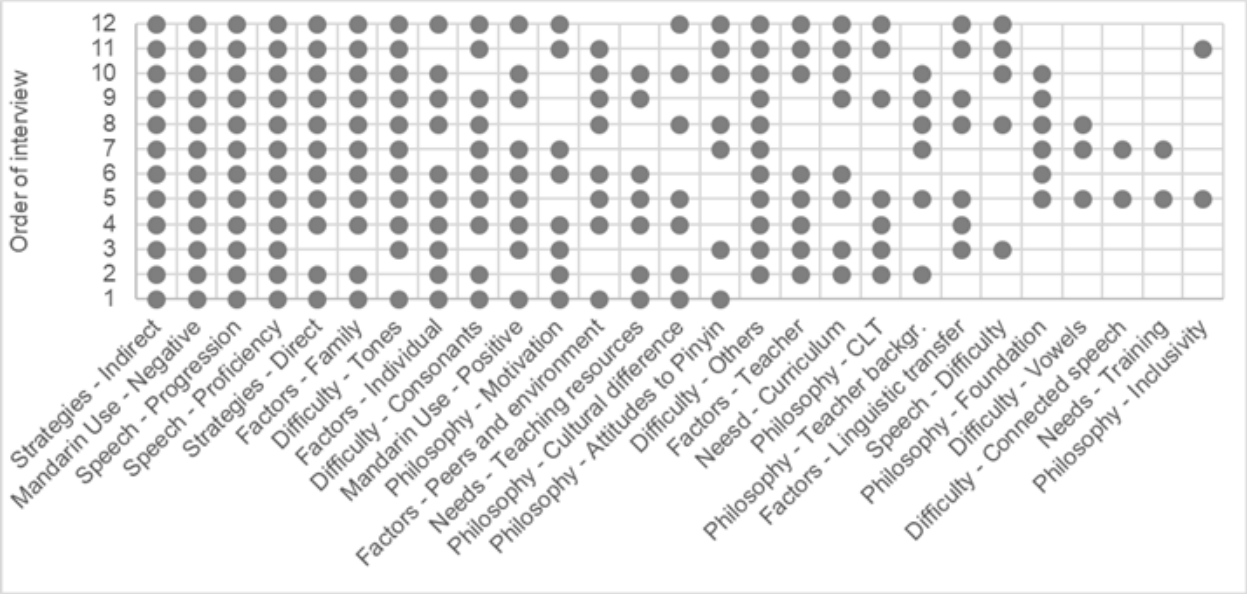


Figure 4.2. Concept emergence and saturation.

Result

Five themes emerged: (1) Mandarin pronunciation learning is difficult but progressive; (2) Pronunciation learning is impacted by multiple individual factors; (3) The societal majority language impacts the bilingual space at school; (4) Teachers incorporate direct and indirect techniques to teach Mandarin pronunciation; (5) Teachers express concerns and needs about teaching pronunciation in bilingual classrooms. The first three themes discussed how teachers

reflected on the impacts of linguistic, individual, and social factors. The last two themes highlighted teachers' powerful roles but also advocated for their needs.

Mandarin pronunciation learning is difficult but progressive

Developmental trajectories of Mandarin speech.

Most of our teachers believed that students' Mandarin skills progressed across grade levels as their pronunciation "gets better (Teacher On)," and "their ability to express themselves is growing (Teacher Yi)." Meanwhile, some teachers pointed out that students' pronunciation progressed slower in higher grades:

> It is certain that their Chinese will improve, but when you leave the Chinese environment, you actually have limited room for improvement … Their growing speed will slow down. (Teacher Hsu)
>
> I think they are better in tones in the lower grades. As they slowly advance to higher grades, it seems that they forget about the tones. (Teacher Feng)

These observations are comparable to the empirical results in French immersion education: Young children are able to acquire new sound categories, but their pronunciation may fossilize due to limited input (Meziane & MacLeod, 2020; Netelenbos et al., 2016). More quantitative research is needed to verify these reflections in Mandarin learning.

Challenging speech units.

Despite the generally progressive learning trajectory, teachers agreed that Mandarin pronunciation was challenging and identified speech units that were difficult to articulate.

Tones were mentioned by most teachers. Tones can be challenging for learners from an English-dominant environment since they cannot be assimilated directly into English categories (Best & Tyler, 2007; Flege & Bohn, 2021). English speakers may perceive tones as

uncategorized speech units, nonlinguistic melodies, or intonations, which do not bear as much linguistic significance (So & Best, 2004), as recognized by teachers:

> Sometimes, they just shrug and tell you that they can't hear anything, there's no difference at all. (Teacher Feng)
>
> [T]he tone, aww … it's hard for them! … they feel like they're saying it, … but they're not … I find it's hard for them to [perceive the difference] because it's not in English. (Teacher Young)

In terms of the specifically challenging tone targets, tone pairs that share more phonetic similarities are more confusable (So & Best, 2010). These were recognized by teachers: "Their Tone 1 (high-level tone) is often exchanged with Tone 4 (high-falling tone) (Teacher Gwok)," and "Especially Tone 3 (falling-rising tone) is very difficult for them. They often make it as Tone 2 (rising tone) (Teacher Hsu)." Moreover, teachers identified tones that are universally challenging. For example, "Their Tone 4 is not emphasized enough (Teacher Liu)." A similar phenomenon was documented in younger Mandarin-L1 children (Wong, 2012), i.e., young children's Tone 4 had a shallower falling slope due to less developed articulatory skills.

In addition to tones, teachers mentioned challenging speech sounds. Teachers' reflections were compared with Lin and Johnson (2010), which investigated Mandarin speech sound patterns in bilingual students in an English immersion program in Taiwan (Appendix E). Teachers nominated challenging consonants that were comparable to the empirical findings, including the alveolo-palatal fricatives and affricates *j q x* /tɕ tɕʰ ɕ/ and retroflex fricatives and affricates *zh ch sh r* /tʂ tʂʰ ʂ ʐ/. In contrast, *j q x* are established before two years of age in Mandarin-L1 children (Zhu, 2002). Bilingual students' speech development does not seem to follow the trajectory of monolingual children, which can be attributed to cross-linguistic influence (Meziane & MacLeod, 2020). Nonetheless, teachers gave fewer examples of challenging vowels, and their observations did not match previous empirical data. For example,

teachers mentioned difficulties in the apical vowel /ɿ/ and the rhotic vowel /ɚ/, which were not documented among bilingual children in a Mandarin-majority context (Lin & Johnson, 2010). Therefore, further investigation into vowel patterns in bilingual children is warranted.

Meanwhile, when asked about challenging speech units for the children to pronounce, teachers mentioned issues that are not pronunciation-related. For example, they reported that students used the English letter "c" to transcribe *k* in *Pīnyīn*. It appeared that teachers had difficulties pinpointing specific speech sounds using phonological or phonetic terms. Instead, in their reflections, pronunciation learning is a "by-product" of learning *Pīnyīn*. Since Mandarin pronunciation was only addressed through *Pīnyīn* in the curriculum, and teachers were not required to receive training in phonology or phonetics, such an association among teachers is not surprising.

In summary, the first theme discussed the development and difficulties in students' Mandarin pronunciation. The developmental trajectories were generally progressive but might also experience plateauing. Specific examples of challenging speech units were discussed, which suggested the impacts of the phonetic (dis)similarities between L1 and L2 speech systems as well as the universal difficulty of the targets. In addition to linguistic factors, bilingual speech is also impacted by individual factors of the learners, discussed in the next theme.

Pronunciation learning is impacted by multiple individual factors

Family language environment.

The effect of family backgrounds was frequently mentioned. Teachers reported that students with no Mandarin background would be influenced by English in their speech learning. Moreover, teachers considered the impacts of Chinese *fāngyán*. Some stressed *fāngyán*'s

differences from Mandarin, and others highlighted its facilitative effect due to the shared linguistic and cultural features. These different foci may be related to teachers' own language backgrounds. Compare the statements of Teacher Gwok and Teacher Hsu who were native speakers of Cantonese and Hokkien, respectively, and Teacher Yi who was native in Mandarin:

> If they already speak Cantonese, … tones are no problem for them. (Teacher Gwok)
>
> Overall, it is Asian children [who are better at Mandarin speech] … He can relate. (Teacher Hsu)
>
> They actually have a lot of Chinese culture components and Chinese language components in their minds, but because they use a *fāngyán* [other than Mandarin] … they are affected by it in … pronunciation. (Teacher Yi)

Although all teachers identified home background as a factor, they did not view it as dichotomous or decisive. In contrast, they considered the family language environment as a continuous variable and discussed the quantity and quality of input (as discussed by Flege & Bohn, 2021). The following excerpts highlight teachers' reflections on the importance of speech input at home:

> He is not likely to learn according to the teacher's accent … It is not like his family, … every day, 24 hours a day, with mom and dad. (Teacher Bai)
>
> [If] his family doesn't speak Chinese, nor does his friends, he has no opportunity to speak Chinese except speaking with Chinese teachers in class every half day. You don't have so many people in the end. (Teacher Jia)

Compared to Teacher Bai, who emphasized the amount of input, Teacher Jia discussed the number of speakers who provided input. Empirical evidence suggested that children's language development in a minority language was predicted by the number of native speakers in their life (Place & Hoff, 2016). A plausible explanation is that input from various speakers helps tease linguistic information apart from speaker information and establish robust representations of speech categories (Werker & Curtin, 2005).

With such diversity in the student population, teachers recognized that their students are "completely different (Teacher Liu)" from students in China, and "overseas Chinese teaching is not the same as domestic Chinese teaching (Teacher Yi)." Teachers had realistic and differentiated expectations for students without a Mandarin background and provided support as needed. "[W]e also allow you to enter the school, and then gradually and continuously help … through teaching. (Teacher Yi)." They often emphasized how students with non-Mandarin backgrounds could also make significant achievements in Mandarin, where motivation and learner aptitudes played a role and sometimes compensated for the lack of home input.

Other factors of Mandarin speech learning.

Teachers highlighted motivation as a factor and used effort and participation as indicators: "If you are not interested, … it is difficult to improve (Teacher Bai)." "The English families are … very motivated to use whatever [activities] I get them (Teacher Young)." "If he takes time to practice, he can speak very well (Teacher Yi)." Language learning motivation can be divided into instrumental and integrative motivations, which reflect the desire to achieve specific goals and to achieve social integration through L2, respectively (Gardner & Lambert, 1972). Based on the teachers' reflections, instrumental motivation was rare in these young students, which is in line with survey-based research (Xu & Case, 2015).

Language learning aptitude was another factor identified by the teachers:

> Unless the children's innate language talent is strong, it is not so easy to learn Chinese here. (Teacher Jia)
>
> For language, … comprehension, I think still there are differences. Some children actually don't have any background to speak Chinese or any other Asian language. However, they have a way to catch the points. (Teacher Hsu)

"Comprehension" in Teacher Hsu's quote should not be interpreted as language comprehension. Instead, it may be more similar to the "innate language talent" in Teacher Jia's quote, which is related to L2 aptitude (Robinson, 2005). Our teachers did not use the terminology or further operationalize the concept, but they had the intuition that Mandarin speech development was impacted by learner-internal factors such as "talents."

In addition to these learner factors, teachers highlighted their own language backgrounds as a factor. Here is a quote from a Mandarin-L1 teacher:

> I think that if they learn from Chinese teachers who are native speakers of Chinese, as the teachers will subconsciously speak Chinese to them a lot, their progress in Chinese will be faster. (Teacher Jia)

However, almost half of our interviewees were not native speakers of Mandarin. A discussion around teachers' nativeness emerged, where some teachers believed that native-speaking teachers could offer authentic input and promote Mandarin use, while others identified unique values of non-native teachers: (1) They provided a wide range of speech input which stabilized the speech representations; (2) They had strong metalinguistic awareness and translated their own learning experiences into teaching; (3) They could relate to the students' bilingual perspectives and serve as a role model. See the following excerpts:

> With a variety of teachers, … [students] will not be influenced by a particular accent of teachers across six years. (Teacher Bai)

> I learned Chinese and English bilingually, so I think [of] how the teacher used to encourage us to learn before. I still remember some homework or some projects … Maybe I can teach them these again. (Teacher Gwok)

> Many of us … used to learn Chinese here … So when seeing us, many parents will say: "You know, in fact, learning Chinese also has a successful side." (Teacher Lee)

This discussion suggests that the impact of (non)native teachers can be complex. It seems intuitive that education programs should consider speech proficiency as a staffing criterion

(Place & Hoff, 2016). However, language proficiency is not defined by the status of nativeness (Derwing et al., 2014). Evidence showed that English learners taught by native and non-native teachers achieved similar speech outcomes (Levis et al., 2016), and bilingual teachers who understood the students' L1 had better resources to support their learning (Copland & Yonetsugi, 2016). No study seems to have addressed non-native teachers' effects on children's pronunciation learning in a minority language (Duff & Doherty, 2019). This study took the initiative to provide qualitative evidence on this topic by avoiding ubiquitous nativism and presenting a variety of opinions.

In this section, we presented several individual factors of pronunciation learning, including family language environment, motivation, and L2 aptitudes, as well as teacher effects. In addition to these individual factors, teachers also reflected on the social factor of language environments, which is discussed in the next theme.

The societal majority language impacts the bilingual space at school

Bilingual education is impacted by the language environments of the school and the broader society (Baker, 2011; MacLeod & Stoel-Gammon, 2010). This section will discuss how a bilingual space is created at school, and how it is affected by the surrounding environment.

The environment at school.

The merit of the two-way bilingual design lies in the additive bilingual conceptualization and the interactions among peers from diverse backgrounds (Dicks & Genesee, 2017; Menke, 2017; Nance, 2020). Teachers suggested that the program provided access to knowledge and social interactions for new immigrant children, promoted heritage speech maintenance, and supported

L2 pronunciation learning of non-Mandarin students, which matched the functions of two-way bilingual programs listed in Wu (2005).

With diverse students fulfilling a variety of linguistic and academic goals at school, a unique bilingual space was established, where students flexibly chose among their languages based on communication situations and provide authentic language environments for each other.

> She came from … Chinese family but not Mandarin-speaking, but she would speak to the other friend in Mandarin because that friend doesn't speak much English. (Teacher On)
>
> They're more willing to [use Mandarin], they feel more comfortable maybe, when there are some Mandarin-speaking kids around. (Teacher Young)

However, Schwartz et al. (2016) suggested that to achieve language balance, the language-majority and -minority students should be around the same number. In this program, "only one student in my class speaks Chinese at home (Teacher Jia)," "many parents are second-generation (Teacher Ding)," and "one of them is from South Korea, and then many of them are mixed-race (Teacher Lee)." With such diversity, students eventually resorted to English as a *lingua franca* at school, which is related to the language status of English in the community.

The environment of society.

English dominance in the community has inevitably sneaked into the school environment. Although 50% of the class content was delivered in Mandarin, the broader school environment was unbalanced: "Not all the lunch supervisors will speak Chinese, so it is difficult to require [Mandarin use]… The assembly of the school, and the announcement of the school, they are all in English (Teacher Cheng)." Also due to English's majority status, "the older siblings will speak English to them," "parents begin to pick up English as well (Teacher Ding)," and "the videos they watch are in English (Teacher Lee)." Consequently, students "still tend to speak English (Teacher Cheng)."

Meanwhile, the social factor interacted with the factors reviewed in the first two themes. Teacher Hsu described, “Once you said to use Chinese, everyone will be silent even though they talked loudly just now,” because Mandarin was linguistically challenging and “they don’t have this language at all.” In addition, students lost motivation for using Mandarin because “I [the student] will use English anyway. Why should I use Chinese?” It appeared that the development of students’ Mandarin pronunciation skills, the willingness to speak Mandarin, and a balanced bilingual environment cyclically impacted each other.

It became clear that tensions existed between the language environments at school and in the community. Two-way bilingual education does not magically promote the minority language. Instead, teachers took active and purposeful efforts to maintain the bilingual space. To improve students’ Mandarin pronunciation skills and encourage Mandarin use, teachers adopted a variety of techniques and strategies, which is reviewed in the next theme.

Teachers incorporated direct and indirect strategies to teach pronunciation

Previous checklist-based surveys suggested that pronunciation teaching techniques lacked innovation and diversity in practice (Foote et al., 2011; Murphy, 2011). Our study used a semi-structured interview method and elicited open-ended discussions to identify the techniques used. The node of “teaching strategies” was further divided into two categories: direct strategies, which directly targeted speech forms (i.e., Form-Focused Instruction or FFI), and indirect strategies, which contextualized speech learning in meaningful activities (i.e., Communicative Language Teaching or CLT). To share teachers’ insights among researchers and fellow educators, full quotes were presented in Appendix F.

Direct strategies of Mandarin speech teaching.

Teachers used a variety of techniques to address speech forms directly. Letting students repeat after models and directly correcting their errors, which were often reported in previous literature (Foote et al., 2016; Murphy, 2011), constituted only a part of their responses. Although teachers emphasized the importance of repetitive practice, they were aware that this was "not rote memorization (Teacher Liu)" in traditional pronunciation teaching (Isaac, 2009). Instead, they highlighted its functional importance, that it can improve fluency (Teacher Jia), familiarize students with the "flow of speech (Teacher Liu)," and reinforce learning outcomes (Teacher Bai). In addition, teachers provided listening materials to exemplify "the correct pronunciation (Teacher Liu)" and exaggerated their models for students "to hear it clearly (Teacher Lee)." Meanwhile, they provided multimodal cues, including but not limited to visual cues (e.g., graphic illustrations of tones), tactile cues (e.g., feeling the articulators), gestural cues (e.g., using hands to demonstrate articulatory gestures), and written cues (e.g., *Pīnyīn*).

The richer inventory of direct strategies in this study may be attributed to three reasons. First, some of our teachers received training in L2 teaching and Mandarin teaching from a wide range of sources, such as university courses and professional development (PD) sessions. Some discussed what they learned from linguistic courses and readings in phonetics and phonology (e.g., Teacher On and Teacher Yi), which is different from Couper's (2021) report on teachers' lack of knowledge in these areas. Therefore, some teachers were equipped with specific knowledge to teach pronunciation explicitly. Second, most of our teachers were L2 learners of English. In East Asia, English is often taught as a foreign language (Tokumoto & Shibata, 2011) with a large proportion of FFI. Teachers might have translated their own learning experiences into teaching skills. Third, the Mandarin speech system is so different from English that they

would not expect it to be "picked up." This forced teachers to adopt explicit teaching techniques. For example, Teacher Feng reflected that students "couldn't hear the difference [between tones] very well," so they had to "use gestures to tell them."

Indirect strategies of Mandarin speech teaching.

Teachers assigned equal emphasis, if not more, to CLT and used indirect strategies to teach pronunciation in meaningful contexts. First, teachers adopted classroom policies and reward systems to encourage Mandarin use. This was different from practicing with adult learners, who usually have stronger instrumental and/or integrative motivations to learn a minority language (Baker et al., 2011). However, the extrinsic requirements and rewards would not be enough to promote continuous learning (Noels et al., 2000), and other strategies were incorporated to promote intrinsic motivations.

Second, teachers used multimedia resources to teach Mandarin speech. These not only provided culturally authentic materials but addressed the young learners' motivations, particularly integrative motivations. This is appropriate for child learners because they are seldom encouraged by instrumental motivations (Xu & Case, 2015) – As Teacher On recognized, "They are not gonna tell me 'I wanna learn Mandarin because it's good for my job.'" Instead, they stated, "They think, 'Oh, Mandarin-speaking, there are cool TV shows.' And they want to understand it." Teacher Gwok let the students "listen to the songs that I used to listen to when I was a child … They like them, really. Then you will see some foreign children … find out on YouTube to sing." Through the long list of songs they provided, it became clear that the songs were not simply a teaching tool but were attached to cultural memories shared amongst Chinese communities around the world, which were passed down intergenerationally and passed around interculturally through teaching practices.

Third, teachers taught Mandarin pronunciation through meaningful activities. Skills were practiced through language and literacy activities such as daily conversations, personal narratives, and reading. In addition, pronunciation was practiced in other subject areas such as health and mathematics, which were taught in Mandarin by the curricular design (Alberta Education, 2006). For example, students watched "videos related to health science" to "learn the subject but also learn the language (Teacher Ding)" and recorded videos in Mandarin to "tell … how to do addition (Teacher Bai)."

In summary, teachers presented powerful toolkits of direct and indirect pronunciation teaching strategies. They flexibly and innovatively negotiated between curriculum-adopted CLT (Alberta Education, 2006) and the efficiency of FFI (Isaacs, 2009) to achieve the expected learning outcomes. Using these strategies, teachers not only addressed specific speech difficulties but also encouraged students' motivation and language use.

Teachers express concerns and needs about teaching pronunciation in bilingual classrooms

Despite the powerful roles teachers played, they were limited by the resources provided by the educational system. First, the optimal amount of immersion should be revisited. "Time is limited" was repetitively brought up. Teacher Bai elaborated, "The Chinese-English bilingual [program] is not the same as … 100% immersive teaching … This is two-way bilingual, 50% in our class, but … after you leave the classroom, your environment is an English world." Results in French immersion suggested that more immersion in the minority language was related to better learning outcomes (Genesee, 2004). More evidence of bilingual education in Mandarin and other minority languages is needed to decide the optimal proportion of immersion.

Second, teachers reflected on the curriculum design. In the curriculum, pronunciation is only addressed through the sound-symbol system, i.e., *Pīnyīn*, in grades one and two, and the focus is switched to Chinese characters in order to facilitate reading without reliance on *Pīnyīn* (Alberta Education, 2006), which is in line with the curriculum for Mandarin-L1 students in Mainland China (MOE P.R.China, 2012). However, given the diverse student population, spoken language proficiency should not be assumed (Duff & Doherty, 2019). Pronunciation should be considered an explicit goal instead of an indirect goal that is often attached to *Pīnyīn*, and students "should … first know how to pronounce (Teacher Liu)." In addition, a continuous focus on pronunciation will help teachers justify the incidental pronunciation instruction in higher grades without feeling reluctant that they are "no longer supposed to learn *Pīnyīn* (Teacher Cheng)."

Third, teachers reflected on their needs in teaching materials. "There's no fixed textbook in North America (Teacher Bai)." Therefore, teachers "could be flexible in teaching (Teacher Yi)" and avoid the cultural inappropriateness of imported textbooks (Duff & Doherty, 2019). However, it also meant that teachers had to design lessons by themselves, and "the workload is relatively large (Teacher Bai)." In addition to textbooks, teachers reported that the multimedia resources online were "not so accurate," "not so appropriate (Teacher Yi)," and sometimes "could only be used in China (Teacher Feng)." Therefore, we advocate for teaching materials that are accessible, accurate, and age- and culturally appropriate for bilingual children in Canada.

Fourth, PD opportunities in pronunciation instruction should be provided considering teachers' diverse backgrounds. Teacher Gwok stated, "Our mother tongue is almost English … So I think if we can have more training classes, let us go every year, we can learn more." Meanwhile, teachers emphasized that instead of theoretical knowledge, they were interested in

"how to apply them in practical use (Teacher Yi)," i.e., "something that I can use right away (Teacher Gwok)."

Discussion

This study presented teachers' discussions on Mandarin pronunciation teaching and learning in a two-way bilingual program in Canada. Five main themes emerged from the interviews (Figure 4.1), which demonstrated how pronunciation teaching is impacted by linguistic, individual, and social factors, how teachers use both direct and indirect strategies to actively address the challenges of Mandarin pronunciation teaching, and how teachers need more teaching time, improved curriculum design, practical PD opportunities, and teaching materials. Through sharing teachers' lived experiences, we hope to raise awareness of bilingual children's pronunciation learning in non-English minority languages among three groups of stakeholders: educators, researchers, and policymakers.

Teachers practice flexibly as policymakers to negotiate between language education policies and the more complex realities (Menken & García, 2010). Students are expected by the curriculum design to achieve increased proficiency in Mandarin (Alberta Education, 2006), but Mandarin phonology is challenging to learn given its linguistic uniqueness. Therefore, teachers utilize a variety of FFI strategies to target challenging speech units and help students improve their skills. In terms of individual factors, the program is designed to provide bilingual education for any students despite their language backgrounds (ECBEA, n.d.), but it is challenging to deliver a bilingual program to students with different levels of proficiency (Schwartz et al., 2016). Therefore, teachers provided gradual and continuous support without assuming the students' proficiency. As for social factors, the curriculum assigns limited class time to Mandarin pronunciation instruction, and the society is English-dominant which reduces students' exposure

to Mandarin outside of the classroom. Therefore, teachers used CLT strategies to balance the bilingual space and encourage communication in Mandarin. It needs to be acknowledged that teachers' strategies in this study emerged from their lived experiences in a bottom-up manner, therefore featuring individual differences. Nonetheless, this study celebrates these insights and initiatives of teachers and shares their successful experiences as practical evidence with bilingual teachers worldwide as an intercultural community who might be faced with similar challenges of teaching pronunciation of a minority language (Fishman, 1976).

For researchers, pronunciation learning in non-English minority languages has been understudied, which leaves out several research gaps to be further explored. Teachers qualitatively reflected on students' pronunciation learning, yet quantitative data on this topic is rare (e.g., Menke, 2017; Nance, 2020). More developmental studies in a variety of minority languages are needed to examine whether bilingual students improve their pronunciation skills and whether the L2 learners can catch up to the L1-speaking students. Similarly, teachers discussed the effects of teachers' language backgrounds. Evidence suggested that English teachers' status of nativeness did not impact speech learning outcomes (Levis et al., 2016), but the learners were university students, and the study was conducted in the United States where English is the majority language. Studies are needed in minority languages and should consider not only the pronunciation outcomes but also the social and cultural values of native and non-native teachers. Moreover, teachers reported using a variety of CLT and FFI techniques, but further research is needed to decide the effectiveness of the integration of these approaches (e.g., Isaacs, 2009).

For policymakers, there are several issues that can be considered to facilitate pronunciation teaching and learning in bilingual education. When designing the curriculum,

policymakers can consider the optimal proportion of immersion and increase the time of exposure to the minority language (Genesee, 2004); They can also address pronunciation as an explicit learning goal, especially for L2 learners, and revisit the timeline of pronunciation goals so that it can be taught early and continuously (c.f., a French immersion curriculum, Alberta Education, 2008). In addition, teaching resources and PD in pronunciation instruction are frequently called for (Couper, 2021; Estrada & Chacón, 2015; Wisecup, 2017). Policymakers can provide accessible and appropriate teaching resources and PD opportunities to facilitate teachers' practice, and the strategies teachers shared in this study can serve as a starting point for PD programs that involve phonological/phonetic knowledge and practical methods to teach pronunciation. After all, the merits of two-way bilingual education and the powerful role of bilingual teachers need to be supported by evidence-based policies and resources, and the development of bilingual education programs should be viewed as a collaborative and dynamic process.

Acknowledgements

The authors would like to thank the schools, principals, and teachers who organized and participated in the interviews. Thanks to the Edmonton Chinese Bilingual Education Association (ECBEA) for their continued dedication to bilingual education and research. Thanks to Ms. Maria Pollock, Ms. Xiaozhu Chen, and Ms. Minjia Tao for their help with interview transcription and coding. Thanks to Dr. Andrea A.N. MacLeod for providing insights during early discussions of this study and providing advice and suggestions for a poster presentation that was related to this manuscript. Thanks to Ms. Yina Liu for suggestions on the manuscript.

Disclosure

The authors report there are no competing interests to declare.

Funding

This work was supported by the Social Sciences and Humanities Research Council (SSHRC) under the Insight Grant [435-2017-1086] and the Vanier Canada Graduate Scholarship through SSHRC (CGV — 163274).

References of Chapter 4 (Paper 2)

Alberta Education. (2006). *Chinese Language Arts Kindergarten to Grade 9.* International Languages, Programs of Study. Retrieved on February 12, 2023, from https://education.alberta.ca/international-languages-k-6/programs-of-study/

Alberta Education. (2008). *Programme d'éducation pour la maternelle*. French Language Arts (M à 12), Programmes d'études. Retrieved on February 12, 2023, from https://education.alberta.ca/french-language-arts-m-%C3%A0-12/programmes-d%C3%A9tudes/

Asia Pacific Foundation of Canada. (2013). *Canada's Asia Challenge: Creating Competence for the Next Generation of Canadians*. Retrieved on February 12, 2023, from http://www.asiapacific.ca/sites/default/files/filefield/asia_competence_tf_-_final_revised_report.pdf

Austin, Z., & Sutton, J. (2014). Qualitative research: getting started. *The Canadian Journal of Hospital Pharmacy, 67*(6), 436-440.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly, 48*(1), 136-163.

Baker, C. (2011). *Foundations of bilingual education and bilingualism (5th ed.).* Multilingual Matters.

Baker, C., Andrews, H., Gruffydd, I., & Lewis, G. (2011). Adult language learning: a survey of Welsh for Adults in the context of language planning. *Evaluation & Research in Education, 24*(1), 41-59.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). John Benjamins.

Çagatay, S. (2021). Identity clashes of EFL instructors in Turkey with regard to pronunciation and intonation in English. *International Journal of Curriculum and Instruction, 13*(3), 2557-2584.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge University Press.

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America, 129*(6), 3964-3980.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique. 8*(45), 24-27.

Chen, X., Anderson, R. C., Li, W., Hao, M., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology, 96*(1), 142-151.

Copland, F., & Yonetsugi, E. (2016). Teaching English to young learners: Supporting the case for the bilingual native English speaker teacher. *Classroom Discourse, 7*(3), 221-238.

Couper, G. (2021). Pronunciation Teaching Issues: Answering Teachers' Questions. *RELC Journal, 52*(1), 128-143.

Derwing, T. M. (2020). Integrating multiple views and multiple disciplines in the understanding of child bilingualism and second language learning. In F. Li, K. E. Pollock, & R. Gibb

(Eds.), *Child bilingualism and second language learning: Multidisciplinary perspectives* (pp. 29-41). John Benjamins Publishing Company.

Derwing, T. M., Fraser, H., Kang, O., & Thomson, R. I. (2014). L2 accent and ethics: Issues that merit attention. In A. Mahboob & L. Barratt (Eds.), *Englishes in multilingual contexts* (pp. 63-80). Springer, Dordrecht.

Dicks, J., & Genesee, F. (2017). Bilingual education in Canada. *Bilingual and Multilingual Education,* 453-467.

Duff, P., & Doherty, L. (2019). Learning Chinese as a heritage language: Challenges, issues and ways forward. In C-R. Huang, Z. Jing-Schmidt, & B. Meisterernst (Eds.), *The Routledge handbook of Chinese applied linguistics* (pp. 149-164). Routledge.

Estrada, P. L., & Chacón, A. M. (2015). A teacher' s tensions in a Spanish first grade two-way bilingual immersion program. *Revista Portuguesa de Educação, 28*(2), 211-229.

Fishman, J. A. (1976). Bilingual Education: An International Sociological Perspective.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R.Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*(5), 3125-3134.

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal,* 1-22.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and Motivation in Second-Language Learning.* Newbury House Publishers.

Genesee, F. (2004). What do we know about bilingual education for majority language students?. In T.K. Bhatia & W. Ritchie (Eds.), *Handbook of bilingualism and multiculturalism* (pp. 547 – 576). Blackwell.

Genesee, F., & Lindholm-Leary, K. (2008). Dual language education in Canada and the USA. *Encyclopedia of Language and Education, 2*(5), 253-263.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal,* 1-12.

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity.* Oxford University Press.

Lee-Kim, S. I. (2014). Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association, 44*(3), 261-282.

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation, 6*(3), 310-328.

Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly, 50*(4), 894-931.

Levis, J., & Sonsaat, S. (2017). Pronunciation teaching in the early CLT era. In O. Kang, R.I. Thomson, & J.M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 267-283). Routledge.

Li, F., & Munson, B. (2016) The development of voiceless sibilant fricatives in Putonghua-speaking children. *Journal of Speech, Language, and Hearing Research,* 59, 699-712.

