THE PERCEPTION AND PRODUCTION DEVELOPMENT OF CANTONESE SYLLABLE-FINAL SEGMENTS BY MANDARIN SPEAKERS

ZHANG CHANGWEI

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

The Perception and Production Development of Cantonese Syllable-final Segments by Mandarin Speakers

ZHANG Changwei

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

July 2019

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

____________________ (Signed)

ZHANG Changwei (Name of student)

Abstract

Previous research on the second language (L2) phonetic learning stresses the influence of learners' native language (L1) experience and the modulation of perception-production interface. In this dissertation, Mandarin speakers' learning of Cantonese syllable-final segments ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅]) is taken as a starting point to disentangle the research questions centering around these two issues.

Under the influence of L1 experience, L2 learners are apt to link L2 speech sounds to those previously existing L1 categories when confronting an L2. Several predominant L2 learning models including the Speech Learning Model (Flege, 1995), the Second Language Linguistic Perception Model (Escudero, 2005), the Perceptual Assimilation Model to Second Language (Best & Tyler, 2007) and the revised Speech Learning Model (Flege & Bohn, 2021) consistently prioritize the important role of cross-L1-L2 similarity in predicting the learning outcome and learning difficulty of L2 speech learning. However, the definitions and measurement methods of the cross-L1-L2 similarity are not always in agreement, which could lead to a distinct interpretation of these theories.

Researchers come up with divided opinions about how perception would interact with production during the development of L2 phonetic learning. To date, the production-precedence pattern (Goto, 1971; Sheldon & Strange, 1982; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Hattori & Iverson, 2009; Ingvalson, McClelland, & Holt, 2011), the perception-precedence pattern

(Flege, 1995), and the perception-production co-evolve pattern (Flege & Bohn, 2021) have been proposed. The relationship between perception and production remains debatable and it could lead to different decisions on training methods of L2 speech learning. However, the above studies seldom specify whether perception would interact diversely with production according to different targets. As remarked by Cheng & Zhang (2009), the perception-production relationship varies depending on target segments. It could be possible that some L2 targets achieve correct perception before successful pronunciation, while some L2 targets might take the alternative path to develop production before perception or to co-evolve perception and production together during the developmental procedure. If so, intervention should be appropriately administered to better facilitate L2 learners' development of different targets.

The research questions were fourfold: **(1)** to measure the cross-language perceptual similarity of syllable-final segments between Mandarin and Cantonese by Mandarin speakers; **(2)** to interpret the development of Cantonese syllable-final segments by Mandarin speakers in the framework of SLM-r. Then, a general discussion would supplement the SLM-r predictions of the current study with alternative PAM-L2 and L2LP hypotheses and advance heuristic amendments to the application of aforementioned L2 models; **(3)** to delineate the perception-production interface during the learning procedure of targets; and **(4)** to testify the effectiveness of perception-only training on promoting Mandarin speakers' learning of targets.

Four experiments were incorporated to disentangle these research questions. As revealed by the results of the Perceptual Assimilation Test (Experiment 1) in **Research Question (1)**, Cantonese [-∅] represented a reasonably good exemplar to Mandarin [-∅]. Moderately fit exemplars were represented by Cantonese [-p̚], [-t̚] and [-k̚] to Mandarin [-∅], Cantonese [-m] to Mandarin [-ŋ], Cantonese [-n] to Mandarin [-n] and Cantonese [-ŋ] to Mandarin [-ŋ]. Accordingly, Cantonese [-∅] was classified as an identical category with the least learning difficulties and the other six targets were categorized as less learnable similar sounds to their Mandarin counterparts in SLM-r.

These SLM-r hypotheses were attested in **Research Question (2)** with the developmental patterns of Cantonese syllable-final segments by an experimental group of Mandarin speakers through the pre-test (Experiment 2), seven sessions of perception training (Experiment 3), and the post-test (Experiment 4). In line with the predictions, identical target [-∅] generally outperformed similar targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] in the pre-test. During the intervention, learnable identical target ([-∅]) was developed earlier and with better performances than those more difficult similar sounds ([-m], [-n], [-ŋ], [-p̚], [-t̚] and [-k̚]) by Mandarin speakers.

The SLM-r hypotheses, however, failed to clarify why Cantonese [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] were all classified as similar sounds but targets [-t̚] and [-k̚] were more difficult than other similar targets for Mandarin speakers. As both Cantonese [-t̚] and [-k̚] were mapped to Mandarin [-∅] without significant differences in the similarity rating scores, Cantonese contrast [-t˺]-[-k˺] can be categorized as Single-category Assimilation in PAM-L2 (Best & Tyler, 2007) or the learning of NEW scenario in L2LP (Escudero, 2005). It is predicted that both cases are the most difficult for Mandarin speakers. Such PAM-L2 and L2LP hypotheses provide a possible account for the least satisfying performances and learning outcomes of targets [-t˺] and [-k˺] by Mandarin speakers in the present study.

Other than the influence of L1 experience on L2 phonetic learning as proposed by the abovementioned theories, it is also possible that the suspension of contrastive distribution of coronal and dorsal targets in Cantonese could pose more learning difficulty in Mandarin speakers' development of targets [-t˺] and [-k˺]. Cantonese participants' performances in Experiment 2 (the pre-test) generally support the coronal-dorsal merger and the alveolarization merging tendency as observed by previous studies (see Bauer, 1979; Yeung, 1981; Chen, 1999; Zee, 1999a & 1999b; Law, Fung, & Bauer, 2001; Wong, 2005; Ding, 2010; Bauer & Benedict, 2011; To, Cheung, & McLeod, 2013; To, McLeod, & Cheung, 2015). The relatively lower learnability of these two targets can result from the immersion of the phonetic variants of [-t˺]-[-k˺] merger in daily Cantonese by Mandarin speakers. Extant studies of theoretical models in L2 phonetic learning all target contrastively distributed L2 sounds with little attention paid to a situation where target L2 sounds might be experiencing a merger. The coronal-dorsal merging tendency of Cantonese syllable-final segments observed in this study opens up a chance to unveil this issue.

Therefore, heuristic amendments to the application of theoretical models in L2 phonetic learning are advanced. It is suggested that a more comprehensive prediction of L2 phonetic learning should integrate the perceived cross-L1-L2 similarity with the distribution patterns of L2 targets.

As concerned by **Research Question (3)**, diverse patterns of perception-production interface as a function of the learnability of different targets in the L2 phonetic learning is pointed out. In the case of less learnable similar targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ], the perception was positively correlated with production and performed a perception-precedence pattern as proposed by Flege (1995). The successful perception of these similar targets was developed before their production. The case of the identical target [-∅] failed to provide overt support to any of the production-precedence pattern (Goto, 1971; Sheldon & Strange, 1982; McCandliss et al., 2002; Hattori & Iverson, 2009; Ingvalson et al., 2011), the perception-precedence pattern (Flege, 1995), and the perception-production co-evolve pattern (Flege & Bohn, 2021). The perception and production of this target were completed with a ceiling accuracy score when the experiments were initiated and showed no correlation during the developmental procedure.

Regarding **Research Question (4)**, perception training was empirically testified to be effective in promoting Mandarin speakers' learning of targets. The perception and production of Cantonese syllable-final segments by Mandarin speakers performed a gradual improvement during the intervention (Experiment 3). This trained group then outperformed the untrained group in Experiment 4 (the post-test).

Collectively, the significant findings of the current study are as follows. Empirically, this dissertation provides developmental data about Mandarin speakers' perception and production learning of seven Cantonese syllable-final segments which are undergoing a coronal-dorsal merging tendency ([-t̚]-[-k̚] and [-n]-[-ŋ]). Theoretically, the notion of cross-L1-L2 similarity in SLM-r and its predictions deduced by the perceptual similarity are generally verified in the current study. Different tenets employed by PAM-L2 and L2LP provide alternative accounts of the observed results beyond the SLM-r predictions. Furthermore, the effects of the distribution patterns of L2 targets are heuristically proposed to supplement the decisive role of cross-L1-L2 similarity in predicting L2 speech learning by the aforementioned L2 learning models. As for the relationship between L2 perception and production, diverse patterns of perception-production interface varying according to the learnability of different targets is advanced to replace one-size-fits-all patterns proposed by previous studies.

Acknowledgments

At the time when this thesis is completed, I would like to give my sincere thanks to those who have kindly offered help to me during my Ph.D. journey. Firstly, I would like to express my appreciation to my supervisors Dr. LEUNG Man-tak and Dr. CHEN Si. During the examination procedure, Prof. CHEN Ying, Prof. MOK Pik Ki, and Dr. YAO Yao gave importantly helpful comments on the theoretical application of this study.

I also have deep gratitude to those who have contributed to this study. During the experiments, employed anonymous participants showed great interest and spent time completing the intervention. Meanwhile, Mr. TONG Kwok Tai and Mr. CHAN Kam Yiu provided technical support in the laboratory. LI Tsz Ying, NG Ka Sin, WU Ka Man, and other student helpers worked professionally in data collection, data input, and data rating. When such results were summed up in this thesis, Dr. CHUN Eunjin, Dr. HOU Renkui, Dr. LAU Kai Yan, Dr. LIU Hongchao, Dr. QIN Zhen, Dr. YANG Yuxiao, Dr. QIU Xueqiang, and YANG Yike gave selfless suggestions to the presentation of the manuscript. Miss. SZE King Wan, Miss CHOW Oi Han, and Miss YIP Nga Sze helped kindly with the submission procedure.

My thanks lastly go to my friends GAO Xuefeng, HUANG Xunan, HONG Yitian, LAU Yan Ping, LI Yanlin, WANG Xia, WEN Yulin, ZHANG Mengting, ZHOU Mi, and my family. Their company always comforts me and supports me to go through any problem. For others whose names are not mentioned here, I thank you all.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background of the Study

As globalization processes, cross-language contact has become increasingly popular and different language users show greater interest in learning a second language (L2) or multiple languages. In this social context, researchers have paid attention to L2 phonetic learning and contributed copious literature in this research field. However, more studies are still necessary to deepen the understanding of L2 phonetic learning by adult learners. During the developmental procedure of an L2, non-native performances would inevitably cause ambiguity and unintelligibility of communication. Therefore, it is an important goal for L2 learners to obtain native-like perception and production of novel speech sounds. However, adult L2 learners rarely achieve this goal (Fabra & Romero, 2012), as many factors impact the outcomes of L2 phonetic learning. Previous research on L2 phonetic learning stresses the influence of learners' native language (L1) experience and the modulation of perception-production interface.

Several predominant L2 learning models including the Speech Learning Model (Flege, 1995), the Second Language Linguistic Perception Model (Escudero, 2005), the Perceptual Assimilation Model to Second Language (Best & Tyler, 2007) and the revised Speech Learning Model (Flege & Bohn, 2021) consistently value the influence of L1 experience on L2 phonetic learning. It is suggested that L2 learners are apt to link L2 speech sounds to those previously existing L1 categories when confronting a second language. The cross-L1-L2 similarity between L2 sounds and corresponding closest L1 counterparts plays an important role in predicting the learning outcome and learning difficulty of L2 phonetic development. Yet, the definitions and measurement methods of the cross-L1-L2 similarity are not always in agreement in the above models, which could lead to different predictions of learning performances towards even the same L2 targets. By now, extant studies have endeavored to compare the validity of these models and to decide which one of them can offer a better account for a particular L2 phonetic learning case. Following the same line, this dissertation attempts to enrich the discussion of theoretical models in L2 speech learning. After reviewing the frameworks of extant L2 learning models and the commonly used measurement methods of cross-L1-L2 similarity in Chapter 2, the current study mainly focuses on the notion of cross-L1-L2 similarity in SLM-r and its predictions deduced by the comparison of perceptual similarity. Different tenets employed by PAM-L2 and L2LP could lead to different predictions supporting or denying the SLM-r hypotheses, which will be involved to supplement the interpretation of SLM-r. More importantly, the present study intends to provide solutions of refinement to the latent inadequacies of SLM-r, PAM-L2, and L2LP when applying these models to our case.

On the other hand, researchers come up with divided opinions about how perception would interact with production during the development of L2 phonetic learning. Goto (1971), Sheldon & Strange (1982), McCandliss, Fiez, Protopapas, Conway, & McClelland (2002), Hattori & Iverson (2009), and Ingvalson, McClelland, & Holt (2011) report that L2 learners can accurately produce learning targets even with poor perceptual ability. Other researchers come up against the above opinion and point out that perception would be developed ahead of production. Flege, Munro, & MacKay (1995) specifically claim that the production of the L2 sounds would be inaccurate without correctly perceiving targets to guide the sensor motor learning of sounds. Flege (1995), Flege et al. (1995), and Flege, Schirru, & MacKay (2003) suggest that the perception accuracy would put a ceiling on the production accuracy of L2 sounds. This opinion has received increasing attention and becomes a more dominant idea in the discussion of the L2 perception-production relationship. Following the same line, the extant literature further verifies this perception-precedence pattern with empirical evidence (Brannen, 2002; Peperkamp & Dupoux, 2003; Cheng & Zhang, 2009; Tremblay, 2011; Hao & de Jong, 2016; Sakai & Moorman, 2018; Thorin, Sadakata, Desain, & McQueen, 2018; Baese-Berk, 2019; Casillas, 2020). However, with the evidence about a strong bi-directional and co-equal connection between production and perception (see Sheldon & Strange, 1982; Perkell, Guenther et al., 2004; Perkell, Matthies et al., 2004; Peperkamp & Bouchon, 2011; Darcy & Krüger, 2012; Shultz, Francis, and Llanos, 2012; Chao, Ochoa, & Daliri, 2019), a co-equal perception-production relationship is proposed more recently. It is suggested that perception and production co-evolve in a bi-directional way without precedence in L2 phonetic learning (Flege & Bohn, 2021). L2 perception is paralleled and interfaced with L2 production as L2 learners' development processes.

The relationship between perception and production remains debatable, and it results in different intervention methods of second language learning. Perception-only training has been adopted often in some newly developed studies. McCandliss et al. (2002), Handley et al. (2009), Baese-Berk (2010 & 2019), Anderson (2011), and Shinohara & Iverson (2018) target on L2 learner's learning of segmental units, while Francis, Ciocca, Ma, & Fenn (2008) and Lu, Wayland, & Kaan (2015) look into the development of super-segmental units. Indicated by the training results of the above studies, L2 learners can achieve significant perception gains which have been further generalized to the improvement of production for all the targets through perception-only training.

However, the above studies seldom specify whether perception would interact diversely with production according to different targets. As remarked by Cheng & Zhang (2009) and Cho (2021), the perception-production relationship varies depending on target segments. As deduced from the abovementioned patterns of L2 perception-production interface, it could be possible that the learning of some L2 targets tends to rely more on correct perception before successful pronunciation (perception-precedence pattern), while some L2 targets might take the alternative path and develop production before perception (production-precedence pattern) or develop perception and production co-equally (perception-production co-evolve pattern). If so, training methodologies should be adjusted to better facilitate L2 learners' development of different targets. The perception-production interface and the effectiveness of perception-only training on promoting L2 phonetic learning will be testified in the current study.

1.2 Objective of the Study

In this dissertation, Mandarin speakers' learning of Cantonese syllable-final segments ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]) is taken as a starting point to disentangle the following research questions centering around the impact of cross-L1-L2 similarity and perception-production interface on L2 phonetic learning.

(1) What is the cross-language perceptual similarity of syllable-final segments between Mandarin and Cantonese by Mandarin speakers?