Lin, L. C., & Johnson, C. J. (2010). Phonological patterns in Mandarin–English bilingual children. *Clinical Linguistics & Phonetics, 24*(4-5), 369-386.

Littlewood, W. (2011). Communicative language teaching: an expanding concept for a changing world. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2). Routledge.

Liu, W. (2020). Success factors for a Mandarin bilingual program: An autoethnographic case study. *Chinese as a Second Language (漢語教學研究—美國中文教師學會學報), 55*(3), 208-229.

MacDonald, S. (2002). Pronunciation-views and practices of reluctant teachers. *Prospect 17*(3), 3-18.

MacLeod, A. A., & Stoel-Gammon, C. (2010). What is the impact of age of second language acquisition on the production of consonants and vowels among childhood bilinguals?. *International Journal of Bilingualism, 14*(4), 400-421.

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

Menke, M. R. (2017). Phonological development in two-way bilingual immersion: The case of Spanish vowels. *Journal of Second Language Pronunciation, 3*(1), 80-108.

Menken, K., & García, O. (2010). *Negotiating language education policies: Educators as policymakers.* Routledge.

Meziane, R. S. & MacLeod, A. A. N. (2020). Integrating multiple views and multiple disciplines in the understanding of child bilingualism and second language learning. In F. Li, K.E. Pollock, & R. Gibb (Eds.), *Child bilingualism and second language learning: Multidisciplinary perspectives* (pp. 9-27). John Benjamins Publishing Company.

Ministry of Education of the People's Republic of China (MOE P.R.China). (2012). *The Elementary School Chinese Curriculum in Compulsory Education Period – Revised*. People's Education Press. [In Chinese]

Murphy, D. (2011). An investigation of English pronunciation teaching in Ireland. *English Today, 27*(4), 10-18.

Nance, C. (2021). Scottish Gaelic revitalisation: Progress and aspiration. *Journal of Sociolinguistics, 25*(4), 617-627.

Netelenbos, N., Li, F., & Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism, 20*(3), 346-357.

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning, 50*(1), 57-85.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism, 1*(3), 213-237.

Place, S., & Hoff, E. (2016). Effects and non-effects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition, 19*(5), 1023-1041.

Richter, K. (2019). *English-medium instruction and pronunciation: Exposure and skills development.* Multilingual Matters.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46-73.

Schwartz, M., Wee Koh, P., Xi Chen, B., Sinke, M., & Geva, E. (2016). Through the lens of teachers in two bilingual programmes: a look at early bilingual education. *Language, Culture and Curriculum, 29*(2), 141-168.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*(2), 273-293.

Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly, 42*(2), 181-207.

The Edmonton Chinese Bilingual Education Association (ECBEA). (n.d.). *About the program. Edmonton Chinese Bilingual Education Association*. https://www.ecbea.org/about-the-program/

Tokumoto, M., & Shibata, M. (2011). Asian varieties of English: Attitudes towards pronunciation. *World Englishes, 30*(3), 392-408.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349-357.

Trotter, R. T. (2012). Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives. *Preventive Medicine, 55*(5), 398-400.

Werker, J. F. & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development, 1*(2), 197-234.

Wisecup, P. A. (2017). Teacher Voices Regarding a Bilingual Education Program: A Qualitative Inquiry (Publication No. 10616769) [Doctoral dissertation, Northcentral University]. ProQuest Dissertations Publishing.

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*(1), 141-151.

Wu, J. Z. (2005). Bilingual education in western Canada and Chinese language minority students' self perceptions of their citizenship and ethnicity. *Canadian & International Education, 34*(1), 23-30.

Xu, W., & Case, R. E. (2015). Age-related differences in motivation in learning English among Mainland Chinese students. *International Journal of Applied Linguistics, 25*(1), 67-82.

Zhu, H. (2002). *Phonological Development in Specific Contexts: Studies of Chinese-Speaking Children.* Multilingual Matters.

Chapter 4 (Paper 2) presented teachers' perspectives on bilingual students' Mandarin pronunciation. Almost all teachers nominated Mandarin lexical tones as a challenging phonological dimension for bilingual children in Canada to learn. Teachers identified specific tone targets that were challenging to produce (e.g., Tone3) and tone pairs that were difficult to differentiate (e.g., Tone2 and Tone3). Meanwhile, teachers discussed how bilingual students' pronunciation learning could be influenced by their home language backgrounds and grade levels in the bilingual program. Teachers also recognized that Mandarin input was in general limited for bilingual students given the English-dominant environment in society. Subsequently, Paper 3 (Chapter 6) uses quantitative methods to verify teachers' observations and intuitions and investigates how bilingual students' Mandarin lexical tone production is influenced by multifaceted factors.

Meanwhile, a focus on tones was justified by quantitative observations of the students' speech production. In a larger project, the full sample of 165 students produced single words spontaneously with an average Percentage of Consonants Correct (Shriberg et al., 1997) of 89.11 (*SD* = 7.42) and a Percentage of Vowels Correct of 93.73 (*SD* = 5.86), whereas their Percentage of Tones Correct was 77.80 (*SD* = 14.77). This suggested that bilingual students' productions of tones featured lower accuracy and higher variability compared to consonants and vowels.

Before proceeding to this empirical study, Chapter 5 introduces some methodological innovations during the work of this dissertation: two functionalities in the software program Phon (Rose & Hedlund, 2020) designed to facilitate the research on Mandarin speech development. This chapter contains details that cannot be fully included in the manuscript in Chapter 6 and will be developed into a tutorial paper for future dissemination.

Chapter 5 Methodologies to Study Tone Productions

This chapter specifies two methodological innovations in studying Mandarin tones using perception-based (transcription) in Phon (Hedlund & Rose, 2020): A Chinese-IPA dictionary and tone match analyses, including the PTC of each speaker, tone patterns, and the match of each token production. These new functionalities can facilitate researchers around the world to investigate tone productions in a variety of tonal languages.

Phon, a software program for phonological data

Phon is a software program for phonological data corpus management that is widely used in the field of child speech development (Rose & Stoel-Gammon, 2015). Phon's user interface (Session Editor) allows researchers to conduct streamlined transcription: First, researchers start with a list of words that are orthographically transcribed. Then, Phon provides dictionaries in 20 languages that can convert the orthography into broad transcriptions of the target words using the International Phonetic Alphabet (IPA). Subsequently, researchers can modify the transcription of the speaker's production based on audio or video recordings and optionally with the assistance of acoustic information (Figure 5.1).

With a transcribed corpus, Phon provides rule-based syllabifiers and aligners in more than 20 languages (e.g., Cree, English, French, Urdu) to syllabify and align transcriptions. Therefore, Phon is able to take syllabic information into account and provide solutions for phone-by-phone match/mismatch analysis by comparing a transcription of a speech sample against the target IPA transcription (Figure 5.2).

Building on such functionality, Phon provides solutions for researchers to run queries on certain speech sounds based on their phonological features and word positions (e.g., final stops,

nasal onsets) and obtain basic psychometric scores such as Percentage of Consonants Correct (PCC, Shriberg et al., 1997). However, until 2019, Phon was unable to auto-transcribe Mandarin speech samples or analyze tone match. Therefore, I worked with the developers of Phon (Dr. Yvan Rose and Mr. Gregory Hedlund) to develop two functionalities: (1) a Chinese dictionary to convert Chinese characters or *Pīnyīn* into IPA and (2) an accuracy analysis for tones.

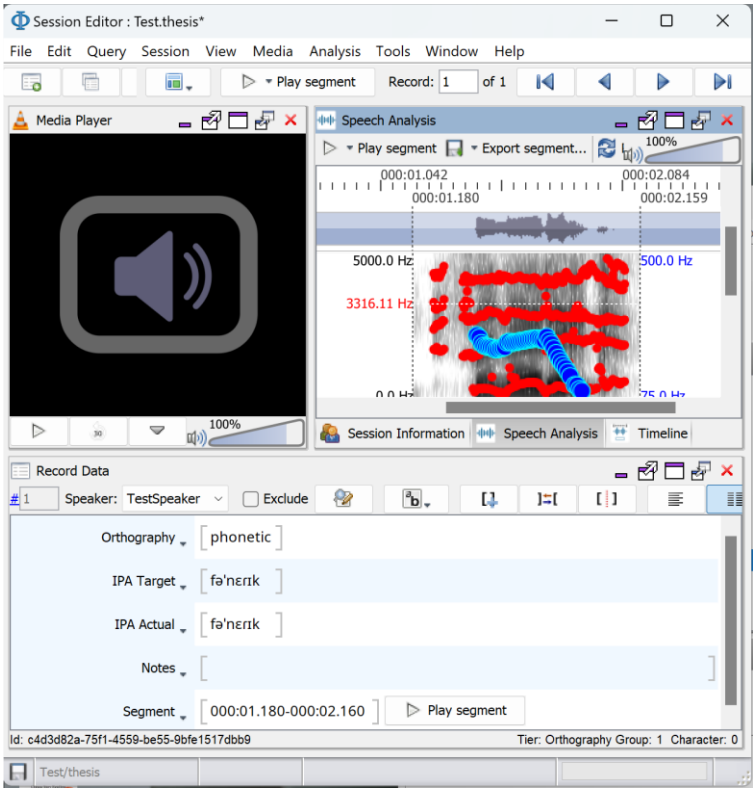


Figure 5.1. Phon's Session Editor with a media player on the top-left, a speech analysis window on the top-right, and a session data window on the bottom for streamlined IPA transcription.

Figure 5.2. The production "phonetic /fəˈnɛɾɪk/ → [ˈnɛɾɪt]" with incorrect (left) and correct (right) syllabification and alignment in Phon.

A Chinese IPA dictionary

A *Pīnyīn*-to-IPA dictionary was first developed based on the exhaustive list of more than 400 toneless syllables in Mandarin (Appendix G). The IPA transcription referred to common conventions in Chinese linguistics (The Office of Modern Chinese PKU, 2006; Table 3.1) but transcribed the first phone as a glide in a diphthong or triphthong that begins with *i* [i], *u* [u], or *ü* [y] (i.e., *ia* is transcribed as [ja], and *üan* is transcribed as [ɥɛn], Duanmu, 2007; Lin et al., 2020) and transcribed other diphthongs and the last two phones in triphthongs with ligatures (i.e., *ou* is transcribed as [o͡ʊ], and *iou* is transcribed as [jo͡ʊ]). This allows cross-linguistic comparison between Mandarin and English (e.g., sway [swe͡ɪ] and 碎 *sui* [swe͡ɪ] as opposed to [suei]). With the understanding that there are multiple ways to phonetically transcribe Mandarin, this operation maximizes the differentiation among the three phones in a triphthong and allows researchers to adapt the transcriptions to their preferred convention through text replacing (e.g., replace [w] with [u]).

The dictionary can also convert tone types in *Pīnyīn* into the 5-scale convention that is commonly used in Chinese linguistics (Chao, 1930): Tone1 [55], Tone2 [35], Tone3 [214], and Tone4 [51], where [5] represents the highest pitch and [1] represents the lowest pitch in the speaker's tonal range. The dictionary requires that in *Pīnyīn*, tones are marked as an Arabic numeral after each syllable (e.g., *yu3yin1* for *yǔyīn*), which is a common convention in natural language processing in Chinese. The dictionary also allows the tone type of 0, which indicates the neutral tone on a stressless syllable (Duanmu, 2007) and is transcribed as [0] in IPA.

Furthermore, the dictionary implements basic tone sandhis (phonological and contextual changes of tones) in Mandarin based on phonological rules. Mandarin has two major types of phonological sandhis on Tone3: full sandhi and semi-sandhi (Yang, 2016). In a full sandhi,

Tone3 /214/ is realized as [35] when followed by another Tone3, which is phonetically identical to a Tone2 /35/ (e.g., 水果 *shui3guo3* [ʂweɪ̯$^{\mathbf{35}}$kwo^{214}]). In a semi-sandhi, Tone3 /214/ is realized as [21], a "half" Tone3, when it is followed by a syllable that is not Tone3 (e.g., 水杯 *shui3bei1* [ʂweɪ̯$^{\mathbf{21}}$peɪ̯^{55}]). Semi-sandhi can be a free variation when Tone3 is produced at the end of an utterance if the utterance is not monosyllabic (e.g., 喝水 *he1shui3* [xɤ^{55}ʂweɪ̯$^{\mathbf{214}}$] or [xɤ^{55}ʂweɪ̯$^{\mathbf{21}}$]). However, these are only the sandhi rules within a word. The sandhi rules apply first within word boundaries and then across words. When a word-final Tone3 is followed by another word that starts with Tone3, it also experiences the full sandhi (e.g., 水果好 *shui3guo3 hao3* [ʂweɪ̯^{35}kwo$^{\mathbf{35}}$ xɑʊ̯^{214}]). On the contrary, if a word-initial Tone3 has already experienced full sandhi, a word-final Tone3 preceding it does not go through the full sandhi because the following Tone3 has already surfaced as [35] (e.g., 买水果 *mai3 shui3guo3* [maɪ̯$^{\mathbf{21}}$ ʂweɪ̯^{35}kwo^{214}]. These were achieved in the dictionary through the phonological rules in Appendix H.

Finally, this *Pīnyīn*-to-IPA dictionary was coupled with an existing, open-access Chinese-to-*Pīnyīn* dictionary, CC-CEDICT (MDBG, 2023). Parsed words in Chinese characters are converted into *Pīnyīn* using CC-CEDICT and then transcribed as IPA. Since CC-CEDICT provides a dictionary in both Simplified and Traditional Chinese, these two orthographical options are made available in Phon. The *Pīnyīn*-to-IPA and Chinese-to-IPA dictionaries were published in Phon in 2021, freely available for researchers around the world.

Tone match (accuracy) analyses

To investigate bilingual students' tone production and facilitate research on Chinese speech, an accuracy analysis, Percentage of Number Tones Correct (PNTC), was developed in Phon. The term "Number Tones" was used because this function applies to tones transcribed using tone

numbers, the superscript numbers that appear after a syllable (e.g., 音 *yīn* [in^{55}]), as opposed to other symbols of tones (e.g., [in˥] or [ǐn]). The term "Correct" was used to be consistent with the other percent correct measurements such as PCC (Shriberg et al., 1997), acknowledging that correct/incorrect was judged according to the match/mismatch between the target pronunciation and transcribers' perceived pronunciation with no direct relations to speech errors or disorders. PNTC is calculated using Equation 5.1 (Hedlund & Rose, 2020), and a report like Figure 5.3 is generated after running the PNTC analysis on one or multiple sessions (files), each with one or multiple target speakers. Like any other percent correct analyses in Phon, the PNTC analysis allows researchers to select their words, syllables, and speakers of interest based on certain criteria. For simplicity, this function is referred to as "PTC" below as this dissertation adopts Chao's (1930) numeric tone letters and always uses the tone numbers in Phon to transcribe tones.

$$\begin{aligned} PNTC &= correct \,/\, (correct + incorrect) \times 100 \\ &= correct \,/\, (correct + deletion + substitution) \times 100 \end{aligned} \tag{5.1}$$

Speaker	# Target	# Actual	Syllables Deleted	Syllables Epenthesized	# Correct	# Substituted	# Deleted	PNTC
TestSpeaker1	4	4	0	0	2	2	0	50
TestSpeaker2	2	2	0	0	2	0	0	100
	6	6	0	0	4	2	0	66.667

Figure 5.3. A PNTC report in Phon.

Moreover, the PTC function is able to generate a tone pattern summary (Figure 5.4) that facilitates pattern analyses. For example, in Figure 5.4, it is clear that Tone2 /35/ was transcribed as Tone1 [55] once, and Tone3 /214/ was transcribed as Tone2 [35] once, whereas Tone4 /51/ was always perceived as the intended target in this test speech sample. Moreover, the numeric convention allows finer phonetic transcription. For example, a production of target Tone1 /55/ can be transcribed by listeners as [54], which may not match any tone categories in Mandarin. In this case, although the production is marked as "substituted" in the report, it is not necessarily

perceived as a different phonemic category. Researchers may choose the level of transcription narrowness based on their needs.

Target Tone	Actual Tone	Test.thesis	# Correct	# Substituted	# Deleted
35	55	1	0	1	0
51	51	2	2	0	0
55	55	2	2	0	0
55	54	1	0	1	0
214	35	1	0	1	0

Figure 5.4. A PTC tone pattern summary in Phon.

Perhaps more relevant to the current study, the PTC function can generate a detailed "PTC by Word" list which includes the file name of the speech sample, the speaker ID, the record number, the orthographical and IPA transcriptions, the number of target tone(s), the target and transcribed tones, and the match (correctness) of each production (Figure 5.5). This list is especially useful for token-level analyses. Compared to an aggregated PTC score for each speaker, a token analysis allows statistical models to control the effects of speakers and word tokens and examine the effects of specific tone targets, which fits this dissertation's goal to identify targets that are challenging due to their phonetic features (Best & So, 2010; Flege & Bohn, 2021).

Session	Speaker	Age	Record #	Group #	IPA Target	IPA Actual	Orthography (Word)	# Target	# Actual	Target Tone	Actual Tone	Syllables Deleted	Syllables Epenthesized	# Correct	# Substituted	# Deleted
Test.thesis	TestSpeaker1		1	1	ta^{55}	ta^{55}	搭	1	1	55	55	0	0	1	0	0
Test.thesis	TestSpeaker1		2	1	$ta^{55}an^{51}$	$ta^{55}an^{51}$	答案	2	2	35,51	55,51	0	0	1	1	0
Test.thesis	TestSpeaker1		3	1	ta^{214}	ta^{35}	打	1	1	214	35	0	0	0	1	0
Test.thesis	TestSpeaker2		4	1	$ta^{51}tɕja^{55}$	$ta^{51}tɕja^{55}$	大家	2	2	51,55	51,55	0	0	2	0	0

Figure 5.5. A PTC token analysis report in Phon.

Chapter summary

Two methodological innovations to transcribe and analyze tone productions were introduced in Chapter 5: A Chinese-IPA dictionary and tone match analyses, including the PTC of each speaker, tone patterns, and the match of each token production. In the next chapter, these transcription-based methods are used in conjunction with acoustic measurements to quantitatively analyze bilingual students' Mandarin tone productions.

Chapter 6 Bilingual Students' Production of Mandarin Lexical Tones

Lin, Y., Pollock, K. E., & Li, F. (in preparation). Speech production of Mandarin lexical tones among Canadian elementary students enrolled in Chinese-English bilingual schools.

Speech production of Mandarin lexical tones among Canadian elementary students enrolled in Chinese-English bilingual schools

Abstract

Purpose: This study investigates how Mandarin-English bilingual students in Canada produce Mandarin tones and how this is influenced by intra-language factors (e.g., phonemic and phonetic complexity of speech targets), inter-language factors (e.g., cross-linguistic influences), and extra-language factors (e.g., home input and school input).

Method: Participants were 82 students enrolled in a Chinese bilingual program in Western Canada. A cross-sectional design was used, and students were recruited from Grades 1, 3, and 5. Students were divided into two groups based on their home language backgrounds: The heritage language (HL) group (N = 38) had early and strong home input in Mandarin, and the second language (L2) group (N = 44) received mostly English input at home. Tone productions were audio-recorded through a single-word elicitation task. Samples were then transcribed by Mandarin-native listeners for match (accuracy) and pattern analyses. In addition, acoustic measurements were extracted to validate the transcriptions and provide phonetic detail.

Results: First, Tone3 was challenging across groups due to its complexity, suggesting the effects of intra-language factors. Second, because English does not assign as much linguistic significance to pitch characteristics, L2 students' tone learning was impacted by this inter-language factor and showed more signs of categorical confusion. Third, increased tone match rates were related to both home input and school input, but bilingual students did not achieve ceiling match rates in Grade5. Instead, L2 students produced certain phonetic features less

accurately in higher grades. This was attributed to reduced pronunciation instruction and limited home and community input.

Conclusions: Results in this study suggest that bilingual students' speech development in a minority language has similarities to monolingual children's speech development documented in previous studies but also shows unique influences of language transfers and speech input. This study provides evidence for bilingual speech theories in the suprasegmental domain and has implications for pronunciation learning and teaching of a minority language in the context of bilingual education.

Keywords: bilingual education, pronunciation, tone, children, acoustic

Introduction

Bilingual speech development is often different from monolingual speech development due to (1) interactions between two phonological systems and (2) varied quantity and quality of input in both languages (Baker & Trofimovich, 2005). There are several theoretical frameworks to account for the difficulties in second language (L2) speech learning or bilingual speech development, for example, the Perceptual Assimilation Model (PAM, Best & Tyler, 2007) and Speech Learning Model – Revised (SLM-r, Flege & Bohn, 2021).

Both theories discuss interactions between learners' first language (L1) and L2 phonological systems. Specifically, they address how learners' perception is attuned by their L1 which influences the perception and production of L2 phonological contrasts (Best & Tyler, 2007) and phonetic categories (Flege & Bohn, 2021). However, these theories primarily focus on the relationships between segmental categories (speech sounds) in L1 and L2 and provide limited accounts for how suprasegmental categories are mapped between L1 and L2 (So & Best, 2010).

In particular, it remains unclear how children in a non-tonal language environment learn tones in L2. Lexical tones are critical phonological categories in tonal languages such as Mandarin. In tonal languages, pitch and other suprasegmental features are used to contrast word meanings. Despite their early acquisition in native speakers (Holm & Dodd, 2006; Zhu, 2002), tones are challenging for L2 learners with a non-tonal background. In a non-tonal language environment, infants learn to ignore suprasegmental information in word recognition at 9 months of age for more efficient lexical processing (Singh et al., 2008). When adult speakers of a non-tonal language have to re-attend to suprasegmental information in an L2, their tone productions are prone to errors, and their speech can be harder to understand (Hao, 2012; Yang, 2016). Research on English-speaking children's production of Mandarin tones can provide insight into the ongoing process of speech development in tonal-non-tonal language pairs among younger learners and expand L2 speech theories into the suprasegmental domain.

L2 speech learning theories stress the important roles of L2 input. Such input may include previous, cumulative input and recent/current input (Bedore et al., 2016). In the field of language learning (e.g., grammar and vocabulary), some studies found that cumulative input (quantified by age of acquisition and length of exposure) could explain most variations in children's L2 learning outcomes (Birdsong, 2005), but others found that children's current L2 input and output were more explanatory (Unsworth et al., 2014). Less is known about the relationship between cumulative and current input in speech (pronunciation) learning. There is reason to believe that early input shapes learners' speech perception (Best & Tyler, 2007; Flege, 1995). On the other hand, theorists believe that L2 learning is a lifelong, dynamic process influenced by recent input (Flege & Bohn, 2021). There is evidence indicating that school-aged L2 learners can catch up with their L1 peers given that they currently receive intensive, high-

quality L2 input (Menke, 2017; Nance, 2020). Comparisons between heritage speakers and L2 learners can help understand the relationship between cumulative and current L2 input: Heritage speakers receive early speech input from their home environment, which is unavailable for L2 learners. However, the proportion of home language input may become less as they are more exposed to the societal majority language. With limited input, heritage speakers may experience L1 attrition, and their speech performance may become similar to their L2 peers, who have received increased current input (Chang et al., 2011).

The current study presents a unique population of school-aged children who are enrolled in a Chinese-English two-way bilingual program in Western Canada, including heritage speakers of Mandarin who received both early Mandarin input at home and current input in school and L2 learners of Mandarin who recently started receiving Mandarin input at school age. This study examines bilingual students' lexical tone productions, a unique phonological dimension that is non-existent in English. Its results provide evidence of the relationship between home and school input and shed light on the learning of suprasegmental features in a societal minority language. The following subsections review the literature on bilingual children's tone learning, summarize factors that impact such learning, and introduce the Chinese bilingual program as a testing field of theoretical accounts for bilingual tone learning. At the end of this section, specific research questions are posed for the current study.

Evidence on bilingual children's tone development

Research evidence on bilingual children's tone learning is less available than that of speech sound learning and shows mixed results. On one hand, research documented similar development between bilingual and monolingual children. Holm and Dodd (2006) found that young Cantonese-L1 children in Australia (ages 2;0 to 5;7) who sequentially learned English had

similar accuracy rates compared with L1 peers in Hong Kong. Mok and Lee (2018) also found that simultaneous Cantonese-English bilingual children in Hong Kong (ages 2;0 to 2;6) had comparable tone match rates with monolingual peers, and the two rising tones in Cantonese with similar *f0* contours were the most challenging in spite of language backgrounds. On the other hand, different developmental patterns of tones were found in bilingual children compared to monolingual peers. Despite the comparable match rates, bilingual children showed a high-low template in disyllabic tone productions, indicating the influences of English stress patterns (Mok & Lee, 2018). Yao et al. (2020) found that Urdu-Cantonese bilingual children in Hong Kong (ages 4;5 to 6;6) were more prone to tone mismatches than monolingual peers. Kan and Schmid (2019) found that school-aged (ages 5 to 11) Cantonese-English bilingual children in the US scored lower than monolingual peers in tone perception. These varied results may be due to the variety of factors that can impact bilingual speech learning, such as language environment (e.g., Australia versus Hong Kong, home versus school environments), L1 transfer (e.g., Urdu-L1 versus English-L1), and specific tone targets (e.g., monosyllabic versus disyllabic tones). These factors will be reviewed in the next section and guide the current study.

Studies reviewed above focused on Cantonese. Mok and Lee (2018) advocated for more child studies involving a variety of tonal languages, including Mandarin, to expand the understanding of cross-linguistic interactions in the suprasegmental domain. Studies as such mostly focused on perception. School-aged bilingual children exhibited categorical perceptions for tonal continua similar to Mandarin monolingual children and as opposed to the continuous perception of English monolingual children (Yang & Liu, 2012) and showed similar skills of tone discrimination to their monolingual peers (Marinova-Todd et al., 2010). However, accurate

categorization does not necessarily imply accurate production of phonetic characteristics. It remains unclear how school-aged children further develop their tone productions in Mandarin.

Factors affecting bilingual children's learning of Mandarin tones

Synthesizing the aforementioned evidence and theories of L2 speech development, bilingual children's tone learning may be impacted by three levels of factors: (1) intra-language factors, i.e., phonemic and phonetic complexity of the targets, (2) inter-language factors, i.e., transfer effects between L1 and L2, and (3) extra-language factors, such as learners' language experiences, including the quantity and quality of input they receive across different periods of time and social contexts (Bedore et al., 2016; Paradis, 2011). This means bilingual children's pronunciation learning is similar to that of monolingual children to some extent, but there are unique developmental patterns related to cross-linguistic effects and language experiences.

Intra-language factors: Phonemic and phonetic complexities of targets

Evidence in monolingual children's tone acquisition provides insights into the complexity of learning targets. Mandarin has four citation tones in isolated, monosyllabic productions: a high-level tone (Tone1), a mid-rising tone (Tone2), a low-dipping or falling-rising tone (Tone3), and a high-falling tone (Tone4). In Chinese linguistics, tones are often transcribed in the 5-scale convention (Chao, 1930): Tone1 [55], Tone2 [35], Tone3 [214], and Tone4 [51], with [5] representing the highest pitch and [1] representing the lowest pitch in the speaker's tonal pitch range. In addition to the pitch contrasts, Mandarin tones are produced with different durations: Tone3 is usually produced with the longest duration, and Tone4 is the shortest (Xu, 1997; Yang, 2016). Meanwhile, the low pitch in Tone3 [214] often co-occurs with a creaky voice (Kuang, 2017). All these phonetic characteristics of tones were shown to contribute to listeners'

perception (Blicher et al., 1990; Rhee et al., 2020). Due to space limitations, this study focuses on pitch and duration information in tone productions. Figure 6.1 presents the four citation tones produced by a female Mandarin-L1 speaker.

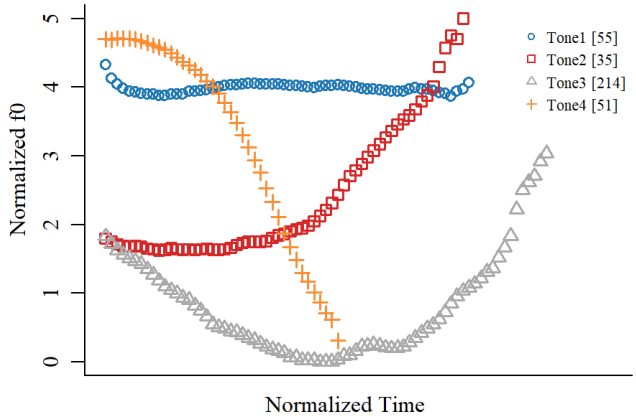


Figure 6.1. Mandarin citation tones produced by a female native speaker. The x-axis represents time (ms), and the y-axis represents normalized fundamental frequency (*f0*) using Equation 6.1.

In terms of the age of establishment, Mandarin citation tones as a whole category are established as early as 1.5 years of age in monolingual children (Zhu, 2002). Zhu (2002) believed that tone contrasts are mastered early because they are phonologically salient (i.e., tones are compulsory and differentiate lexical meanings) and perceptually salient. Moreover, the contrasts between tones are simpler than those between speech sounds. However, when examined through acoustic analyses, researchers found that monolingual children's tone productions were not yet adult-like at the age of five. Such acoustic differences existed despite the use of normalization methods to account for age differences (Xu Rattanasone et al., 2018) and were related to lower match rates (Wong, 2012). Some researchers believed that the protracted refinement of tone productions was related to the development of muscle physiology and motor skills in young children (Wong, 2012), and others believed that it was a process of learning to integrate phonetic

cues in perception and production (Rhee et al., 2020). In either case, the protracted refinement of tone productions indicated the complexity of Mandarin tones as phonetic targets, despite the early development of the phonemic categories. Therefore, it is important to integrate transcription-based analyses and acoustic measurements when examining tone learning.

In terms of developmental order, Tone1 and Tone4 are early-developmental among monolingual children, and Tone2 and Tone3 are late-developmental (Zhu, 2002; Wong, 2012). Evidence in adult L2 learners is comparable, suggesting that Tone3 productions had the lowest match and were frequently perceived as Tone2 (Yang, 2016; Wang et al., 2003). The challenge of producing Tone3 can be attributed to both phonetic and phonological reasons: First, Tone3 is difficult to articulate with its compound *f0* contour (falling-rising) and extremely low pitch (Wong, 2012). Second, Tone3 is phonetically confusable with Tone2 because both involve a rising section in their contours (Blicher et al., 1990). Third, to make Tone3 and Tone2 more confusable, Tone3 experiences a full sandhi when followed by another Tone3, realizing as a mid-rising tone that is phonetically identical to Tone2 [35]. Fourth, Tone3's allophonic realization is the most complex among the four citation tones. In addition to the full sandhi, it experiences semi-sandhi when followed by a syllable that is not Tone3 and becomes a short, low-falling tone [21] (Xu Rattanasone et al., 2018; Yang, 2016). In terms of the other late-developmental tone, Tone2, Wang et al. (2003) showed that Tone2 was responsive to perceptual training in L2 learners. Its match rate increased from 64% to 85% after a two-week training program, which made it the second most accurate tone. Therefore, the current study hypothesizes that in bilingual students' productions, Tone3 has the lowest match, and Tone1, Tone2, and Tone4 have similarly moderate to high match.

Inter-language factors: Cross-linguistic influences

For bilingual learners, another level of complexity of tone learning is cross-linguistic influences. Among L2 speech theories, PAM made the most explicit effort to specify L1 influences in the suprasegmental domain (PAM-S; So & Best, 2010). PAM-S proposed two hypotheses of English-L1 listeners' perception of Mandarin tones. On one hand, tones may be perceived as prosodic categories in English, such as sentential intonations or lexical stress patterns. For example, Tone2 [35] may be assimilated to question intonation or iambic stress patterns, and Tone4 [51] may be assimilated to statement intonation or trochaic stress patterns (Hallé et al., 2004; So & Best, 2010). On the other hand, tones may be perceived as nonlinguistic melodies (Hallé et al., 2004), and the learning depends on their phonetic properties. Consequently, tones that share phonetic similarities are more confusable, for example, Tone1-Tone4 ([55]-[51]), Tone1-Tone2 ([55]-[35]), and Tone2-Tone3 ([35]-[214]) (So & Best, 2010). Production studies in adult L2 learners confirmed that Tone2-Tone3 was the most confusable pair (Wang et al., 2003). Based on such evidence, it is hypothesized that bilingual students do not assign as much phonemic significance to tones due to the lack of lexical tones in English, so their tone productions have a lower match. Furthermore, tone pairs that can be assimilated into different prosodic categories in English are produced with less confusion (e.g., Tone2-Tone4, which can be assimilated into question and statement intonations, respectively), whereas pairs that are phonetically similar are subjected to substitution (e.g., Tone2-Tone3).

Extra-language factors: Effects of speech input

Different from PAM which focuses on the L1 assimilation effects, SLM-r (Flege & Bohn, 2021) highlights the role of L2 input. SLM-r views speech learning as a life-long, dynamic process, and it is not solely decided by the age of learning (Flege, 1995). The revised SLM states that L2

speech learning is a function of the quantity and quality of L2 input. Such a revision poses questions to the notion "earlier is better," as bilingual pronunciation learning seems to be related to more complex factors such as language status, learners' motivation, and learners' output (see MacLeod & Stoel-Gammon, 2010; Netelenbos et al., 2016), which are then related to the quantity and quality of input.

Although SLM-r did not provide an operational definition of L2 speech input, previous studies on language development suggested examining both early, cumulative input and recent/current input (Bedore, et al., 2016). Accordingly, the current study hypothesizes that bilingual students with early home input in Mandarin (i.e., Mandarin-dominant households) and longer recent earning experiences in school (i.e., higher grade levels) will have higher match rates in their tone productions.

A Mandarin-English two-way bilingual program in Western Canada

English is the majority language in Alberta, Canada, but publicly-funded two-way bilingual education programs are available in many international languages. Among them, the Mandarin-English bilingual program enjoys a 40-year history and attracts not only students from Chinese-heritage backgrounds but also students who learn Mandarin as an L2 in kindergarten through Grade 12. At the elementary level (Grades 1-6), half of the class content is delivered in Mandarin Chinese, and the other half in English. Through the two-way language immersion, students are expected to develop or maintain functional proficiency in the societal minority language, i.e., Mandarin, at no cost of their development of English language skills and academic skills (Alberta Education, 2006).

Evidence has shown that one-way immersion education was insufficient for L2 learners to acquire native-like pronunciation, despite an early age of onset. This was attributed to the limited amount and reduced quality of L2 input in a foreign-language learning setting (Harada, 2007; Netelenbos et al., 2016). On the contrary, the two-way immersion design was considered to be of merit as it provided authentic input from native-speaking teachers and peers of both languages (Cummins, 1979). This was supported by empirical evidence from Spanish-English and Gaelic-English bilingual programs that no articulatory differences were found between heritage learners and L2 learners after a few years of immersion (Menke, 2017; Nance, 2020).

However, teachers of the Mandarin program reported that students had difficulties learning Mandarin tones (Lin et al., 2022). Therefore, gaps existed between learning outcomes as perceived by teachers and the belief in the effectiveness of two-way bilingual education. The multifaceted factors of bilingual pronunciation learning should be taken into account, including the phonemic and phonetic complexity of tone targets, the dissimilarities and cross-linguistic influences between the majority language (i.e., English) and minority language (i.e., Mandarin), and students' early and current speech input in Mandarin at home and in school.

The current study

This study investigates the development of Mandarin lexical tones in students who are enrolled in a Mandarin-English two-way bilingual program in Western Canada. Both transcription-based analyses and acoustic analyses are used to examine students' tone productions. Research questions are guided by the measurements and the multifaceted factors reviewed above: (1) What are bilingual students' tone match rates in general and across specific tone targets? (2) What are the major tone patterns? (3) How do bilingual students produce the multidimensional phonetic

characteristics of Mandarin tones? (4) Meanwhile, what roles do the intra-, inter-, and extra-language factors play in bilingual students' tone production?

To date, no study has examined Mandarin tone productions in school-aged children in a context where Mandarin is a minority language. Thus, the current study will not only provide empirical evidence for L2 speech learning theories in the suprasegmental domain but also has evidence-based implications for pronunciation teaching and learning in bilingual education.

Method

Participants

Students in Grades 1, 3, and 5 at Mandarin bilingual schools were voluntarily registered by their parents. Parents reported either no diagnoses of hearing, speech, language, or learning problems, or previous diagnoses of mild disorders pre-kindergarten (N = 5). All but two participants passed a hearing screening in both ears at the frequencies of 1, 2, and 4 kHz. The two participants who did not pass the screening failed at one or two frequencies in one ear and were referred to their pediatricians for follow-up. Since parents and participants reported no concerns about hearing, these children's speech samples were included in the study.

Among the students, 38 were heritage language speakers of Mandarin (group name "HL") with early and strong Mandarin home input, and 44 were L2 learners (group name "L2") with mostly English home input and late onset of Mandarin. Participant numbers were balanced across grade levels. The two groups' profiles are depicted in Table 6.1. Even within the L2 group, there were usually extended family members who spoke Mandarin or another Chinese language. Within the HL group, most students spoke English as their current dominant language. However, it is still clear that these two groups had different home language environments: HL had early

onset of exposure to Mandarin (0-1 month of age), whereas L2 had late onset of Mandarin (older than 3 years of age) and early onset of English (0-1 months of age); HL had at least one parent who spoke Mandarin ≥ 50% of the time, whereas L2 had both parents who spoke English ≥ 80% of the time; HL had at least one parent who self-reported high proficiency in Mandarin (4 or 5 out of the 0-5 scale), whereas L2 had parents with high proficiency in English and low proficiency in Mandarin. As a result, HL had higher proficiency in Mandarin than L2, indicated by raw scores of the Chinese Peabody Picture Vocabulary Test (CPPVT, Lu & Liu, 1998).

Table 6.1. Demographic information and home language environment of HL and L2 students (Mean (SD) [Minimum, Maximum] for numeric measurements and Median [Minimum, Maximum] for ordinal measurements).

	Heritage language (HL)	Second language (L2)
Participant numbers	Grade1 (G1) = 15 G3 = 11 G5 = 12	G1 = 16 G3 = 14 G5 = 14
Chronological age (months)	G1: 77 (3) [72, 83] G3: 104 (4) [97, 110] G5: 124 (3) [120, 129]	G1: 78 (4) [71, 86] G3: 103 (4) [97, 111] G5: 126 (3) [122, 130]
Proportion of students whose current dominant language is English (%)	G1: 53; G3: 82; G5: 83	100
Onset age of regular exposure to Mandarin (months)*	0 (0) [0, 1]	63 (11) [37, 96]
Onset age of regular exposure to English (months)*	28 (26) [0, 93]	0 (0) [0, 1]
Parent % of time speaking Mandarin to the child**	72 (33) [0, 100] 89 (13) [50, 100]	0 (0) [0, 1] 2 (4) [0, 15]
Parent % time speaking English to the child**	10 (14) [0, 50] 22 (28) [0, 100]	96 (7) [80, 100] 99 (2) [90, 100]
Parent self-reported Mandarin proficiency (scale 0-5)**	5 [0, 5] 5 [4, 5]	0 [0, 1] 1 [0, 3]
Parent self reported English proficiency (scale 0-5)**	3 [0, 4] 3 [2, 5]	5 [4, 5] 5 [5, 5]
CPPVT Scores (full mark = 99)	55 (21) [18, 83]	11 (6) [0, 25]

* When parents did not report an onset for the exposure to that language at home, it was assumed that English exposure started at 60 months (kindergarten) and Mandarin started at 72 months (grade one).
** The data of two parents (when applicable) are presented in the order of the larger and smaller numbers between the two parents.

In addition, 12 Chinese teachers provided speech samples. Among them, seven were L1 speakers of Mandarin, five were L1 speakers of another Chinese language and started learning Mandarin in school, and two were born in Canada and graduated from the bilingual program. These teachers formed a representative sample to understand the Mandarin input students receive at school, especially for the L2 students whose main source of Mandarin input is the teachers.

Procedures

Questionnaire. Parents filled out a questionnaire to quantify their language environment and language experiences. It was adapted from the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007) and the Alberta Language Environment Questionnaire (ALEQ, Paradis, 2011) and made available in both English and Chinese.

Speech sample collection. A picture-based single-word elicitation test was adapted from Zhao and Bernhardt (2012) and Zhu (2002). The test included 72 target words. Among them, 32 were citation tone targets (monosyllabic words), which were the focus of the current study (see Appendix I for a list of these 32 words). Examples of eliciting questions included “What is this?” and “What is this person doing?” Three examiners, who were Mandarin-L1 speakers and proficient in English, elicited productions that were as spontaneous as possible, but imitative models were provided as needed. Speech samples were audio-recorded using a Zoom H1n digital recorder with a Pro Lavalier JK MIC-J 055 unidirectional cardioid condenser microphone

positioned in front of the child's chest. The recordings were mono audios at a 48 kHz sampling rate and 24-bit resolution.

Phonetic transcription. Speech samples were transcribed by four Mandarin-L1 researchers. Transcribers coded whether each production was spontaneous or imitative. Subsequently, each tone production was transcribed as either one of the four citation tones, the semi-sandhi of Tone3 which is inappropriate in the monosyllabic context (Xu Rattanasone et al., 2018), or an uncategorizable production. In the whole wordlist, 23% of the samples were transcribed by a second transcriber and reached 90% inter-transcriber reliability. This was interpreted as acceptable since allophonic variations and uncategorizable productions were considered (Shriberg et al., 1997). Therefore, the first transcribers' transcriptions were adopted.

Spontaneity. Both groups of students made a considerable number of imitative productions. Imitative models can significantly increase tone match in bilingual students (Yang et al., 2021). Moreover, a Mann-Whitney U test showed a significant group effect on the number of spontaneous productions ($U = 105.5$, $p = 0.000$): The L2 group ($M = 18.386$, $SD = 6.721$) tended to produce more imitative tones than HL ($M = 29.763$, $SD = 4.863$), which will compound the results. Therefore, only spontaneous productions were analyzed. This left 1131 productions by HL, 809 productions by L2, and 387 productions by teachers (all spontaneous).

Tone match analysis. In this study, tone analysis was fulfilled in Phon (Hedlund & Rose, 2020), a software program for phonological data corpus that is commonly used in the field of child speech development. New functionality was developed to compare the transcribed tone against the target on each syllable. A full list of individual word productions can be generated with detailed information about the speaker, spontaneity, word token, target tones, and transcriber perceived tones to support token-level match analyses and mismatch pattern analyses.

Tone labelling. In Mandarin, a syllable includes an onset, a rime, and a tone. A rime can comprise a pre-nucleus glide, a nucleus, and a coda, and only the nucleus is mandatory (Table 3.1). Although a voiced onset and a glide can carry an *f0* contour (Chao, 1968), Howie (1974) finds that these segments contain erratic *f0* patterns. Therefore, tones should be labelled over the nucleus and any voiced segments after it (Wong, 2012; Xu, 1998; see Chen & Benjamin, 2013, however, for evidence of perceptual cues located in sonorant onset *f0*). Tones were labelled in Praat (Boersma & Weenink, 2022) using TextGrids. The onset of a vowel after a voiceless consonant was operationalized as the first zero-crossing point after the regular voicing pulses start, and the onset of a vowel after a sonorant (e.g., nasal, glide) was indicated by a shift in formants, a drop in intensity, and increased regularity in the waveform. The end of a tone was marked at the last zero-crossing point with a clear *f0* contour and formant structure. Four cycles at the beginning and at the end are excluded to eliminate irregular pitch patterns (Wong, 2012).

Acoustic measurements. ProsodyPro was used to extract *f0*, where automatically recognized voicing pulses were examined and manually adjusted (Xu, 2013). The smoothing function was disabled to obtain raw *f0* values. Ten *f0* values with equal time intervals were extracted for each syllable. In addition, the durations of tones were measured using ProsodyPro (Xu, 2012) by subtracting the end point and the starting point of the TextGrid interval.

f0 normalization. The *f0* values were normalized into T-values within each speaker according to Function (6.1) (Shi & Wang, 2006). This equation compresses *f0* differences in the higher pitch range through a logarithmic transformation and compresses a speaker's *f0* by their own extreme values (i.e., this speaker's minimal and maximal *f0*s). It then converts the logarithmic values into T values which range from 0 to 5. This is consistent with Chao's (1930) 5-scale transcription of Chinese tones: T values between 0 and 0.99 are Chao's tone letter [1],

and T values between 4 and 5 are Chao's tone letter [5]. This *f0* normalization method is able to minimize anatomical variation and sociolinguistic variation across speakers and preserve the phonemic distinctions between tone categories (Zhang, 2018). However, it is influenced by extreme values within a speaker's production. For example, if a speaker produced one *f0* value of 500 Hz, but all other *f0* values ranged between 150 Hz and 300 Hz, most of their T values will be low (ranging between 0 and 2.88), with only the 500 Hz value being converted to T = 5.

$$T = 5 \times (\log_{10}(f0) - \log_{10}(f0_{min})) / (\log_{10}(f0_{max}) - \log_{10}(f0_{min})) \quad (6.1)$$

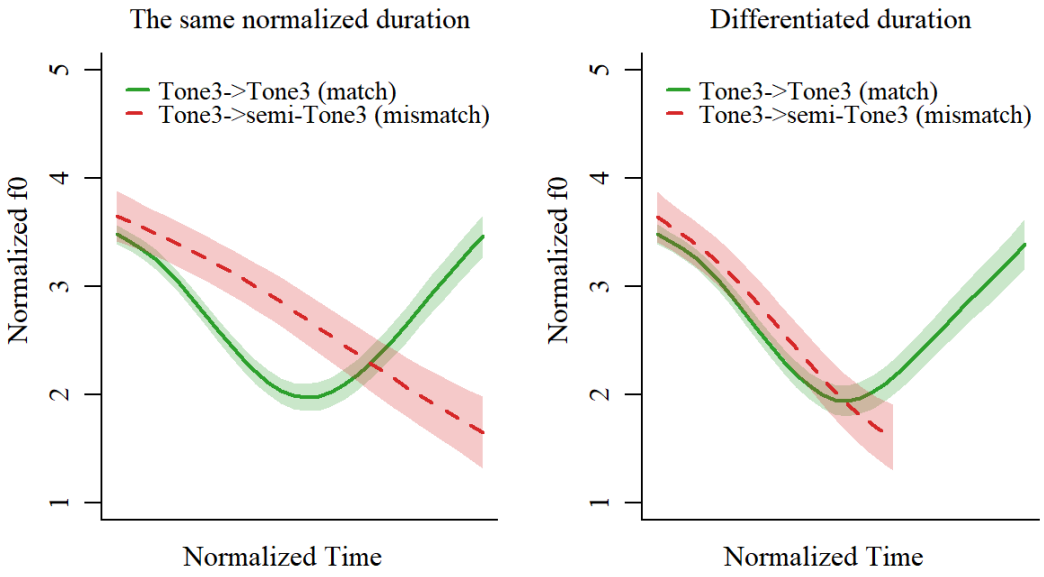


Figure 6.2. Plots of Tone3->Tone3 productions and Tone3->semi-Tone3 productions with equal durations (left) and differentiated durations (right).

Differentiated durations. Duration covaries with tones' pitch features (Gandour, 1977) and serves as a secondary cue for tone perceptions (Blicher et al., 1990). Therefore, it is important to incorporate duration information into *f0* contours. Furthermore, if all *f0* contours are plotted with the same normalized time, the contours can be misleading. For example, in Figure

6.2, some of the Tone3 /214/ are transcribed as a match, the others as a shorter, semi-sandhi of Tone3 [21]. The left figure suggests that when the two types of productions are plotted with equal durations, they barely overlap. However, when their average durations are considered, semi-Tone3 overlaps with the falling section of Tone3, which is a better reflection of the phonetic reality. Therefore, instead of using a normalized time (0–9, 10 points), the time domain of *f0* contours was differentiated by the average duration of a certain subgroup of tone productions. The subgroups were defined by factors of target tone, transcribed tone, speakers' grade level, and speakers' language background. For example, when Grade3 L2-learning students produced Tone4 as Tone2, their average duration was the longest (434 ms), so the time range was maintained as 0 to 9; When Grade1 HL students produced Tone4 that matched the target, the average duration was 241 ms, which is 0.56 times of the former type of productions (241 ms / 434 ms = 0.56), so the time range was 0 to 5.04 (9 × 0.56 = 5.04). A full list of the average duration of each type of production is available in Appendix J.

Statistical analysis. Statistical analyses were conducted in `R` 4.2.3 (R Core Team, 2023). Descriptive statistics were illustrated using `R`'s base plot functions. Logistic mixed models are conducted to model tone match using the `glmer()` function in the `lme4` package (Bates et al., 2015). Generalized additive mixed models (GAMMs) are conducted to model *f0* contours using `mgcv` and `itsadug` in R (van Rij et al., 2022; Wood, 2017). Mixed linear models are conducted to model tone duration using the `lmer()` function in the `lme4` package. Full codes with the full dataset are not publicly available at this moment, but pseudo-codes for all statistics in this paper are available in Appendix K.

Result

Tone match (accuracy)

Match analyses can help understand bilingual students' tone learning outcomes and the factors that might have impacted such learning. Students' and teachers' match rates of spontaneously produced citation tones were plotted by tone target, group, and grade (Figure 6.3). Visual inspection suggests that Tone3 was the most challenging for both HL and L2 groups, and even for teachers. HL students' tone match rates were similar to teachers', with the exception of Tone3 which has a lower match rate. On the other hand, L2 students' match rates were generally lower than teachers and HL, with the exception of Tone2, which was comparable with HL's match rates. Both HL and L2 achieved high match rates in Grade3, but there seemed to be a trend for L2's match to be lower in Grade5, whereas HL seemed resistant to this trend.

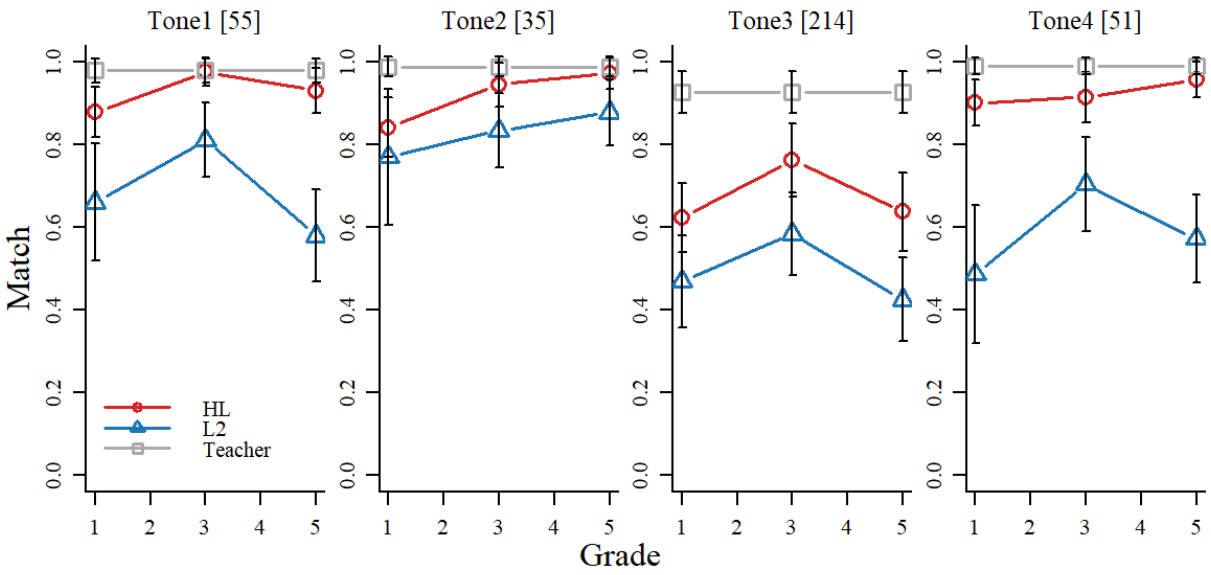


Figure 6.3. Raw data of average match rates of different tones by group and grade with teachers' data as a reference across grade levels. Error bars mark 95% confidence intervals.

To verify these observations, a logistic mixed model was built to investigate the match rates of tones. The target was a binomial variable of the match of each production (match or

mismatch). Fixed effects included group (HL or L2), grade level (Grade1, 3, or 5), and tone target (Tone1, 2, 3, and 4). Random effects included word token and speaker. Teacher productions were not included in the model because they did not have the dimensions of grade or group, but their means and confidence intervals were calculated as a reference. Fixed effects and interactions were added to the model in a stepwise manner, and AIC (criterion = 2) was used to compare models. The finally selected model included fixed effects of group, grade, target, and interactions of target × group and target × grade. The fixed effects abovementioned were presented in Figure 6.4, with raw data of teachers' match rates plotted as a reference when appropriate. The *p*-values were Bonferroni adjusted for contrasts of interest. A threshold value of probability $\alpha = 0.05$ was used to evaluate significant difference, $\alpha = 0.1$ was used to evaluate marginal difference, and $p > 0.1$ indicated no significant difference.

The model indicated the effect of language background group on tone match. HL had a significantly higher probability of tone match than L2 ($MD = 0.238$, $z = 7.205$, $p < 0.05$). Furthermore, the model indicated the effect of grade level. The difference was significant between Grade1 and Grade3, with a higher probability in Grade3 ($MD = 0.130$, $t = 3.317$, $p < 0.05$). The model supported the visual inspection of raw data (Figure 6.3) that HL had higher match rates, and both groups had higher match rates in Grade3 than in Grade1. But it did not support the observed trend of lower match of L2 in Grade5 as there was no significant interaction between group and grade.

Meanwhile, the model indicated a target effect, with Tone3's match rate lower than the other tones (*ps* < 0.05). Moreover, there was an interaction between group and target: Although L2 produced all four tones with lower match rates than HL, the differences were larger in Tone1 ($MD = 0.249$, $z = 5.894$, $p < 0.05$) and Tone4 ($MD = 0.347$, $z = 7.004$, $p < 0.05$) and smaller in

Tone2 ($MD = 0.103$, $z = 2.954$, $p < 0.05$) and Tone3 ($MD = 0.192$, $z = 3.254$, $p < 0.05$). These results are in line with the visual inspections that Tone3 was similarly challenging for both groups, and Tone2 seemed to be an easier target for L2 to achieve a match rate similar to their HL peers. On the other hand, there was an interaction between grade and target: Tone2 was the only target that had continually higher average match rates in higher grades and had a marginally significant difference between Grade5 and Grade1 ($MD = 0.138$, $z = 3.000$, $p \approx 0.1$). On the contrary, the other tones had the highest average match rates in Grade3, although the differences between grades were not significant ($ps > 0.1$).

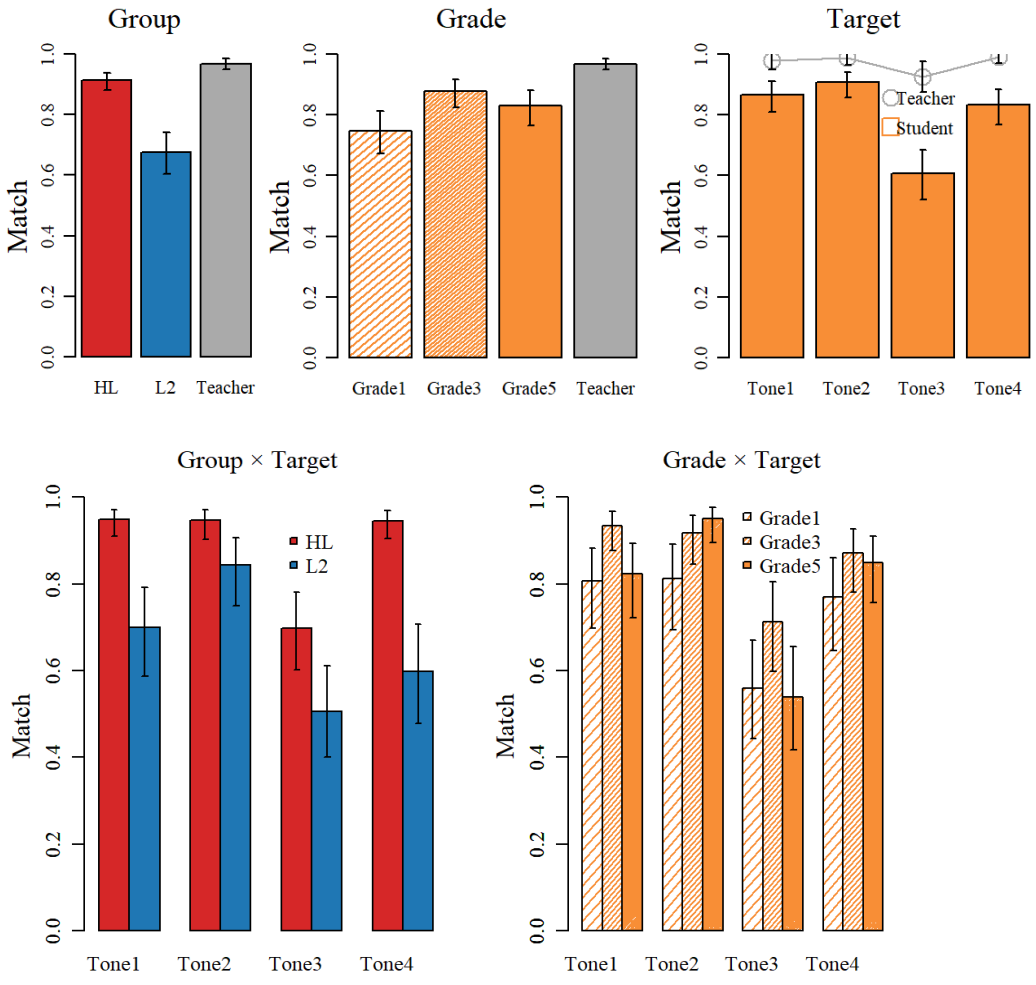


Figure 6.4. Logistic mixed model-fitted probabilities of tone match with the main effects of group, grade, target, and the interaction effects of group × target and grade × target. Teachers' match rate was plotted in grey as a reference when appropriate. Error bars mark 95% confidence intervals.

In summary, the difference in tone match rates related to language backgrounds was not levelled out through bilingual education across the grade levels observed in this study. On the contrary, the HL group produced tones with a higher match than their L2 counterparts, but both groups achieved a similarly high match for Tone2 and a similarly low match for Tone3. Students achieved a higher match rate in Grade3, but their tone match rate was not higher in Grade5. Meanwhile, match rates differed across targets, Tone3 being the most challenging across groups and grades.

Mismatch types

Table 6.2. Confusion matrices of HL and L2 groups' production of citation tones. Each row presents a target tone, and each column presents the transcribed production. Transcribed productions include the four citation tones, the semi-sandhi of Tone3 ("sT3"), and uncategorizable productions ("U"). Shading represents the percentage of this pattern among all patterns of this target.

	HL group's transcribed production						L2 group's transcribed production					
N %	T1	T2	T3	T4	sT3	U	T1	T2	T3	T4	sT3	U
T1	–	10 45.5%	0 0%	1 4.5%	0 0%	11 50%	–	24 38.7%	7 11.3%	10 16.1%	0 0%	21 33.9%
T2	1 4%	–	13 52%	0 0%	3 12%	8 32%	4 16%	–	14 56%	1 4%	2 8%	4 16%
T3	9 8.5%	12 11.3%	–	13 12.3%	36 34%	36 34%	13 9.6%	71 52.2%	–	7 5.1%	11 8.1%	34 25%

T4	4 17.4%	2 8.7%	2 8.7%	–	1 4.3%	14 60.9%	17 23.6%	29 40.3%	7 9.7%	–	0 0%	19 26.4%

In addition to tone match rates, we were interested in tone patterns produced by students from different backgrounds. These can provide insights into the intra-language factors (e.g., which tone pairs are confusable) and inter-language factors (e.g., whether the patterns are related to English transfer). Therefore, confusion matrices were generated for both groups (Table 6.2). Only mismatched productions were included, and the cells were shaded by the percentage of this pattern among all patterns of this target to indicate how dominant this pattern was. Notice that a darker shade did not indicate more mismatches, since the target might have few mismatches in total (e.g., Tone2->Tone3 pattern was dominant in both groups, but the numbers of this pattern were small).

Both groups produced Tone2->Tone3 mismatches, probably related to their phonetic similarities, as their *f0* contours both include a rising trend. Tone2-Tone3 confusion was documented among adult L2 learners, but it remained inconclusive whether the confusion was bidirectional or unidirectional (Hao, 2012; Wang et al., 2003). Results of the current study suggest bidirectional confusion among L2 students since Tone3->Tone2 was dominant in their Tone3 patterns. A GAMM model was used to model the *f0* contours of Tone2-Tone3 confusion and verify the transcribers' judgement (Figure 6.5). Transcribers' judgement was validated by acoustic measurements since the *f0* contours were similar based on transcribed tones in spite of the intended targets. If anything, transcribers were lenient when recognizing mismatches – Productions had to deviate from the target a lot to be perceived as another category. Specifically, a Tone3 target had to show a strong rising trend to be recognized as Tone2, and a Tone2 target had to show a strong dipping contour to be recognized as Tone3. Unlike the Tone2-Tone3

confusion which can be explained by phonetic similarity, mismatches in the L2 group violated distinctive features of the targets (Xu, 1997), such as Tone1->Tone3 (high-level to low-dipping tone) and Tone2->Tone4 (rising to falling tone). These patterns may be better explained by limited linguistic significance assigned to tone categories due to English influences (So & Best, 2010) and an immature phonetic representation of tone categories (Flege & Bohn, 2021).

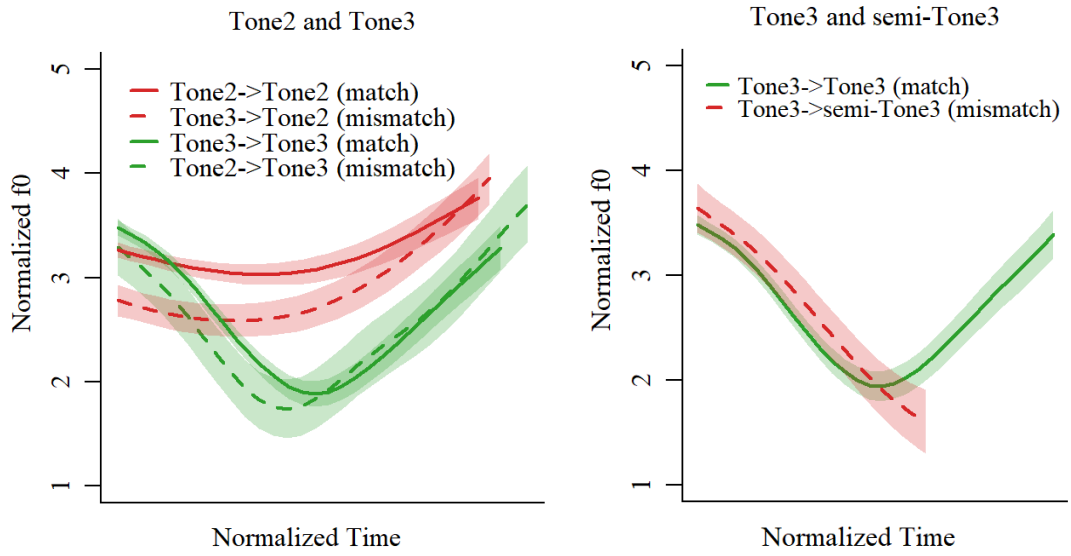


Figure 6.5. GAMMs-fitted *f0* contours of productions related to Tone2-Tone3 confusion (left) and Tone3-semiTone3 confusion (right). Durations are set as the average duration of each type of production. Matched productions are coded in solid lines, and mismatched productions are coded in dashed lines. Ribbons describe 95% confidence intervals.