(2) How can SLM-r be applied to interpret the learning difficulty and learning outcome during the development of Cantonese syllable-final segments by Mandarin speakers?

(3) How would L2 perception interact with production during the learning procedure of targets?

(4) Is perception-only training effective in promoting Mandarin speakers' learning of targets?

Four experiments are arranged to tackle these four research questions. Experiment 1 attempts to measure the cross-Mandarin-Cantonese perceptual similarity of syllable-final segments by Mandarin speakers through the Perceptual Assimilation Test. Based on this, corresponding predictions about the learning difficulties and learning outcomes of Cantonese syllable-final segments by Mandarin speakers are proposed according to SLM-r (Flege & Bohn, 2021). The development of Cantonese syllable-final segments by an experimental group of Mandarin speakers through pre-test (Experiment 2), seven sessions of perception training (Experiment 3), and post-test (Experiment 4) would eventually verify or deny such predictions. The L2 perception-production relationship is manifested through the interface between perception and production during the intervention (Experiment 3). The developmental trend of perception and production during the intervention can verify the effect of perception-only training which is then brought up again by comparing the performances between the trained and untrained group in Experiment 4.

1.3 Organization of the Study

This thesis is divided into eight main chapters.

After introducing the background, objective, and organization of this study in Chapter 1, the research questions of the current study are clarified. Chapter 2 sets the scene for the current study by reviewing theoretical models in L2 speech learning, measurement methods of cross-L1-L2 similarity, and Mandarin speakers' learning of Cantonese syllable-final segments. Research gaps are identified by reviewing previous studies.

In Chapter 3, Experiment 1 is about the Perceptual Assimilation Test. Methodological issues including participants, stimuli of this experiment are described. Through Four-alternative Forced-choice Identification Test and Goodness-of-fit Rating Test, the cross-language similarity of Cantonese syllable-final segments by twenty Mandarin participants is measured. Based on the observed perceptual patterns, corresponding hypotheses of learning difficulty and learning outcome of targets are proposed. Experiment 2 is discussed in Chapter 4. Through the Two-alternative Forced-choice Identification Test and Monosyllable Repetition Test, differences in the perception and production of Cantonese syllable-final segments between the control group (sixteen Cantonese native speakers) and the experimental group (thirty-two Mandarin participants) are compared. As reported in Chapter 5, only sixteen of the above thirty-two Mandarin participants are randomly selected as the experimental group for the intervention of Experiment 3. The experimental group goes through seven sessions of perception-only training. SLM-r (Flege & Bohn, 2021) is applied to interpret Mandarin speakers' learning procedure of Cantonese syllable-final segments. The deficiency of SLM-r predictions (if any) in discussing the learning difficulties of different targets during this developmental procedure is pointed out. The patterns of perception-production interface are also discussed in this chapter. Chapter 6 elaborates the obtained results of Experiment 4. The experimental group is assessed together with the control group (the rest of the sixteen Mandarin participants) via the same tasks utilized in Experiment 2. With Experiment 2 as the pre-test and Experiment 4 as the post-test, the differences between the experimental group and the control group are compared, which helps to reveal the effects of perception training. The deficiency of SLM-r predictions (if any) in discussing the learning outcomes of different targets by these two groups is proposed.

Returning to the research questions of the current study, Chapter 7 administers a general discussion about the learning of Cantonese syllable-final segments by Mandarin speakers from the perspectives of perception and production. The empirical results of the above four experiments are incorporated with theoretical implications in this section. Alternative PAM-L2 (Best & Tyler, 2007)/L2LP (Escudero, 2005) hypotheses to the possible deficiency of above SLM-r predictions and heuristic amendments to current L2 learning models are advanced to supplement the discussion of the present study. The last part of this study (Chapter 8) comes up with a conclusion first. The reflection of the current study and directions for future studies are then presented.

Chapter 2 Literature Review

2.1 Theoretical model in L2 Speech Learning

Research on L2 phonetic learning stresses the influence of learners' L1 experience and the modulation of perception-production interface. With different purposes, assumptions, and working frameworks, several well-known theoretical models in L2 speech learning are proposed as the Perceptual Assimilation Model (Best, 1995), the Perceptual Assimilation Model to Second Language (Best & Tyler, 2007), the Speech Learning Model (Flege, 1995), the revised Speech Learning Model (Flege & Bohn, 2021), and the Second Language Linguistic Perception Model (Escudero, 2005; van Leussen & Escudero, 2015).

2.1.1 The Perceptual Assimilation Model and the Perceptual Assimilation Model to Second Language

The Perceptual Assimilation Model (Best, 1995) is originally developed to account for the perception of the non-native contrasts by naive listeners who are first exposed to a foreign language.

Influenced by naive listeners' L1 experience, foreign sounds are perceptually assimilated to the most articulatory similar native sounds in three ways. Firstly, a non-native sound can be perceptually assimilated to a native category, a cluster, or a string through listeners' linguistic-phonetic knowledge (Best, 1995). In this case, a categorized non-native sound can vary in the assimilation fitness. It could be categorized as a good exemplar or a deviant exemplar of a native sound. Best (1995) exemplifies a good-deviant categorization case via English listeners' perception of velar contrast /k/-/k'/ in Thompson Salish. Both these two sounds are assimilated to English [k^h] but the Salish /k/ is less deviant than /k'/ from their English counterpart (Best, 1995). The Salish /k/ is thus categorized as a good exemplar, while Salish /k'/ refers to a deviant category of English [k^h]. Secondly, novel sounds are recognized as speech sounds but cannot be classified as any of the native categories (namely uncategorized sounds). Listeners process the assimilation of these uncategorizable speech sounds by combining their linguistic and auditory knowledge (Best, 1995). For instance, there is no liquid contrast /l/-/r/ in the phonological system of Japanese. Many Japanese listeners regard English /l/ and /r/ as natural sounds of human speech but cannot categorize them as any Japanese phoneme categories, so it is difficult for them to correctly discriminate English contrast /l/-/r/ (Best, 1995). Lastly, the non-native segment fails to assimilate into the native phonological space and is regarded as a non-speech sound in some cases (Best, 1995). For example, English speakers perceive the clicks in Bantu of Africa as non-speech sounds instead of any English phoneme categories (Best, 1995).

These categorized, uncategorized, and non-assimilated foreign sounds can form up six possible assimilation contrasts as a function of the cross-L1-L2 similarity towards their L1 counterparts. According to Best (1995), there are Two-category Assimilation (TC type), Category-goodness Difference Assimilation (CG type), Single-category Assimilation (SC type), Uncategorized-uncategorized Assimilation (UU type), Uncategorized-categorized Assimilation (UC type), and Non-assimilable to Non-assimilable Assimilation (NA type). L2 novices' discrimination performances towards these six assimilation patterns can range from the poor to the excellent (Best, 1995).

Best & Tyler (2007) further generalizes the tenets of PAM to the second language and formulates PAM-L2. Compared with PAM which focuses on the perception of non-native sounds by naive listeners at the initial exposure to a foreign language, PAM-L2 (Best & Tyler, 2007) intends to investigate the speech perception by active beginning L2 learners. The abovementioned six assimilation types of non-native contrasts in PAM are further elaborated with testable developmental trajectories and learning difficulties in PAM-L2 as below.

If L2 perceivers assimilate one L2 sound as a good exemplar to an L1 phonological category at the phonetic level, they will confront either Two-category Assimilation or Uncategorized-categorized Assimilation formed up by this L2 sound and other L2 categories (Best & Tyler, 2007). The discrimination of these two assimilation patterns will be less challenging for the L2 perceivers.

L2 learners are also able to successfully discriminate those L2 phonological categories which are assimilated to one L1 category but with deviant goodness-of-fit (Category-goodness Difference Assimilation in PAM). For the deviant L2 phone, a new L2 category will be eventually developed with L2 learners' accumulating perception of the lexical functional differences between these L2 contrasts (Best & Tyler, 2007). For the better-fitting L2 phone which is perceived as an equivalence to its L1 counterpart, no new category is likely to be learned (Best & Tyler, 2007).

When two L2 phonological categories are assimilated to one L1 category with equal goodness-of-fit (termed as Single-category Assimilation in PAM), the L2 learners will find it most difficult to differentiate these two L2 phones at the beginning (Best & Tyler, 2007). It can advantage the formation of a new L2 category of Single-category Assimilation if the two phones of this assimilation type are frequently-used contrasting words or contain many minimally contrasting words in their phonological context (Best & Tyler, 2007). Once the L2 learners perceptually develop a new category for at least one of these L2 phones, they are able to further obtain the other new phonological category or categories.

Uncategorized-uncategorized Assimilation addresses no L1-L2 phonological assimilation and neither of the two L2 categories of this assimilation can be assimilated clearly to any L1 segment (Best & Tyler, 2007). PAM-L2 (Best & Tyler, 2007) hypothesizes that it is relatively easy to perceptually learn these two L2 phones if each of them is assimilated to distantly different sets of L1 phones. Alternatively, L2 learners' perceptual learning will be hindered when these two uncategorized L2 phones are assimilated to the same set of L1 phonemes (Best & Tyler, 2007).

Under some circumstances, L2 sounds are recognized as non-speech sounds and are non-assimilable to any existing L1 phonological categories (Best & Tyler, 2007). If L2 learners ultimately classify non-assimilable novel targets into uncategorized tokens, they will initiate the perceptual learning of these sounds (Best & Tyler, 2007). Otherwise, some non-assimilable sounds might never be integrated into L2 learners' phonological knowledge.

The perception and production are tightly synchronized in PAM and PAM-L2. Under the direct realism framework, these two models collectively posit that articulatory gestures of production are the direct primitives of speech perception. As perception and production share the same articulatory metric with compatible information, it is possible for listeners to perceive the articulatory gestural properties of production from the ambient language (Best, 1995; Best & Tyler, 2007). The perceived articulatory similarities in the production between listeners' native language and the foreign language determine the assimilation patterns and the perceptual learning of target sounds. It is predicted that experienced listeners can detect the produced gestural discrepancies between native and novel segments, while naïve perceivers would consider native and non-native segments to be similar in terms of articulatory gestures and encounter more discrimination difficulties. However, both PAM and PAM-L2 focus primarily on non-native perception instead of production and neither of them explicitly addresses the relationship between these two modalities (Ingham, 2014; Nagle, 2018a; Tyler, 2019; Nagle & Baese-Berk, 2021). Therefore, previous studies mainly focus on perception in their discussion of these two models (e.g., Best, Halle, Bohn, & Faber, 2003; Francis, Ciocca, & Ma, 2004; Francis et al., 2008; So & Best, 2010; Tyler, Best, Faber, & Levitt, 2014; Pilus, 2016; Cebrian, Carlet, Gorba, & Gavaldà, 2019; Tuninetti, Whang, & Escudero, 2019; Luo, Li, & Mok, 2020). Although some other studies by Hao (2012), Ingham (2014), Cheng & Zhang (2015), Mokari & Werner (2017), Thorin et al. (2018), Baese-Berk (2019) interpret PAM and PAM-L2 in both perception and production, most of them seldom explicitly address the relationship between these two modalities. In Mokari & Werner (2017), both perception test and production test are employed to investigate Azerbaijani learners' development of English vowels. The validity of PAM is generally testified when predicting Azerbaijani learners' performances of targets. This study also aims at exploring the perception-production link in L2 phonetic learning. However, PAM is not directly applied to the discussion of perception-production interface in Mokari & Werner (2017) and the obtained results point to an uncorrelated perception-production link. What's more, neither PAM nor PAM-L2 includes an explicit account of the developmental process in L2 speech learning, which makes it difficult to address a time-course perception-production interaction (Nagle & Baese-Berk, 2021). On the other hand, these two models target at non-native contrasts rather than individual non-native sounds. Thus, the above studies often apply discrimination tasks, which assess non-native listeners' abilities to distinguish sound contrasts, in their experiments. The targeted modality (non-native perception), targeted segment (non-native contrasts), and adopted task (discrimination tests) of PAM/PAM-L2 will be considered when deciding the theoretical application of the current study.

2.1.2 The Speech Learning Model and the Revised Speech Learning Model

Compared with the focus on the initial state of L2 perception by naive listeners/beginning L2 learners in PAM/PAM-L2, Flege (1995) and Flege et al. (2003) instead pay attention to experienced learners' ultimate achievement of L2 perception and production in the Speech Learning Model (SLM). Since allophones of phonemes vary across languages (e.g., the contrastive similarities between two sets of allophones of phonemes /p, t, k/ and /b, d, g/ are different in French and German by Kohler, 1981) and acoustic cues of the same allophones in a language could be different according to positions (e.g., English word-medial and word-final allophones of the same stop phonemes are different in their perceptual cues by Dmitrieva, 2019), the learning targets in PAM/ PAM-L2 (L2 contrasts) are replaced by individual position-sensitive allophones in SLM.

Coexisting in a common phonetic space, the position-sensitive allophones of L1 and L2 are perceptually linked to each other through a subconsciously interlingual identification in L2 learners' cognitive process (Flege, 1995). Decided by the detected cross-L1-L2 phonetic similarity, SLM (Flege, 1995) proposes that L2 sounds can be categorized as identical, similar, and new sounds relative to their closest L1 counterparts. Accordingly, different learning outcomes are predicted. If

L2 sounds are regarded as identical to their L1 counterparts, L2 learners can take advantage of a positive L1-L2 transfer and would find it relatively easy to develop identical L2 categories (Flege, 1995). When confronting new sounds which cannot find counterparts in learners' L1, the formation of new L2 categories is processive with cumulating L2 experience and the learning of such sounds will be ultimately successful (Flege, 1995). The learning of similar sounds is the most demanding for L2 learners. In this case of SLM (Flege, 1995), related L1 and L2 categories are assimilated into a merged L1-L2 category (equivalence classification). L2 listeners are unable to extract the phoneme correctly from the equivalence classification, which prevents the successful formation of similar categories (Flege, 1995). The native-like L2 performance will be more likely to be attained by learners who have earlier exposure to the L2 environment and who use L2 more frequently (Flege, MacKay, & Meador 1999; Piske, MacKay, & Flege 2001, Flege & MacKay 2004).

Driven by previous empirical studies of SLM, Chu, Yang, & Liu (2019) points out a paradox in this framework. The detected cross-L1-L2 phonetic similarity, which plays a crucial role in the formation of an L2 category, sometimes happens at the beginning of L2 learning when L2 learners initially confront a non-native sound. This, however, contradicts the applicability of SLM to advanced L2 learners who are with years of exposure. It is also a drawback that the framework of SLM is not applicable to the early stage of L2 learning. Guion, Flege, Akahane-Yamada, & Pruitt (2000) and Mayr & Escudero (2010) thus suggest that the original framework of SLM should be revised to accommodate the initial L2 learning stage.

In response to the above drawbacks of SLM proposed by previous studies, the revised Speech Learning Model (Flege & Bohn, 2021) is developed to accommodate both beginning and advanced learners with an attempt to better understand the process of L2 phonetic category formation during naturalistic L2 learning. The SLM-r retains the main ingredients and hypotheses from SLM. Two models both target on position-sensitive allophones of L2 sounds and rely on perceived cross-L1-L2 phonetic similarity when predicting the learning of L2 sounds. SLM-r (Flege & Bohn, 2021) further suggests that perceptual assessment instead of acoustic measurement should be taken to estimate the cross-L1-L2 phonetic similarity since studies by Johnson, Flemming, & Wright (1993) and Levy & Strange (2008) point out that listeners could perceive divergently from a prediction by acoustic analysis.

The SLM and the SLM-r are, however, different in the interpretation of the perception-production relationship. In SLM (Flege, 1995), it is predicted that the development of perception will be achieved ahead of successful production. Flege (1995), Flege, Munro et al. (1995), and Flege, Schirru et al. (2003) suggest that the perception accuracy would put a ceiling on the production accuracy of L2 sounds. Several observations raise inconsistent opinions on this SLM statement and present evidence about a strong bi-directional and co-equal connection between perception and production (see Sheldon & Strange, 1982; Perkell, Guenther et al., 2004, Perkell, Matthies et al., 2004; Peperkamp & Bouchon, 2011; Darcy & Krüger, 2012; Shultz et al., 2012; Chao et al., 2019). Taking these previous studies into consideration, SLM-r acknowledges the perception-production interface but points out a Co-evolve Hypothesis that there is a co-equal relationship between perception and production. These two modalities co-evolve in a bi-directional way without precedence in L2 learning (Flege & Bohn, 2021). The Perception-precedence Hypothesis postulated by SLM has been widely discussed in the studies by Cheng & Zhang (2009), Cardoso (2011), Hao & de Jong (2016), Sakai & Moorman (2018), Thorin et al. (2018), Baese-Berk (2019), Nagle (2019), Casillas (2020), Melnik-Leroy, Turnbull, & Peperkamp (2021) and so on. Some of these studies (Cardoso, 2011; Hao & de Jong, 2016; Thorin et al., 2018; Nagle, 2019; Casillas, 2020; Melnik-Leroy et al., 2021) have observed the perception-precedence pattern and explicitly support that perception develops before production in the L2 phonetic learning. For instance, in the study by Melnik-Leroy et al. (2021), the perception-production interface in the performances of a French minimal pair /u/-/y/ by English natives with sufficient learning experience in French are assessed. The obtained results in this study point to a robust perception-production link and it is suggested that success in the perception is the prerequisite for the achievement of production. Casillas (2020) come up with similar conclusions about the production-perception relationship. In this work, the performances of Spanish obstruent voicing pairs by English-speaking novices of Spanish are documented. The obtained results suggest that the category formation in L2 phonetic learning performs a perception-driven pattern at the initial stage. English speakers' perception of targets is achieved first to guide the development of production. In contrast, other studies (Cheng & Zhang, 2009; Sakai & Moorman, 2018; Baese-Berk, 2019) point out that perception does not always precede production and the perception-production interface could vary in the L2 phonetic learning. By measuring experienced Chinese-speaking English learners' perception and production of English consonants and vowels, Cheng & Zhang (2009) intends to testify the perception-production relationship. Diverse perception-production patterns varying according to different tested targets are observed. Precisely, perception and production are positively correlated with each other in participants' performances of English consonants, while there is no significant relationship between these two modalities in any of the targeted vowels. On the other hand, the co-evolve relationship between perception and production by the recently proposed SLM-r remains to be empirically testified. This leaves room for further debate about the perception-production relationship of L2 speech learning in the current study.

2.1.3 The Second Language Linguistic Perception Model

Based on Linguistic Perception Model (Escudero & Boersma, 2004; Escudero, 2006), the Second Language Linguistic Perception Model (L2LP) is developed by Escudero (2005) and van Leussen & Escudero (2015). Compared with its predecessor, L2LP aims to model the entire developmental process from the initial to the end stage of L2 speech perception within the computational learning framework of the Stochastic Optimality Theory (Boersma, 1998). To explicate the effects of L1 experience and perception-production interface in L2 perception learning, the Full Copying Hypothesis and the Optimal Perception Hypothesis are proposed in this model.

The Full Copying Hypothesis (Escudero, 2005 & 2009) of L2LP glimpses the influence of previously existing L1 knowledge on L2 perceptual learning. According to Escudero (2005 & 2009), L2 learners duplicate their L1 perception grammar when initiating the L2 learning process. The perceived L2 contrasts are directly mapped into the sounds and categories of L1 grammar. As learning processes, L2 learners would gradually attune their perception of L2 to become more native-like with accumulating experience. Different L1-L2 mapping patterns can lead to three learning scenarios (the NEW scenario, the SIMILAR scenario, and the SUBSET scenario) in L2LP. As outlined below, Escudero (2005) and van Leussen & Escudero (2015) provides precise learning tasks, developmental trajectories, and learning difficulties of these three learning scenarios for L2 learners.

If learners encounter more L2 sounds than those in their L1 perception, the NEW scenario occurs and the learning of new L2 categories is involved. Two L2 sounds are often realized as one L1 category in the NEW scenario of L2LP (Escudero, 2005). In this case, the learning of new L2 sounds relies on creating new L2 categories or splitting previously existing L1 categories, which is predicted to be the most difficult for L2 learners (Escudero, 2005).

The SIMILAR scenario of L2LP refers to the situation that L2 novices classify two novel targets separately as two L1 categories (Escudero, 2005). These L1 and L2 categories overlap in their acoustic features but perform delicately phonetic differences in their productions (Escudero, 2005). According to the prediction of L2LP, it is less problematic to tackle a SIMILAR scenario than a NEW scenario. L2 learners can simply duplicate the previously existing L1 categories and adjust the L1 boundaries to fit those of the L2 categories (Escudero, 2005).

Opposite to the situation in the NEW scenario, learners confront fewer L2 categories than those in their L1 perception grammar in the SUBSET scenario. In this case, an L2 segment is perceived as more than one L1 category (Escudero, 2005). L2 learners do not need to create new L2 contrasts in the perception of this scenario and thus will find it less difficult than the learning of a NEW scenario (Escudero, 2005).

The Optimal Perception Hypothesis (Escudero, 2005 & 2009) of L2LP addresses the perception-production interface in L2 perception learning. It asserts that L2 learners are optimal perceivers who will favor the most reliable acoustic cues of production in their ambient language environment to correctly perceive the L2 sound contrasts (Escudero, 2005 & 2009). Since L2 learners are equipped with a duplicated L1 perception grammar (refers to the aforementioned Full Copying Hypothesis), they will initially perceive the L2 contrasts as a function of how these sounds are produced in their L1 at the beginning of L2 learning. The learning purpose of L2 perception development is to obtain the optimal perception of the target language. As learning processes, L2 listeners will perceive L2 targets as a function of how these sounds are produced in the target L2 and gradually attune their developing perceptual abilities to become more native-like. Escudero (2005) demonstrates this hypothesis via the perception of Spanish vowel contrast /i/-/e/ which differs in their F1 values. According to the variations of the average productions of these two vowels by Spanish native speakers, the perception boundary of their F1 values is computed as 407 Hz (The F1 of /i/ < 407 Hz; The F1 of /e/ > 407 Hz). To lower the possibility of misunderstanding, optimal Spanish L2 learners would perceive these two vowels according to how they are produced in the ambient Spanish environment. Therefore, variants of these two vowels with an F1 value lower than 407 Hz are perceived as /i/, while those with an F1 value higher than 407 Hz are recognized as /e/ by Spanish L2 learners. However, similar to PAM and PAM-L2, L2LP focuses mainly on the L2 perception instead of production and does not account for a direct L2 perception-production link (Nagle & Baese-Berk, 2021).

Although L2LP is still a working model and has not been well-documented like PAM, PAM-L2, and SLM, researchers have applied L2LP to some perceptual studies of vowels (see Mayr & Escudero, 2010; Escudero, Simon, & Mitterer, 2012; Elvin, 2016; Elvin, Williams, & Escudero, 2016; Yazawa, Kondo, & Escudero, 2017; Chappell, 2019; Luo et al., 2020). The aforementioned ingredients of L2LP are illustrated by these studies. Mayr & Escudero (2010) adopts cross-language perceptual similarity to predict most of the developmental scenarios in the perception of German rounded vowels by English-speaking learners of German. Escudero et al. (2012), instead, points out that acoustic similarity can predict the cross-language perceptual assimilation patterns and L2 perception scenarios through North Holland and Flemish listeners' perception of English front vowels. In Elvin et al. (2016), the Full Copying Hypothesis and the perception-production relationship of L2LP are discussed. This study suggests that L2 learners initiate the L2 learning process with a duplicated L1 grammar. They initially perceive and produce L2 sounds according to the acoustic features of their L1 counterparts (Elvin et al., 2016). Following the same line, Chappell (2019) also supports the Full Copying Hypothesis of L2LP through the L2 perception of reduced Spanish vowels. It is suggested that L2 learners perceive L2 contrast according to a duplicate L1 perceptual system then adjust their L2 perception to become more native over time.

2.2 Measurement of Cross-L1-L2 Similarity

All the aforementioned models acknowledge the important influence of L1 experience on L2 speech learning and come up with corresponding predictions as a function of cross-L1-L2 similarity. The crucial role of the cross-L1-L2 similarity has been widely recognized while the appropriate measurement to gauge the cross-language distance remains to be determined (Flege, 1995; Strange, 1999; Strange, Yamada, Kubo, Trent, & Nishi, 2001; Bohn, 2002; Flege, 2007; Strange, 2007). The following measurement methods (the comparison of ① IPA symbol, ② acoustic similarity, ③ articulatory similarity, and ④ perceptual similarity) of cross-L1-L2 similarity are mainly adopted in the extant studies.

2.2.1 IPA Symbol

Investigations related to SLM (Flege, 1987 & 1988a & 1992) take the comparison of IPA symbol as a preliminary method to assess the cross-L1-L2 similarity. The categorizations of identical, new, and similar L2 sounds are contingent on the IPA transcription of L2 learning targets and their L1 counterparts. An identical L2 sound shares the same IPA transcription with its L1 counterparts, while a new L2 sound is transcribed by a dissimilar IPA symbol that is not used for any of the L1 categories. A similar L2 sound also uses the same IPA symbol with its corresponding L1 sounds. To better distinguish between the identical sounds and similar sounds, additional evidence of acoustic similarities or perceptual features is necessary.

The IPA symbol metric has the advantage of providing straightforward evidence of cross-L1-L2 similarity and has been applied in the verification of SLM and PAM/PAM-L2 predictions in studies by Flege (1987 & 1988a & 1992), Larson-Hall (2004), Ingham (2014), Milenova (2015), Kitikanan (2017), Yang, Chen, & Xiao (2020) and so on. However, it is suggested that this criterion should be used only as a provisional measure for the following reasons. The first problem of using this measure is that many phonetic transcription systems in use are not always consistent (Flege, 1992). Second, broad and narrow transcription of IPA symbols, which can be different in some cases, are employed by different researchers. These choices can lead to distinct predictions of cross-L1-L2 similarity (Yang et al., 2020). Lastly, this method is insufficient when being applied to distinguish between identical sounds and similar sounds. Extra evidence by other measurements should be conducted as a supplement since both identical sounds and similar sounds are transcribed via the same IPA symbols.

2.2.2 Acoustic Similarity

Compared with the relatively superficial IPA symbol criterion, acoustic measurement of target sounds can provide more objective and concrete evidence about the cross-L1-L2 similarity.

Both Flege's (1995) SLM and Escudero's (2005) L2LP model explicitly propose that different acoustic realizations of L1 and L2 sounds are the important indicator of cross-L2-L1 similarity and can be used to predict learners' perception and production of L2 sounds. In SLM (Flege, 1987 & 1992), the new, similar and identical L2 sounds represent increasing acoustic similarities to their L1 rivals. Precisely, the acoustic features of new L2 phones are largely distinct from those of corresponding L1 sounds. Smaller but statistically significant acoustic differences can be identified between similar L1 and L2 sounds. By contrast, the identical L2 sounds bear the biggest acoustic similarities to their L1 counterparts. L2LP (Escudero, 2005 & 2009; van Leussen & Escudero, 2015) also considers detailed acoustic comparisons between L1 and L2 categories as a reliable predictor of discrimination performance and category formation in the learning of NEW, SIMILAR, and SUBSET scenarios. In the NEW scenario, the acoustic features of two L2 targets largely approximate to those of one L1 counterpart. The corresponding L1-L2 contrasts of a SIMILAR scenario are similar in their acoustic properties. They overlap in terms of acoustic features but perform delicately phonetic differences in the production. The SUBSET scenario refers to the case that the acoustic features of an L2 sound overlap with those of several L1 sounds but such L2 acoustic features cannot be processed in the same way as those in L1 by L2 learners.

The application of acoustic similarity criterion is mainly observed in the studies under the framework of SLM and L2LP (Flege, Munro, & Skelton, 1992; Flege, 1995; Escudero & Boersma, 2004; Strange, Bohn, Trent, & Nishi, 2004; Escudero, Paola, Benders, & Lipski, 2009; Escudero & Vasiliev 2011; Escudero et al., 2012; Escudero & Williams, 2012; Escudero, Sisinni, & Grimaldi, 2014; van Leussen & Escudero, 2015; Elvin, 2016; Elvin et al., 2016; Yang et al., 2020). Different acoustic parameters are measured according to the inherent features of L2 learning targets. Studies concerning cross-linguistic similarity in vowels compare L1 and L2 categories in terms of spectrum and duration (Flege, 1992; Strange et al., 2004; Strange, Bohn, Nishi, & Trent, 2005; Gilichinskaya & Strange, 2010; Escudero & Vasiliev, 2011; Escudero et al., 2012; van Leussen & Escudero, 2015; Elvin, 2016; Elvin et al., 2016). Studies targeting L2 consonants gauge the cross-L1-L2 similarity by measuring the voice onset time (VOT) in Flege & Eefting (1987), Riney & Takagi (1999), van Alphen & Smits (2004), Antonioua, Best, Tylera, & Kroosa (2011), Piccinini & Arvaniti (2015), Bennett, Tang, & Ajsivinac Sian (2018), Yang et al. (2020) and so on.

Although acoustic similarity successfully predicts L2 perception and production in the studies by Flege (1995), Riney & Takagi (1999), Gilichinskaya & Strange (2010), Escudero et al. (2012), Elvin (2016), and so on, some other studies provide contrary results (Strange et al., 2004; Strange et al., 2005; Levy & Strange, 2008; Yang et al., 2020). It is suggested that the comparison of acoustic similarity is not always aligned with the perceptual patterns by L2 learners and it is necessary to investigate the perceptual similarities between L1 and L2 directly.

2.2.3 Articulatory Similarity

By retaining the central tenet of Articulatory Phonology (Browman & Goldstein, 1989 & 1992 & 1993), PAM (Best, 1995) and PAM-L2 (Best & Tyler, 2007) consider the dynamic actions of articulatory gestures in human beings' vocal tract as the basic units of production, perception, and mental representation of speech. Such articulatory gestures represent not only the primitive actions of the vocal tract but also the phonological information units of different phones (Browman & Goldstein, 1992). The comparison of articulatory similarity, as a matter of course, is adopted as an important method to assess the cross-L1-L2 similarity by some PAM and PAM-L2 studies (Best & McRoberts, 2003; Best, Goldstein, Nam, & Tyler, 2016; Best & Tyler, 2007; Tyler, Best, Goldstein, & Antoniou, 2014).