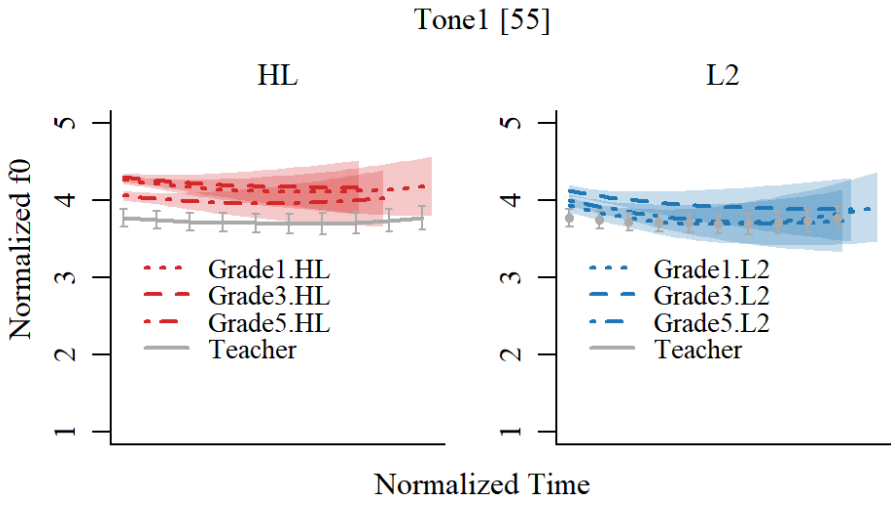
On the other hand, HL made more uncategorizable productions and produced Tone3 as its semi-sandhi more frequently. It should be mentioned that since tones are mandatory on Mandarin monosyllabic words, the uncategorizable code did not indicate productions of "no tone." Instead, this code was often used when the production had perceivable differences from the expected tone target but could not be assigned to a different tone category or when the production could be assigned to more than one tone category ambiguously as perceived by

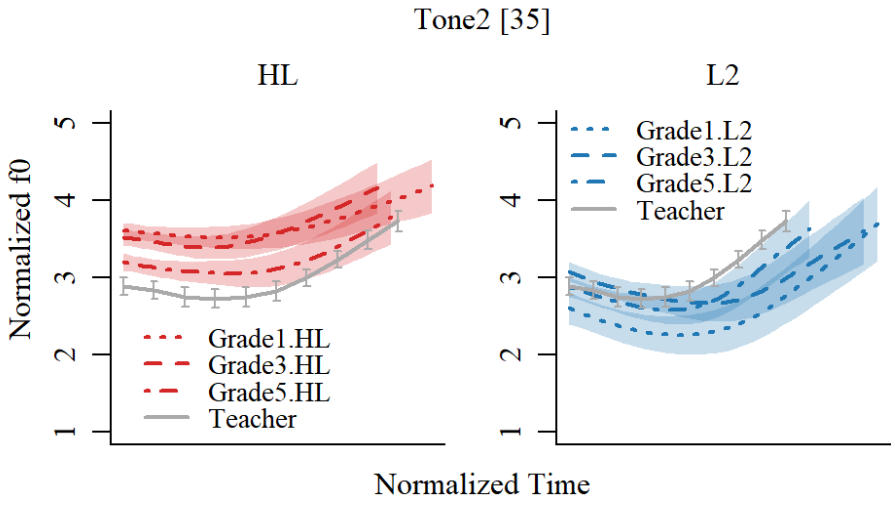
transcribers. Therefore, the uncategorizable productions might indicate a more mature development of tones, for the students produced tones that were phonetically different from the intended targets but were not misrecognized as other categories. This is similar to Mandarin tone development in monolingual children at a younger age, where they first established the categories and key features and then took years to refine the phonetic specifications (Wong, 2012; Xu Rattanasone et al., 2018). Such continual refinement at the phonetic level will be further depicted in the next sections. The Tone3->semi-Tone3 pattern is worth more discussion. Tone3 is realized as [214] in citation tones and as [21] in multisyllabic productions (Duanmu, 2007). To validate this allophonic distinction, we asked 12 linguistically naive Mandarin-L1 listeners to listen to children's Tone3->semiTone3 productions. We found that a monosyllabic semi-Tone3 production out of context was often perceived as Tone4 (6 out of 12 listeners) or uncategorizable (4 out of 12 listeners). Listeners commented that the productions sounded "like a half tone," which is supported by the *f0* contours plotted in Figure 6.5. Therefore, we chose to transcribe [21] productions as semi-Tone3 and analyze them as "mismatches" to preserve the perceivable phonetic differences. However, although HL students applied the sandhi rule incorrectly, such a pattern suggested that they were exposed to multisyllabic speech materials and were able to produce the distinctive feature of "low pitch" for Tone3, which probably indicated an intermediate level of tone learning (Wong, 2012; Yang, 2016).

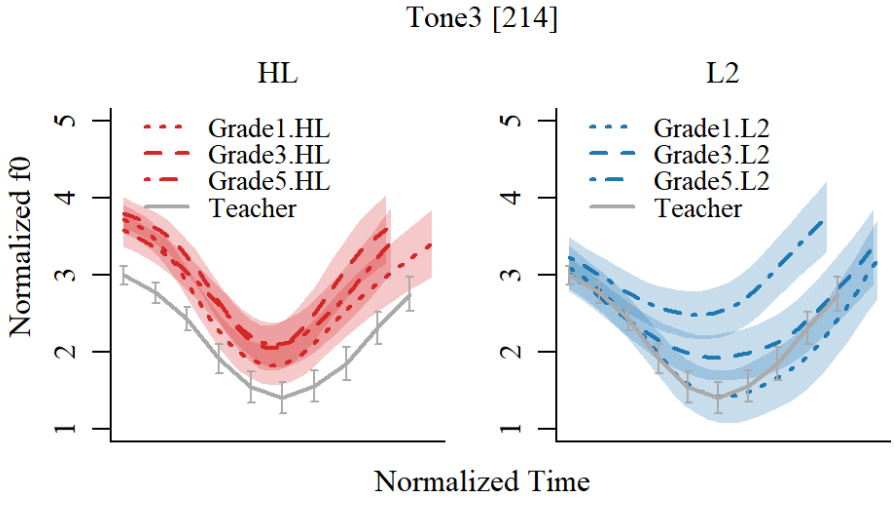
In summary, transcribers' perceptual judgements were supported by acoustic measurements (Figure 6.5). The L2 group showed a bidirectional confusion between Tone2 and Tone3. It seemed that the HL group exhibited mismatches that deviated from the targets but still showed signs of understanding of the categories, whereas the L2 group produced mismatches that violated the distinctive features of the categories.

f0 contour

Acoustic examinations of matched productions (perceived as accurate) can help understand how bilingual students learn the phonetic specifications of tones. Therefore, *f0* contours of matched productions were modelled with GAMMs. This method was chosen because its non-linear terms are especially suitable to model *f0* contours of tones. The model used normalized *f0* values as the dependent variable and normalized time as the independent variable, with the fixed effects of group (HL and L2), grade (1, 3, and 5), and tone target (Tone1, 2, 3, 4), as well as the random effects of speaker and word token. Interaction effects were implemented using indexed coding. The `compareML()` function was used to compare between models, and AR(1) models were built to control autocorrelation effects (van Rij et al., 2022). The model with the lowest AIC was selected, which included parametric and smooth terms of group × grade × tone interaction (adjusted R^2 = 0.548). This suggested that *f0* contours were influenced by not only targets but also students' language backgrounds and grade levels. A separate model was built for each tone to present group and grade effects, with teachers' raw data plotted in grey (Figure 6.6).




Tone2 [35]
HL
L2
Normalized f0
Grade1.HL
Grade3.HL
Grade5.HL
Teacher
Grade1.L2
Grade3.L2
Grade5.L2
Teacher
Normalized Time


Tone3 [214]
HL
L2
Normalized f0
Grade1.HL
Grade3.HL
Grade5.HL
Teacher
Grade1.L2
Grade3.L2
Grade5.L2
Teacher
Normalized Time

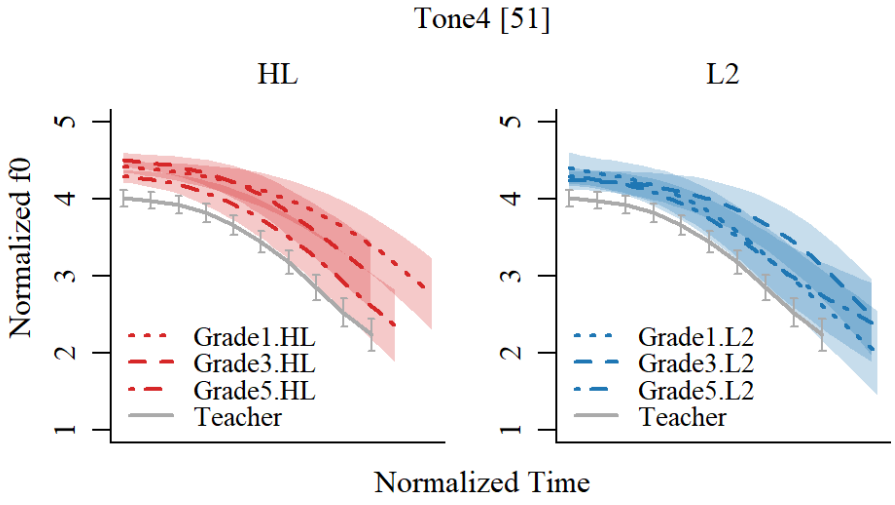

Tone4 [51]
HL
L2
Normalized f0
Grade1.HL
Grade3.HL
Grade5.HL
Teacher
Grade1.L2
Grade3.L2
Grade5.L2
Teacher
Normalized Time

Figure 6.6. GAMM-fitted *f0* contours of each tone target by group and grade with teachers' contours plotted in grey as references. Ribbons and error bars mark 95% confidence intervals.

Through visual inspection, most of the important features of citation tones were produced even in Grade1. A general trend is that students' productions resembled teachers' in higher grades. For example, HL produced a steeper rising contour for Tone2 and a steeper falling contour for Tone4 in Grade5. However, L2's Tone3 productions did not seem to follow this progressive trajectory. Instead, in higher grades, L2 produced Tone3 with higher *f0* and a shallower dip, thus exhibiting a stronger rising trend. Therefore, although the productions were transcribed as matching the target, they were more subjected to being misrecognized as Tone2.

Duration

Duration is an important secondary cue for tone perception (Blicher et al., 1990). Although durations may vary across and within individuals depending on speech rates, the patterns we detected from the data were convincing and can provide insights into students' learning of secondary cues of tones. A plot of raw data by target and group in each grade level is presented in Figure 6.7, with teachers' data plotted in grey as a reference. It seems that the teachers' duration pattern was comparable with that in L1 literature: Tone3 had the longest duration, and Tone4 had the shortest (Xu, 1997). However, only HL in Grade3 and Grade5 produced a similar pattern, whereas L2 consistently produced Tone2 with a similar duration to Tone3, if not longer.

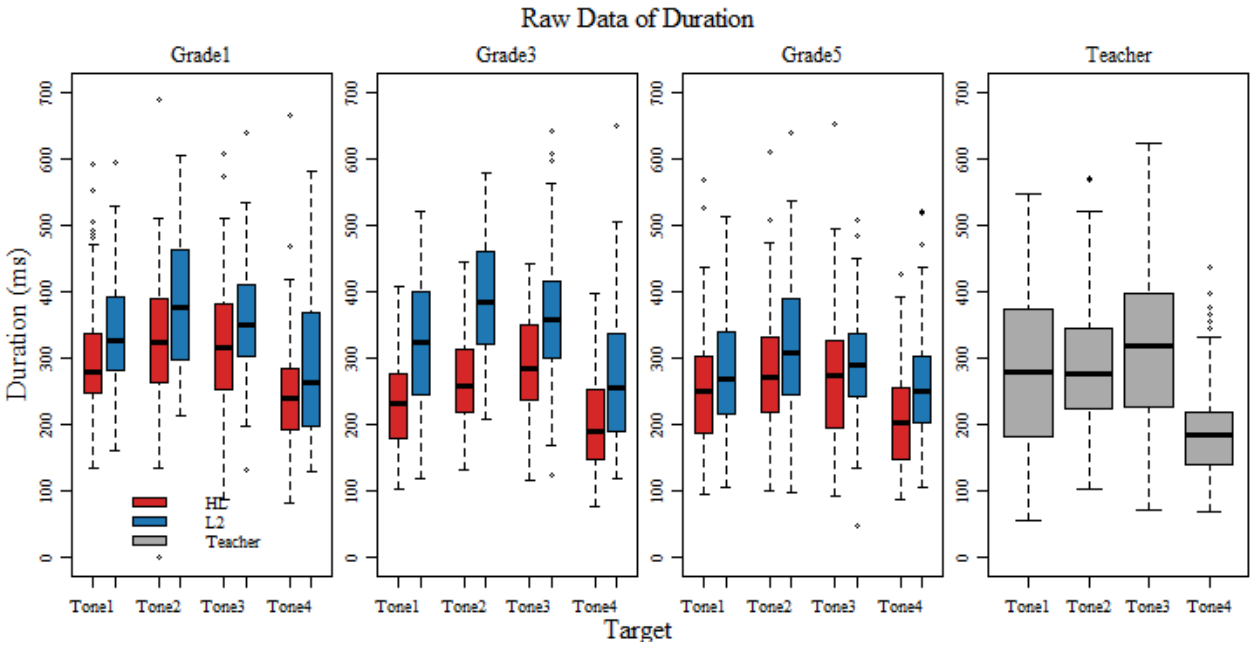


Figure 6.7. Raw data of the duration of each target by group in each grade level, with teachers' productions plotted in grey as a reference. The boxes and whiskers mark quantiles and the scattered points mark outliers.

To verify the observed patterns, a linear mixed model was built to investigate the duration of tones. The target was a numeric variable of the duration (ms) of each production. Fixed effects included group, grade, and target. Random effects included word token and speaker. Models were compared and selected based on AIC. The selected model suggested a significant group effect: HL produced tones with significantly shorter durations ($MD = 45.981$, $t = 4.664$, $p < 0.05$). Meanwhile, the model suggested a significant grade effect: Grade5 produced a shorter duration than Grade1 ($MD = 43.405$, $t = 3.475$, $p < 0.05$). Such trends are unsurprising as temporal characteristics are related to proficiency in the language (Trofimovich & Baker, 2006).

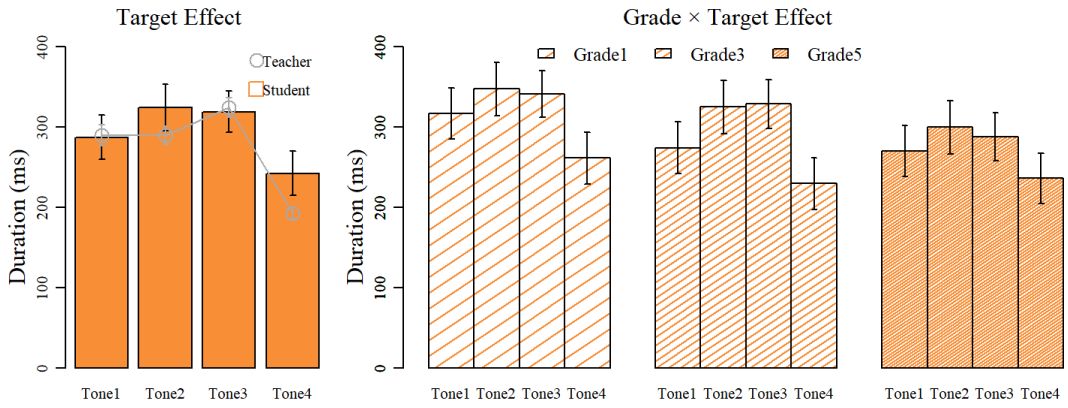


Figure 6.8. Linear mixed model-fitted duration with tone target effect (left) with teachers' productions plotted in grey as a reference and target × grade effect (right). Error bars mark 95% confidence intervals.

More relevant to the current study, bilingual students' durations differed across targets (Figure 6.8). First, the tone target had a significant effect: The means indicated an order of Tone2 > Tone3 > Tone1 > Tone4. Tone4 was significantly shorter than Tone2 (MD = 81.555, t = 4.434, $p < 0.05$) and Tone3 (MD = 76.700, t = 4.476, $p < 0.05$) but not shorter than Tone1 (MD = 44.662, t = 2.498, $p > 0.1$). Tone1, Tone2, and Tone3 were not significantly different ($ps > 0.1$). That is to say, the model indicated a duration order of Tone3 = Tone2 = Tone1 ≥ Tone4, which does not match the patterns in L1 speakers, i.e., Tone3 is usually the longest (Xu, 1997). Such a pattern among bilingual students left Tone2 and Tone3 prone to confusion in the time domain. Moreover, there was a significant interaction between tone and grade. Tone3 was marginally longer than Tone1 in Grade3 (MD = 54.447, t = 2.957, $p < 0.1$), but it was no longer than Tone1 in Grade5 (MD = 17.522, t = 0.958, $p > 0.1$). Meanwhile, Tone3 was always longer than Tone4 but never longer than Tone2. Therefore, using Tone1's duration as a reference, bilingual students produced Tone3 with a relatively long duration in Grade3, but this feature was not maintained in Grade5.

In summary, bilingual students did not pick up the duration difference between Tone2 and Tone3 in the time window between Grade1 and Grade5. If anything, the long-duration feature of Tone3 seemed reduced in Grade5. Moreover, L2 was more subjected to confusing these two tones, as they produced Tone2 with a longer duration on average according to our visual inspection of the raw data.

Discussion

The current study presented evidence on how elementary students in a Mandarin-English bilingual program in Western Canada produced Mandarin citation tones. Both transcription-based and acoustic analyses were included to examine the match rates, mismatch patterns, and phonetic realizations of tone productions. Match analyses suggest that HL students had advantages over their L2 peers, although neither group reached teachers' ceiling level of match. Students achieved higher match rates in Grade3 than Grade1 but not higher match rates in Grade5. Moreover, both groups achieved high average match rates in Tone2 but had low match rates in Tone3. In terms of mismatches, those in the HL group tended to partly demonstrate the features of the targets (e.g., Tone3->semiTone3), but those in the L2 group violated the target features (e.g., Tone4->Tone2). Meanwhile, both groups produced Tone2-Tone3 confusion, but L2 more frequently simplified Tone3 to Tone2. In terms of acoustic measurements, both groups produced *f0* contours that resembled teacher and L1 speakers, but L2 did not emphasize the low-dipping contour of Tone3 in higher grades. Neither group of students fully acquired the duration features of Mandarin tones up to Grade5 compared to teachers' and L1 speakers' patterns.

The next sections will discuss the results in relation to factors of bilingual pronunciation learning and present implications for researchers, theorists, and educators.

Impacts of intra-, inter-, and extra-language factors

Bilingual students' tone learning showed influences of intra-language factors, which were especially salient in Tone3. Tone3 has unstable phonological realizations (Xu Rattanasone et al., 2018), and advanced motor skills are required to produce its low pitch and compound *f0* contour (Wong, 2012). Therefore, Tone3 is the latest-developmental citation tone in Mandarin and is challenging for L2 learners (Wang et al., 2003; Wong, 2012). In the current study, Tone3 had the lowest match rates for both HL and L2 students. Meanwhile, Tone3 involves complex phonetic characteristics, such as a low, dipping *f0* contour, a creaky voice that often co-occurs with the low pitch (Kuang, 2017), and the longest duration (Blicher et al., 1990; Xu, 1997). Evidence suggests that monolingual children spend years refining these phonetic specifications (Rhee et al., 2020; Wong, 2012). These cues also appeared challenging for bilingual students to pick up – Neither group produced Tone3 with a longer duration than Tone2 at any of the three grade levels we observed. In this sense, bilingual children's tone production depends on the characteristics of the tone targets within the language and thus shows consistency across groups and similarity to monolingual peers.

Meanwhile, bilingual students' tone learning was influenced by inter-language factors, i.e., cross-linguistic influences from English. According to PAM-S, since English does not assign as much linguistic significance to pitch information, bilingual learners, especially L2 students, may have difficulty perceiving and contrasting Mandarin tone categories. They may assimilate Mandarin tones into other prosodic categories in English. They may also produce more categorical substitutions, especially between the categories that are phonetically similar (So & Best, 2010). Results in the current study support PAM-S's theoretical accounts. Tone2 was the only tone where L2 and HL students reached similar levels of high match rates. According to

teachers' reflections, they used the question intonation in English as a prompt to facilitate the learning of Tone2 (Lin et al., 2022). With this strategy, L2 students learned Tone2 efficiently, which is in line with Wang et al. (2003) that L2 learners showed the most improvement in Tone2 after a short period of perceptual training. Moreover, L2 produced bidirectional substitutions between Tone2 and Tone3 due to their phonetic similarities. Particularly, they produced Tone3->Tone2 mismatches more frequently, which can be attributed to both categorical collapsion and phonetic simplification (i.e., falling-rising->rising). Therefore, the substitution patterns may be related to not only perceptual assimilation but also difficulties in producing the motor patterns. Furthermore, L2 produced more tone substitutions that violated the features of the tone targets (e.g., Tone4->Tone2, even though they could be assimilated into two different prosodic categories in English), which indicated limited sensitivity to tone contrasts.

Finally, bilingual students' tone learning was influenced by extra-language factors such as Mandarin input in learners' environments (Flege & Bohn, 2021), which were operationalized as home language backgrounds and grade levels in the Mandarin program. Results of the current study showed that even for HL students in Grade5, their tone match did not reach the ceiling level like their teachers. This indicated that bilingual students' tone learning was affected by the limited speech input in the English-dominant society and was more protracted compared to monolingual children (Wong, 2012). SLM-r (Flege & Bohn, 2021) did not make specific hypotheses about how (early) home input and (current) school input would impact speech learning, but the results of the current study suggested that the effects are not simply additive. The pitch and duration features of Tone3 were reduced for the L2 group in higher grades, which means their developmental trajectories were not parallel to HL peers, nor did their trajectories converge after receiving five years of Mandarin input at school. Such developmental trajectories

may be related to L2 students' limited home input in Mandarin at early ages. This is in line with evidence in infant speech development that infant in non-tonal language environments learn to ignore suprasegmental information in word recognition (Singh et al., 2008), which can explain why it was challenging for L2 students in this study to assign linguistic significance to tones. Such evidence did not support Wong's (2012) hypothesis that the protracted refinement of tones was related to immature physiology. The loss of certain phonetic features in higher grades among L2 students suggested that tone learning is more related to integrating phonetic cues through perceptual learning and social interactions (Kuhl et al., 2008; Rhee et al., 2020).

Implications and future directions

In addition to providing the evidence aforementioned, the current study makes several implications for research methods, bilingual speech theories, and pedagogical practices.

In terms of research methods, the current study did not use overall tone match (e.g., PTC) as a single psychometric score. Instead, we examined the match of each production as a function of speaker, word token, and tone target. It is not uncommon to observe specific targets separately in language and speech development, as each category and even item may have different levels of complexity (McMillen et al., 2020; Zhu, 2002). The current study again proved the merit of such analyses, since Tone3 was significantly more challenging across groups and grade levels. However, it should be noted that general accuracy scores such as PTC can evaluate learners' speech production and be related to other psychometric measurements (Shriberg et al., 1997). Therefore, future studies will continue to investigate PTC and its relationships with speech input and other language outcomes. In addition, the current study investigated both phonemic categories and phonetic characteristics of tones, which revealed more details in the learning process. Instead of forcing transcribers to choose a phonemic category (e.g., Zhu, 2002), the

current study allowed transcribers to indicate that a tone production was an inappropriate allophonic realization or was uncategorizable (Wang et al., 2003). With such allophonic considerations, students' match rates did not reach the ceiling level, even though the HL group was exposed to Mandarin at home before school age. This is different from the evidence in monolingual and bilingual children that tones are mastered early (Holm & Dodd, 2006; Mok & Lee, 2018; Zhu, 2002). Indeed, both groups reached a functional level of match, but the confusion matrices further revealed that the two groups had different tone patterns. Meanwhile, even among matched productions, acoustic analyses suggested that bilingual students might not utilize certain phonetic features. This supports Wong's (2012) argument that phonemic transcriptions may overestimate children's tonal skills (Zhu, 2002), and children's tone development would appear more protracted when phonetic details are considered. It is noteworthy, however, that acoustic analyses should be linked to listener perception to provide functional implications (Lindemann & Subtirelu, 2013; Munro & Derwing, 2020). Therefore, future research will continue to include other acoustic measurements such as phonation measurements and *f0* contour parameters (e.g., *f0* shift and *f0* range) (Kuang, 2017; Wong, 2012) and relate them to functional perceptual judgements such as intelligibility and accentedness of speech (Munro & Derwing, 2020).

In terms of theoretical implications, the current study supported PAM-S and SLM-r: Bilingual students' tone productions were related to English transfer (So & Best, 2010) and Mandarin speech input (Flege & Bohn, 2021). Results also expanded these theories in the suprasegmental domain. First, the results emphasized the effects of universal challenges in speech production (Wong, 2012) by showcasing the difficulty of producing Tone3 across groups. Second, the current study suggests the unique value of home input. At least in the context of

learning a tonal language in a non-tonal environment, HL students who received home input in Mandarin were able to achieve high match rates and maintain important phonetic features in their production, whereas their L2 peers seemed to be less sensitive to categorical differences and phonetic details of tones despite their current exposure to Mandarin at school. Thus, it seems that "earlier is better." However, it needs to be pointed out that HL's advantage might also be related to the high-quality and continual input at home and in the community (Flege & Bohn, 2021; Liu, 2020). Therefore, our results did not support a decisive effect of early exposure (Bedore et al., 2016). In adult L2 learners, short-term, intensive perceptual training is effective for English speakers to improve their Mandarin tone production (Wang et al., 2003). Meanwhile, it is argued that effective learning of phonetic cues is related to cognitive learning in functional contexts (Blicher et al., 1990; Rhee et al., 2020). Nonetheless, the cross-sectional, observational design of the current study could not provide conclusive evidence. Longitudinal, experimental studies should be conducted to understand the effects of improved quantity and quality of speech input among L2 learners, especially the speech input that involves meaningful social interactions.

Finally, the current study provides implications for pedagogical practices. First, it provides valuable evidence on pronunciation learning, which is understudied in the context of bilingual education. Results showed that both HL and L2 students were able to produce citation tones with core phonetic features in Grade1 and achieve match rates that were well-above chance levels. In addition, both groups produced tones with increased match rates in Grade3. Furthermore, the HL group maintained an advantage in tone production and did not seem to experience attrition in Mandarin tone production. These findings showcased the effectiveness of two-way bilingual education design in fulfilling its functions of L2 learning and heritage language maintenance. But on the other hand, the differences related to home language

backgrounds were not levelled out through bilingual education at the elementary level (c.f., Menke, 2017; Nance, 2020). In fact, both groups' progress seemed to have plateaued between Grade3 and Grade5, especially in terms of further learning of phonetic specifications of tone productions, although their learning outcomes in higher grades and as adult learners remained unknown. Such a result can be attributed to typological differences between Mandarin and English and generally limited Mandarin input in the community, and it also seems to be related to the curriculum design. Since Mandarin pronunciation was mainly addressed through the instruction on sound-letter relationships in lower grades (Alberta, 2006), teachers were hesitant to teach pronunciation in higher grades. Meanwhile, peer input and community input were limited according to teachers' report (Lin et al., 2022). Therefore, policymakers and educators should consider making pronunciation a long-term goal, especially for societal minority languages. Second, the current study can help identify practical strategies for Mandarin pronunciation instruction. All measurements clearly indicated that Tone3 was challenging and was often confused with Tone2. Therefore, educators may benefit from professional development programs to understand the phonetic features of these tone targets (Lin et al., 2022). For example, educators can emphasize that Tone3 is not only dipping but also low, often accompanied by a creaky voice (Kuang, 2017). They can also stress that Tone3 is produced with a longer duration, which allows it to be distinguished from Tone2 (Blicher et al., 1990) and gives students more time to articulate its complex *f0* contour (Wong, 2012).

The current study is among the first to document school-aged children's learning of tones in a non-tonal language environment, investigate the pronunciation learning outcomes of students enrolled in bilingual education programs, and use both phonemic and phonetic analyses to examine such learning. It draws attention to the multifaceted factors that influence bilingual

speech development and calls for researchers and educators to continually support bilingual children's pronunciation learning in a minority-language context.

References of Chapter 6 (Paper 3)

Alberta Education. (2006). Chinese Language Arts Kindergarten to Grade 9. International Languages, Programs of Study. https://education.alberta.ca/international-languages-k-6/programs-of-study/

Baker, W., & Trofimovich, P. (2005). Interaction of native-and second-language vowel system (s) in early and late bilinguals. *Language and Speech, 48*(1), 1-27.

Bates. D, Mächler. M, Bolker. B, Walker. S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01.

Bedore, L.M., Peña, E. D., Griffin, Z., & Hixon, G. (2016). Effects of age of English exposure, current input/output, and grade on bilingual language performance. *Journal of Child Language, 43*(3), 687-706

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (eds.), *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. DeGroot (eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109-127). Oxford University Press.

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

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America, 129*(6), 3964-3980.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique. 8*(45), 24-27.

Chen, T. Y., & Tucker, B. V. (2013). Sonorant onset pitch as a perceptual cue of lexical tones in Mandarin. *Phonetica, 70*(3), 207-239.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251.

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed). OUP Oxford.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge University Press.

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics, 29*(1), 41-65.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*(2), 269-279.

Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics in Language Teaching, 45*, 353-378.

Hedlund, G. & Rose, Y. (2020). Phon 3.1 [Computer Software]. Retrieved from https://phon.ca.

Holm, A., & Dodd, B. (2006). Phonological development and disorder of bilingual children acquiring Cantonese and English. In H. Zhu & B. Dodd (eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 286-325). De Gruyter.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica, 30*(3), 129-148.

Kan, R. T., & Schmid, M. S. (2019). Development of tonal discrimination in young heritage speakers of Cantonese. *Journal of Phonetics, 73*, 40-54.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America, 142*(3), 1693-1706.

Lin, Y., Pollock, K. E., & Li, F. (2022). Teaching Mandarin pronunciation in Chinese-English bilingual schools in Canada. *The 13th Annual Pronunciation in Second Language Learning and Teaching Conference (PSLLT2022)*, Brock University, ON, June 16-18.

Lindemann, S., and Subtirelu, N. (2013). Reliably biased: the role of listener expectation in the perception of second language speech. *Language Learning, 63*, 567–594.

Liu, W. (2020). Success factors for a Mandarin bilingual program: An autoethnographic case study. *Chinese as a Second Language, 55*(3), 208-229.

Lu, L., & Liu, H. S. (1998). *The Peabody picture vocabulary test-Revised in Chinese*. Taipei, Taiwan: Psychological Publishing.

MacLeod, A. A., & Stoel-Gammon, C. (2010). What is the impact of age of second language acquisition on the production of consonants and vowels among childhood bilinguals? *International Journal of Bilingualism, 14*(4), 400-421.

Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin–English bilingual children. *Clinical Linguistics and Phonetics, 24*(4-5), 387-400.

McMillen, S., Anaya, J. B., Peña, E. D., Bedore, L. M., & Barquin, E. (2022). That's hard! Item difficulty and word characteristics for bilinguals with and without developmental language disorder. *International Journal of Bilingual Education and Bilingualism, 25*(5), 1838-1856.

Menke, M. R. (2017). Phonological development in two-way bilingual immersion: The case of Spanish vowels. *Journal of Second Language Pronunciation, 3*(1), 80-108.

Mok, P. P. K., & Lee, A. (2018). The acquisition of lexical tones by Cantonese–English bilingual children. *Journal of Child Language, 45*(6), 1357-1376.