Based on Articulatory Phonology (Browman & Goldstein, 1989 & 1992 & 1993), the articulatory gestures can be categorized in terms of ① Articulatory Organ (tongue body, tongue tip, lips, glottis, velum), ② Constriction Location (labial, dental, alveolar, post-alveolar, palatal, velar, uvular, pharyngeal), and ③ Constriction Degree (closed, critical, narrow, mid, wide). Accordingly, PAM and PAM-L2 studies apply these three dimensions and their parameters to gauge the articulatory similarity of target sounds between L1 and L2. Best & McRoberts (2003), Kochetov (2005), Best et al. (2016), and Flemming (2016) conjecture that more distinct articulatory gestures indicate greater articulatory differences between two target segments of L1 and L2.

2.2.4 Perceptual Similarity

The PAM (Best, 1994 & 1995), SLM (Flege, 1995), PAM-L2 (Best & Tyler, 2007), SLM-r (Flege & Bohn, 2021), and L2LP (Escudero, 2005 & 2006 & 2009) consistently agree that the cross-L1-L2 perceptual patterns are a valuable determiner of learning difficulties in L2 phonetic development and suggest that the perceptual similarity should be empirically measured through cross-language mapping experiments.

The measurement methods of perceptual similarity have experienced gradual adjustments before a widely accepted paradigm is settled. Flege, Munro, & Fox (1994) firstly adopt a quantitative technique to gauge how L2 learners would perceive the similarities/differences between L1 and L2 sounds. The productions of L1 and L2 tokens are presented as a pair in each trial. Listeners are required to directly compare two tokens and rate their similarities/differences via a Likert scale. Flege (2005) and Flege & Bohn (2021) also touch upon this pairwise goodness-of-fit rating method. As this method requires a large number of high-variability L1-L2 sound pairs and it could be time-consuming sometimes, it has seldomly been used in previous studies of L2 phonetic learning although it can achieve an adequate measure of perceived cross-L1-L2 similarity (Flege, 2021). On the other hand, Best, McRoberts, & Goodell (2001) turns to an alternative transcription technique that requests listeners to categorize L2 sounds in terms of L1 categories. The obtained categorization patterns can be used to predict L2 phonetic learning. Mayr & Escudero (2010) and Escudero & Vasiliev (2011) employ similar transcription techniques for the assessments of perceptual similarity.

Another measurement method of perceptual similarities (Perceptual Assimilation Test) is proposed and has been widely used in more recent studies (Flege, 1991; Guion et al., 2000; Flege & MacKay, 2004; Best & Tyler, 2007; Strange, et al., 2004; Strange, 2007; Levy, 2009; Mayr & Escudero, 2010; Escudero & Vasiliev, 2011; Escudero et al., 2012; Tyler et al., 2014; Mokari & Werner, 2017; Cebrian et al., 2019; Tyler, 2019; Yang et al., 2020). This test combines the methods of Flege et al. (1994) and Best et al. (2001) with slight adjustments and entails two sequential steps: ① categorization of L2 tokens into L1 phones and ② goodness-of-fit rating of L1-L2 mapping pairs after categorization. The first step is normally realized as an N-alternative forced-choice identification task. L2 perceivers are instructed to identify the most similar L1 sound from several presented L1 categories to an L2 token that they hear during the task. Afterward, they rate the similarity between the classified L1-L2 pairs with interval rating scales. Some studies (Guion et al., 2000; Tyler, 2019) propose that it is essential to conduct the goodness-of-fit rating with categorization task. In the case that different L2 tokens obtain the same/similar categorization percentages, the rating scores can be used to weight categorization results and to help further decide which L2 token is better classified.

Different indicators are employed to interpret the results of categorization and rating. Researchers mainly use a 50% or 70% categorization rate to indicate that an L2 token is categorized if it is assimilated to a given L1 category more than 50% or 70% of the time (see Tyler et al., 2014; Mokari & Werner, 2017 and so on for a 50% threshold, and Bundgaard-Nielsen, Best, & Tyler, 2011; Antoniou, Tyler, & Best, 2012; Faris, Best, & Tyler, 2018; Yang et al., 2020 and so on for a 70% threshold). Meanwhile, the goodness-of-fit rating is indicated through a 5-point scale (Flege & MacKay, 2004; Chan, 2012; Escudero et al., 2014; Garibaldia & Bohnb, 2015; Yang et al., 2020), 7-point scale (Guion et al., 2000; Strange et al., 2004; Hao, 2012; Cebrian et al., 2019) or 9-point scale (Flege et al., 1994; Levy, 2009; Bohn & Ellegaard, 2019). The results of categorization and rating can also be calculated via fit index (multiplication of categorization rate and goodness-of-fit rating). A higher fit index indicates that an L2 sound is perceptually mapped onto an L1 category, while lower fit indexes refer to that L2 sounds are perceived as dissimilar foreign sounds. The fit index is applied by Guion et al. (2000) and Bohn & Ellegaard (2019) to measure perceptual similarities.

With an attempt to predict L2 perception and production through the perceptual patterns between L1 and L2 categories, the perceptual similarity criterion is widely applied by previous studies of PAM, PAM-L2, SLM, and L2LP (Flege, 1991; Guion et al., 2000; Flege & MacKay, 2004; Best & Tyler, 2007; Strange, et al., 2004; Strange, 2007; Levy, 2009; Mayr & Escudero, 2010; Escudero & Vasiliev, 2011;

Escudero et al., 2012; Tyler et al., 2014; Mokari & Werner, 2017; Cebrian et al., 2019; Tyler, 2019; Yang et al., 2020). The newly proposed SLM-r (Flege & Bohn, 2021) also prioritizes the perceptual similarity criterion in its discussion. It is explicitly pointed out that cross-L1-L2 similarity should be measured through the perceptual assessment rather than the acoustical method.

2.3 Mandarin Speakers' Learning of Cantonese Syllable-final Segments

Cantonese is a Chinese dialect spoken by some 40 million speakers worldwide (Bauer & Benedict, 2011). Mandarin immigrants to Hong Kong represent a unique population of Cantonese users. Compared with other Cantonese syllable structures encountered by Mandarin speakers in Hong Kong, the percentage of CVC syllables is up to 32.23% (Leung, Law, & Fung, 2004). Mandarin speakers' learning of syllable-final segments in Cantonese CVC syllables is taken as a case to elaborate the influence of L1 experience and perception-production interface proposed by the aforementioned models.

CV syllable is the most unmarked syllable structure across languages (Jakobson, 1971; Clements & Keyser, 1983). Mandarin and Cantonese both involve this [CV∅] syllable structure and therefore [-∅] is the shared sound in their phonological systems. Apart from this, Mandarin and Cantonese employ different patterns of syllable-final segments respectively. As can be seen from Table 1, Mandarin comprises contrastively distributed /-n/ and /-ŋ/ as syllable-final consonants (Li & Thompson, 1989). In Cantonese, there are six permissible syllable-final consonants [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ] (Cheung, 1986; Yip, 1997; Bauer & Benedict, 2011). Among them, syllable-final obstruents [-p˺], [-t˺] and [-k˺] are unreleased (Flege & Wang, 1989; Edge, 1991; Cichocki, House, Kinloch, & Lister, 1993).

	PoA/MoA	Nasal	Obstruent
Mandarin	Labial	--	--
	Coronal	[-n]	--
	Dorsal	[-ŋ]	--
Cantonese	Labial	[-m]	[-p˺]
	Coronal	[-n]	[-t˺]
	Dorsal	[-ŋ]	[-k˺]

Table 1. Syllable-final Segments in Cantonese and Mandarin

It is worth noting that the contrastive distribution of six syllable-final consonants in Cantonese is undergoing a merger. To, Cheung, & McLeod (2013) proposes that four variants related to final consonants (/-ŋ/→/-n/, /-n/→/-ŋ/, /-k˺/→/-t˺/ and /-t˺/→/-k˺/) are commonly used by Cantonese adult speakers. Chen (1999), Zee (1999a), and To, McLeod, & Cheung (2015) point out that [-t˺] and [-k˺] tend to merge and become phonetic variations especially when the nucleus is [a]/[ɐ]. More studies attempt to identify the merging tendency between the coronal and dorsal segments. In Zee (1999b) and Chen (1999), both /-ŋ/→/-n/ and /-k˺/→/-t˺/ are documented. From the data of Wong (2005) and Bauer & Benedict (2011), the merging tendency from /-ŋ/ to /-n/ is reported. In Law, Fung, & Bauer (2001)'s findings based on the young generation of Cantonese native speakers, merging occurs more in stop endings [-t̚], [-k̚] than those in nasal endings [-n], [-ŋ]. Stop endings [-t̚], [-k̚] are interchanged into each other at about the same rate, while nasal ending [-ŋ] tend to merge into [-n].

Only a few empirical studies touch upon Mandarin speakers' perception and production of Cantonese syllable-final segments. Chu (2009) investigates Mandarin speakers' discrimination of Cantonese coda contrasts (/p/ - /t/, /p/ - /k/, /t/ - /k/, /m/ - /n/, /m/ - /ŋ/, /n/ - /ŋ/) based on the predictions of PAM (Best, 1995). His results generally verify the effect of L1 phonemic categories on the perception of L2 contrasts. On the other hand, Cantonese subjects in Chu's (2009) study have significant difficulties in discriminating coronal-dorsal contrast, which is in line with the coronal-dorsal merging pattern identified in Zee (1999a). It is suggested that the effect of L2 input of such merging tendency should be considered together with the influence of L1 phonemic categories in the investigation of L2 perception. Zhang, Chen, & Leung (2019) focuses more specifically on Cantonese syllable-final stops. According to the results, Mandarin speakers' perception and production accuracy of the above three targets follow the ranking as being [-p̚] >> [-t̚], [-k̚]. It is assumed that daily exposure to Cantonese merging [-t̚]-[-k̚] could pose difficulties for Mandarin speakers to distinguish these two segments. Their perception and production of these two segments, as a result, both refer to the lowest accuracy. Gao (2005) assesses the discrimination and identification of Cantonese /-p̚/, /-t̚/, /-k̚/ by Hong Kong Cantonese native speakers and Mandarin speakers.

Cantonese speakers performed significantly better than their Mandarin counterparts with an accuracy score of 93% vs. 78% for discrimination and 82% vs. 69% for identification. In other studies of Mandarin speakers' perception of English unreleased syllable-final stops (Flege, 1989; Flege & Wang, 1989; Flege & Liu, 2001), it is suggested that the absence of a release burst results in less accurate identification of such targets. It remains unclear whether Mandarin speakers would confront the same difficulties when identifying Cantonese unreleased syllable-final stops.

The above studies mainly provide synchronic evidence. However, few studies have probed into Mandarin speakers' learning of Cantonese syllable-final segments from a developmental perspective. Besides, the coronal-dorsal merger of syllable-final segments by Hong Kong Cantonese native speakers is well-documented, while it remains unclear how other populations of Cantonese users would respond to such sound merger. The current study attempts to supplement these topics by looking into Mandarin speakers' perception and production development of seven Cantonese syllable-final segments which are undergoing a coronal-dorsal merging tendency.

2.4 The Present Study

To investigate the influence of L1 experience and L2 perception-production interface on L2 phonetic learning, the current study intends to investigate the perception and production development of Cantonese syllable-final segments by Mandarin speakers with no Cantonese experience. Working frameworks of the afore-reviewed L2 learning models are recapped here to decide an appropriate theoretical explanation of the present study.

The development of L2 phonetic learning involves language systems of L2 learners' native language, the language used in the transition process of learning, and the target language. Interlanguage is proposed to refer to the transition language system used by adult L2 learners during their learning procedure (Selinker, 1972; Eckman, 1981; Lakshmanan & Selinker, 2001; Tarone, 2012; Selinker & Rutherford, 2013). The aforementioned L2 learning models provide different interpretations of the perception-production interface which are applicable to different language systems of L2 phonetic learning. As reviewed in Section 2.1.1, PAM (Best, 1995) and PAM-L2 (Best & Tyler, 2007) primarily focus on non-native perception rather than production. These two models posit that the performances of non-native perception are decided by listeners' abilities to perceive the similarities of articulatory gestures in the production between listeners' native language and the target language (Best, 1995; Best & Tyler, 2007). As can be seen in Section 2.1.3, L2LP (Escudero, 2005; van Leussen & Escudero, 2015) also mainly addresses the issues of second language perception. In this framework, L2 learners will favor the most reliable acoustic cues of production in their ambient language environment to achieve successful perception. They initiate L2 perception learning by perceiving L2 contrasts according to how these sounds are produced in their L1 and then gradually polish their developing perception by making use of the produced acoustic cues in ambient L2 environment (Escudero, 2005; van Leussen &

Escudero, 2015). Thus, these models involve the interface between interlanguage perception and the production of the native language/target language. By contrast, SLM and SLM-r deal with perception and production in L2 phonetic learning. The Perception-precedence Hypothesis of SLM (Flege, 1995) suggests that L2 perception would develop in advance to guide the development of L2 production, while the Perception-production Co-evolve Hypothesis of SLM-r (Flege & Bohn, 2021) negates the precedence of either of these two modalities and points out a co-equal and bi-directional relationship between perception and production during L2 speech learning. Thus, these two hypotheses both consider the relationship between perception and production of L2 learners' interlanguage. SLM/SLM-r can provide a more overt account for the L2 perception-production interface during the gradual learning procedure of Mandarin-Cantonese interlanguage in the current study.

From the perspective of different developmental stages assessed in these models, PAM/PAM-L2 and SLM make predictions on the perception/production difficulties of non-native sounds by beginning listeners/advanced L2 learners in the initial/final stage of L2 phonetic learning. By contrast, SLM-r and L2LP not only incorporate the initial and final stage but also touch upon the developmental process of L2 phonetic learning. Mandarin speakers' gradual learning of targets can be better described in either SLM-r or L2LP.

On the other hand, the aforementioned L2 learning models have two different focuses. PAM/PAM-L2 and L2LP both target at novel sound contrasts. In contrast, SLM/SLM-r considers individual position-sensitive allophones which match the target token (seven individual syllable-final segments in Cantonese) of the current study.

Taking the above factors into consideration, the SLM-r is, therefore, more relevant to the research goals and experimental design of the present study and will be mainly discussed below. This, however, does not mean to deny the validity of other L2 models. The predictions of PAM-L2 and L2LP will be supplemented in the general discussion for a better understanding of the theoretical application of L2 phonetic learning in the current study. More concrete SLM-r predictions about Mandarin speakers' learning of Cantonese syllable-final segments will be specified after the comparison of the cross-Mandarin-Cantonese perceptual similarity of targets.

Chapter 3 Experiment 1: Perceptual Similarity of Syllable-final Segments between Mandarin and Cantonese

Experiment 1 aims at assessing the perceptual similarity of seven Cantonese syllable-final segments ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ], [-∅]) by Mandarin speakers through Perceptual Assimilation Test, based on which, the SLM-r predictions of Mandarin speakers' learning of above targets are proposed.

3.1 Methodology

3.2.1 Participants

With the approval of the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University, twenty Mandarin speakers were recruited after signing consent forms for their participation. Their personal information and language background were collected via a self-report questionnaire and summarized in Appendix 1.

According to the summary of the questionnaire, these twenty participants included 9 males and 11 females. Their age range was 18-30 years and their mean age was 22.85 ± 4.09 years. All of them were from northern China and using Mandarin as their first language. On average, they had been immersed in Hong Kong Cantonese for 12.9 ± 7.87 months without over-exposure to other Chinese dialects. As reported, all participants mainly used Mandarin in their daily lives, while they seldom or never used Hong Kong Cantonese. None of the participants had been exposed to a naturalistic English environment and some of them only used English in limited circumstances in Hong Kong (for example, in school for classes). None of the participants had a history of speaking, hearing, or language difficulties.

3.2.2 Stimuli and Manipulation

Cantonese CVC syllables and corresponding CV∅ syllables were targeted. In Cantonese phonology (Cheung, 2002), only [ɛ], [a] and [ɐ] can be collocated with all syllable-final consonants ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ]). Since vowels /ɛm/, /ɛn/, /ɛp/ and /ɛt/ can only be used in limited samples of loan words, onomatopoeic words and spoken language, such vowels were not be considered. Instead, four sets of tokens with /a/ or /ɐ/ as the nucleus and seven syllable-final segments as the coda were selected from ‘Chinese Character Database: With Word-formations

Phonologically Disambiguated According to the Cantonese Dialect'[1]. Set 1 included [tʃap˺ 3], [tʃat˺ 3], [tʃak˺ 3], [tʃam 33], [tʃan 33], [tʃaŋ 33] and [tʃa 33]. Based on Set 1, Set 2 was manipulated into different onset as [kap˺ 3], [kat˺ 3], [kak˺ 3], [kam 33], [kan 33], [kaŋ 33] and [ka 33]. Set 3 was with different nucleus as [tʃɐp˺ 3], [tʃɐt˺ 3], [tʃɐk˺ 3], [tʃɐm 33], [tʃɐn 33] and [tʃɐŋ 33] (no CVø syllables with nucleus [ɐ] in Cantonese). Set 4 was modified by different tone into [tʃap˺ 2], [tʃat˺ 2], [tʃak˺ 2], [tʃam 22], [tʃan 22], [tʃaŋ 22] and [tʃa 22]. All four sets of stimuli were employed to testify whether participants' responses to different targets would be affected by different contexts (different onsets, nucleus, and tone) of the syllable. According to Leung et al.'s (2004) report about the type and token frequency of linguistic units in Hong Kong Cantonese Adult Language Corpus (Leung & Law, 2001), these selected stimuli encountered by Mandarin speakers in their daily immerse of Cantonese are with different type frequencies ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81). This will be considered to identify the effect of type frequency of the targets on Mandarin speakers' performance.

These 27 tokens were produced by a 28-year-old male native Cantonese speaker without any known speaking, hearing, or language difficulties. Stimuli were recorded in a soundproof booth of the phonetics lab via an Audio-Technical AT2020 microphone. Every token was produced three times and the fittest one was

---

[1] Chinese Character Database: With Word-formations Phonologically Disambiguated According to the Cantonese Dialect was developed by The Chinese University of Hong Kong and can be retrieved from http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/.

selected for further normalization via Praat (Boersma & Weenink, 2018). All syllables were digitized at 22050 Hz (Tsukada et al., 2004) and 75 dB then normalized for peak intensity (99% of the full scale) and duration (mean duration of all syllables by the speaker: 390.46 ms). It should be noted that the syllabic length of the CVC syllable is relatively shorter than other syllable types (e.g., CV∅) in Cantonese (Zhu, Jiao, Yan, & Hong, 2008). After targeted CVC syllables and corresponding CV∅ syllables were standardized with the same duration, the sample speaker's pronunciation was still assessed as natural by two phoneticians who also used Hong Kong Cantonese as their first language. After selected stimuli were manipulated, they were employed in the experiment.

3.2.3 Experimental Apparatus and Procedure

In SLM (Flege, 1995) and SLM-r (Flege, 2021), L2 learners' proficiency in the L1 phonetic category could impact L2 speech learning. Although the employed Mandarin participants' L1 language backgrounds were generally well-controlled as reflected in the summarized results of the questionnaire, a quiz including Three-alternative Forced-choice Identification Test and Monosyllable Repetition Test was administered via PowerPoint to further gauge their proficiency in Mandarin syllable-final segments. Three Mandarin syllables ([p^han 55], [p^haŋ 55] and [p^ha 55]) produced by a 28-year-old male native Mandarin speaker were targeted as stimuli. In the identification test (3 targets * 10 repetitions = 30 trials), only one audio stimulus was played and participants were required to name the stimulus from three-alternative categories in each trial. The Pinyin representations of three categories were presented on the slide for 1500 ms. Meanwhile, the recording of the target stimulus was played. Participants were instructed to identify which of the Pinyin representations matched the audio signal. In the production test (3 targets * 10 repetitions = 30 trials), the recording of the target stimulus was played in each trial. Participants were instructed to repeat each of three target syllables at a normal speaking rate and a comfortable loudness level after they heard the recording. Participants' performances in this quiz were judged immediately by the examiner. All the employed participants achieved a full score in their L1 category proficiency.

According to Alispahic, Mulak, & Escudero (2017), Cebrian et al. (2019), Guion et al. (2000), Strange et al. (2004), Yang et al. (2020), and so on, the common procedure of Perceptual Assimilation Test encompasses the categorization of an L2 token in terms of L1 categories and the goodness-of-fit rating of the tested L1-L2 sound pairs. The current study adopted the same paradigm in Experiment 1 (see Appendix 2 for stimuli and trials of Experiment 1).

3.2.3.1 Four-alternative Forced-choice Categorization Test

In the Four-alternative Forced-choice Categorization Test, only one Cantonese audio stimulus was played for 391 ms (the approximate duration of the standardized syllables) in each trial. The Pinyin representations of three syllables with permissible Mandarin syllable-final segments ([-n], [-ŋ], and [-∅]) were presented on the screen for 1500 ms. These three syllables are minimal pairs different only in their syllable-final segments. Besides, the option of 'wu (none)' was adopted by some previous studies (Flege, 1991; Yang, 2020; Yang et al., 2020) to deal with the situation where the perceived L2 targets cannot be categorized as any provided L1 categories. In case participants would arbitrarily respond to the stimuli if they were forced to categorize an L2 target with clearly perceived differences as any provide L1 choices, the option of ‘wu (none)’ was presented as the fourth choice here. After listening to the recording, participants were instructed to categorize the Cantonese audio stimulus in terms of four-alternative choices by pressing button A, B, C or D (stood for syllable A [CVn], syllable B [CVŋ], syllable C [CV∅], or option ‘wu’ respectively). The above selected 27 tokens were repetitively tested four times. Therefore, there were a total of 108 trials (27 targets * 4 repetitions).

3.2.3.2 Goodness-of-fit Rating Test

In the Goodness-of-fit Rating Test, participants were required to decide the degree of similarity between the Cantonese audio stimuli and the displayed Mandarin syllables through a 5-point rating scale ranging from 1 (least similar) to 5 (very similar). In each trial, a Cantonese token was played for 391 ms and the Pinyin representation of a syllable with one of the three permissible Mandarin syllable-final segments ([-n], [-ŋ] or [-∅]) was presented on the screen for 1500 ms. This test entailed 81 trials (27 targets * 3 types of Mandarin syllables).

To activate the L1 phonological knowledge by Mandarin participants and yield a better measurement of perceptual similarity of targets between Mandarin and Cantonese, this Perceptual Assimilation Test was conducted in Mandarin. Tests in Experiment 1 were administered through E-prime 2.0 (Psychology Software Tools Inc., 2016) in a soundproof phonetic lab via a high-quality headphone (Sony MDR-

EX32LP). Participants made responses at a self-paced rate. After they completed a trial, the next trial was played automatically. The order of stimuli and tests was randomized.

3.2 Data Analysis

3.2.1 Statistical Analysis of Four-alternative Forced-choice Categorization Test

The collected categorization data were fit in a mixed-effects multinomial logistic regression model (Hedeker, 2003) using the 'multinom' function (Venables & Ripley, 2002) of 'nnet' package (Ripley & Venables, 2021) in R (R Core Team, 2018). Participants' keyboard responses (dependent variable) were analyzed with three fixed factors: Target ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅]; reference level = [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone; reference level = Set 1 syllable) and Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81; reference level = [-p˺]: 31) in this model.

3.2.2 Statistical Analysis of Goodness-of-fit Rating Test

The collected rating data were fit in an ordinal regression model using the 'ordinal' R package (Christensen, 2019) in R (R Core Team, 2018). With random effects from Item and Subject, the dependent variable (rating score) was analyzed with three fixed factors: Cantonese-Mandarin target pair (each of Cantonese syllable-final segments: [-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅] paired with each of Mandarin syllable-final segments: [-n], [-ŋ] and [-ø]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone) and Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81) in this model.

3.2.3 Statistical Analysis of Fit Index

The raw categorization data and rating data were transformed into mean categorization percentages and mean rating scores according to seven Cantonese syllable-final segments and related four Mandarin category choices. Following Guion et al. (2000), Bohn & Ellegaard (2019), and Cebrian et al. (2019), the fit index was calculated by multiplying the mean categorization percentages and mean rating scores.

3.3 Results

3.3.1 Results of Four-alternative Forced-choice Categorization Test

The mixed-effects multinomial logistic regression model for categorization data tested three factors: Target ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-ø]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone) and Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81). Results suggested significant contribution of Target [χ^2 (1) = 2006.25, $p < 0.001$] and Condition of syllable [χ^2 (1) = 300.891, $p < 0.001$] to categorization responses. No valid effects were identified from Type frequency of syllable [χ^2 (1) = -

5.104179e-05, p > 0.05]. The final model was summed up as Table 2 according to Hlavac (2018).

(Significant codes: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$)

	Dependent variable: categorization responses		
	[-n]	**[-ŋ]**	**Wu (none)**
Target: [-p̚]	0.150	0.970	2.624***
	(0.513)	(0.853)	(0.359)
Target: [-t̚]	1.226***	0.639	1.525***
	(0.394)	(0.853)	(0.371)
Target: [-k̚]	2.363***	2.056***	1.669***
	(0.371)	(0.770)	(0.377)
Target: [-m]	6.075***	8.133***	5.285***
	(0.511)	(0.812)	(0.515)
Target: [-n]	7.083***	6.940***	4.635***
	(0.504)	(0.819)	(0.528)
Target: [-ŋ]	19.893***	21.487***	17.693***
	(0.352)	(0.507)	(0.374)
Condition: Set 2	1.913***	2.131***	1.861***
	(0.238)	(0.281)	(0.223)
Condition: Set 3	-0.075	2.152***	1.660***
	(0.275)	(0.286)	(0.222)
Condition: Set 4	-0.060	0.556*	0.144

	(0.254)	(0.287)	(0.240)
Constant	-4.065***	-5.928***	-4.056***
	(0.381)	(0.749)	(0.374)
Akaike Inf. Crit.	3,487.009	3,487.009	3,487.009

Table 2. Statistical Model of Categorization Data in PAT

3.3.2 Results of Goodness-of-fit Rating Test

The ordinal regression model for rating data took Cantonese-Mandarin target pair (each of Cantonese syllable-final segments: [-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅] paired with each of Mandarin syllable-final segments: [-n], [-ŋ] and [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone) and Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81) as three fix effects. Results indicated that Cantonese-Mandarin target pair [χ^2 (1) = 91.943, p < 0.001] alone can significantly predict the dependent variable (rating data) with random effects of Item and Subject in the model, whereas no such contribution was spotted in Condition of syllable [χ^2 (1) = 1.984, p > 0.05] and Type frequency of syllable [χ^2 (1) = 0.656, p > 0.05]. See Table 3 for the final model. (Significant codes: ****p* < 0.001, ***p* < 0.01, **p* < 0.05)

Coefficients	**Estimate**	**SE**	***z***	***p***
Pair.L	1.595	0.319	4.995	< 0.001 ***
Pair.Q	-1.543	0.321	-4.806	< 0.001 ***
Pair.C	-1.198	0.318	-3.764	< 0.001 ***

Pair^4	-1.997	0.319	-6.262	< 0.001 ***
Pair^5	-0.457	0.317	-1.441	> 0.05
Pair^6	0.140	0.311	0.451	> 0.05
Pair^7	0.029	0.315	0.095	> 0.05
Pair^8	2.459	0.320	7.679	< 0.001 ***
Pair^9	-2.494	0.319	-7.817	< 0.001 ***
Pair^10	-1.723	0.319	-5.387	< 0.001 ***
Pair^11	1.115	0.314	3.551	< 0.001 ***
Random effects	**Variance**	**SD**		
Subject	0.338	0.58		
Item	0.208	0.456		

Table 3. Statistical Model of Rating Data in PAT

3.3.3 Results of Fit Index

According to seven Cantonese syllable-final segments and related four Mandarin category choices, the mean categorization percentages, rating scores, and fit indexes are calculated and summarized in Table 4.

Cantonese Category		**Mandarin Category**			
		[-n]	[-ŋ]	[-∅]	None
[-p˺]	Percent	2.19%	1.56%	58.75%	37.50%
	Rating	1	1	3.438	0
	Fit index	0.022	0.016	2.02	0
[-t˺]	Percent	8.44%	1.56%	72.81%	17.19%

	Rating	2	1	3.925	0
	Fit index	0.169	0.016	2.858	0
[-k̚]	Percent	19.38%	5%	60.31%	15.31%
	Rating	2	1	3.975	0
	Fit index	0.388	0.05	2.4	0
[-m]	Percent	25.31%	56.56%	2.50%	15.63%
	Rating	2	3.863	1	0
	Fit index	0.506	2.185	0.025	0
[-n]	Percent	66.25%	21.56%	2.50%	9.69%
	Rating	3.55	3	1	0
	Fit index	2.352	0.647	0.025	0
[-ŋ]	Percent	38.13%	55.94%	0%	5.94%
	Rating	3	3.738	1	0
	Fit index	1.144	2.091	0	0
[-∅]	Percent	4.17%	0.83%	90.83%	4.17%
	Rating	1	1	4.583	0
	Fit index	0.042	0.008	4.163	0

Table 4. Categorization Percent, Rating Similarity, and Fit Index in PAT

According to Tyler et al. (2014), Mokari & Werner (2017), and so on, an L2 token is categorized if it is assimilated to a given L1 category more than 50% of the time. The 50% categorization percentage is adopted as a cut-off standard to interpret Mandarin participants' perceptual assimilation pattern of Cantonese syllable-final segments. As can be seen from Table 4, Cantonese [-p˺], [-t˺] and [-k˺] were perceived as Mandarin [-∅] at 58.75%, 72.81%, and 60.31% with a mean similarity rating score at 3.438, 3.925 and 3.975 respectively. Mandarin participants assimilated Cantonese [-n] and [-ŋ] as Mandarin [-n] and [-ŋ] separately with a categorization percentage at 66.25% (similarity rating: 3.55) and 55.94% (similarity rating: 3.738). Around 56.56% of Cantonese [-m] were perceived as Mandarin [-ŋ] with a goodness-of-fit rating score at 3.863. The [-∅] between Mandarin and Cantonese was highly assimilated. Mandarin participants perceived Cantonese [-∅] as the Mandarin counterpart 90.83% of the time and scored them as the most similar pair (4.583).

Following Guion et al. (2000), Bohn & Ellegaard (2019), and Cebrian et al. (2019), fit index (multiplication of categorization rate and goodness-of-fit rating) is applied to decide the perceived similarity in the current study. If an L2 token represents a categorization percentage close to 100% and a goodness-of-fit rating approximates to 5, it is assumed that this L2 token is consistently accepted as a good exemplar of the corresponding L1 category. In this case, a good L2 exemplar to L1 categories will obtain a fit index around 5 (100% * 5). By contrast, those L2 tokens with relatively low fit indexes close to 0 will be considered as distant instances of L1 counterparts. According to Table 3, Cantonese [-∅] represented a reasonably good exemplar to Mandarin [-∅] with the highest fit index at 4.163. Cantonese [-p˺], [-t˺], and [-k˺] were accepted as moderately fit exemplars to Mandarin [-∅] with a fit index at 2.02, 2.858, and 2.4 separately. Similar moderately fit exemplars were represented by Cantonese [-m] to Mandarin [-ŋ] (2.185), Cantonese [-n] to Mandarin [-n] (2.352) and Cantonese [-ŋ] to Mandarin [-ŋ] (2.091).