Munro, M. J., and Derwing, T. M.. (2020). Collecting Data in L2 Pronunciation Research. In *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference* (pp. 2380–9566).

Nance, C. (2021). Scottish Gaelic revitalisation: Progress and aspiration. *Journal of Sociolinguistics, 25*(4), 617-627.

Netelenbos, N., Li, F., & Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism, 20*(3), 346-357.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism, 1*(3), 213-237.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Rhee, N., Chen, A., & Kuang, J. (2020). Integration of spectral cues in the development of Mandarin tone production. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 3135-3138).

Shi, F., & Wang, P. (2006). A statistic analysis of tone groups in Beijing Mandarin. *Contemporary Linguistics, 8*(4), 324-333.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708-722.

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development, 4*(2), 157-178.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*(2), 273-293.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1-30.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics, 35*(4), 765-805.

van Rij, J., Wieling, M., Baayen, R., van Rijn, H. (2022). itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.4.1.

Wang, Y., Jongman, A., & 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.

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*(1), 141-151.

Wood, S. N. (2017). Generalized additive models: an introduction with R. CRC press.

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology, 9*, 817.

Xu, Y. (1997). Contextual tonal variations in Mandarin. Journal of phonetics, 25(1), 61-83.

Xu, Y. (2013). ProsodyPro-A Tool for Large-scale Systematic Prosody Analysis. *TRASP* Aix-en-Provence, 7–10.

Yang, C. (2016). *The acquisition of L2 Mandarin prosody: From experimental studies to pedagogical practice* (Vol. 1). John Benjamins Publishing Company.

Yang, J., & Liu, C. (2012). Categorical perception of lexical tone in 6 to 8-year-old monolingual and bilingual children. *International Journal of Asian Language Processing, 22*(2), 49-62.

Yang, L., Lin, Y., Pollock, K. E., & Li, F. (2021). Accuracy of spontaneous and imitative productions in children learning Mandarin in a bilingual program. *The International Child Phonology Conference (ICPC2021)*, Lethbridge University, AB, June 17 - 19.

Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Multilingual Matters.

Chapter 7 Conclusion and Contributions

This dissertation sought to understand the speech learning of a minority language, Mandarin, by bilingual students who were enrolled in a Chinese-English two-way bilingual program in Western Canada. Despite the well-acknowledged merits of two-way bilingual education design, it remained unknown how bilingual students learned the pronunciation of Mandarin, a minority language in Canada whose phonological system is distinctive from English. This dissertation presented three manuscripts that used different scopes, took multiple perspectives, and adopted various research methods to approach the topic. This final chapter revisits the research questions and findings, discusses research gaps that were addressed by this dissertation, identifies the key contributions and limitations of this work, and recognizes other research directions that are related to this dissertation.

Revisiting research purposes and the structure of this dissertation

This section reviews the research purposes stated in Chapter 1 and the relevant findings. Subsequently, a visualization is presented to revisit the structure of this dissertation, which is used to identify future research directions in a later section.

In Chapter 1, three purposes of this dissertation were listed: (1) review theories and evidence in pronunciation learning to provide insights into speech learning in a bilingual program, (2) understand how the pronunciation of a minority language is taught in our bilingual program of interest, and (3) investigate how students from diverse backgrounds learn Mandarin tones in the bilingual education context in Canada.

To address the first purpose, I proposed an encompassing conceptual model in Chapter 2 (Paper 1) to review theories and evidence in multiple research disciplines, which guided the two

following chapters. I first reviewed the literature on interlocutors' attitudes toward L2 pronunciation and education's role in forging such language ideologies. This became vivid in Chapter 3 when I reviewed the bilingualism and multiculturalism of Canada and in Chapter 4 (Paper 2) when I presented teachers' diverse views on the nativeness of students' pronunciation and the values of their own (non)nativeness. Next, I reviewed the literature on how interlocutors' speech systems interact, which impacts their perceptual learning as a learner or a listener. This was later demonstrated in Chapter 6 (Paper 3), that bilingual students' Mandarin tone production was influenced by the majority language, i.e., English. Then, I reviewed the literature on perceptual and productive (e.g., acoustic) measurements of L2 pronunciation. Such knowledge guided my practice in Paper 3. I held a cautious view on native listeners' judgements: The match and mismatch between an expected target and transcribers' perception were not referred to as "correct" or "incorrect," as I acknowledged the subjectivity of listeners' judgements and the arbitrariness of an idealized, standardized phonology. I used acoustic measurements to validate transcribers' judgements (Figure 6.5) and linked *f0* (Figure 6.6) and duration (Figure 6.8) measurements to the perceived accuracy of tones. Finally, I stated that language learners are faced with real challenges in L2 pronunciation, and the educational system plays a crucial role in fostering positive attitudes and improving pronunciation. This led me to the second research purpose of understanding how pronunciation was taught in bilingual classrooms.

To address the second purpose, I first reviewed the Chinese-English bilingual program of interest in Chapter 2. I introduced the speech system taught in the program and the type of speech input and pronunciation instruction specified by the program design: The target language, Mandarin, has a phonological system that is uniquely different from English, and the program design was supposed to provide peer input from L1 speakers of both English and Mandarin to

facilitate the development of functional bilingualism (ECBEA, n.d.). However, upon review of the curriculum design, I learned that pronunciation was never specified as a goal, and the indirect instruction through sound-symbol systems was less attended to in higher grades (Alberta Education, 2006). Therefore, I was curious how teachers' practices responded to these challenges and went beyond. In Chapter 4 (Paper 2), I presented teachers' reflections on multifaceted factors that influenced students' speech learning, not limited to their home language backgrounds, but also including their motivation, language aptitudes, teacher input, and meaningful interactions in school and in society. Teachers shared a variety of teaching strategies, including CLT and FFI approaches, to facilitate students' perception, production, and interaction in Mandarin. In this study, most teachers pointed out that Mandarin tones were challenging targets, which led to the third paper on bilingual students' Mandarin tone learning.

To address the third purpose, in Chapter 6 (Paper 3), I presented a quantitative research study on bilingual students' Mandarin tone productions. The key findings suggested that bilingual speech development was influenced by intra-language factors (e.g., the complexity of target speech units), inter-language factors (e.g., English's influences on Mandarin pronunciation), and extra-language factors (e.g., home input and school input). Although the bilingual program was effective in helping students with diverse language backgrounds learn the key features and phonemic contrasts of tones, the students who did not have strong Mandarin input at home might still need more intensive and long-term support to refine their phonetic specifications of Mandarin tones.

Figure 7.1 visualizes the structure of this dissertation. I started with a wide overview of L2 pronunciation (Chapter 2) and the backgrounds of bilingual education (Chapter 3), zoomed into teachers' reflections on teaching Mandarin pronunciation (Chapter 4), and took a deep dive

into the learning outcomes of a unique aspect of Mandarin pronunciation, lexical tones (Chapters 5 and 6). Each study was motivated by the previous one and had a slightly different visual field, from broad to detailed, to examine the same research topic of bilingual pronunciation learning. Meanwhile, from the review above, it became clear that the empirical studies (Papers 2 and 3) fitted in the framework proposed in Paper 1 (color-coded in Figure 7.1), which demonstrated the value of the encompassing review of the multidisciplinary literature.

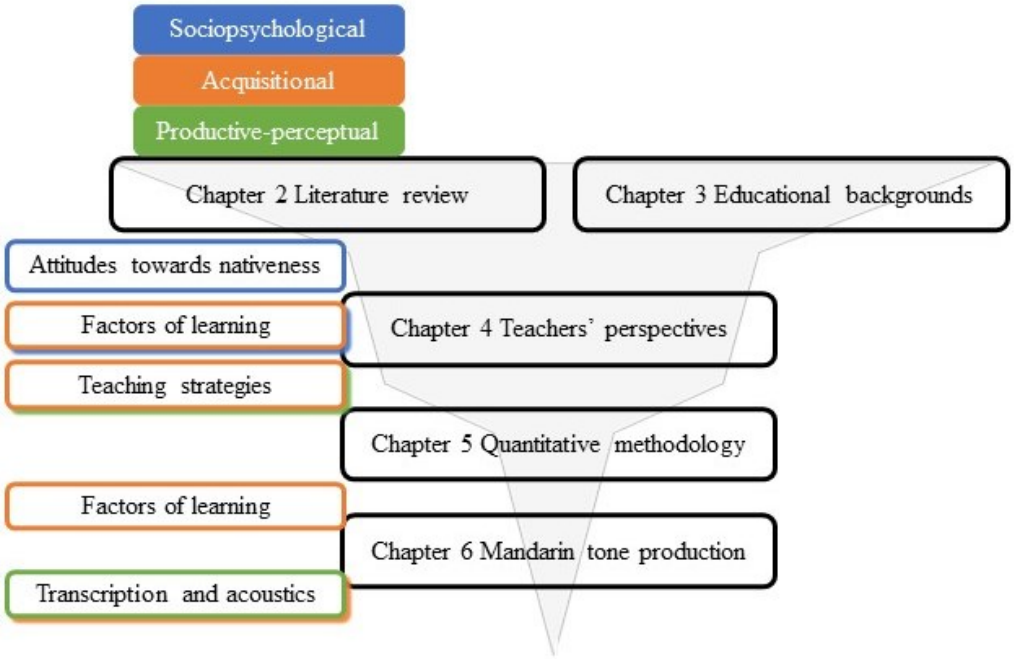


Figure 7.1. A visualization of the structure of this paper-based dissertation. The chapters proceed from broad overviews to detailed analyses of one specific phonological dimension within the same topic. Themes in each study are color-coded using the three layers of conceptual models proposed in Chapter 2 (Paper 1).

Research gaps addressed by this dissertation

This section contrasts several long-standing policies, ideologies, and theories versus the lack of research in the corresponding areas and demonstrates how this dissertation addresses some of the research gaps.

It has been 52 years since Canada adopted multiculturalism as an official policy, the first country doing so in the world (Wright, 2021). Whereas the government recognizes multiculturalism as one of the great strengths of Canada (Government of Canada, 2023), and the world recognizes multiculturalism and multilingualism as key features of Canada (Liu, 2020), the learning of a minority language is still often related to the maintenance of heritage languages and cultures (Dicks & Genesee, 2017). In contrast, researchers in bilingual education have been advocating for a language-as-resource orientation and bilingual education for not only minority groups but also the whole society (Cummins, 2017; Ruiz, 1984). There are several signs of the shift of orientation from local heritage culture maintenance to actively participating in globalization: First, in the 1990s, the Alberta government revised the terminology from "heritage" to "international" language programs. Second, since Edmonton had twinned with the city of Harbin, China in the 1980s, each of the Chinese bilingual schools in Edmonton became twinned with a school in Harbin (ECBEA, 2009). Meanwhile, the Edmonton Public School Board established a partnership with the Confucius Institute of China in 2007 to facilitate language and cultural communication (Liu, 2020). Third, an increasing number of students are enrolling in French immersion and international language programs to learn a third language, in addition to English and their own home languages, as their economic and cultural capital (Dagenais, 2003; Liu, 2020; Wu, 2005). However, compared to the rich literature on immigrant children learning English (e.g., Paradis, 2011) or English-speaking children learning French (see Dicks & Genesee, 2017 for a review), there is less research on bilingual students, especially L2

students, learning a minority language in Canada (Cummins & Danesi, 1990). In the current dissertation, Chapter 4 (Paper 2) presented how teachers used a variety of strategies to help both HL students and L2 students to learn Mandarin pronunciation, and Chapter 6 (Paper 3) suggested that L2 students' learning of Mandarin tones had some similar patterns with their HL peers, but they might need further support to attend to the phonetic details of tones.

It has been 40 years since the Chinese bilingual program was established in Edmonton. This program is considered the largest Chinese bilingual program in North America and among the best Chinese programs outside of China (ECBEA, n.d.). Despite its time-honored history, large scale, and high quality, there are few research studies on this program. Wu & Bilash (2000) presented the program as an example of multifunctional bilingual education design, and Liu (2020) analyzed the factors that contributed to the success of the program. Wu (2005) presented interview data on bilingual students' self-perceived language abilities and identities, and Sun (2010) presented observational and interview data on bilingual students' language and literacy practices. All of these studies used a qualitative approach. This dissertation is the first to our knowledge to use both qualitative and quantitative methodologies to study the teaching and learning in the Chinese bilingual program. It showcased the multifunctionality, effectiveness, and practical strategies of this unique program as a successful example among the world's bilingual programs. Meanwhile, it constructively identified areas of need and improvement. For example, Chapter 4 (Paper 2) advocated for revisiting the curriculum and providing teachers with resources and PD opportunities, and Chapter 6 (Paper 3) pointed out the challenges for L2 students to learn Mandarin tones in an English-dominant environment.

It has been almost 30 years since Celce-Murcia et al. (1996) raised the "Cinderella" analogue to advocate for re-attention to pronunciation in the era of CLT. Researchers have

differentiated the goals of nativeness and intelligibility (Levis, 2020; Munro & Derwing, 1995) and justified the latter as an appropriate principle in pronunciation teaching. However, for many, explicit instruction on pronunciation symbolizes tedious rote approaches, whose demonic shadows should have been dismissed with an emphasis on *meaning* instead of *form* (Littlewood, 1980). This was reflected in the Chinese bilingual program's curriculum that pronunciation was never specified as a goal (Alberta Education, 2006). This was also manifested in teachers' reflections in Chapter 4 (Paper 2) that many were not prepared to teach pronunciation or believed that pronunciation should be picked up implicitly. This dissertation believes that meaning and form, or CLT and FFI, do not play a zero-sum game (Isaacs, 2009). Instead, through Chapter 2 (Paper 1), I argued that pronunciation marks speakers' identity and facilitates speakers' intercultural communication. Researchers and educators should not feel reluctant to address pronunciation, especially when the learners are young and are naïve about the language's speech system. Therefore, in Chapter 4 (Paper 2), I shared teachers' strategies that focused on both CLT and FFI. In Chapter 6 (Paper 3), I advocated for improved pronunciation instruction for bilingual students, especially those L2 students who were learning Mandarin mainly from school.

It has been almost 30 years since several influential L2 speech theories were proposed (e.g., PAM, 1995; SLM, 1995). There was plenty of empirical evidence for these theories on consonants and vowels (Best et al., 2001; Guion et al., 2000; Piske et al., 2002; Tyler et al., 2014). In addition, there were some studies that applied these theories to adult L2 learners' production and perception of tones (Hao, 2012; So & Best, 2010; Wang et al., 2003). However, very few studies have applied L2 speech learning theories to school-aged bilingual learners of a tonal language. This population is unique as (1) one language involves a phonological dimension that does not exist in the other language, and (2) their speech input includes shared school input

and a wide variety in the quantity and quality of home input. In this dissertation, Chapters 4 and 6 (Papers 2 and 3) discussed the influence of English on learning Mandarin tones and the complex relationship between early home input and current school input, which had theoretical contributions to PAM-S (Best & Tyler, 2004) and SLM-r (Flege & Bohn, 2021). Both HL and L2 groups were able to produce tone categories at a moderate to high match (accuracy) rate depending on the universal difficulties of the targets. But L2 students had a lower match rate in general and did not pick up or maintain certain phonetic cues in higher grades, possibly due to the lack of early exposure and continual input in Mandarin. By analyzing these multifaceted factors that influence bilingual speech learning, this dissertation (1) acknowledged the diversity of world languages and demonstrated the procedures of investigating phonologically distinctive language pairs (e.g., tonal and non-tonal languages) and (2) acknowledged the diversity in learner populations and demonstrated the procedures of investigating learners with a variety of backgrounds (e.g., HL and L2 learners).

Highlights and contributions

Now that I have summarized the findings of this dissertation and reviewed the research purposes and research gaps it addressed, I would like to identify its highlights and key contributions:

First, the dissertation provided a very *ambitious overview of the literature on L2 pronunciation and the bilingual education contexts*. This made sure I had a good understanding of this multidisciplinary field and also reminded readers of the wide scope of the topic.

Second, this dissertation presented *diverse teacher and student populations* in the bilingual program (Wu, 2005). I included a representative sample of twelve teachers with a variety of language and training backgrounds. I valued their insights and tried to share their

teaching strategies as powerful experiences among researchers and educators. I also included both HL students and L2 students in the study, which partly reflected the complex language backgrounds and the wide range of Mandarin proficiency of students in the program. Before making any claims or statements, I hoped to first document and celebrate the diversity in the program and bring the complex reality alive in my studies.

Third, the dissertation embraced a wide range of *methodological approaches*, including qualitative and quantitative methods, transcription and acoustic analyses, and multiple statistical models to take the findings beyond data descriptions. This dissertation made a few innovations in methodology: A Chinese dictionary and tone analysis functionalities were developed in Phon (Hedlund & Rose, 2020). This showcased how researchers can and should adopt a variety of methods flexibly to investigate one topic in depth.

Fourth, this dissertation provided evidence for *L2 speech theories* by examining data at different scales. Specifically, I examined students' language experiences, especially their home language input and schooling experiences (i.e., grade levels), which allowed me to explore the influence of speech input on tone learning: Findings in this dissertation supported SLM-r's hypothesis that L2 speech development is a function of quantity and quality of L2 speech input (Flege & Bohn, 2021). In addition, they suggested potentially different roles of early home input and recent school input in bilingual students' pronunciation learning and the risk of attrition without continual input. I examined the mismatch patterns of bilingual students, which allowed me to explore the potential influences of English transfer on Mandarin tone production: Findings in this dissertation supported PAM-S's hypothesis that tone categories are assimilated into other prosodic categories in a non-tonal L1 (So & Best, 2010). However, L2 learners still confused tone categories that could be assimilated into different prosodic categories in their L1, indicating

a general unfamiliarity with the tonal dimension. I examined tone match at the tone category and item level, which allowed me to investigate the role of universal difficulty of specific targets: Bilingual students had difficulties producing Tone3 due to its phonological and phonetic complexity despite their language experiences. This was in line with PAM-S (So & Best, 2010) that when a tone category is unassimilable into L1 prosodic categories, its learning depends on its phonetic features. I examined the acoustic details of bilingual students' tone productions, which allowed me to explore how students continued to refine the phonetic specifications of tones: Although tone categories tend to be established early, their phonetic refinement is a long-term learning process (Wong, 2012). This seems especially challenging for L2 learners, suggesting that the slow progression in phonetic learning may be related to not only physiological maturation (Wong, 2012) but also the integration of phonetic cues in production and perception through meaningful social interactions (Kuhl et al., 2008; Rhee et al., 2020).

Fifth, this dissertation also made constructive, evidence-based *recommendations to improve educational policies and practices*. I was motivated by an educator's question about bilingual students' non-native Mandarin pronunciation despite their long-term immersion in the program. Previous evidence on pronunciation learning in bilingual programs suggested that home language differences could be levelled out through immersive education and peer interactions (Menke, 2017; Nance, 2020). However, the findings in this dissertation were in line with the educators' observation: Both teachers' reflections and data on students' tone productions indicated that Mandarin lexical tones were challenging to learn, especially for L2 learners. This reminded researchers and educators that the effectiveness of two-way bilingual education might be influenced by the specific pair of target languages, for example, how phonologically distinctive they are. Furthermore, I found that pronunciation was given limited

attention in higher grades, and the bilingual students resorted to English as their common language at school. Consequently, L2 students continued to have difficulty producing tones accurately with refined phonetic cues. This reminded researchers and educators that the two-way bilingual design did not magically eliminate group differences. Instead, bilingual students, especially L2 students, will benefit from extra support from the program since they cannot access the same speech input in the minority language as their HL peers from their households and communities (Liu, 2020).

Limitations

I acknowledge there were several limitations of this paper-based dissertation:

First, with a very ambitious scope of literature review, Chapter 2 (Paper 1) could only overview the field briefly and could not include all the available evidence (e.g., language ideology discussions such as Levis, 2020 and neurological evidence on bilingual pronunciation such as Reiterer, 2019).

Second, since the dissertation studies were part of a larger project, it was not possible to design speech materials to serve an overly specific purpose, such as syllables with controlled speech sounds to investigate tone productions. However, fortunately, the token differences were controlled by random effects in mixed modeling in Chapter 6 (Paper 3).

Third, although all Mandarin transcribers received phonetic training and attended weekly sessions to calibrate their transcription, it was practically challenging to coordinate multiple transcribers to process a large speech dataset and reach a full consensus. Therefore, because the inter-transcription reliability reached 90%, the first transcriber's judgements were adopted to improve efficiency.

In addition, the cross-sectional design restrained this dissertation from interpreting results from a developmental perspective. It needs to be reminded that although the study in Chapter 6 (Paper 3) presented data from three different grade levels, the differences could not indicate individuals' progress over time. Meanwhile, the dissertation adopted an observational study design to compare groups of students with various language experiences. Such a design is common among bilingual speech development studies (e.g., Baker & Trofimovich, 2005; MacLeod & Stoel-Gammon, 2010), as it is often not feasible to manipulate children's language experiences on a large scale. Therefore, causal relationships could not be assumed for any interpreted influences of the quantity and quality of Mandarin speech input.

Other current and future directions

Synthesizing this dissertation's structure (Figure 7.1) and its limitations, there are many research directions that can be further explored. In fact, many of such directions are currently ongoing in the larger SSHRC-funded project with my participation but could not be included in this dissertation due to time and space constraints. Some research directions that are relevant to this dissertation are presented in Figure 7.2 using the same structure as Figure 7.1, with ongoing directions marked with solid-line shapes and future directions marked with dashed-line shapes.

A first direction for research addresses how the quantity and quality of Mandarin speech input influence bilingual students' learning (Flege & Bohn, 2021). A large questionnaire (Marian et al., 2007; Paradis, 2011) was used to collect rich information about students' language backgrounds and language experiences. However, to make sure the research questions and models were concise in Chapter 6 (Paper 3), only a few columns of responses in the questionnaire were used to select relatively homogenous groups from the heterogeneous sample. Future research can explore different methods to quantify bilingual experiences (see Luk &

Esposito, 2020 for a mini-series on tools to document bilingual experiences) and how different methods fulfill certain research objectives. For example, I explored using Principal Component Analysis (PCA) to reduce the dimensions of parent questionnaire responses and obtain component scores as a numeric representation of students' language experiences (Mady, 2017). Furthermore, measurements of language experiences, whether categorical or numeric, can be related to psychometric scores of bilingual students' speech learning outcomes (e.g., PPC scores, Shriberg et al., 1997). Such models can help researchers understand how language experiences influence bilingual speech learning in general, and how specific components of language experiences (e.g., quantity versus quality of input [Flege & Bohn, 2021], cumulative versus current input [Bedore et al., 2016]) make different influences or interact with each other. Such studies will address the social experiences of the learners and understand how learners' L1 and L2 experiences influence their speech acquisition, which fits in the sociopsychological and acquisitional layers of the conceptual model (Figure 7.2).

A second direction for research involves different measurements of bilingual students' speech production. In addition to tones, research can analyze consonant and vowel productions using transcription and acoustic measurements. This will explore how bilingual students learn shared and unshared speech sounds and contrasts and will be comparable with other bilingual speech research cross-linguistically (e.g., Meziane & MacLeod, 2020). For example, an ongoing study found that L2 students' production of Mandarin fricatives was more influenced by English in lower grades, but they had the potential to develop Mandarin-unique fricative contrasts in higher grades (Yang et al., 2023). Such findings were in line with SLM-r's hypotheses that bilingual speech learning is a function of speech input (Flege & Bohn, 2021), including home and school input, and indicated that the learning processes may differ in shared categories (e.g.,

consonants) compared to unshared categories (e.g., tones). Moreover, research should investigate the production of connected speech. Connected speech samples may better represent the functional use of language in daily communication (McLeod & Baker, 2017). They may also reveal challenges that are not as obvious in single-word samples as they are more complex and demand higher cognitive or academic language proficiency (CALP, Cummins, 1981). An ongoing study found that the match rates of phones in connected speech in English were positively related to students' English input and unrelated to Mandarin input. However, the match rates of phones and tones in connected speech in Mandarin were positively related to Mandarin input and negatively related to English input (Bishop et al., 2023). These findings suggested that bilingual speech learning outcomes may be influenced by language status. These research directions will address how learners' linguistic experiences and the interactions between their two (or more) languages influence their speech learning, which fits in the acquisitional layer of the conceptual model.

A third direction for research is to apply more functional methods to measure listeners' perception, such as perceived intelligibility, comprehensibility, and accentedness (Munro & Derwing, 1995). A further step would be to measure the listeners' attitudes and linguistic experiences and map these to their ratings to capture individual differences (Lin et al., 2023). Meanwhile, research may link acoustic deviances to listeners' perceptual ratings in order to understand the acoustic sources of perceived unintelligibility or accentedness (Idemaru et al., 2019; Lindemann & Subtirelu, 2013). Such studies will use perceptual and productive methods to measure bilingual learners' speech productions and understand the relationships between these measurements, which fit in the productive-perceptual layer of the conceptual model.

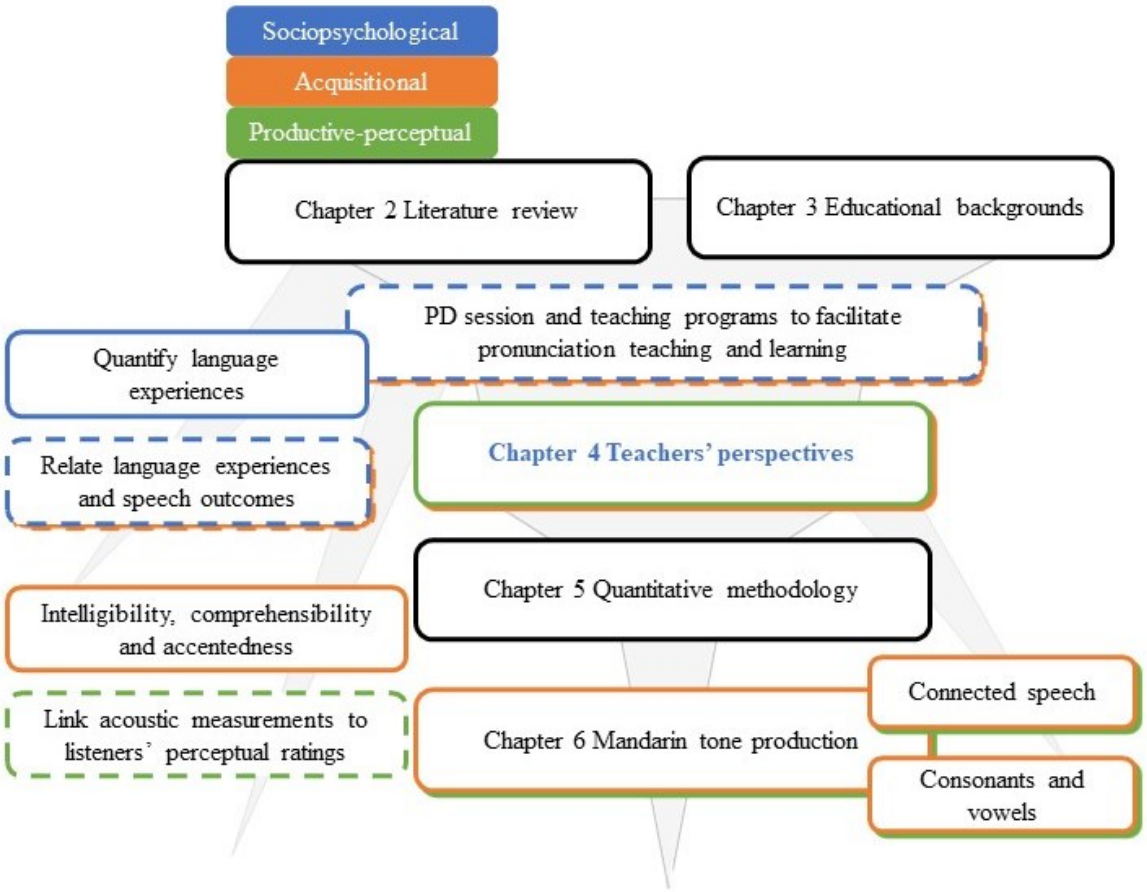


Figure 7.2 A visualization of current and future directions related to this dissertation using the same broad-to-specific structure in Figure 7.1. Themes in each direction are color-coded using the three layers of conceptual models proposed in Chapter 2 (Paper 1). Currently ongoing research is marked in solid-line shapes, and future directions are marked in dashed-line shapes.

A fourth direction for research is to implement teacher training programs and tone teaching programs. Currently, we are reaching out to educators, parents, and the general public to share our research findings and bring awareness to pronunciation learning. It is a future direction to design intervention programs based on findings in this dissertation and test the effectiveness of such programs. Such studies will address educators' and learners' cultural and linguistic awareness and improve the quality of speech input, which fits in the sociopsychological and acquisitional layers of the conceptual model.

Concluding remarks

Mandarin is the most spoken non-official language in Canada (Statistics Canada, 2022) and the most spoken language in the world. This dissertation was the first to examine the speech development of Mandarin as a minority language in a Chinese-English two-way bilingual education program. It took a large-scope overview of the relevant theoretical and educational contexts and an in-depth investigation of the details related to the pedagogical practices and learning outcomes.

The results from this case example of a Chinese bilingual program support previous findings that a two-way bilingual education design is effective in facilitating high-quality peer input from L1 speakers of both languages (Cummins, 1979; 2004), but it does not magically guarantee to level out students' differences in language backgrounds and render native-like speech in L2 learners. Instead, to facilitate the development of functional oral proficiency in the minority language for students with diverse backgrounds, it requires a continual focus on pronunciation in the curriculum, improved teaching resources and PD opportunities for teachers, enhanced bilingual environment at school (especially increased use of the minority language), and increased support for L2 students who cannot access speech input from their home or community. In addition, when generalizing the evidence from a bilingual education program, researchers and educators should understand that the speech learning outcomes in a bilingual program are related to the specific pair of languages involved in that program. Mandarin in this dissertation provided an example: It contains a phonological dimension that is uniquely distinctive from English (i.e., lexical tones) and some of the tone targets are universally challenging (i.e., Tone3). Therefore, the learning outcomes of tones were not fully comparable to

the evidence from other bilingual programs which focused on the learning of speech sounds (Menke, 2017; Nance, 2020).

This dissertation adds to the literature on bilingual education, bilingual speech development, and Mandarin speech acquisition. It brings awareness of these research topics and encourages more evidence to support pronunciation teaching and learning in bilingual contexts, the development of bilingual speech theories in the suprasegmental domain, and the learning of Mandarin and other world languages as heritage languages or second languages in this globalized world.

Final Bibliography

Alberta Education. (2006). Chinese Language Arts Kindergarten to Grade 9. International Languages, Programs of Study. Retrieved from https://education.alberta.ca/international-languages-k-6/programs-of-study/.

Alberta Education. (2008). Programme d'éducation pour la maternelle. French Language Arts (M à 12), Programmes d'études. Retrieved from https://education.alberta.ca/french-language-arts-m-%C3%A0-12/programmes-d%C3%A9tudes/.

Asia Pacific Foundation of Canada. (2013). Canada's Asia Challenge: Creating Competence for the Next Generation of Canadians. Retrieved from http://www.asiapacific.ca/sites/default/files/filefield/asia_competence_tf_-_final_revised_report.pdf.

Ates, B., Eslami, Z. R., & Wright, K. L. (2015). Incorporating world Englishes into undergraduate ESL education courses. *World Englishes, 34*(3), 485-501.

Austin, Z., & Sutton, J. (2014). Qualitative research: getting started. *The Canadian Journal of Hospital Pharmacy, 67*(6), 436-440.

Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America, 137*(5), 2823-2833.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174-EL180.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly, 48*(1), 136-163.

Baker, C. (2007). Becoming bilingual through bilingual education. *Handbook of Multilingualism and Multilingual Communication, 5,* 131-152.

Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed.). Multilingual Matters.

Baker, C., Andrews, H., Gruffydd, I., & Lewis, G. (2011). Adult language learning: a survey of Welsh for Adults in the context of language planning. *Evaluation & Research in Education, 24*(1), 41-59.

Baker, W., & Trofimovich, P. (2005). Interaction of native-and second-language vowel system (s) in early and late bilinguals. *Language and Speech, 48*(1), 1-27.

Baran-Łucarz, M. (2014). The link between pronunciation anxiety and willingness to communicate in the foreign-language classroom: The Polish EFL context. *Canadian Modern Language Review, 70*(4), 445-473.

Barrett DeWiele, C. E., & Edgerton, J. D. (2020). Opportunity or inequality? The paradox of French immersion education in Canada. *Journal of Multilingual and Multicultural Development*, 1-15.

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy, 9*(1), 95-114.

Bates. D, Mächler. M, Bolker. B, Walker. S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bedore, L.M., Peña, E. D., Griffin, Z., & Hixon, G. (2016). Effects of age of English exposure, current input/output, and grade on bilingual language performance. *Journal of Child Language, 43*(3), 687-706

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica, 161*, 25-35.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp.171-204). York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). John Benjamins.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775-794.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. DeGroot (eds), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109-127). Oxford University Press.

Bishop, S., Stone, M., Lin, Y., Li, F., & Pollock, K. E. (2023). English speech sound production accuracy in children attending a Mandarin-English bilingual program. *Bilingualism and Multilingualism Conference (BAM)*, Kelowna, May 13 - 15.

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

Boersma, P., & Weenink, D. (2022). Praat 6.3.03 [Computer Software]. Retrieved from https://www.fon.hum.uva.nl/praat/.

Byun, T. M., & Rose, Y. (2016). Analyzing clinical phonological data using Phon. *Seminars in Speech and Language, 37*(2), 085-105.

Çagatay, S. (2021). Identity clashes of EFL instructors in Turkey with regard to pronunciation and intonation in English. *International Journal of Curriculum and Instruction, 13*(3), 2557-2584.

Carrie, E. (2017). 'British is professional, American is urban': attitudes towards English reference accents in Spain. *International Journal of Applied Linguistics, 27*(2), 427-447.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge University Press.

Chan, K. Y., Hall, M. D., & Assgari, A. A. (2017). The role of vowel formant frequencies and duration in the perception of foreign accent. *Journal of Cognitive Psychology, 29*(1), 23-34.

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America, 129*(6), 3964-3980.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique. 8*(45), 24-27.

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

Chen, T. Y., & Tucker, B. V. (2013). Sonorant onset pitch as a perceptual cue of lexical tones in Mandarin. *Phonetica, 70*(3), 207-239.

Chen, X., Anderson, R. C., Li, W., Hao, M., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology, 96*(1), 142-151.

Chong, R. H. H., & Tan, Y. Y. (2013). Attitudes toward accents of Mandarin in Singapore. *Chinese Language and Discourse, 4*(1), 120-140.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1-39.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116*(6), 3647-3658.

Commission of Written Language Reformation of the People's Republic of China. (1958). *The Scheme of Chinese Phonetic Alphabet*. Written Language Reformation Press.

Cooper, A., & Bradlow, A. R. (2016). Linguistically guided adaptation to foreign-accented speech. *Journal of the Acoustical Society of America, 140*(5), EL378-EL384.

Cooper, B., Payne, G., Hu, X., Dixon, Q., & Kuo, L. (2020). The impact of linguistic diversity education on L1 English speakers' ideologies, attitudes, and perceptions of international teaching assistants. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference* (pp. 2380-9566).

Copland, F., & Yonetsugi, E. (2016). Teaching English to young learners: Supporting the case for the bilingual native English speaker teacher. *Classroom Discourse, 7*(3), 221-238.

Couper, G. (2021). Pronunciation Teaching Issues: Answering Teachers' Questions. *RELC Journal, 52*(1), 128-143.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251.

Cummins, J. (1981). Empirical and theoretical underpinnings of bilingual education. *Journal of Education, 163*(1), 16-29.

Cummins, J. (2014). To what extent are Canadian second language policies evidence-based? Reflections on the intersections of research and policy. *Frontiers in Psychology, 5*, 358.

Cummins, J. (2017). Teaching for transfer in multilingual school contexts. *Bilingual and Multilingual Education, 1*, 103-116.

Cummins, J., & Danesi, M. (1990). *Heritage languages: The development and denial of Canada's linguistic resources*. Garamond Press.

Dagenais, D. (2003). Accessing imagined communities through multilingualism and immersion education. *Journal of Language, Identity, and Education, 2*(4), 269-283.

de Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition, 13*(1), 33-40.

de Saussure, F. (1959). *Course in General Linguistics*. Philosophical Library.

Derwing, T. M. (2020). Integrating multiple views and multiple disciplines in the understanding of child bilingualism and second language learning. In F. Li, K. E. Pollock, & R. Gibb (eds.), *Child bilingualism and second language learning: Multidisciplinary perspectives* (pp. 29-41). John Benjamins Publishing Company.

Derwing, T. M. (2020). Issues for second language pronunciation in children. In F. Li, K. E. Pollock, & R. Gibb (eds.), *Child Bilingualism and Second Language Learning: Multidisciplinary perspectives* (pp. 29-44). John Benjamins Publishing Company.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Benjamins.

Derwing, T. M., Fraser, H., Kang, O., & Thomson, R. I. (2014). L2 accent and ethics: Issues that merit attention. In A. Mahboob & L. Barratt (eds.), *Englishes in multilingual contexts* (pp. 63-80). Springer, Dordrecht.

Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development, 23*(4), 245-259.

Dewaele, J. M., & McCloskey, J. (2015). Attitudes towards foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development, 36*(3), 221-238.

Dicks, J., & Genesee, F. (2017). Bilingual education in Canada. In O. García, A. Lin, and S. May (eds.), *Bilingual and Multilingual Education* (pp. 453-467). Springer.

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed). OUP Oxford.

Duff, P., & Doherty, L. (2019). Learning Chinese as a heritage language: Challenges, issues and ways forward. In C-R. Huang, Z. Jing-Schmidt, & B. Meisterernst (eds.), *The Routledge handbook of Chinese applied linguistics* (pp. 149-164). Routledge.

Edmonton Chinese Bilingual Education Association (ECBEA). (n.d.). About the program. Edmonton Chinese Bilingual Education Association. https://www.ecbea.org/about-the-program/.

Estrada, P. L., & Chacón, A. M. (2015). A teacher's tensions in a Spanish first grade two-way bilingual immersion program. *Revista Portuguesa de Educação, 28*(2), 211-229.

Farris, C., Trofimovich, P., Segalowitz, N., & Gatbonton, E. (2008). Air traffic communication in a second language: Implications of cognitive factors for training and assessment. *Tesol Quarterly, 42*(3), 397-410.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*(3), 313-326.

Fishman, J. A. (1976). *Bilingual education: An international sociological perspective*. ERIC.

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America, 76*(3), 692-707.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). York Press.

Flege, J. E. (2021). New methods for second language (L2) speech research. In R. Wayland (ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 119-156). Cambridge University Press.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R.Wayland (ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3-83). Cambridge University Press.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*(5), 3125-3134.

Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication, 40*(4), 467-491.

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal,* 1-22.

Gandour, J. (1977). On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica, 34*(1), 54-65.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and Motivation in Second-Language Learning*. Newbury House Publishers.

Gatbonton, E., & Trofimovich, P. (2008). The ethnic group affiliation and L2 proficiency link: Empirical evidence. *Language Awareness, 17*(3), 229-248.

Genesee, F. (2004). What do we know about bilingual education for minority language students? In T. K Batia & W. Ritchie (eds.), *Handbook of bilingualism and multiculturalism* (pp. 547-576). Blackwell.

Genesee, F., & Lindholm-Leary, K. (2008). Dual language education in Canada and the USA. *Encyclopedia of Language and Education, 2*(5), 253-263.

Giles, H. (1970). Evaluative reactions to accents. *Educational Review, 22*(3), 211-227.

Golberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied psycholinguistics, 29*(1), 41-65.

Government of Canada. (2023). Multiculturalism. Canadian identity and society. Retrieved from https://www.canada.ca/en/services/culture/canadian-identity-society/multiculturalism.html.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107*(5), 2711-2724.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

Han, W., Brebner, C., & McAllister, S. (2016). Redefining 'Chinese' L1 in SLP: Considerations for the assessment of Chinese bilingual/bidialectal language skills. *International Journal of Speech-Language Pathology, 18*(2), 135-146.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*(2), 269-279.

Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics in Language Teaching, 45*, 353-378.

Hau, J. A., Holt, C. M., Finch, S., & Dowell, R. C. (2020). The adaptation to Mandarin-accented English by older, hearing-impaired listeners following brief exposure to the accent. *Journal of Speech, Language, and Hearing Research, 63*(3), 858-871.

Hedlund, G., & Rose, Y. (2020). Phon 3.1 [Computer Software]. Retrieved from https://phon.ca.

Holm, A., & Dodd, B. (2006). Phonological development and disorder of bilingual children acquiring Cantonese and English. In H. Zhu and B. Dodd (eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 286-325). Multilingual Matters.

Howatt, A. (1984). *A history of English language teaching*. Oxford University Press.

Howie, J. M. (1974). On the domain of tone in Mandarin. *Phonetica, 30*(3), 129-148.

Hu, C. F. (2021). Adaptation to an Unfamiliar Accent by Child L2 Listeners. *Language and Speech, 64*(3), 491-514.

Hult, F. M., & Hornberger, N. H. (2016). Revisiting orientations in language planning: Problem, right, and resource as an analytical heuristic. *The Bilingual Review, 33*(3), 30.

Idemaru, K., Wei, P., & Gubbins, L. (2019). Acoustic sources of accent in second language Japanese speech. *Language and Speech, 62*(2), 333-357.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal,* 1-12.

Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford University Press.

Kachru, B. B. (1990). World Englishes and applied linguistics. *World Englishes, 9*(1), 3-20.

Kan, R. T., & Schmid, M. S. (2019). Development of tonal discrimination in young heritage speakers of Cantonese. *Journal of Phonetics, 73*, 40-54.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology, 28*(4), 441-456.

Kang, O., Rubin, D., & Lindemann, S. (2015). Mitigating US undergraduates' attitudes toward international teaching assistants. *TESOL Quarterly, 49*(4), 681-706.

Kennedy, S., & Trofimovich, P. (2008). I. *Canadian Modern Language Review, 64*(3), 459-489.

Kennedy, S., & Trofimovich, P. (2017). Pronunciation acquisition. In S. Loewen & M. Sato (eds.), *The Routledge handbook of instructed second language acquisition* (pp. 260-279). Routledge.

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14*(1), 106-111.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America, 142*(3), 1693-1706.

Kuang, J., & Liberman, M. (2016). The effect of vocal fry on pitch perception. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5260-5264. Shanghai, China.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & 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.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.

Lambert, W. E., & Tucker, G. R. (1972). *Bilingual education of children: The St. Lambert experiment*. Newbury House Publishers.

Lee-Kim, S. I. (2014). Revisiting Mandarin 'apical vowels': An articulatory and acoustic study. *Journal of the International Phonetic Association, 44*(3), 261-282.

Lee, J. S. (2019). Teacher as change agent: Attitude change toward varieties of English through teaching English as an international language. *Asian Englishes, 21*(1), 87-102.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice, 2*(12), 59-67.

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation, 6*(3), 310-328.

Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly, 50*(4), 894-931.

Levis, J., & Sonsaat, S. (2017). Pronunciation teaching in the early CLT era. In O. Kang, R.I. Thomson, & J.M. Murphy (eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 267-283). Routledge.

Li, F., & Munson, B. (2016) The development of voiceless sibilant fricatives in Putonghua-speaking children. *Journal of Speech, Language, and Hearing Research, 59*, 699-712.

Li, R. (1989). Languages and Fangyan in China: instruction of Language Atlas of China. *Fangyan, 3,* 161-167.

Li, S., & Huang C-R. (2010). Chinese word segmentation based on word boundary decision. *Journal of Chinese Information Processing, 24*(1), 3-7.

Lin, L. C., & Johnson, C. J. (2010). Phonological patterns in Mandarin-English bilingual children. *Clinical Linguistics & Phonetics, 24*(4-5), 369-386.

Lin, Y., Lin, Y. J., Wang, F., Wu, X., & Kong, J. (2020). The development of phonological awareness and Pinyin knowledge in Mandarin-speaking school-aged children. *International Journal of Speech-Language Pathology, 22*(6), 660-668.

Lin, Y., Pollock, K. E., & Li, F. (2022). Teaching Mandarin pronunciation in Chinese-English bilingual schools in Canada. *The 13th Annual Pronunciation in Second Language Learning and Teaching Conference (PSLLT2022)*, Brock University, ON.

Lin, Y., Pollock, K. E., MacLeod, A. A. N., & Li, F. (2023). A conceptual model of second language pronunciation in communicative contexts: Implications for children's bilingual education. *Frontiers of Psychology, 14*, 1125157.

Lindberg, R., and Trofimovich, P. (2020). Second language learners' attitudes toward French varieties: The roles of learning experience and social networks. *Modern Language Journal, 104*(4), 822-841. doi: 10.1111/modl.12674

Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society, 31*(3), 419-441.

Lindemann, S., & Subtirelu, N. (2013). Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning, 63*(3), 567-594.

Lippi-Green, R. (2011). *English with an Accent: Language, Ideology, and Discrimination in the United States* (2nd ed). Routledge.

Littlewood, W. (2011). Communicative language teaching: an expanding concept for a changing world. In E. Hinkel (ed.), *Handbook of research in second language teaching and learning* (Vol. 2). Routledge.

Littlewood, W. T. (1980). Form and meaning in language-teaching methodology. *Modern Language Journal*, 441-445.

Liu, C., Jin, S. H., & Chen, C. T. (2014). Durations of American English vowels by native and non-native speakers: Acoustic analyses and perceptual effects. *Language and Speech, 57*(2), 238-253.

Liu, W. (2020). Success factors for a Mandarin bilingual program: An autoethnographic case study. *Chinese as a Second Language (漢語教學研究—美國中文教師學會學報), 55*(3), 208-229.

Lu, L., & Liu, H. S. (1998). *The Peabody picture vocabulary test-Revised in Chinese*. Psychological Publishing.

Luk, G., & Esposito, A. G. (2020). BLC mini-series: Tools to document bilingual experiences. *Bilingualism: Language and Cognition, 23*(5), 927-928.

MacDonald, S. (2002). Pronunciation-views and practices of reluctant teachers. *Prospect 17*(3), 3-18.

MacLeod, A. A., & Stoel-Gammon, C. (2010). What is the impact of age of second language acquisition on the production of consonants and vowels among childhood bilinguals?. *International Journal of Bilingualism, 14*(4), 400-421.

Mady, C. (2017). The bilingual advantage for immigrant students in French immersion in Canada: Linking advantages to contextual variables. *International Journal of bilingual education and bilingualism, 20*(3), 235-251.

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

Marinova-Todd, S. H., Zhao, J., & Bernhardt, M. (2010). Phonological awareness skills in the two languages of Mandarin-English bilingual children. *Clinical Linguistics and Phonetics, 24*(4-5), 387-400.

Marx, N. (2002). Never quite a 'native speaker': Accent and identity in the L2 - and the L1. *Canadian Modern Language Review, 59*(2), 264-281.

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

McLeod, S., & Baker, E. (2017). *Children's speech: An evidence-based approach to assessment and intervention*. Pearson.

McMillen, S., Anaya, J. B., Peña, E. D., Bedore, L. M., & Barquin, E. (2022). That's hard! Item difficulty and word characteristics for bilinguals with and without developmental language disorder. *International Journal of Bilingual Education and Bilingualism, 25*(5), 1838-1856.

MDBG. (2023). CC-CEDICT. Retrieved from https://www.mdbg.net/chinese/dictionary?page=cedict.

Menke, M. R. (2017). Phonological development in two-way bilingual immersion: The case of Spanish vowels. *Journal of Second Language Pronunciation, 3*(1), 80-108.

Menken, K., & García, O. (2010). *Negotiating language education policies: Educators as policymakers*. Routledge.

Meziane, R. S. & MacLeod, A. A. N. (2020). Integrating multiple views and multiple disciplines in the understanding of child bilingualism and second language learning. In F. Li, K.E. Pollock, & R. Gibb (eds.), *Child bilingualism and second language learning: Multidisciplinary perspectives* (pp. 9-27). John Benjamins Publishing Company.

Ministry of Education of the People's Republic of China. (2012). *The Elementary School Chinese Curriculum in Compulsory Education Period - Revised*. People's Education Press.

Mok, P. P. K., & Lee, A. (2018). The acquisition of lexical tones by Cantonese-English bilingual children. *Journal of Child Language, 45*(6), 1357-1376.

Monfared, A. (2019). Ownership of English in the Outer and Expanding Circles: teachers' attitudes toward pronunciation in ESL/EFL teaching contexts. *Asian Englishes, 21*(2), 207-222.

Mougeon, R., & Rehner, K. (2017). The influence of classroom input and community exposure on the learning of variable grammar. *Bilingualism, 20*(1), 21.

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Learning, 20*, 139-154.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73-97.

Munro, M. J., & Derwing, T. M. (2020). Collecting data in L2 pronunciation research. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching conference* (pp. 2380-9566). Ames, Iowa.

Munro, M. J., Derwing, T. M., & Sato, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers. *Prospect, 21*(1), 67-79.

Murphy, D. (2011). An investigation of English pronunciation teaching in Ireland. *English Today, 27*(4), 10-18.

Nance, C. (2020). Bilingual language exposure and the peer group: Acquiring phonetics and phonology in Gaelic Medium Education. *International Journal of Bilingualism, 24*(2), 360-375.

Netelenbos, N., Li, F., & Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism, 20*(3), 346-357.

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning, 50*(1), 57-85.

Olson, L. L., & Jay Samuels, S. (1973). The relationship between age and accuracy of foreign language pronunciation. *The Journal of Educational Research, 66*(6), 263-268.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism, 1*(3), 213-237.

Piske, T., Flege, J. E., MacKay, I. R., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica, 59*(1), 49-71.

Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition, 19*(5), 1023-1041.

Porretta, V., Kyröläinen, A. J., & Tucker, B. V. (2015). Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception, & Psychophysics, 77*(7), 2438-2451.

R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reiterer, S. M. (2019). Neuro-psycho-cognitive markers for pronunciation/speech imitation as language aptitude. In Z. Wen, P. Skehan, A. Biedroń, S. Li, and R. Sparks (eds.), *Language Aptitude* (pp. 277-298). Routledge.

Rhee, N., Chen, A., & Kuang, J. (2020). Integration of spectral cues in the development of Mandarin tone production. In *Proceedings of the 19th International Congress of Phonetic Science*s, 3135-3138. Melbourne, Australia.

Richter, K. (2019). *English-medium instruction and pronunciation: Exposure and skills development.* Multilingual Matters.

Rivers, D. J., & Ross, A. S. (2020). L1/L2 communication self-efficacy beliefs and the contribution of personality. *The Language Learning Journal, 48*(6), 700-714.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46-73.

Roper Public Affairs. (2006). National Geographic-Roper public affairs 2006 geographic literacy study. National Geographic Foundation. https://media.nationalgeographic.org/assets/file/NGS-Roper-2006-Report.pdf.

Rose, Y., & Stoel-Gammon, C. (2015). Using PhonBank and Phon in studies of phonological development and disorders. *Clinical Linguistics and Phonetics, 29*(8-10), 686-700.

Ruiz, R. (1984). Orientations in language planning. *Bilingual Research Journal: The Journal of the National Association for Bilingual Education, 8*(2), 15-34.

Ryan, E. B. (1973). Subjective reactions toward accented speech. In R. W. Shuy & R. W. Fasold (eds.), *Language attitudes: Current trends and prospects* (pp. 60-73). Georgetown University Press.

Saito, K., Dewaele, J. M., & Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech, 60*(4), 614-632.

Sardegna, V. G., Lee, J., & Kusey, C. (2014). Development and validation of the learner attitudes and motivations for pronunciation (LAMP) inventory. *System, 47*, 162-175.

Sardegna, V. G., Lee, J., & Kusey, C. (2018). Self-efficacy, attitudes, and choice of strategies for English pronunciation learning. *Language Learning, 68*(1), 83-114.

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science, 11*(2), e1521.

Schwartz, M., Wee Koh, P., Xi Chen, B., Sinke, M., & Geva, E. (2016). Through the lens of teachers in two bilingual programmes: a look at early bilingual education. *Language, Culture and Curriculum, 29*(2), 141-168.

Shi, F., & Wang, P. (2006). A statistic analysis of tone groups in Beijing Mandarin. *Contemporary Linguistics, 8*(4), 324-333.

Shintani, N., Saito, K., & Koizumi, R. (2019). The relationship between multilingual raters' language background and their perceptions of accentedness and comprehensibility of

second language speech. *International Journal of Bilingual Education and Bilingualism, 22*(7), 849-869.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708-722.

Simmons, R. V. (2016). The evolution of the Chinese 四呼 concept of syllable classification. *Historiographia Linguistica, 43*(3), 251-284.

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development, 4*(2), 157-178.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*(2), 273-293.

Southwood, M. H., & Flege, J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics and Phonetics, 13*(5), 335-349.

Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly, 42*(2), 181-207.

Statistics Canada. (2022). More than half a million people speak predominantly Mandarin or Punjabi at home in Canada. Retrieved from https://www150.statcan.gc.ca/n1/daily-quotidien/220817/cg-a001-png-eng.htm.

Subtirelu, N. C., & Lindemann, S. (2016). Teaching first language speakers to communicate across linguistic difference: Addressing attitudes, comprehension, and strategies. *Applied Linguistics, 37*(6), 765-783.

Sun, M. (2011). The language and literacy practices of English-Chinese bilingual students in Western Canada. [Doctoral Dissertation, University of Alberta]. University of Alberta Library.

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning, 26*(2), 233-253.

Suzukida, Y. (2021). The contribution of individual differences to L2 pronunciation learning: Insights from research and pedagogical implications. *RELC Journal, 52*(1), 48-61.

Tavares, A. J. (2000). From heritage to international languages: Globalism and Western Canadian trends in heritage language education. *Canadian Ethnic Studies Journal, 32*(1), 156-182.

The Edmonton Chinese Bilingual Education Association (ECBEA). (2007). *Talking to the World: Edmonton's Chinese (Mandarin) bilingual program: the first 25 years*. High Speed Printing.

The Edmonton Chinese Bilingual Education Association (ECBEA). (n.d.). About the program. Edmonton Chinese Bilingual Education Association. https://www.ecbea.org/about-the-program/

The Office of Modern Chinese at Peking University. (2006). *Modern Chinese (revised)*. Commercial Press.

Tokumoto, M., & Shibata, M. (2011). Asian varieties of English: Attitudes towards pronunciation. *World Englishes, 30*(3), 392-408.

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349-357.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1-30.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*(4), 905-916.

Trotter, R. T. (2012). Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives. *Preventive Medicine, 55*(5), 398-400.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2003). The perception and production of English /ɛ/and /æ/ by Korean children and adults living in North America. In *15th International Congress of Phonetic Sciences* (pp. 1589-1592). Barcelona, Spain

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica, 71*(1), 4-21.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., & Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics, 35*(4), 765-805.

van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology, 6*, 1000.

van Rij, J., Wieling, M., Baayen, R., van Rijn, H. (2022). itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.4.1.

Wang, Y., Jongman, A., & 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.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly, 4*(2), 123-130.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics, 18*(3), 345-373.

Werker, J.F. & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197-234.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86-116.

Wisecup, P. A. (2017). Teacher Voices Regarding a Bilingual Education Program: A Qualitative Inquiry (Publication No. 10616769). [Doctoral dissertation, Northcentral University]. ProQuest Dissertations Publishing.

Wong, P. (2012). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*(1), 141-151.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC Press.

Wright, D. (2021). Fifty years on, Canada's multiculturalism policy remains a pillar of its diversity. The Global and Mail. Retrieved from https://www.theglobeandmail.com/opinion/article-50-years-on-canadas-multiculturalism-policy-remains-a-pillar-of-its/.

Wu, J. Z. (2005). Bilingual education in western Canada and Chinese language minority students' self perceptions of their citizenship and ethnicity. *Canadian and International Education, 34*(1), 23-30.

Wu, J., & Bilash, O. (2000). *Bilingual Education in practice: A multifunctional model of minority language programs in Western Canada*. University of Alberta.

Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., & Demuth, K. (2018). Five-year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology, 9*, 817.

Xu, W., & Case, R. E. (2015). Age-related differences in motivation in learning English among Mainland Chinese students. *International Journal of Applied Linguistics, 25*(1), 67-82.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61-83.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*, 179-203.

Xu, Y. (2013). ProsodyPro-A Tool for Large-scale Systematic Prosody Analysis. *TRASP*, Aix-en-Provence, 7-10.

Yang, C. (2016). *The acquisition of L2 Mandarin prosody: From experimental studies to pedagogical practice* (Vol. 1). John Benjamins Publishing Company.

Yang, J., & Fox, R. A. (2017). L1-L2 interactions of vowel systems in young bilingual Mandarin-English children. *Journal of Phonetics, 65*, 60-76.

Yang, J., & Liu, C. (2012). Categorical Perception of Lexical Tone in 6 to 8-year-old Monolingual and Bilingual Children. *International Journal of Asian Language Processing, 22*(2), 49-62.

Yang, L., Lin, Y., Pollock, K. E., & Li, F. (2021). Accuracy of spontaneous and imitative productions in children learning Mandarin in a bilingual program. *The International Child Phonology Conference (ICPC2021)*, Lethbridge, June 17-19.

Yang, L., Pollock, K. E., Lin, Y., Tucker, B. V., & Li, F. (2023). Fricative production of children learning Mandarin in a Chinese-English bilingual program. *Bilingualism and Multilingualism Conference (BAM)*, Kelowna, May 13-15.

Zhang, J. (2018). A comparison of tone normalization methods for language variation research. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*. Hong Kong.

Zheng, X., & Gao, Y. (2017). Language attitude education as ICC facilitation: An explorative class. *Chinese Journal of Applied Linguistics, 40*(1), 3-20.

Zhu, H. (2002). *Phonological development in specific contexts: Studies of Chinese-speaking children*. Multilingual Matters.

Appendices

Appendix A. Scheme of the Chinese Phonetic Alphabet (Commission of Written Language Reformation, 1958) in Chapter 3.

1. 字母表 *zìmǔ biǎo*, Table of the Alphabet

Pinyin alphabet is the same as the English alphabet, therefore is omitted here.

2. 声母表 *shēngmǔ biǎo*, Table of Onsets

b [p]* 玻 [po]**	p [p^h] 坡 [p^ho]	m [m] 摸 [mo]	f [f] 佛 [fo]	d [t] 得 [tɤ]	t [t^h] 特 [t^hɤ]	n [n] 讷 [nɤ]	l [l] 勒 [lɤ]
g [k] 哥 [kɤ]	k [k^h] 科 [k^hɤ]	h [x] 喝 [xɤ]		j [tɕ] 基 [tɕi]	q [tɕ^h] 欺 [tɕ^{h}i]	x [ɕ] 希 [ɕi]	
zh [tʂ] 知 [tʂɻ]	ch [tʂ^h] 蚩 [tʂ^{h}ɻ]	sh [ʂ] 诗 [ʂɻ]	r [ʐ/ɻ] 日 [ʐɻ]	z [ts] 资 [tsɻ]	c [ts^h] 雌 [ts^{h}ɻ]	s [s] 思 [sɻ]	

* The IPA transcriptions follow the convention of 北京大学中文系现代汉语教研室 The Office of Modern Chinese PKU, 2006. Differences will be indicated.

** The 呼读音 *hūdúyīn* (sound names) were indicated in the scheme with Chinese characters and 注音符号 *zhùyīn fúhào*, a.k.a. Bopomofo, which was used before Pinyin and is still used in Taiwan. Vowels were added to fulfill the phonotactic rules of Mandarin. Here the Chinese characters are transcribed into IPA for readers' convenience.

3. 韵母表 *yùnmǔ biǎo*, Table of Rimes

	i 衣 [i]*	u 乌 [u]	ü 迂 [y]
a 啊 [a]	ia 呀 [ja] **	ua 蛙 [wa]	
o 喔 [o]		uo 窝 [wo]	
e 鹅 [ɤ]	ie 耶 [jɛ]		üe 约 [ɥɛ]
ai 哀 [aɪ] ***		uai 歪 [waɪ]	

ei 欸 [eɪ]		uei 威 [weɪ]	
ao 熬 [ɑʊ]	iao 腰 [jɑʊ]		
ou 欧 [oʊ]	iou 忧 [joʊ]		
an 安 [an]	ian 烟 [jɛn]	uan 弯 [uan]	üan 冤 [ɥæn]
en 恩 [ən]	in 因 [in]	uen 温 [wən]	ün 晕 [yn]
ang 昂 [ɑŋ]	iang 央 [jɑŋ]	uang 汪 [wɑŋ]	
eng [əŋ] 亨的韵母 The rime of [xəŋ]	ing 英 [iŋ]	ueng 翁 [wəŋ]	
ong [uŋ] 轰的韵母 The rime of [xuŋ]	iong 雍 [juŋ]		

* The Chinese characters represent the rimes. These are the characters for syllables with zero consonant, with a couple of exceptions. The original scheme also included Bopomofo in the rime table. The rimes were transcribed into IPA for the readers' convenience.

** Again, it remains controversial whether the "medial" is a vowel or a glide. We use glides to transcribe medials in this paper.

*** Similarly, the transcription convention of coda vowels remains controversial. We use lax vowels.

4. 声调符号 *shēngdiào fúhào*, Marks for Tones

阴平* yīnpíng	阳平 yángpíng	上声 shǎngshēng	去声 qùshēng
ˉ	ˊ	ˇ	ˋ

* These are the tone categories defined by historical phonology. In Standard Mandarin, these four categories can be described as high level, high rising, low dipping, and high falling tones according to the shapes of the tonal contours.

The tonal marks should be on top of the nucleus vowel. Neutral tones are not marked. For

example:

妈 mā	麻 má	马 mǎ	骂 mà	吗 ma

5. 隔音符号 *géyīn fúhào*, Mark for Syllable Boundaries

When a syllable that starts with a, o, or e is preceded by another syllable, and the syllable boundary is ambiguous, use a syllable boundary mark to demarcate, for example, pi’ao (皮袄 [pi.ɑʊ]).

Appendix B. COREQ 32-item checklist (Tong et al., 2007) in Chapter 4 (Paper 2).

No.	Item	Presence/Absence in this study
Domain 1: Research team and reflexivity		
Personal Characteristics		
1.	Interviewer	Reported in Method: The interviewer was the first author.
2.	Credentials	Reported in Method: The interview guidelines were designed by the PI of the project, who is a professor; the interviewer is a Ph.D. student.
3.	Occupation	Reported in Method: The first author is a Ph.D. student.
4.	Gender	The interviewer is female.
5.	Experience and training	Reported in Method: The interviewer had backgrounds in Chinese linguistics and phonetics.
Relationship with participants		
6.	Relationship established	The authors had been working in two of the schools with students prior to interviewing the teachers. In the third school, interviews took place before working with students.
7.	Participant knowledge of the interviewer	Participants knew the interviewer through the research project.
8.	Interviewer characteristics	Reported in Method: The interviewer was proficient in both Mandarin and English and was able to conduct the interview in the teachers' preferred language.
Domain 2: study design		
Theoretical framework		
9.	Methodological orientation	Reported in Method: thematic analysis.
Participant selection		
10.	Sampling	Reported in Method: voluntary and snowball.
11.	Method of approach	Flyer, face-to-face, email.
12.	Sample size	Reported in Method: twelve Chinese teachers.
13.	Non-participation	Teachers may choose to not sign up, usually due to time conflicts. There was no drop-out among the teachers who signed up.
Setting		
14.	Setting of data collection	Reported in Method: in classrooms or study rooms
15.	Presence of non-participants	Teacher On (pseudonym) was interviewed during lunch break with students present occasionally. Teacher Ding (pseudonym) was interviewed with Teacher Cheng (pseudonym) present.
16.	Description of sample	Reported in Method.