3.4 Discussion

In SLM-r (Flege & Bohn, 2021), L2 sounds could be categorized as identical, similar, and new sounds relative to the closest L1 sounds. As discussed in Section 2.2, four measurement methods (the comparison of ① IPA symbol, ② acoustic similarity, ③ articulatory similarity, and ④ perceptual similarity) of cross-L1-L2 similarity are mainly adopted to decide the categorization of L2 sounds in the extant studies. However, no ready-made acoustic or articulatory features about the cross-Mandarin-Cantonese similarity of syllable-final segments are available from previous studies. Before further scrutinizing the actual perceptual pattern of Cantonese syllable-final segments by Mandarin speakers in Experiment 1, this section starts with a preliminary discussion about the cross-L1-L2 similarity by comparing the IPA transcription of Cantonese targets and their Mandarin counterparts.

According to Flege (1987 & 1988a & 1992), both identical L2 sounds and similar L2 sounds share the same IPA transcription with their L1 counterparts. Segments [m], [n], [ŋ], and [∅] are shared sounds between Mandarin and Cantonese. These four segments are transcribed by the same IPA symbols as their Mandarin counterparts and can be categorized as identical/similar sounds to Mandarin speakers. However, only the comparison of IPA symbols is not sufficient enough to further decide whether targets [m], [n], [ŋ], and [∅] are identical sounds or similar sounds. Additional evidence of acoustic, articulatory, or perceptual similarities is necessary. Targets [p̚], [t̚] and [k̚] perform position-sensitive articulation features in Cantonese. When these three tokens are in the syllable-final positions, they are pronounced without released burst and are acoustically less salient relative to their released syllable-initial counterparts [p], [t], and [k]. The broad or narrow transcription of targets can lead to different categorizations (Yang et al., 2020). In the case of narrow transcription, the IPA symbols of Cantonese [p̚], [t̚] and [k̚] are not used for any Mandarin categories. According to Flege (1987 & 1988a & 1992), a new L2 sound is transcribed by a dissimilar IPA symbol. These three Cantonese targets thus can be categorized as new sounds to Mandarin speakers. In the case of broad transcription, these three Cantonese sounds are transcribed as [p], [t] and [k], which could result in a misleading impression that they are identical/similar sounds relative to the Mandarin syllable-initial counterparts [p], [t] and [k]. However, such cross-syllable-position mapping is not necessarily transferred from syllable-initial to syllable-final positions (Flege & Davidian, 1984; Flege, McCutcheon, & Smith, 1987; Flege, 1988b; Flege, 1992).

As can be seen from the above discussion, there is uncertainty about the cross-language similarity of syllable-final segments between Mandarin and Cantonese when applying the metric of IPA symbols. More reliable indicators for the cross-L1-L2 similarity of targets are suggested instead of a provisional measure of IPA transcription.

Alternatively, the observed perceptual pattern in Experiment 1 of the current study provides different hints to gauge the similarity of syllable-final segments between Mandarin and Cantonese. The employed Mandarin speakers noticed the cross-Mandarin-Cantonese phonetic dissimilarity of seven Cantonese syllable-final segments in some cases. However, none of these seven targets were categorized as new sounds more than 50% of the time. Cantonese [-∅] can be classified as an identical target since Mandarin speakers consistently recognized Cantonese [-∅] as Mandarin [-∅]. Other six targets ([-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ]) were categorized as similar sounds.

This perceptual assimilation pattern of Cantonese syllable-final segments by Mandarin speakers underlays the SLM-hypotheses of Mandarin speakers' learning difficulties and learning outcomes towards those targets. It is suggested that identical target [-∅] between Cantonese and Mandarin refers to the least learning difficulties. With a positive L1-L2 transfer, Mandarin speakers would find it relatively easy to form the category of Cantonese [-∅] and achieve satisfying performance in the perception and production of this target. However, Mandarin speakers would confront much more challenges in learning similar sounds [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ]. These Cantonese sounds and their closest Mandarin counterparts would be assimilated into corresponding merged categories (Cantonese [-p˺], [-t˺], [-k˺] to Mandarin [-∅]; Cantonese [-m], [-ŋ] to Mandarin [-ŋ]; Cantonese [-n] to Mandarin [-n]). It is difficult for Mandarin speakers to extract the correct phoneme from such equivalence classifications, which prevents the successful formation of new categories and leads to unsatisfying perception and production of these similar targets.

Apart from the effect of cross-L1-L2 similarity on L2 phonetic learning, SLM and SLM-r also consider the influence of the perception-production relationship. The Co-evolve Hypothesis of SLM-r suggests that perception and production co-evolve in a bi-directional way without precedence in L2 learning (Flege & Bohn, 2021). Based on this hypothesis, it is predicted that the perception would interact with production during Mandarin speakers' learning of targets and one of these two modalities would not develop significantly before the other one. By contrast, the Perception-precedence Hypothesis of SLM proposes that successful perception will develop before satisfying production during the L2 learning procedure (Flege, 1995). It is then expected the perception would be achieved successfully before production as Mandarin speakers' development of targets processes. The investigation about Mandarin speakers' perception and production learning of Cantonese syllable-final segments helps to unveil the patterns of perception-production interface.

3.5 Summary

In this chapter, the perceived similarities of syllable-final segments between Mandarin and Cantonese by Mandarin speakers are discussed. The perceptual assimilation pattern suggests that Cantonese [-∅] was accepted as a good exemplar, while other targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] were considered as moderately fit exemplars towards their Mandarin counterparts. Based on these observed results about cross-L1-L2 similarity, SLM-r hypotheses about the learning difficulty and learning outcome of Cantonese syllable-final segments by Mandarin speakers are deduced. Cantonese [-∅] can be classified as an identical category to Mandarin [-∅] and Mandarin speakers would find it relatively easy to form the category of this sound. However, Mandarin speakers would find it difficult to learn similar sounds [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] as these Cantonese sounds and their closest Mandarin counterparts are assimilated into corresponding merged categories (Cantonese [-p̚], [-t̚], [-k̚] to Mandarin [-∅]; Cantonese [-m], [-ŋ] to Mandarin [-ŋ]; Cantonese [-n] to Mandarin [-n]). It is predicted that the successful learning of Cantonese [-∅] would be developed relatively earlier than that of Cantonese [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ]. As suggested by the Co-evolve Hypothesis of perception-production interface in SLM-r, Mandarin speakers' perception and production of Cantonese syllable-final segments would co-evolve in a bi-directional way without precedence during the learning procedure. Their perception of Cantonese syllable-final segments would develop almost concurrently and interactively with their production of these targets. On the other hand, the Perception-precedence Hypothesis of SLM suggests that Mandarin speakers' perception of targets would develop before production. Successful perception of seven Cantonese targets would be expected earlier than their production during Mandarin speakers' learning process. The above hypotheses are testified in the following chapters.

Chapter 4 Experiment 2: Pre-test

4.1 Methodology

4.1.1 Participants

Other forty-eight participants were recruited with consent and ethical approval in Experiment 2. One experimental group had thirty-two Mandarin speakers and one control group included sixteen Cantonese native speakers. According to the summary of their personal information and language background via a self-report questionnaire (see Appendix 1), the thirty-two Mandarin participants in the experimental group included 15 males and 17 females. Their age range was 19-30 years, and their mean age was 24.84 ± 2.7 years. All of them were born in northern China and used Mandarin as their native language. On average, they had been exposed to Hong Kong Cantonese for 9.48 ± 6.7 months without over-exposure to other Chinese dialects. These participants reported that they mainly used Mandarin in their daily lives, while they seldom or never used Hong Kong Cantonese. The sixteen native Cantonese speakers (6 males, 10 females; mean age ± SD: 22.75 ± 3.54 years; age range: 18 ~ 30 years) of the control group were born in Hong Kong and had been using Hong Kong Cantonese as their mother language. Participants in both groups reported limited use of English in school for classes in their daily language use. None of them had a history of speaking, hearing, or language difficulties.

4.1.2 Stimuli and Manipulation

To testify whether the observed perceptual pattern in PAT of Experiment 1 can successfully predict Mandarin speakers' perception and production of Cantonese syllable-final segments, stimuli, and manipulation in Experiment 1 were consistently adopted in Experiment 2 but randomly presented to participants in a new experimental procedure.

4.1.3 Experimental Apparatus and Procedure

The category proficiency in Mandarin syllable-final segments by 32 Mandarin participants was assessed via the same quiz used in Experiment 1 but presented to subjects in a new randomized order. These participants achieved an accuracy rate of 100% in this quiz. Then all 48 participants were instructed to focus on the coda of experimental syllables and trained to read the phonological representation (IPA transcription) of targets in a preliminary session. After they were familiarized with the procedure, they went through the formal test (see Appendix 3 for stimuli and trials of Experiment 2). Two tasks (Two-alternative Forced-choice Identification Test and Monosyllable Repetition Test) were utilized in the formal test to measure participants' perception and production abilities of targets respectively. The procedure of Experiment 2 was administered through E-prime 2.0 (Psychology Software Tools Inc., 2016) in a soundproof phonetic lab via a high-quality headphone (Sony MDR-EX32LP). Participants made responses at a self-paced rate. After they completed a trial, the next trial was played automatically. The stimuli and tests were randomly displayed for every subject.

4.1.3.1 Two-alternative Forced-choice Identification Test

Identification tests assess not only listeners' abilities to discern the acoustic differences of provided choices but also the internalized phonetic categories through listeners' decisions when they are required to label the stimuli (Strange & Shafer, 2008). It is suggested that this test reconstructs a way similar to the real-world language perception by listeners. The identification test, thus, is widely employed in the studies of speech perception. Based on the paradigm of the Two-alternative Forced-choice Identification Test in Wayland & Li (2008), only one audio stimulus was played and participants were required to identify the stimulus from two-alternative categories in each trial. The phonological representations (IPA transcription) of two categories in the form of "_ _ coda A" and "_ _ coda B" were presented on the screen for 1500 ms. Meanwhile, the recording of the target stimulus was played for 391 ms (the approximate duration of the standardized syllables). Participants were instructed to focus on the coda of experimental syllables and to identify which of the phonological representations matched the audio signal by pressing button A or B (stood for 'coda A' or 'coda B' respectively). The above selected 27 tokens formed up 78 minimal pairs for two-alternative categories in this test. Each one of the two tokens from the 78 minimal pairs served as the played signal alternately. Therefore, there were a total of 156 trials.

4.1.3.2 Monosyllable Repetition Test

An imitation test investigates not only speech perception and production but also the coordination between these two modalities through a repeat-after action

(Hao & de Jong, 2016). Some previous studies consider imitation as an important bridge between speech perception and production (Gambi & Pickering, 2013; Reiterer, Hu, Sumathi, & Singh, 2013; Nagle, 2019). Following Flege & Eefting (1987 & 1988), this test was adopted in the current study with an attempt to reveal the perception-production interface in Mandarin speakers' perception and production of Cantonese syllable-final segments. For Monosyllable Repetition Test, the recording of the target stimulus was played for 391 ms (the approximate duration of the standardized syllables) in each trial. Participants were instructed to repeat each of 27 target syllables at a normal speaking rate and a comfortable loudness level through a microphone after they heard the recording. The mouth-to-microphone distance was fixed at 5 cm. During this test, the subjects' pronunciations were recorded for further analysis.

4.1.4 Rating Scheme

For Monosyllable Repetition Test, a total of 1296 tokens (27 targets * 48 subjects) were recorded. All these 1296 tokens were played to two trained phoneticians randomly in a soundproof phonetic lab via a headphone (Sony MDR-EX32LP) using Praat (Boersma & Weenink, 2018). Participants' productions were scored as target-like production (1) or non-target-like production (0). To lower raters' bias against different participants, these two trained phoneticians conducted ratings without knowing participants' information of each produced token. The grading scores were settled after they obtained consent. Inter-rater reliability of two raters' rating scores was measured by the 'irr' R package (Gamer, Lemon, &

Fellows, 2007) in R (R Core Team, 2018). Two raters obtained a substantial agreement (Cohen's Kappa Coefficient = 0.806, $p < 0.001$) when rating participants' production in the Monosyllable Repetition Test independently. Following Ingham (2014), 70% of accuracy rate was adopted in the current study as an acquisition criterion to indicate participants' perception and production abilities of targets.

4.2 Data Analysis

Two models of between-group (experimental vs. control group) comparison were administered to compare two groups' accuracy data in the Identification Test and Monosyllable Repetition Test of Experiment 2 separately.

4.2.1 Statistical Analysis of Two-alternative Forced-choice Identification Test

For Identification Test, the collected raw identification data were transformed into a response of either 0 (wrong) or 1 (correct). The transformed identification scores then were fit in a mixed-effects model (Gries, 2009) using the 'lme4' R package (Bates et al., 2020) in R (R Core Team, 2018) for the between-group (experimental vs. control Group) comparison. This model included four factors: Target ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]; reference level = [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone; reference level = Set 1 syllable), Type frequency of syllable ([-p̚]: 31, [-t̚]: 60, [-k̚]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81; reference level = [-p̚]: 31) and Group (the experiment group and the control group; reference level = the control group) with random effects of Item and Subject. To testify the respective influence of the above four factors on identification accuracy, post hoc pairwise comparisons among different parameters of each fixed effect were conducted by releveling the reference level in the same model.

4.2.2 Statistical Analysis of Monosyllable Repetition Test

With potential random effects (Item and Subject), a mixed-effects model took Target ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅]; reference level = [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone; reference level = Set 1 syllable), Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81; reference level = [-p˺]: 31) and Group (the experiment group and the control group; reference level = the control group) as four fixed effects for the between-group (experimental vs. control Group) comparison of production data in Monosyllable Repetition Test. Post hoc tests were administered to compare pairwise parameters of each fixed factor separately.

4.3 Results

The results of statistical models and related post hoc tests of the Identification Test and Monosyllable Repetition Test are summarized and discussed in this section.

4.3.1 Results of Two-alternative Forced-choice Identification Test

The mixed-effects model for identification data tested four factors: Target ([-

p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-ø]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone), Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81) and Group (the experiment group and the control group) with random effects of Item and Subject. Results suggested significant contribution of Target [χ^2 (1) = 92.684, p < 0.001] to identification accuracy. No valid effects were identified from Condition of syllable [χ^2 (1) = 1.475, p > 0.05], Type frequency of syllable [χ^2 (1) = 0.146, p > 0.05] and Group [χ^2 (1) = 0.24, p > 0.05] in this model. See Table 5 for the final model.