Data collection		
17.	Interview guide	Reported in Appendix B.
18.	Repeat interviews	N/A.
19.	Audio/visual recording	Reported in Method: audio recorded.
20.	Field notes	Yes.
21.	Duration	Reported in Method: 15-45 minutes.
22.	Data saturation	Reported in Method.
23.	Transcripts returned	No.
Domain 3: analysis and findings		
Data analysis		
24.	Number of data coders	Reported in Method: two coders.
25.	Coding tree	Reported in Appendix D.
26.	Derivation of themes	Reported in Method: derived from the data.
27.	Software	NVivo 12.
28.	Participant checking	Yes. We sent a drafted manuscript to the teachers for feedback.
Reporting		
29.	Quotations presented	Yes. See Result. All themes and subthemes were supported by quotes.
30.	Data and findings consistent	See Result and Discussion.
31.	Clarity of major themes	See Result.
32.	Clarity of minor themes	See Result.

Appendix C. Teacher interview guidelines in Chapter 4 (Paper 2).

1. What is your name?
2. Where were you born?
3. Are you a native speaker of Mandarin?
4. On a scale of 1-7 (1: nonnative; 7; proficient), can you rate your proficiency as a Mandarin speaker?
5. How long have you worked in the Mandarin-English bilingual program, and in what capacity (i.e. grade level, subject content areas)?
6. *Note: Teachers will be told that we are not asking for specific information on individual children, but more general impressions based on their experience teaching in the program.* What is your overall impression of your students' Mandarin pronunciation? Do they typically make progress in their production of Mandarin sounds and tones? In general, what are the most difficult sounds for children to articulate?
7. Do children communicate in Mandarin in class, after class, and in other occasions during school time?
8. What are the challenges that you have faced in teaching children new Mandarin sounds or those sounds that are confusable to English?
9. Do you have any tips for teaching Mandarin sounds or tones?
10. In your view, what is the best way to improve pronunciation instruction in a second language in a classroom setting?

Appendix D. Codebook in Chapter 4 (Paper 2).

Name	Description	Notes
Mandarin speech	General impression of children's speech.	
Difficulty	Mandarin speech development is difficult or challenging. Or the opposite.	
Progression	The trajectory can be progressive, regressive, or remain the same.	
Speech proficiency	Comments on whether their speech is fluent, intelligible, accented/standard, etc.	
Difficult phonetic units	Sounds that are hard to pronounce, easily confused with English sounds, and hard to teach.	
Connected speech	Larger than words, or interactions between sounds, tones, and larger units.	
Consonants	Consonants.	
Others	Non-phonetic difficulties mentioned under the question (e.g. symbol-sound relationship).	
Tones	Tones.	
Vowels	Vowels.	
Factors	Factors that impact speech development.	
Family	Family background and environments.	
Linguistic transfer	Specific impacts of native languages or the dominant languages.	Split from “family background.”
Individual	Individual motivation, interest, talent, cognition, and efforts.	
Peers and environment	Friends, classmates, other communication partners, and society.	Used to be “other.”
Teacher	Teacher's accent or teaching methods.	
Use of Mandarin	Under the question: Do the students use Mandarin to communicate at school?	
Negative	Children do NOT use Mandarin to communicate, the scenarios or situations, and possible reasons.	Merged with a previous code “English dominance”
Positive	Children DO use Mandarin to communicate, scenarios or situations, and possible reasons.	
Teaching strategies	Under the question of teaching strategies.	
Direct strategies	Strategies that directly target the pronunciation of specific sounds.	“Practice” is under this if it addresses specific speech elements.

Indirect strategies	Strategies that do not directly address speech elements.	"Practice" is under this if it addresses language use.
Philosophy	What teachers think is important, would be helpful, or should happen.	Philosophies of L2 teaching and Chinese teaching are merged since they can overlap.
Attitudes to *Pīnyīn*	Broad opinions or statements of *Pīnyīn*.	
Foundation	The importance of lower grades.	
Motivation and interests	Teachers should encourage motivation and interest.	As opposed to motivation as a factor of speech development.
CLT	Intelligibility, language use, and communication needs.	"Practice" is under CLT if it addresses language use.
Attitudes to teachers' backgrounds	Opinions about how teachers' language backgrounds impact their teaching.	
Cultural differences	Cultural differences in teaching styles, teaching goals, and teaching materials.	
Inclusivity	Support multilevel children with diverse backgrounds.	
Needs	"We don't have…" "... is not good/ideal" "We need …" "Can you …"	
Curriculum	Statements about the curriculum design	
Teaching resources	Software, website, books, videos, human resources, more time, etc.	
Training	Training on teaching methods or on Chinese.	

Appendix E. A comparison between teachers' reflections on challenging speech sounds in Mandarin and the phonological patterns in bilingual children in Taiwan (Lin & Johnson, 2010) in Chapter 4 (Paper 2).

Our study	**Lin & Johnson (2010)**
Qualitative study based on teacher interviews	Quantitative study based on speech tests
Students in kindergarten to grade 5	Students in preschool
A Chinese-English bilingual program in Canada, where English is the majority ambient language	An English immersion school in Taiwan, where Mandarin is the majority ambient language
Speech sounds that are challenging to learn	Phonological processes based on Hua (2002)
Consonants (number of mentionings)	
Alveolo-palatal "j q x" /tɕ tɕʰ ɕ/ (11)	Fronting (e.g., /ɕ/ -> [s], /ɕ/ -> [ʂ]) Stopping (e.g., /tɕ/ -> [t]) Velarization (e.g., /ɕi/ -> [xi]) Affrication and deaffrication (e.g., /ɕ/ <-> [tɕ]) Aspiration and deaspiration (e.g., /tɕ/ <-> [tɕʰ])
Retroflex "zh ch sh r" /tʂ tʂʰ ʂ ʐ/ (11)	Fronting (e.g., /ʂ/ -> [s]) Stopping (e.g., /tʂ/ -> [t]) Velarization (e.g., /ʂu/ -> [xu]) De-retroflexion (e.g., /ʂ/ -> [s] or [ʃ]) Palatalization (e.g., /ʂ/ -> [ɕ])
Alveolar "z c s" /ts tsʰ s/ (5)	Backing (e.g., /s/ -> [ʃ]) Stridency deletion (e.g., /s/ -> [θ])
Coda "n" and "ng" (3)	/n/ -> [ŋ]
Not mentioned by the teachers	Consonant deletion Assimilation Liquidization (e.g., /n/ -> [l]) Denasalization (e.g., /n/ -> [t])
Vowels (number of mentionings)	
Front-high rounded vowel /y/ (1)	/y/ deviation
Rhotic vowel /ɚ/ (1)	
Apical vowel /ɿ/ -> [ɚ] (1)	
Diphthong /a͡ʊ/ -> [o͡ʊ] (1)	
Not mentioned by teachers	Diphthong and triphthong reduction

Appendix F. Strategies used to teach Mandarin speech reported by teachers in Chapter 4 (Paper 2).

Direct strategies that addressed pronunciation as an object		
Practice	Repetition (massed practice, practicing multiple times in a row)	Over and over again. (Teacher On) It's not enough to teach the students once or twice, at least 7 times. (Teacher Bai) If it is not correct, I will play it again … Repetition is the most important. (Teacher Ding) Let them try it all the time. (Teacher Hsu) One story, he reads it for the first time and the second time, there is a big difference between his fluency, pronunciation, and his recognition of Chinese characters. (Teacher Jia)
	Review (distributed practice, practicing with intervals)	I have to reinforce them ceaselessly … If I find that there are some words that most kids didn't get, I will take the word out again, and then practice with them. (Teacher Feng) Practice and listen to it every week. (Teacher Hsu) Just continually reinforce the tones. (Teacher Jia)
	"Practice"	It's to practice. They have to speak more … Non-stop practice … I did flashcards. (Teacher Feng) There is nothing better than more practice. (Teacher Jia) I think it's just practice. (Teacher Young) Read aloud or elocute … It's not rote memorization. It's to make them familiar with the flow of the speech and the sounds. (Teacher Liu)
Auditory models	Teacher models	Teach him that Tone2 is produced like this and Tone3 is produced like this. (Teacher Ding) I let them follow me one sentence at a time. (Teacher Yi) The teacher demonstrates it. (Teacher Lee) The group read after me. (Teacher Jia)
	Peer models	There are always some students whose pronunciation is very good. That is my model. You read it again. Everyone read it. (Teacher Ding) There are three students who can speak Chinese very well … They are our little counsellors. Sometimes my students come to ask me questions … I will tell them… "you can ask [Student Name]." … If he doesn't know, I can help him; if he speaks well, I will confirm that your pronunciation is very good. You can help classmates in this way in the future. (Teacher Yi)
Perceptual training	Listening (no specific focus, no production required)	Watch Chinese … to improve his listening ability. (Teacher Bai) We will watch videos and do listening exercises … continuously listening until getting the feeling. (Teacher Hsu) I think they'd better listen more. (Teacher Jia) The first thing is to listen, to make him familiar … with the correct pronunciation. (Teacher Liu)
	Emphasis or	I will exaggerate the tones … So every time I say "lai2," I will focus on the latter part … Do you feel the tone is rising? …

	exaggeration of certain features	Then Tone3, I will do "lai3—" … to emphasize the tones. (Teacher Feng) We will use the exaggerated pronunciation method … Especially when we talk about Tone3. (Teacher Hsu) Therefore, "ma3" (horse) will be "ma3—" which is very exaggerated. In this case, the students can hear it clearly. (Teacher Lee) I will say "mi4—feng1—." (Teacher Liu)
	Contrasting confusable categories	Comparing in contrast is always really good for kids. (Teacher On) You listen, this is Tone2, this is Tone3. (Teacher Ding) Tell the difference … I laughed at them, then I said, "you see, when you can not pronounce the tones clearly, you're making jokes." (Teacher Feng) I will deliberately compare the two [Tone2 and Tone3] … Let them listen to see if they hear the difference. (Teacher Hsu)
	Error identification	Then let them read to each other … What do you think is wrong with him? Picking someone else's mistakes is very powerful … They will pick out the mistakes like nitpicking. (Teacher Ding) I told them to pretend to be a teacher … but the problem was that if some children couldn't hear the difference … they weren't actually learning anything, they just reinforced the wrong thing. (Teacher Feng)
Feedback	Direct and explicit correction	If the child is reading, I will correct his accent. (Teacher Bai) We often have problems, and we must correct them. (Teacher Yi) When I heard that they were off, I would correct them immediately. (Teacher Feng) If their pronunciation is wrong, I will always correct them. (Teacher Hsu) If I am reading with them, if there are some words which are the most basic ones and produced wrong, I will definitely ask them to correct and say it again. (Teacher Lee) "I like da2gou3," I'll say, "Tone4, da4 … Say it again." (Teacher Liu)
	Recasting (repeating what the student said with correct models)	Some children bravely speak Mandarin with you, but they will have many errors in tones. You have to tirelessly repeat. (Teacher Yi) I'll repeat, "I also like to eat big (da4) apples." They will realize that they were wrong. (Teacher Liu)
Cueing	Visual cues	Look at my teeth. Do I bite it? … I like visuals. (Teacher On) I will draw the tone contours and show them that this is not moving, while this is. (Teacher Gwok)
	Tactile cues	Feel the back when you swallow. (Teacher On)
	Gestural cues	His body has to follow the movement to show that he is speaking Tone3. (Teacher Ding) I will use gestures to tell them … adding the action, so they can see. (Teacher Feng) Use the hand to demonstrate, to show that this is Tone1, this is Tone2, this is Tone3, and this is Tone4 (gesturing). (Teacher Gwok) I regard my left hand as your palate, the front is your teeth, this is your tongue, and what your tongue should be now, for

		example, the tongue may be retroflex like this or contact the alveolar ridge. (Teacher Hsu) I use gestures. (Teacher Liu)
	Written cues	I will put the *Pīnyīn* there for them when teaching the words. (Teacher Liu)
Explicit instruction	Rules	I will tell him … the tone sandhi rules of "yi1 (one)" and "pu4 (no)." These are… frequently used words. (Teacher Yi) I will explain a little [phonotactic] rules, but they don't understand it if I explain with too many details. (Teacher Hsu) I can only explain that it's the tone of Chinese, because there are Tones 1234. (Teacher Jia)
	Articulation	I tell them to focus on certain parts of their mouth. (Teacher On) You imagine that the tongue can be retroflex or flat. (Teacher Hsu) This tone is relatively long. (Teacher Hsu)
	Phonological awareness	I will say, I am saying "sheng. You will hear 'eng–' there." (Teacher Feng) In the first semester, we focus on individual sounds … In the second semester, we will test the combined sound. (Teacher Hsu) I will use words to teach, such as "爸爸 ba4ba0," "菠菜 bo1cai4," and "萝卜 luo2bo0." I'll finish these first, and then I will say that they are all "b." (Teacher Liu)
	Language transfer	I would sometimes tell them some sounds that are in English. (Teacher On) For Tone2, it's like you are asking questions. (Teacher Feng)
Indirect strategies that facilitated pronunciation learning		
Language environment to promote Mandarin use	Teacher Use	It should be 90%, more than 95% … The teacher uses Chinese as much as possible. (Teacher Bai) "[In Mandarin] Turn the lights off" – I am saying this every day. (Teacher Hsu) We will speak Chinese to students as much as possible … For all the learning in the classroom, I use Chinese. (Teacher Lee) Like, "[In Mandarin] Today, the weather is very nice." Just say a little more. (Teacher Liu)
	Language policies	In class … you must use Chinese. This is … the discipline of the whole school. (Teacher Bai) I require that Chinese must be spoken in the classroom. Not only speak Chinese with me, but also they must speak Chinese with their classmates. (Teacher Cheng) You still need to encourage students, in their daily conversations, let them [speak Mandarin]. (Teacher Yi) Using Chinese is for sure. (Teacher Gwok) When they are eating snacks, we will tell them to use Chinese. (Teacher Hsu) I have always stressed that … although you can say it in English if you don't know the word, for the basic words, … you should know how to say those words [in Mandarin]. (Teacher Lee) If he doesn't use it, I will ignore him. I just look at them … If they still don't speak [Chines], I say "I don't understand. I don't speak… English." (Teacher Liu) We try to create as much environment as possible to let them speak. (Teacher Liu)

	Encouragement and rewards	He said English when he spoke to others … Then I asked him to … write down the sentence … then write it again in Chinese. (Teacher Cheng) Every time I heard them speaking Chinese or something, I will make it a big deal! I have a system of reward, so every day or every couple of days, I will go through every child, then I will say, "did you speak Chinese?" … Every time they reach certain points, I will reward them. (Teacher Feng) I will hand out a table to them recording "I have spoken Chinese," and then I will tick on the table before and after the rest every day. Then there is a weekly draw, and for the winning persons … Winners can have a small gift. (Teacher Lee)
	Peer effect	Among the students, they monitor each other … Them checking each other is much more powerful than me. (Teacher Cheng) You should encourage students who have already had a good foundation to talk with students who are not good enough … Because there is a distance between the teacher and the student. (Teacher Yi) The ones that do know Mandarin that are from China. I use them as role models … When the other kids hear them say it, they'll be like, "I'll say it because she said it, and she's my friend." They feel more comfortable when there are some Mandarin-speaking kids around. (Teacher Young)
Activity-based learning	Language activities	In higher grades, there will be an application stage … I will share a Google slide with the children, then he … will apply his knowledge … and then he presents it. (Teacher Bai) We have a speech every day … It is very short. … It is like how many people there are in my family, mom and dad, what dad likes, what mom likes, or what activities you like to do … It's simple. (Teacher Ding) I would read for them. (Teacher Feng) When I do guided reading, we will have a group to read or practice speaking with friends, they can look at the photos, and then they make their own stories. (Teacher Gwok) We do a lot of routines in Chinese … I'll ask them things like "[In Mandarin] What colours do you like)", and then we'll talk about other colours, and then they'll learn to say it in a sentence. (Teacher Young) They had a play that they were doing … If they have something to read, and then the other person has something to read back, then that might give them ideas of … what they can say to their friends. (Teacher Young) Often in reading, such as stories, what characters there are, and which character you like. We ask these questions verbally. While as for writing … I emphasize saying it first … Especially the third-year children, … they are all doing verbal writing. (Teacher Liu) In September, there is a show-and-talk … Each student brings a toy … I'll give them 5 questions … Let them add to it. Then do they like it or not, why? They will say, I like my toy because it looks good, or cute. (Teacher Liu)
	Other subjects	Like mathematics, … record a video and tell me how to do addition … He would say, "[In Mandarin] Five plus five equals ten. Write zero and carry the one to the next digit." (Teacher Bai) I wanted to find some videos related to our health science … They not only learned the subject but also learned the language. (Teacher Ding) When I teach multiplication, I tell him multiplication in Chinese is "cheng2fa3." (Teacher Jia)
Multi-	TV shows and	Sometimes I play 喜羊羊 (a Chinese cartoon) … They find it to be cool … and they wanna understand it … I hope that

media	videos	motivates some to learn more. (Teacher On) 喜羊羊 is OK. It has a plot … It is easier to understand … It does not have fancy vocabulary … YouTube has a lot of dramas lasting for a few minutes. (Teacher Ding) I'll show them videos. (Teacher Jia) I video them so then the parents can see what it's like. So then they can practice with them at home when they're not saying "I don't speak Chinese, I can't practice with them." (Teacher Young)
	Songs	I am going to teach them the song 中国话. There are a lot of … tongue twisters. (Teacher Cheng) They listen to the songs that I used to listen to when I was a child, 对面的女孩看过来, 甜蜜蜜, 月亮代表我的心, and 朋友. They like them, really. Then you will see some foreign children, they will find out on YouTube to sing. (Teacher Gwok) There are lots of YouTube videos of singing for them … I sent home the songs and then to help the parents, I put *Pīnyīn* on the songs and then the parents can sing with them … I can say, "sing it to Twinkle Twinkle Little Star." so then they'll know how to sing it because the tones don't really matter when you sing. (Teacher Young)
	Internet	Another point is the internet. You just follow the teacher to speak which is more standard than me. (Teacher Ding) I like to use some Internet resources. (Teacher Yi) We use a lot of textbooks on the Internet. (Teacher Gwok) If students read the books online, many books have this function, it will read it out for you … If students don't know the word, they can press it, and the software will automatically read it out … The software iChinese Reader has a lot of books …They can read at home or on the iPad … They also have a quiz afterwards … They can collect points, then use the points to … create the clothes or hair for the characters. (Teacher Lee) They used to use iChinese Reader given by the school, that is, Chinese native speakers helped them read stories … Because not all the teachers here are Chinese native speakers … I like to give them the most standard thing. (Teacher Jia) There's another one … iChinese Reader maybe? … It tells you the story … You answer questions … I haven't used it much with kinder[garten] because it's too difficult for them. (Teacher Young)
	Games	Sometimes, we would do … Bingo … Sometimes I drew small prizes for them, then just clicked on them, and then read them to me … then they think that learning Chinese is not so boring. (Teacher Feng) I would also make a Bingo game … I will tell them the pronunciation … and then they have to find it out from the options. (Teacher Gwok)

Appendix G. A *Pīnyīn* to IPA dictionary in Mandarin in Chapter 5.

Pinyin	**IPA**	**Pinyin**	**IPA**	**Pinyin**	**IPA**	**Pinyin**	**IPA**
a	a	gou	koʊ̯	mo	mwo	song	suŋ
ai	aɪ̯	gu	ku	mou	moʊ̯	sou	soʊ̯
an	an	gua	kwa	mu	mu	su	su
ang	aŋ	guai	kwaɪ̯	na	na	suan	swan
ao	aʊ̯	guan	kwan	nai	naɪ̯	sui	sweɪ̯
ba	pa	guang	kwaŋ	nan	nan	sun	swən
bai	paɪ̯	gui	kweɪ̯	nang	naŋ	suo	swo
ban	pan	gun	kwən	nao	naʊ̯	ta	tʰa
bang	paŋ	guo	kwo	ne	nɤ	tai	tʰaɪ̯
bao	paʊ̯	ha	xa	nei	neɪ̯	tan	tʰan
bei	peɪ̯	hai	xaɪ̯	nen	nən	tang	tʰaŋ
ben	pən	han	xan	neng	nəŋ	tao	tʰaʊ̯
beng	pəŋ	hang	xaŋ	ni	ni	te	tʰɤ
bi	pi	hao	xaʊ̯	nian	njɛn	tei	tʰeɪ̯
bian	pjɛn	he	xɤ	niang	njaŋ	teng	tʰəŋ
biao	pjaʊ̯	hei	xeɪ̯	niao	njaʊ̯	ti	tʰi
bie	pjɛ	hen	xən	nie	njɛ	tian	tʰjɛn
bin	pin	heng	xəŋ	nin	nin	tiao	tʰjaʊ̯
bing	piŋ	hong	xuŋ	ning	niŋ	tie	tʰjɛ
bo	pwo	hou	xoʊ̯	niu	njoʊ̯	ting	tʰiŋ
bu	pu	hu	xu	nong	nuŋ	tong	tʰuŋ
ca	tsʰa	hua	xwa	nou	noʊ̯	tou	tʰoʊ̯
cai	tsʰaɪ̯	huai	xwaɪ̯	nu	nu	tu	tʰu
can	tsʰan	huan	xwan	nuan	nwan	tuan	tʰwan
cang	tsʰaŋ	huang	xwaŋ	nve	nɥɛ	tui	tʰweɪ̯
cao	tsʰaʊ̯	hui	xweɪ̯	nvn	nyn	tun	tʰwən

ce	tsʰɤ	hun	xwən	nuo	nwo	tuo	tʰwo
cen	tsʰən	huo	xwo	nv	ny	wa	wa
ceng	tsʰəŋ	ji	tɕi	o	o	wai	waɪ̯
cha	tʂʰa	jia	tɕja	ou	oʊ̯	wan	wan
chai	tʂʰaɪ̯	jian	tɕjɛn	pa	pʰa	wang	waŋ
chan	tʂʰan	jiang	tɕjaŋ	pai	pʰaɪ̯	wei	weɪ̯
chang	tʂʰaŋ	jiao	tɕjaʊ̯	pan	pʰan	wen	wən
chao	tʂʰaʊ̯	jie	tɕjɛ	pang	pʰaŋ	weng	wəŋ
che	tʂʰɤ	jin	tɕin	pao	pʰaʊ̯	wo	wo
chen	tʂʰən	jing	tɕiŋ	pei	pʰeɪ̯	wu	u
cheng	tʂʰəŋ	jiong	tɕjuŋ	pen	pʰən	xi	ɕi
chi	tʂʰɻ	jiu	tɕjoʊ̯	peng	pʰəŋ	xia	ɕja
chong	tʂʰuŋ	ju	tɕy	pi	pʰi	xian	ɕjɛn
chou	tʂʰoʊ̯	juan	tɕɥɛn	pian	pʰjɛn	xiang	ɕjaŋ
chu	tʂʰu	jue	tɕɥɛ	piao	pʰjaʊ̯	xiao	ɕjaʊ̯
chua	tʂʰwa	jun	tɕɥən	pie	pʰjɛ	xie	ɕjɛ
chuai	tʂʰwaɪ̯	ka	kʰa	pin	pʰin	xin	ɕin
chuan	tʂʰwan	kai	kʰaɪ̯	ping	pʰiŋ	xing	ɕiŋ
chuang	tʂʰwaŋ	kan	kʰan	po	pʰwo	xiong	ɕjuŋ
chui	tʂʰweɪ̯	kang	kʰaŋ	pou	pʰoʊ̯	xiu	ɕjoʊ̯
chun	tʂʰwən	kao	kʰaʊ̯	pu	pʰu	xu	ɕy
chuo	tʂʰwo	ke	kʰɤ	qi	tɕʰi	xuan	ɕɥɛn
ci	tsʰɿ	kei	kʰeɪ̯	qia	tɕʰja	xue	ɕɥɛ
cong	tsʰuŋ	ken	kʰən	qian	tɕʰjɛn	xun	ɕyn
cou	tsʰoʊ̯	keng	kʰəŋ	qiang	tɕʰjaŋ	ya	ja
cu	tsʰu	kong	kʰuŋ	qiao	tɕʰjaʊ̯	yan	jɛn
cuan	tsʰwan	kou	kʰoʊ̯	qie	tɕʰjɛ	yang	jaŋ
cui	tswеɪ̯	ku	kʰu	qin	tɕʰin	yao	jaʊ̯
cun	tsʰwən	kua	kʰwa	qing	tɕʰiŋ	ye	jɛ

cuo	tsʰwo	kuai	kʰwa͡ɪ	qiong	tɕʰjuŋ	yi	i
da	ta	kuan	kwan	qiu	tɕʰjo͡ʊ	yin	in
dai	ta͡ɪ	kuang	kʰwaŋ	qu	tɕʰy	ying	iŋ
dan	tan	kui	kʰwe͡ɪ	quan	tɕʰyɛn	yo	jo
dang	taŋ	kun	kʰwən	que	tɕʰɥɛ	yong	juŋ
dao	ta͡ʊ	kuo	kʰwo	qun	tɕʰyn	you	jo͡ʊ
de	tɤ	la	la	ran	ʐan	yu	y
dei	te͡ɪ	lai	la͡ɪ	rang	ʐaŋ	yuan	ɥɛn
den	tən	lan	lan	rao	ʐa͡ʊ	yue	ɥɛ
deng	təŋ	lang	laŋ	re	ʐɤ	yun	yn
di	ti	lao	la͡ʊ	ren	ʐən	za	tsa
dia	tja	le	lɤ	reng	ʐəŋ	zai	tsa͡ɪ
dian	tjɛn	lei	le͡ɪ	ri	ʐɻ	zan	tsan
diao	tja͡ʊ	leng	ləŋ	rong	ʐuŋ	zang	tsaŋ
die	tjɛ	li	li	rou	ʐo͡ʊ	zao	tsa͡ʊ
ding	tiŋ	lia	lja	ru	ʐu	ze	tsɤ
diu	tjo͡ʊ	lian	ljɛn	ruan	ʐwan	zei	tse͡ɪ
dong	tuŋ	liang	ljaŋ	rui	ʐwe͡ɪ	zen	tsən
dou	to͡ʊ	liao	lja͡ʊ	run	ʐwən	zeng	tsəŋ
du	tu	lie	ljɛ	ruo	ʐwo	zha	tsʂa
duan	twan	lin	lin	sa	sa	zhai	tsʂa͡ɪ
dui	twe͡ɪ	ling	liŋ	sai	sa͡ɪ	zhan	tsʂan
dun	twən	liu	ljo͡ʊ	san	san	zhang	tsʂaŋ
duo	two	lo	lo	sang	saŋ	zhao	tsʂa͡ʊ
e	ɤ	long	luŋ	sao	sa͡ʊ	zhe	tsʂɤ
ei	e͡ɪ	lou	lo͡ʊ	se	sɤ	zhei	tsʂe͡ɪ
en	ən	lu	lu	sen	sən	zhen	tsʂən
eng	əŋ	luan	lwan	seng	səŋ	zheng	tsʂəŋ
er	ɚ	lve	lɥɛ	sha	ʂa	zhi	tsʅ

fa	fa	lun	lwən	shai	ʂa͡ɪ	zhong	tʂʊŋ
fan	fan	luo	lwo	shan	ʂan	zhou	tʂo͡ʊ
fang	faŋ	lv	ly	shang	ʂaŋ	zhu	tʂu
fei	fe͡ɪ	ma	ma	shao	ʂa͡ʊ	zhua	tʂwa
fen	fən	mai	ma͡ɪ	she	ʂɤ	zhuai	tʂwa͡ɪ
feng	fəŋ	man	man	shei	ʂe͡ɪ	zhuan	tʂwan
fo	fwo	mang	maŋ	shen	ʂən	zhuang	tʂwaŋ
fou	fo͡ʊ	mao	ma͡ʊ	sheng	ʂəŋ	zhui	tʂwe͡ɪ
fu	fu	me	mə	shi	ʂɻ	zhun	tʂwən
ga	ka	mei	me͡ɪ	shou	ʂo͡ʊ	zhuo	tʂwo
gai	ka͡ɪ	men	mən	shu	ʂu	zi	tsɻ
gan	kan	meng	məŋ	shua	ʂwa	zong	tsuŋ
gang	kaŋ	mi	mi	shuai	ʂwa͡ɪ	zou	tso͡ʊ
gao	ka͡ʊ	mian	mjɛn	shuan	ʂwan	zu	tsu
ge	kɤ	miao	mja͡ʊ	shuang	ʂwaŋ	zuan	tswan
gei	ke͡ɪ	mie	mjɛ	shui	ʂwe͡ɪ	zui	tswe͡ɪ
gen	kən	min	min	shun	ʂwən	zun	tswən
geng	kəŋ	ming	miŋ	shuo	ʂwo	zuo	tswo
gong	kuŋ	miu	mjo͡ʊ	si	sɻ		

Appendix H. Tone3 sandhi rules in Phon in Chapter 5.

1) σ^{214} σ^{214}/ ___]group

Real language: “Tone 214 remains 214 at the end of a word group.”

2) σ^{214} σ^{35}/ ___ $\sigma^{214}\sigma_0$]word

Real language: “Tone 214 becomes 35 whenever it is followed by 214 within the word, but not at the end of a group, as per Rule 1.”

3) σ^{214} σ^{35}/ ___]word [σ^{214}]word

Real language: “Tone 214 becomes 35 whenever it is followed by 214 across word boundaries”

4) σ^{214} σ^{21}/ elsewhere

Real language: “Tone 214 becomes 21 across all positions except when group final, as per Rule 1, and after Rules 2 and 3 have applied.”

Rule application examples:

Input	σ^{214}][$\sigma^{214}\sigma^{214}$	$\sigma^{214}\sigma^{214}$][σ^{214}
Rule 1	σ^{214}][$\sigma^{214}\sigma^{214}$	$\sigma^{214}\sigma^{214}$][σ^{214}
Rule 2	σ^{214}][$\sigma^{35}\sigma^{214}$	$\sigma^{35}\sigma^{214}$][σ^{214}
Rule 3	σ^{214}][$\sigma^{35}\sigma^{214}$	$\sigma^{35}\sigma^{35}$][σ^{214}
Rule 4	σ^{21}][$\sigma^{35}\sigma^{214}$	$\sigma^{35}\sigma^{35}$][σ^{214}
Output	σ^{21}][$\sigma^{35}\sigma^{214}$	$\sigma^{35}\sigma^{35}$][σ^{214}
Real word example	买水果 [maɪ^{21} ʂweɪ^{35}kwo^{214}]	水果好 [ʂweɪ^{35}kwo$^{\mathbf{35}}$ xaʊ^{214}]

Appendix I. The 32 citation tone targets (monosyllabic words) in Chapter 6 (Paper 3) (sorted from the most spontaneous to the least spontaneous within each tone category)

Tone	Word	Gloss	Pinyin	IPA	Spontaneity
Tone1	三	Three	san1	/san^{55}/	98%
Tone1	八	Eight	ba1	/pa^{55}/	98%
Tone1	书	Book	shu1	/ʂu^{55}/	81%
Tone1	吃	Eat	chi1	/tʂʰʅ^{55}/	79%
Tone1	车	Car	che1	/tʂʰɤ^{55}/	79%
Tone1	山	Mountain	shan1	/ʂan^{55}/	69%
Tone1	灯	Light	deng1	/təŋ^{55}/	53%
Tone1	虾	Shrimp	xia1	/ɕja^{55}/	47%
Tone2	蓝	Blue	lan2	/lan^{35}/	93%
Tone2	鱼	Fish	yu2	/y^{35}/	85%
Tone2	球	Ball	qiu2	/tɕʰjo͡ʊ^{35}/	74%
Tone2	糖	Candy	tang2	/tʰɑŋ^{35}/	72%
Tone2	门	Door	men2	/mən^{35}/	65%
Tone2	圆	Circle	yuan2	/ɥɛn^{35}/	64%
Tone2	床	Bed	chuang2	/tʂʰwɑŋ^{35}/	62%
Tone3	五	Five	wu3	/u^{214}/	99%
Tone3	手	Hand	shou3	/ʂo͡ʊ^{214}/	95%
Tone3	水	Water	shui3	/ʂwe͡ɪ^{214}/	94%
Tone3	狗	Dog	gou3	/ko͡ʊ^{214}/	94%
Tone3	马	Horse	ma3	/ma^{214}/	90%
Tone3	紫	Purple	zi3	/tsɿ^{214}/	76%
Tone3	雨	Rain	yu3	/y^{214}/	73%
Tone3	脚	Foot	jiao3	/tɕjɑo͡ʊ^{214}/	65%

Tone3	碗	Bowl	wan3	/wan^{214}/	49%
Tone4	二	Two	er4	/ɐ^{51}/	98%
Tone4	绿	Green	lv4 (lü4)	/ly^{51}/	96%
Tone4	饭	Rice	fan4	/fan^{51}/	72%
Tone4	肉	Meat	rou4	/ʐo͡ʊ^{51}/	71%
Tone4	饿	Hungry	e4	/ɤ^{51}/	68%
Tone4	热	Hot	re4	/ʐɤ^{51}/	68%
Tone4	站	Stand	zhan4	/tʂan^{51}/	62%
Tone4	菜	Vegetable	cai4	/tsʰa͡ɪ^{51}/	57%

Appendix J. Mean durations of tone productions by subgroups* and MATLAB script to convert equal normalized durations to differentiated normalized durations in Chapter 6 (Paper 3).