(Significant codes: ****p* < 0.001, ***p* < 0.01, **p* < 0.05)

Fixed effects	**Estimate**	**SE**	***z***	***p***
(Intercept)	2.669	0.233	11.436	< 0.001***
Target: [-p˺]	-0.678	0.265	-2.560	< 0.01*
Target: [-t˺]	-1.706	0.259	-6.567	< 0.001***
Target: [-k˺]	-1.436	0.261	-5.507	< 0.001***
Target: [-m]	0.128	0.276	0.463	> 0.05
Target: [-n]	-0.739	0.265	-2.788	< 0.01**
Target: [-ŋ]	0.348	0.279	1.246	> 0.05
Random effects	**Variance**	**SD**		
Subject	0.581	0.762		
Item	0.469	0.685		

Table 5. Statistical Model of Two Groups' Identification Data in Pre-test

Two groups' different performances of seven targets in the identification test are drafted in Figure 1. Target segments are labeled on the horizontal axis and the percentage points of the vertical axis stand for the mean identification accuracy. The bar filled in light gray and dark gray refers to an average identification rate of a target by the experimental group and control group respectively. The mean ± standard error of each object is illustrated by the error bar.

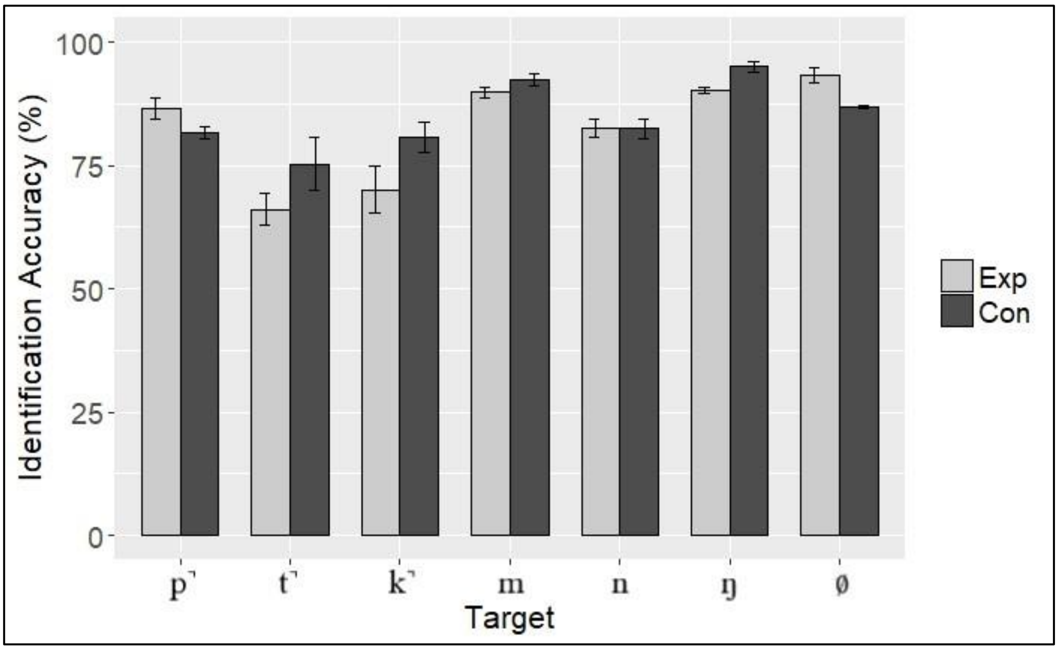

Figure 1. Two Groups' Identification Accuracy in Pre-test

As can be seen from Figure 1, Cantonese speakers' identification scores of all the targets went above 70%, with a relatively lower score in target targets [-t˺] (75%) and [-k˺] (80.71%). In the experimental group, the highest identification rate was achieved in target [-ø] (93.23%), [-m] (92.39%) and [-ŋ] (94.97%) followed by the figures for target [-p˺] and [-n] (86.41% and 82.61% respectively). Target [-t˺] and [-k˺] represented the lowest scores at 65.76% and 69.43%. Although two groups performed similarly across all the targets in the identification test, the control group surpassed the experimental group in targets [-t̚] and [-k̚] with a score over 70%.

4.3.2 Results of Monosyllable Repetition Test

The mixed-effects model for the production data took Target ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone), Type frequency of syllable ([-p̚]: 31, [-t̚]: 60, [-k̚]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81) and Group (the experiment group and the control group) as four fixed effects. Results indicated that Target [χ^2 (1) = 47.017, $p < 0.001$], Group [χ^2 (1) = 46.793, $p < 0.001$], and Target × Group interaction [χ^2 (1) = 187.56, $p < 0.001$] can significantly predict the dependent variable (production responses) with random effects of Item and Subject in the model, whereas no such contribution was spotted in Condition of syllable [χ^2 (1) = 1.569, $p > 0.05$] and Type frequency of syllable [χ^2 (1) = 1.264, $p > 0.05$]. See Table 6 for the final model.

(Significant codes: ****p* < 0.001, ***p* < 0.01, **p* < 0.05)

Fixed effects	**Estimate**	**SE**	**z**	**p**
(Intercept)	4.916	1.041	4.721	< 0.001***
Target: [-p̚]	-1.647	1.119	1.119	> 0.05
Target: [-t̚]	-2.741	1.081	-2.535	< 0.01*
Target: [-k̚]	-4.949	1.067	-4.64	< 0.001***
Target: [-m]	0.298	1.439	0.208	> 0.05

Target: [-n]	-2.669	1.082	-2.467	< 0.01*
Target: [-ŋ]	-4.138	1.067	-3.880	< 0.001***
Group: Experimental	-3.572	1.031	-3.465	< 0.001***
Target: [-p˺]×Group: Experimental	0.225	1.086	0.207	> 0.05
Target: [-t˺]×Group: Experimental	-0.934	1.059	-0.881	> 0.05
Target: [-k˺]×Group: Experimental	1.949	1.035	1.883	> 0.05
Target: [-m]×Group: Experimental	-0.879	1.413	-0.622	> 0.05
Target: [-n]×Group: Experimental	1.652	1.046	1.579	> 0.05
Target: [-ŋ]×Group: Experimental	4.145	1.033	4.012	< 0.001***
Random effects	**Variance**	**SD**		
Subject	0.613	0.783		
Item	0.216	0.464		

Table 6. Statistical Model of Two Group's Production Data in Pre-test

Participants' production performances are illustrated in Figure 2. The horizontal axis represents the targets, and the vertical axis stands for the mean production accuracy. With an error bar representing the mean ± standard error, each light gray bar and dark gray bar refers to an average production rate of a target by the experimental group and control group respectively.

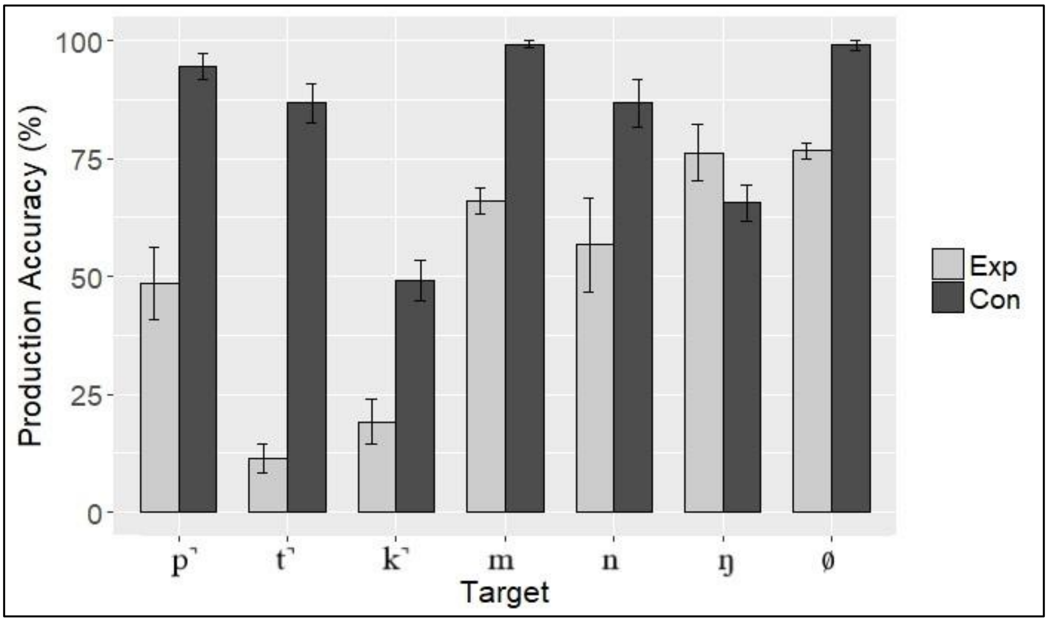

Figure 2. Two Groups' Production Accuracy in Pre-test

As can be seen from Figure 2, Cantonese native speakers' production scores in targets [-ø] (98.96%), [-m] (99.23%), [-p˺] (94.53%), [-n] (86.72%) and [-t˺] (86.72%) were all exceed 70%. By contrast, targets [-ŋ] (65.63%) and [-k˺] (49.22%) referred to least satisfying performances by the control group. With noticeably lower scores than the control group, the experimental group performed deficiency in the production of targets. Participants in this group had the best performance in targets [-ø] and [-ŋ] with an accuracy score at 76.56% and 76.17% separately. Following the figures of targets [-m] (66.02%), [-n] (56.64%) and [-p˺] (48.44%), targets [-t˺] and [-k˺] referred to the lowest production accuracy at around 11.33% and 19.14% respectively.

Except for target [-ŋ], between-group differences were found in other targets. The differences between two groups in target [-ø] was the minimum. The control group surpassed the experimental group with a score at 98.96% versus 76.56% ($p < 0.001$). The scores of the Cantonese group were around 30% higher than those of the Mandarin group in targets [-m] [control-experimental: 99.23%-66.02%, ($p < 0.001$)] and [-n] [control-experimental: 86.72%-56.64%, ($p < 0.001$)]. Two biggest differences in production scores were identified in targets [-p˺] [control-experimental: 94.53%-48.44%, ($p < 0.001$)] and [-t˺] [control-experimental: 86.72%-11.33%, ($p < 0.001$)]. Although the control group achieved the lowest score in target [-k˺], its accuracy rate was still higher than that of the experimental group [control-experimental: 49.22%-19.14%, ($p < 0.001$)].

4.4 Discussion

Together with sixteen Cantonese native speakers of the control group, thirty-two Mandarin participants of the experimental group were assessed in Experiment 2. As supported by the above results, the decisive role of cross-L1-L2 similarity in predicting Mandarin speakers' perception and production of Cantonese syllable-final segments is empirically verified. The observed coronal-dorsal merging tendency of Cantonese syllable-final segments is a by-product of this experiment as deduced from Cantonese participants' performance of targets.

Mandarin participants' perception and production of Cantonese syllable-final segment preliminarily verify that L1 experience has an impact on L2 learners' perception and production of L2 targets according to how these L2 targets are mapped into L1 existing categories. Turning back to the results of Experiment 1 about the perceptual assimilation pattern of Cantonese syllable-final segments by Mandarin speakers, Cantonese [-∅] were consistently categorized as an easier learnable identical segment and other six targets ([-p˺], [-t˺], [-k˺], [-m], [-n] and [-

ŋ]) were classified as more difficult similar sounds. In line with such SLM-r predictions, learnable identical target [-∅] obtained the best performance with an accuracy score over the adopted criterion at 70% in both perception and production. The similar targets [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ] bear more learning difficulties as predicted by the SLM-r hypotheses. Mandarin participants' perception of similar targets [-m], [-ŋ], [-n], and [-p˺] surpassed the adopted standard (70% of accuracy rate), while their perceptual abilities of targets [-t˺] and [-k˺] were still developing. In production, the accuracy scores of almost all the similar sounds ([-m], [-n], [-p˺], [-t˺] and [-k˺]) were below 70%. The unsatisfying performances in the perception and production of similar targets suggest that similar targets are relatively more difficult than identical target for Mandarin speakers, which generally lends support to the SLM-r predictions.

The merging tendency of syllable-final segments in Cantonese is documented by Bauer (1979), Yeung (1981), Chen (1999), Zee (1999a & 1999b), Law et al. (2001), Wong (2005), Ding (2010), Bauer & Benedict (2011), To et al. (2013), To et al. (2015) and so on. All these studies point to similar results that Cantonese syllable-final segments are undergoing a merger in the place of articulation (coronal-dorsal). Some studies further suggest that such coronal-dorsal merger presents an alveolarization tendency from dorsal segments to coronal segments (see Bauer, 1979; Wong, 2005; Bauer & Benedict, 2011; Yeung, 1981; Zee, 1999a; Chen, 1999; Ding, 2010). In the current study, Cantonese participants satisfied the adopted criteria (70% of accuracy rate) in the perception of all the targets but obtained a significantly lower score in targets [-t˺] and [-k˺]. Except for dorsal targets [-ŋ] and [-k˺], Cantonese participants' production scores were over 70% for other targets. As better performance was identified in coronal targets than dorsal targets in the production, these participants showed a remarkable preference for coronal targets ([-n] and [-t˺]) instead of the dorsal ones ([-ŋ] and [-k˺]) in this modality. The obtained results in Experiment 2 empirically supports the coronal-dorsal merger and the merging directionality from dorsal to coronal targets as observed by previous studies.

Until pre-test, the less satisfying performances of the similar targets than that of identical target generally verify the SLM-r predictions of the current study. However, it remains unclear how would the coronal-dorsal merging tendency of syllable-final segments in Cantonese affect Mandarin speakers' performances of these similar targets. Since Cantonese [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ] were all categorized as similar sounds by Mandarin speakers, it is expected that they should refer to comparative learning difficulties and would obtain similar learning outcomes after the same learning procedure. If Cantonese merging segments ([-t˺]-[-k˺] and [-n]-[-ŋ]) perform different developmental patterns from other similar targets, it would provide a chance to dig out the potential effects of Cantonese coronal-dorsal merging tendency on Mandarin speakers' learning of such targets. This issue remains to be discussed in the following chapters.

4.5 Summary

Chapter 4 measured the perception and production of Cantonese syllable-final segments ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-ø]) by Mandarin and Cantonese speakers. Based on Cantonese phonology (Cheung, 2002), four sets of CVC syllables and corresponding CVø syllables, manipulated with different onsets, nuclei, and tones, were selected as stimuli for the Two-alternative Forced-choice Identification Test and Monosyllable Repetition Test. Forty-eight participants from two groups (16 Cantonese participants in the control group and 32 Mandarin participants in the experimental group) were assessed via E-prime 2.0 (Psychology Software Tools Inc., 2016). The accuracy rates of perception and production tests went through statistical analysis in R (R Core Team, 2018). It revealed that identical target /-ø/ was satisfactorily perceived and produced by Mandarin speakers, while less learnable similar targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] posed more difficulties for Mandarin speakers in both perception and production. These findings answered the SLM-r hypotheses about the predicted patterns of identical targets and similar targets in L2 phonetic learning. Additionally, the performances in perception and production by Cantonese native speakers of the control group indicated the coronal-dorsal merging tendency of Cantonese syllable-final segments as proposed by previous studies. However, it remains to be discussed whether there is a potential effect of Cantonese coronal-dorsal merging tendency of syllable-final segments on Mandarin speakers' performances of similar targets. The upcoming results of perception training (Experiment 3) and post-test (Experiment 4) could provide hints for this issue.

Chapter 5 Experiment 3: Perception Training

5.1 Methodology

5.1.1 Participants

Sixteen native Mandarin speakers were randomly selected from those thirty-two Mandarin participants employed in Experiment 2 and served as the experimental group in the intervention of Experiment 3. These 16 participants include 6 males and 10 females whose age range was 20-29 years with a mean age of 24.5 ± 2.21 years. On average, they had been exposed to Hong Kong Cantonese for 10.03 months with a standard deviation of 6.09 months.