Target	Transcription	Grade	Group	Mean (ms)	d**
1	1	0	0	292.7866	0.6762
1	1	1	1	304.098	0.7023
1	1	1	2	337.8952	0.7804
1	1	3	1	233.2812	0.5388
1	1	3	2	307.9042	0.7111
1	1	5	1	256.2571	0.5918
1	1	5	2	299.964	0.6928
1	2	1	1	218.8064	0.5053
1	2	1	2	331.6113	0.7658
1	2	3	1	275.5768	0.6364
1	2	3	2	341.3642	0.7884
1	2	5	1	261.8408	0.6047
1	2	5	2	257.0812	0.5937
1	3	1	2	362.8879	0.8381
1	3	3	2	355.678	0.8214
1	3	5	2	315.6543	0.7290
1	4	0	0	150.4315	0.3474
1	4	1	2	305.0115	0.7044
1	4	3	2	151.3078	0.3494
1	4	5	1	127.0545	0.2934
1	4	5	2	150.3805	0.3473
1	7	1	1	294.9072	0.6811
1	7	1	2	424.9883	0.9815
1	7	3	2	377.3237	0.8714
1	7	5	1	218.6863	0.5050
1	7	5	2	258.7265	0.5975

2	1	1	1	224.9419	0.5195
2	1	3	2	355.5045	0.8210
2	1	5	2	383.1165	0.8848
2	2	0	0	291.0445	0.6722
2	2	1	1	325.6172	0.7520
2	2	1	2	411.9192	0.9513
2	2	3	1	269.6184	0.6227
2	2	3	2	392.6003	0.9067
2	2	5	1	285.074	0.6584
2	2	5	2	324.1729	0.7487
2	3	1	1	356.1816	0.8226
2	3	1	2	372.7262	0.8608
2	3	3	1	279.3431	0.6451
2	3	3	2	394.8265	0.9118
2	3	5	1	331.1964	0.7649
2	3	5	2	352.8528	0.8149
2	4	1	2	295.3422	0.6821
2	6	1	1	135.9296	0.3139
2	6	1	2	346.3776	0.7999
2	6	5	1	101.7302	0.2349
2	6	5	2	99.1276	0.2289
2	7	0	0	183.5777	0.4240
2	7	1	1	344.5162	0.7956
2	7	3	1	214.3274	0.4950
2	7	3	2	288.8446	0.6671
2	7	5	2	227.1053	0.5245
3	1	1	1	266.542	0.6156
3	1	1	2	323.7469	0.7477
3	1	3	2	227.0788	0.5244

3	1	5	1	261.0106	0.6028
3	1	5	2	257.0509	0.5937
3	2	1	1	186.5576	0.4308
3	2	1	2	362.9013	0.8381
3	2	3	1	275.9791	0.6374
3	2	3	2	353.5356	0.8165
3	2	5	1	329.9448	0.7620
3	2	5	2	294.6177	0.6804
3	3	0	0	329.9243	0.7619
3	3	1	1	357.187	0.8249
3	3	1	2	380.215	0.8781
3	3	3	1	304.0069	0.7021
3	3	3	2	375.0078	0.8661
3	3	5	1	305.5773	0.7057
3	3	5	2	316.3496	0.7306
3	4	1	1	266.6402	0.6158
3	4	1	2	281.1253	0.6493
3	4	3	1	229.1353	0.5292
3	4	3	2	206.2982	0.4764
3	4	5	1	210.0216	0.4850
3	4	5	2	234.0256	0.5405
3	6	0	0	249.7958	0.5769
3	6	1	1	233.9535	0.5403
3	6	1	2	236.9061	0.5471
3	6	3	1	221.4159	0.5114
3	6	3	2	296.6629	0.6851
3	6	5	1	172.1528	0.3976
3	6	5	2	171.377	0.3958
3	7	1	1	263.3831	0.6083

3	7	1	2	336.7051	0.7776
3	7	3	1	322.9365	0.7458
3	7	3	2	369.1289	0.8525
3	7	5	1	207.8079	0.4799
3	7	5	2	297.2461	0.6865
4	1	1	1	242.6749	0.5605
4	1	1	2	262.616	0.6065
4	1	3	1	203.2396	0.4694
4	1	3	2	351.9744	0.8129
4	1	5	1	171.0575	0.3951
4	1	5	2	288.24	0.6657
4	2	1	2	344.1634	0.7948
4	2	3	1	302.0503	0.6976
4	2	3	2	433.9326	1.0022
4	2	5	1	359.5275	0.8303
4	2	5	2	334.1975	0.7718
4	3	1	1	284.2208	0.6564
4	3	1	2	426.9365	0.9860
4	3	3	2	397.3319	0.9176
4	3	5	2	230.9049	0.5333
4	4	0	0	192.9268	0.4456
4	4	1	1	241.2322	0.5571
4	4	1	2	237.7363	0.5490
4	4	3	1	193.2204	0.4462
4	4	3	2	232.556	0.5371
4	4	5	1	209.638	0.4842
4	4	5	2	229.8224	0.5308
4	7	0	0	118.616	0.2739
4	7	1	1	264.2917	0.6104

4	7	1	2	312.1487	0.7209
4	7	3	1	281.6596	0.6505
4	7	3	2	381.5296	0.8811
4	7	5	1	357.383	0.8254
4	7	5	2	241.813	0.5585

* Subgroups are defined by target (1 = Tone1, 2 = Tone2, 3 = Tone3, 4 = Tone4), transcription (1 = Tone1, 2 = Tone2, 3 = Tone3, 4 = Tone4, 6 = semi-Tone3, 7 = uncategorizable), grade (0 = Teacher, 1 = Grade1, 3 = Grade3, 5 = Grade5), group = (0 = Teacher, 1 = heritage speakers of Mandarin, 2 = L2 learners of Mandarin).

* d = Subgroup_Mean / $\text{Subgroup\_Mean}_{max}$

The MATLAB script to convert time points based on subgroup means:

```
# read in the original data which contains the same grouping factors
table = readmatrix("Normalized.xlsx");
table(end+1,:) = zeros(1,10);
table(:,11) = zeros(1,23001);

# read in a “dictionary” that contains the list of mean durations as presented above
dic_data = readmatrix("Duration.xlsx", "Sheet", "Dictionary");
dic_rows = size(dic_data, 1);
dic = dictionary;

# define the dictionary, the input is the four grouping factors, the output is d, the ratio
# between the mean duration of the group and the longest mean duration (434 ms)
for r = 1:dic_rows
   dic(mat2str(dic_data(r, 1:4))) = dic_data(r, 6);
end

# multiply the original 0–9 time values by d
for i = 1:23000
   k = mat2str(table(i,[1,2,5,6]));
   d = dic(k);
   table(i,11) = table(i,9) * d;
end
```

Appendix K. Pseudo-codes of statistics in Chapter 6 (Paper 3).

K.1. First and last 6 rows of the dataset for tone match and *f0* contour

	target	actual	correct	word	grade	group	speaker	f0	time	traj	normTime
1	1	1	1	5	3	1	2	4.006189	0	1	0.0000000
2	1	1	1	5	3	1	2	3.827686	1	1	0.5387558
3	1	1	1	5	3	1	2	3.635522	2	1	1.0775115
4	1	1	1	5	3	1	2	3.609722	3	1	1.6162673
5	1	1	1	5	3	1	2	3.626841	4	1	2.1550231
6	1	1	1	5	3	1	2	3.646354	5	1	2.6937788
…											
22995	4	4	1	30	0	0	212	3.480345	4	2300	1.782234
22996	4	4	1	30	0	0	212	3.289683	5	2300	2.227792
22997	4	4	1	30	0	0	212	3.079486	6	2300	2.673350
22998	4	4	1	30	0	0	212	2.868184	7	2300	3.118909
22999	4	4	1	30	0	0	212	2.634388	8	2300	3.564467
23000	4	4	1	30	0	0	212	2.416117	9	2300	4.010026

K.2. R codes for descriptive statistics of tone match (accuracy)

```
# obtain aggregated data from the correctness of each token production
accuracy <- aggregate(data$correct, by = data[c("target","group","grade")], FUN = mean)
names(accuracy)[names(accuracy)=="x"] <- "mean"
N <- aggregate(data$correct,by = data[c("target","group","grade")], FUN = length)
names(N)[names(N)=="x"] <- "N"
accuracy <- merge(accuracy,N)
SD <- aggregate(data$correct,by = data[c("target","group","grade")], FUN = sd)
names(SD)[names(SD)=="x"] <- "sd"
accuracy <- merge(accuracy,SD)
accuracy$se <- accuracy$sd/sqrt(accuracy$N)
# code “target” “group” and “grade” as factors
accuracy$target <- factor(accuracy$target, labels=c("Tone1","Tone2","Tone3","Tone4"))
accuracy$group <- factor(accuracy$group, labels=c("HL","L2"))
accuracy$group <- factor(accuracy$grade, labels=c("Grade1","Grade3","Grade5"))

# set up the plotting layout
layout(matrix(c(1,2,3,4), nrow = 1, ncol = 4))
par(family="serif",oma=c(1,1.2,0.5,0),mar=c(2,2,2,0.1),mgp=c(1.8,0.8,0),font=6,font.axis=6,font.lab=6,fon
t.main=6,cex=0.65)
mtext("Grade", side=1,outer=T,cex=1)
mtext("Match", side=2, outer=T,cex=1)
# plot Tone1 productions in the HL group by grade levels and add error bars
```

```
plot(mean~grade,data=accuracy[accuracy$target=="Tone1" & accuracy$group=="HL",],
        bty="l",ylim=c(0,1),type="b",pch=1,lwd=1.5,col="#d62728",
        ylab="",main="Tone1 [55]",xpd=FALSE,cex=1.5)
segments(...grade,...mean-...se*1.96, …grade,…mean+...se*1.96)
arrows(..grade,...mean-...se*1.96, …grade,…mean+...se*1.96,
        angle = 90, code = 3, length = 0.02)
# add L2 group and teachers' data
lines(... add L2 data in col="#1f77b4")
lines(... add Teacher data in col="#aaaaaa")
# add a legend
legend(1,0.2,c("HL", "L2", "Teacher"),
        col=c('#d62728',"#1f77b4","#aaaaaa"),
        pch=c(1,2,0),lwd=1.5,bty="n",xjust=0,yjust=1,y.intersp=2, cex=1)
```

K.3. R codes for the logistic mixed model of tone match rates

```
# prepare factors
data$target <- factor(data$target,labels=c("Tone1","Tone2","Tone3","Tone4"))
data$group <- factor(data$group,labels=c("HL","L2"))
data$grade <- factor(data$grade,labels=c("Grade1","Grade3","Grade5"))
data$speaker <- as.character(data$speaker)

# build logistic mixed models, starting from a null model
m0.null <- glmer(correct~0+(1|speaker)+(1|word),data=data,family=binomial)
summary(m0.null)
m0.1 <- glmer(correct~0+target+(1|speaker)+(1|word),data=data,family=binomial)
summary(m0.1)
anova(m0.null,m0.1)
# select m0.1
…
# glmerControl() was used to deal with difficulty in convergence
m0.5 <- glmer(correct~0+target+group+grade+target:group+target:grade+
        (1|speaker)+(1|word),
        data=data,family=binomial,control = glmerControl(optimizer = "bobyqa"))
summary(m0.5)
anova(m0.4,m0.5)
# select m0.5
m0.7 <- glmer(correct~0+target*group*grade+(1|speaker)+(1|word),
```

```
        data=data,family=binomial,control = glmerControl(optimizer = "bobyqa"))
summary(m0.7)
anova(m0.5,m0.7)
# select m0.5

# test whether both random effects were necessary
# removed word random effect
m0.5.1 <- glmer(correct~0+target+group+grade+target:group+group:grade+(1|speaker),
        data=data,family=binomial,control = glmerControl(optimizer = "bobyqa"))
anova(m0.5,m0.5.1)
# word random effect is useful
# removed speaker random effect
m0.5.2 <- glmer(correct~0+target+group+grade+target:group+group:grade+(1|word),
        data=data,family=binomial,control = glmerControl(optimizer = "bobyqa"))
anova(m0.5,m0.5.2)
# speaker random effect is useful

# the finally selected model "m"
m <- glmer(correct~0+target+group+grade+target:group+target:grade+
        (1|speaker)+(1|word),
        data=data,family=binomial,
        control = glmerControl(optimizer = "bobyqa"))
summary(m)
      AIC      BIC   logLik deviance df.resid
      1875.0   1913.9   -930.5   1861.0     1916

Scaled residuals:
    Min      1Q  Median      3Q     Max
-5.3046  0.1506  0.3281  0.5283  2.0537

Random effects:
 Groups  Name        Variance Std.Dev.
 speaker (Intercept) 0.6506   0.8066
 word    (Intercept) 0.7281   0.8533
Number of obs: 1923, groups:  speaker, 82; word, 32

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
target        0.81133    0.14351   5.653 1.57e-08 ***
group         0.08522    0.30440   0.280 0.779496
grade         0.25524    0.10819   2.359 0.018316 *
target:group -0.36696    0.10523  -3.487 0.000488 ***
target:grade -0.04645    0.03393  -1.369 0.171063
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Correlation of Fixed Effects:
            target group  grade  trgt:grp
group       -0.188
grade       -0.111 -0.478
target:grop -0.352 -0.739  0.402
target:grad -0.149  0.431 -0.791 -0.395
```

```
# test collinearity
library(car)
vif(m)
# results are all smaller than 10, which means the correlation R² is lower than 0.9
       target         group         grade target:group target:grade
       4.604302     7.600792     5.294686     8.075504     5.064289

# get estimated means and contrasts using emmeans()
group <- emmeans(m,specs=pairwise~group)
group_mean <- summary(group)$emmeans
# the returned values should be converted from log-odds to probabilities
group_mean$emmean <- plogis(group_mean$emmean)
group_mean$asymp.LCL <- plogis(group_mean$asymp.LCL)
group_mean$asymp.UCL <- plogis(group_mean$asymp.UCL)
group_contrast <-summary(group)$contrasts
groupDiff <-
        group_mean[group_mean$group=="HL",]$emmean–
        group_mean[group_mean$group=="L2",]$emmean
# get estimated means and contrasts for the other effects
…

# plot the group effect
layout(matrix(c(1), nrow = 1, ncol = 1))
par(family="serif",oma=c(1.5,1.5,1.5,0),mar=c(2,2,1,0.1),mgp=c(1.8,0.8,0),
        font=6,font.axis=6,font.lab=6,font.main=6,cex=0.65)
group_plot <- barplot(height=group_mean$emmean, names=group_mean$group,
        col=c("#d62728","#1f77b4","#aaaaaa"),
        bty="l",ylim=c(0,1),main="")
segments(group_plot,group_mean$asymp.LCL,group_plot,group_mean$asymp.UCL)
arrows(group_plot,group_mean$asymp.LCL,group_plot,group_mean$asymp.UCL,
        angle = 90, code = 3, length = 0.02)
mtext("Accuracy", side=2, outer=T,cex=1)
mtext("Group", side=3, outer=T,cex=1)
```

```
# plot the other effects
…
```

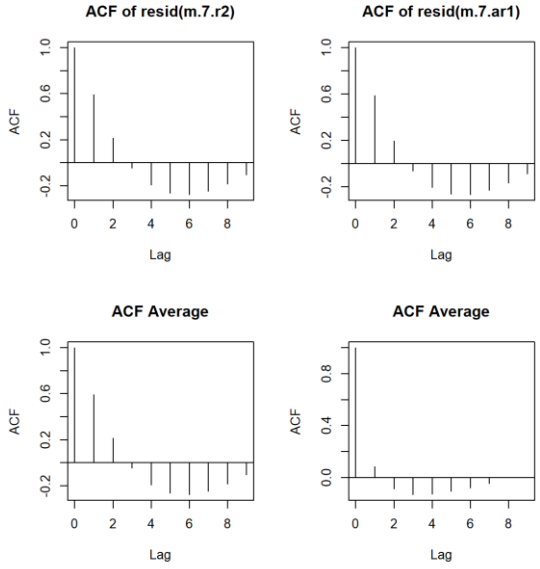
K.4. R codes for the generalized additive mixed model of *f0* contours

```
# import libraries
library(mgcv)
library(itsadug)
source("gamm_hacks.r") # retrieved from https://github.com/soskuthy/gamm_intro

# prepare factors
data$word <- factor(data$word)
data$speaker <- factor(data$speaker)
data$tone <- factor(data$actual, labels=c("Tone1","Tone2","Tone3","Tone4"))
data$grade.order <- ordered(data$grade,
        labels=c("Grade1","Grade3","Grade5"))
contrasts(data$grade.order) <- "contr.treatment"
data$group <- factor(data$group, labels=c("HL","L2"))
# interaction effects were processed using contrastive coding
# to provide a handle to retrieve each group's contour easily
data$tonegrade <- factor(interaction(data$tone, data$grade.order))
data$tonegroup <- factor(interaction(data$tone, data$group))
data$gradegroup <- factor(interaction(data$grade.order, data$group))
data$tonegradegroup <- factor(interaction(data$tone, data$grade.order, data$group))
data$start.event <- data$normTime == 0

# build GAMMs from the simplest model and compare models using compareML()
m.1 <- bam(f0~group+s(normTime,by=group),data=data)
summary(m.1)
# build a model with group and grade main effects
m.2 <- bam(f0 ~ group+grade.order+
        s(normTime,by=group)+s(normTime,by=grade.order),data = data)
summary(m.2)
compareML(m.1, m.2)
# m.2 was selected
…
# the selected model was with three-way interaction and random effects
m.7.r2 <- bam(f0~tonegradegroup+
```

```
                s (normTime, by = tonegradegroup)+
                s (normTime, word, bs = "re")+
                s (normTime, speaker, bs = "re"),data = data)
# an AR(1) model was build to control the effects of autocorrelation within contours
r1 <- start_value_rho(m.7.r2)
m.7.ar1 <- bam(f0~tonegradegroup+
                s (normTime, by = tonegradegroup)+
                s (normTime, word, bs = "re")+
                s (normTime, speaker, bs = "re"),
                data = data, method="fREML", rho=r1,AR.start=data$start.event)
summary(m.7.ar1)
# visualize the autocorrelation effects for the original model and the AR(1) model
par(mfrow=c(2,2))
acf_plot(resid(m.7.r2), split_by=list(data$traj))
acf_plot(resid(m.7.ar1), split_by=list(data$traj))
acf_resid(m.7.r2, split_pred=list(data$traj))
acf_resid(m.7.ar1, split_pred=list(data$traj))
```




```
# the structure in residuals within contours were partly reduced

# check concurvity
concurvity(m.7.ar1)
# concurvity was high (est. ≥ 0.8) since different subgroups shared contours
# within the same tone target
```

```
# build a separate group*grade model for each tone (e.g., Tone1)
data1 <- filter(data, tone == "Tone1")
m1<- bam(f0~gradegroup+
                s (normTime,by=gradegroup)+
                s (normTime, word, bs = "re") +
                s (normTime, speaker, bs = "re"),data = data1)
m1.ar1 <- bam(f0 ~ gradegroup+
          s(normTime,by=gradegroup)+
          s(normTime,word,bs="re")+
          s(normTime,speaker,bs="re"),data=data1,method="fREML",
          rho=start_value_rho(m1),AR.start=data1$start.event)
summary(m1.ar1)
concurvity(m1.ar1)
# concurvity was lowered in the by-tone model
# (est. < 0.8 except for HL.Grade3 and HL.Grade5, random effects < 0.25)

# teachers' raw data will be plotted as a reference
# first get teachers' aggregated data
dataT <- filter(data, group == 0)
Teacher <- dataT %>%group_by(target,time) %>%
          summarize(f0_mean = mean(f0),f0_SD = sd(f0), f0_SE = sd(f0)/sqrt(length(f0)))
# respecify the time using the mean duration of each tone target in teachers' productions
Teacher[Teacher$target==1,]$time=Teacher[Teacher$target==1,]$time*0.6762
…

# plot the f0 contours for this tone (e.g., Tone1) by group and grade using plot_smooth()
layout(matrix(c(1,2), byrow=TRUE, ncol = 2, nrow = 1))
par(family="serif",oma=c(1,1.2,1.2,0),mar=c(0.5,2,2,1),mgp=c(1.8,0.8,0),
          font=6,font.axis=6,font.lab=6,font.main=6,cex.main=1)
mtext("Normalized Time", side=1,outer=T,cex=1,at=0.5)
mtext("Normalized f0", side=2, outer=T,cex=1)
mtext("Tone1 [55]", side=3, outer=T,cex=1)
plot_smooth(m1.ar1, view="normTime", cond=list('gradegroup'="Grade1.HL"),
          lwd=2, bty="l",xlim=c(0,6.3), ylim=c(1,5),rug=F,
          xaxt="n",xlab="",ylab="",main="HL",hide.label=TRUE,
          col='#d62728', lty=3, print.summary=F)
# plot each subgroup by using "add=T"
```

```
...
# add teachers' raw data as a line chart
lines(f0_mean~time,data=Teacher[Teacher$target==1,],
        lwd=2,lty=1,pch=1,col="#aaaaaa",ylab="",xpd=FALSE,cex=0.5)
# add 95% confidence intervals
segments(...time,…f0_mean-...f0_SE*1.96,time,f0_mean+SE*1.96,col="#aaaaaa")
arrows(time,f0_mean-f0_SE*1.96,time,f0_mean+SE*1.96,
        angle = 90, code = 3, length = 0.02,col="#aaaaaa")
legend(0,3.4,c("Grade1.HL", "Grade3.HL", "Grade5.HL","Teacher"),
    col=c('#d62728','#d62728','#d62728','#aaaaaa'),
    lty=c(3,2,4,1), lwd=2,bty="n",xjust=0,yjust=1,y.intersp=1.5, cex=0.9)
# plot the same for the L2 group
# plot the same for the other tone targets
```

K.5. The first and last 6 rows of the dataset of duration

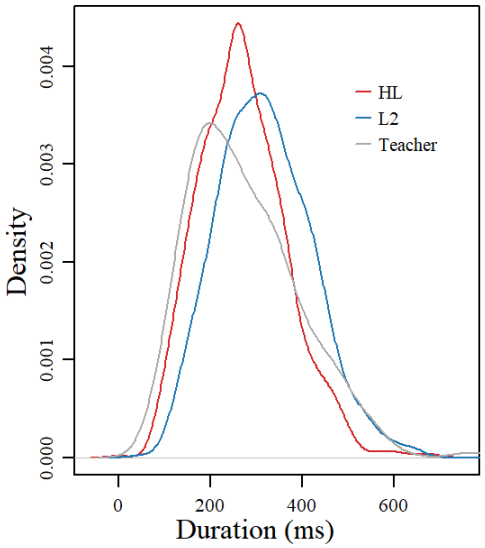
```
      target actual word grade group speaker duration
   1       1      1  che     1     2      11 324.4814
   2       1      1   ba     1     2      11 233.6745
   3       4      1   er     1     2      11 340.7763
   4       1      1  che     5     2      12 184.2635
   5       1      1 shan     5     2      12 209.3065
   6       1      1  san     1     2      15 219.9783
                                          …
1919       3      7 jiao     5     1     149 136.36609
1920       3      7 shou     5     1     149 107.95762
1921       3      7  gou     5     1     149  91.96053
1922       3      7   zi     5     1     156 192.66330
1923       3      7 shou     5     1     156 127.40759
1924       2      7  qiu     3     1     169 214.32742
```

K.6. R codes for the linear mixed model of duration

```
# first, plot the raw data of duration
layout(matrix(c(1,2,3,4), nrow = 1, ncol = 4))
par(family="serif",oma=c(1,1.2,2,0),mar=c(2,2,2,0.1),mgp=c(1.8,0.8,0),
        font=6,font.axis=6,font.lab=6,font.main=6,cex=0.65)
mtext("Raw Data of Duration", side=3, outer=T,cex=1)
mtext("Duration (ms)", side=2, outer=T,cex=1)
mtext("Target", side=1, outer=T,cex=1)
boxplot(duration~group+target,data=data[data$grade==1,],col=c("#d62728","#1f77b4"),
        at=c(1:2,4:5,7:8,10:11),names=c("Tone1","","Tone2","","Tone3","","Tone4",""),
        xlab="",ylab="",main="Grade1",ylim=c(0,700))
```

```
# plot the same for the other subgroups

# check data distribution
plot(density(data[data$group==1,]$duration),col="#d62728",main="",ylab="")
lines(density(data[data$group==2,]$duration),col="#1f77b4")
lines(density(data[data$group==0,]$duration),col="#aaaaaa")
legend(600,0.003,c("HL", "L2", "Teacher"),
      col=c('#d62728',"#1f77b4","#aaaaaa"), lty=1,
      bty="n",xjust=0,yjust=0,y.intersp=1.8, cex=1)
mtext("Density", side=2, outer=T,cex=1)
mtext("Duration (ms)", side=1, outer=T,cex=1)
```




```
# build a linear mixed model, starting with a null model
data$target <- factor(data$target,labels=c("Tone1","Tone2","Tone3","Tone4"))
data$group <- factor(data$group,labels=c("HL","L2"))
data$grade <- factor(data$grade,labels=c("Grade1","Grade3","Grade5"))
data$speaker <- as.character(data$speaker)

m0.null <- lmer(duration~0+(1|speaker)+(1|word),data=data,REML=FALSE)
m0.1 <- lmer(duration~0+target+(1|speaker)+(1|word),data=data,REML=FALSE)
anova(m0.null,m0.1)
# select m0.1
…
m0.5 <- lmer(duration~0+target+group+grade+target:grade+
```

```
        (1|speaker)+(1|word),data=data,REML=FALSE)
anova(m0.3,m0.5)
# select m0.5
m0.6 <- lmer(duration~0+target+group+grade+target:grade+group:grade+
        (1|speaker)+(1|word),data=data,REML=FALSE)
anova(m0.5,m0.6)
# select m0.5

# fit m0.5 with REML to compare random effects
m0.5 <- lmer(duration~0+target+group+grade+target:grade+
        (1|speaker)+(1|word),data=data,REML=TRUE)
m0.5.1 <- lmer(duration~0+target+group+grade+target:grade+
        (1|speaker),data=data,REML=TRUE)
anova(m0.5,m0.5.1)
m0.5.2 <- lmer(duration~0+target+group+grade+target:grade+
        (1|word),data=data,REML=TRUE)
anova(m0.5,m0.5.2)
# both word and speaker effects were necessary


# the selected model was m
m <- lmer(duration~0+target+group+grade+target:grade+
        (1|speaker)+(1|word),data=data,REML=TRUE)
summary(m)
REML criterion at convergence: 22343

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.0881 -0.5966 -0.0182  0.5725  7.3677

Random effects:
 Groups   Name        Variance Std.Dev.
 speaker  (Intercept)  2052     45.30
 word     (Intercept) 13023    114.12
 Residual              5475     73.99
Number of obs: 1924, groups:  speaker, 93; word, 33

Fixed effects:
             Estimate Std. Error        df t value Pr(>|t|)
target         70.5426     9.1063   58.3284   7.747 1.58e-10 ***
group          61.5328     9.5218  124.6719   6.462 2.10e-09 ***
grade          -9.5639     3.8940  215.7595  -2.456   0.0148 *
target:grade    0.7060     0.9267 1851.8744   0.762   0.4463
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Correlation of Fixed Effects:
            target group  grade
group       -0.487
grade       -0.175 -0.100
target:grad -0.081  0.042 -0.604
```

test collinearity

vif(m)

```
  target         group         grade target:grade
1.533307      1.417235      1.849346      1.720546
```

there's no obvious collinearity

normal distribution of residuals

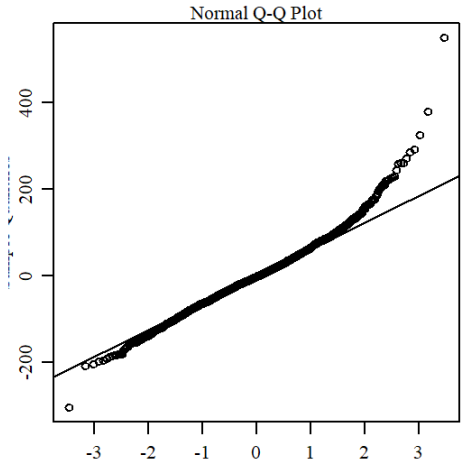
qqnorm(residuals(m))

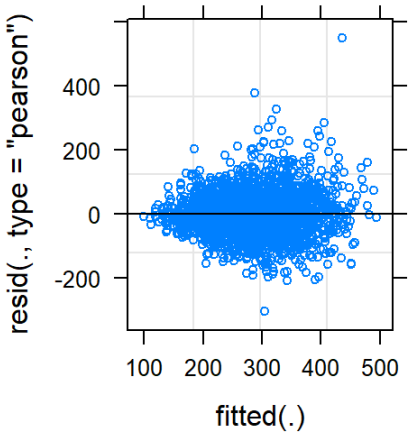
qqline(residuals(m))

slightly more variation on the higher-duration tail

residuals are independent across the fitted line

plot(m)





use emmeans() to get estimated subgroup means and contrasts (e.g., targets)

target <- emmeans(m,specs=pairwise~target)

target_mean <- summary(target)$emmeans

target_contrast <-summary(target)$contrasts

plot fixed effects (e.g., targets)

layout(matrix(c(1), nrow = 1, ncol = 1))

par(family="serif",oma=c(1.5,1.5,1.5,0),mar=c(2,2,1,0.1),mgp=c(1.8,0.8,0),
font=6,font.axis=6,font.lab=6,font.main=6,cex=0.65)

target_plot <- barplot(target_mean$emmean, names=target_mean$target,
col="#f88e36",bty="l",ylim=c(0,400),main="")

```
segments(target_plot,target_mean$lower.CL,target_plot,target_mean$upper.CL)
arrows(target_plot,target_mean$lower.CL,target_plot,target_mean$upper.CL,
      angle = 90, code = 3, length = 0.02)
mtext("Duration (ms)", side=2, outer=T,cex=1)
mtext("Target Effect", side=3, outer=T,cex=1)
```