5.1.2 Stimuli and Manipulation

High Variability Phonetic Training (HVPT), as one of the most state-of-the-art methods in the field of phonetic training, has been widely applied in training-elicited speech learning. This training paradigm targets on stimuli with multiple features including different talkers, varying phonetic contexts, multiple tokens, and so on (Cebrian et al., 2019). Preceding studies demonstrate the effectiveness of HVPT on promoting phonetic learning (Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Iverson, Hazan, & Bannister, 2005). New evidence in more recent studies casts doubt on the efficiency of HVPT and points out that including high variability can sometimes hinder the learning of non-native speech (Antoniou & Wong, 2016; Fuhrmeister & Myers, 2017), especially for those weak-ability learners

(Perrachione, Lee, Ha, & Wong, 2011). The employed Mandarin-speaking participants in the intervention of the current study were inexperienced and weak-ability in Cantonese. To lower the cognitive load resulted from high variability in perception training on inexperienced learners (Antoniou & Wong, 2015), HVPT was not applied in the current study. Four sets of stimuli were employed in Experiment 2, the results indicated that participants' responses to different targets were not influenced by the frequency and the context of syllable (different onsets, nucleus, and tone). It can be assumed that selected stimuli were well-controlled and there were no significant differences among four sets of syllables. In this case, Set 1 stimuli ([tʃap˺ 3], [tʃat˺ 3], [tʃak˺ 3], [tʃam 33], [tʃan 33], [tʃaŋ 33] and [tʃa 33]) recorded by one talker (a 28-year-old male native Cantonese speaker) was randomly selected for the perception training of Experiment 3.

5.1.3 Experimental Apparatus and Procedure

Adopted from Dinnsen & Gierut (2008), seven hour-long sessions of perception training, which lasted for three weeks, were designed for the experimental group. Additionally, fourteen sub-tests were assigned before and after every training session respectively to better describe the learning procedure of training targets.

As summed up in Table 7, different tasks were utilized in each training session (see Appendix 4 for stimuli and trials of Experiment 3). Regarding training methodology, Sakai & Moorman (2018) and Shinohara & Iverson (2018) point out that the combination of discrimination and identification tasks can better facilitate participants' learning than adopting only one task. To better promote the experimental group's learning progress, Experiment 3 utilized both discrimination task and identification task in the training. In each session, there were the Pre-session Sub-test (Two-alternative Forced-choice Identification Test and Monosyllable Repetition Test), Discrimination Training, Identification Training, and the Post-session Sub-test (Two-alternative Forced-choice Identification Test and Monosyllable Repetition Test).

Test	**Session**	**Trial**
Identification Test	Pre-session Sub-test	42
Monosyllable Repetition Test		7
Discrimination Training	Training	168
Identification Training		168
Identification Test	Post-session Sub-test	42
Monosyllable Repetition Test		7

Table 7. Experimental Procedure of Intervention

5.1.3.1 Pre-session Sub-test

Pre-session Sub-test adopted the same methodology from Experiment 2 but with only Set 1 syllables as the stimuli.

In the Two-alternative Forced-choice Identification Test, participants were instructed to name the audio signal of stimulus from two categories (coda A and B) by pressing button A or B (stood for 'coda A' or 'coda B' respectively). In each trial, the phonological representations (IPA transcription) of two categories in the form of "_ _ coda A" and "_ _ coda B" were presented on the screen for 1500 ms. Meanwhile, the recording of the stimulus was played.

In Monosyllable Repetition Test, the audio signal of the target segment was played without phonological representation being displayed on the screen. Participants were instructed to repeat each of the 7 target syllables at a normal speaking rate and a comfortable loudness level through a microphone. The mouth-to-microphone distance was fixed at 5 cm. During this test, the subjects' pronunciations were recorded for further analysis.

5.1.3.2 Discrimination Training

The Discrimination Training was a forced-choice (same/different) discrimination task and consisted of 168 trials. Among them, 84 were test items (21 minimal pairs of different tokens and 21 pairs of the same tokens repeated twice respectively) and the other 84 trials were replay items. Upon listening to two tokens with 500 ms ISI, participants were required to press key 1 (for 'same') or 2 (for 'different') to indicate whether the two tokens were the same. Instant feedback in the format of 'Your answer is correct/wrong. These two tokens are the same/different.' was presented on the screen for 2000 ms after participants' responses. To enhance participants' perception of target contrast, every trial was compulsorily replayed after feedback. The procedure of Discrimination Training is demonstrated below:

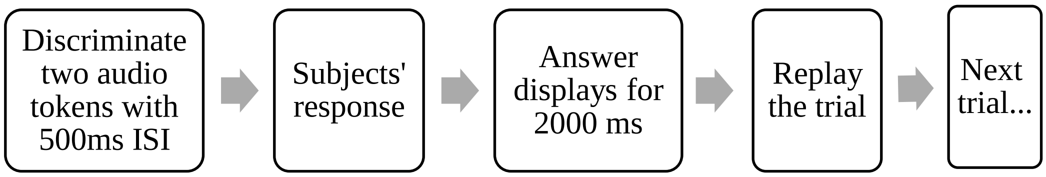

Figure 3. The Procedure of Discrimination Training in Intervention

5.1.3.3 Identification Training

In the Identification Training, there were 168 trials made up of 84 test items and 84 replay items. From two categories (coda A and coda B) displayed on the screen for 1500 ms, participants were instructed to name the one corresponding to the audio signal. Upon the participant's answer, feedback in the format of 'Your answer is correct/wrong. The correct answer in coda A/coda B.' was displayed on the screen for 2000 ms. In the follow-up playback, the audio signal was replayed with only the correct phonological representation shown on the screen for 1500 ms. The procedure of Identification Training is illustrated in Figure 4.

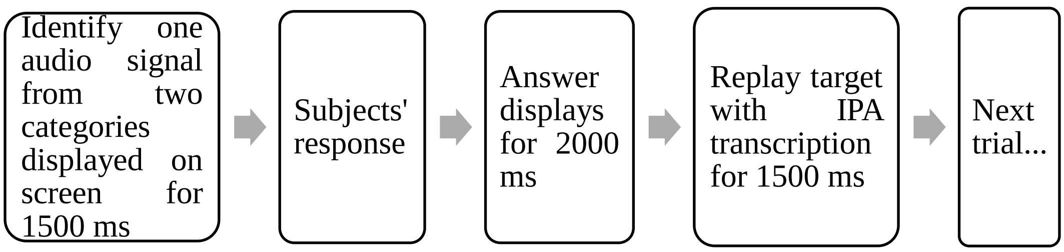

Figure 4. The Procedure of Identification Training in Intervention

5.1.3.4 Post-session Sub-test

To testify training effects before and after each session, the design of identical Pre-session Sub-test and Post-session Sub-test was adopted. Experimental stimuli and settings in the Post-session Sub-test were copied from those in the Pre-session

Sub-test but randomly presented to subjects.

The intervention was administered individually for every participant of the experimental group by using E-prime 2.0 (Psychology Software Tools Inc., 2016) in a soundproof phonetic lab via a high-quality headphone (Sony MDR-EX32LP). Participants made responses at a self-paced rate. After they completed a trial, the next trial was played automatically. The stimuli and tests were randomly displayed for every participant.

5.1.4 Rating Scheme

Only the responses in the Pre-session Sub-tests and the Post-session Sub-tests were collected for rating, while those in the Discrimination Training and Identification Training were not included. The collected raw identification data were transformed into a response of either 0 (wrong) or 1 (correct) for statistical analysis. For production tests, there were a total of 1568 tokens (7 targets*16 subjects*14 sub-tests) from the experimental group. All these 1568 tokens were independently rated by two trained phoneticians as target-like production (1) or non-target-like production (0). To lower raters' bias against different subjects and different tokens collected in different sessions, these two trained phoneticians conducted ratings without knowing the subject information and session number behind each token. The grades of tokens were settled after two raters obtained consent. By using the 'irr' R package (Gamer et al., 2007) in R (R Core Team, 2018), inter-rater reliability of two raters' rating scores in fourteen sub-tests was measured by Cohen's Kappa Coefficient Test. Their ratings in fourteen subtests were significantly correlated as shown in Table 8.

Session	Test	Cohen's Kappa Coefficient	*p*
1	Pre-session Sub-test	0.874	< 0.001
	Post-session Sub-test	0.861	< 0.001
2	Pre-session Sub-test	0.75	< 0.001
	Post-session Sub-test	0.72	< 0.001
3	Pre-session Sub-test	0.83	< 0.001
	Post-session Sub-test	0.806	< 0.001
4	Pre-session Sub-test	0.863	< 0.001
	Post-session Sub-test	0.814	< 0.001
5	Pre-session Sub-test	0.835	< 0.001
	Post-session Sub-test	0.658	< 0.001
6	Pre-session Sub-test	0.832	< 0.001
	Post-session Sub-test	0.806	< 0.001
7	Pre-session Sub-test	0.81	< 0.001
	Post-session Sub-test	0.757	< 0.001

Table 8. Inter-rater Reliability of Fourteen Sub-tests in Intervention

Following Ingham (2014), 70% of the accuracy rate was adopted as an acquisition indicator. If participants reach this criterion for a target in two successive sessions during the intervention, it is assumed that the learning of a target is settled.

5.2 Data Analysis

The identification and production accuracy collected in fourteen subtests reflected participants' development of targets in perception and production. The correlation test of identification scores and production scores revealed the perception-production interface.

With random effects of Subject and Item, fixed effects in the model of identification accuracy included Target ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-ø]; reference level = [-ø]), Type frequency of syllable ([-p̚]: 31, [-t̚]: 60, [-k̚]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81; reference level = [-p̚]: 31) and Session (Sub-test 1 to Sub-test 14; reference level = Sub-test 14). Follow-up comparisons within the parameters of each factor were conducted by releveling the reference level in the same model.

With random effects of Subject and Item, a mixed-effects model for the production accuracy took Target ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-ø]; reference level = [-ø]), Type frequency of syllable ([-p̚]: 31, [-t̚]: 60, [-k̚]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81; reference level = [-p̚]: 31), and Session (Sub-test 1 to Sub-test 14; reference level = Sub-test 14) as fixed effects. Post hoc tests were administered to compare between pairwise parameters of each factor.

For the correlation between identification and production, the mean identification percentages and the mean production percentages were analyzed through a non-parametric correlation test (Spearman's rank correlation coefficient) by using the cor. test (x, y, method="spearman") function in R (R Core Team, 2018).

According to Yeon (2004), raw scores of targets ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅]) in perception test and production test of each sub-test were converted into percentages for this analysis.

5.3 Results

5.3.1 Results of Two-alternative Forced-choice Identification Test during Intervention

Model for identification accuracy took Subject and Item as random effects. Fixed effects included Target ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅]), Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81) and Session (Sub-test 1 to Sub-test 14). With random effects of Subject and Item, the basic model was incrementally augmented after adding Target [$\chi^2(1) = 11.547$, $p < 0.1$] and Session [$\chi^2(1) = 234.78$, $p < 0.001$]. There is no significant contribution from the Type frequency of syllable [$\chi^2(1) = 0.1646$, $p > 0.05$] to this model. See Table 9 for the final model.

(Significant codes: ****p* < 0.001, ***p* < 0.01, **p* < 0.05)

Fixed effects	Estimate	SE	z	p
(Intercept)	3.281	0.543	6.046	< 0.001***
Target: [-p˺]	-0.681	0.670	-1.016	> 0.05
Target: [-t˺]	-1.700	0.658	-2.582	< 0.01**
Target: [-k˺]	-1.818	0.657	-2.767	< 0.01**
Target: [-m]	-0.880	0.668	-1.318	> 0.05

Target: [-n]	-0.884	0.669	-1.322	> 0.05
Target: [-ŋ]	-0.298	0.681	-0.437	> 0.05
Session: 1	-2.252	0.290	-7.763	< 0.001***
Session: 2	-1.677	0.299	-5.617	< 0.001***
Session: 3	-1.307	0.307	-4.255	< 0.001***
Session: 4	-0.847	0.322	-2.627	< 0.01**
Session: 5	-0.883	0.321	-2.751	< 0.01**
Session: 6	-0.470	0.340	-1.382	> 0.05
Session: 7	-0.421	0.343	-1.226	> 0.05
Session: 8	-0.421	0.343	-1.226	> 0.05
Session: 9	-0.069	0.366	-0.191	> 0.05
Session: 10	-0.199	0.357	-0.556	> 0.05
Session: 11	0.332	0.402	0.826	> 0.05
Session: 12	0.074	0.378	0.197	> 0.05
Session: 13	0.332	0.402	0.826	> 0.05
Random effects	**Variance**	**SD**		
Subject	0.602	0.776		
Item	1.027	1.014		

Table 9. Statistical Model of Identification Data in Intervention

The development of perception accuracy during intervention is depicted below. In Figure 5, the horizontal axis labels the sequence of sub-tests and the vertical axis stands for the identification accuracy. Each line refers to the identification rate of a target.

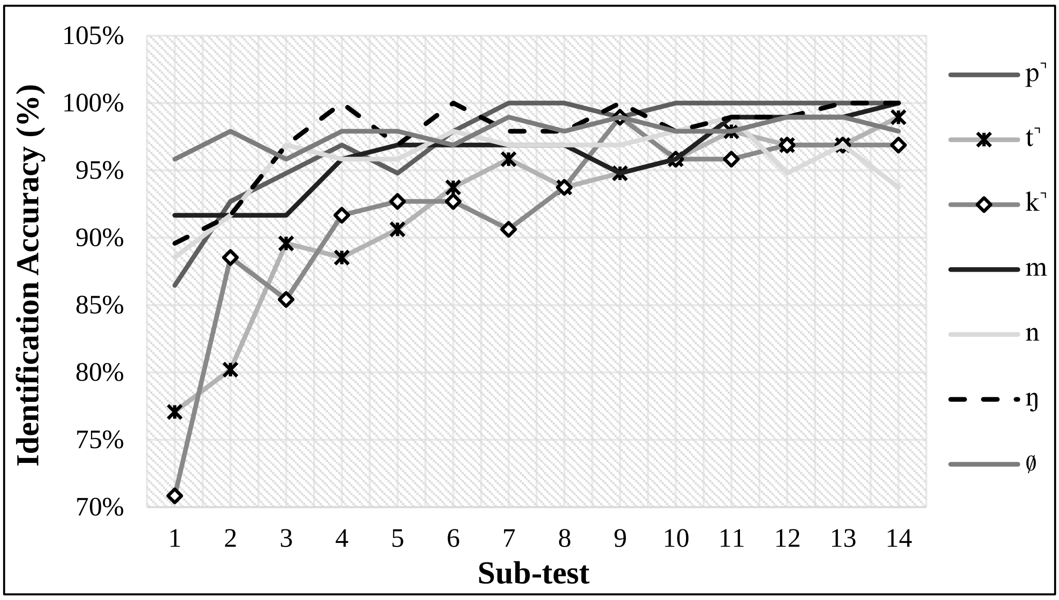

Figure 5. Development of Identification Accuracy in Intervention

It is assumed that the learning of a target is successful if participants meet the basic criterion (an accuracy rate of 70%) for this target in two successive sessions (Dinnsen & Gierut, 2008). As can be seen in Figure 5, the perception of seven targets satisfied this standard in Sub-test 2 after the first session of training. Among these targets, target [-∅] maintained the best identification score across all these fourteen sub-tests. Starting with the highest score at 95.83% in Sub-test 1, its figure ranged between 95.83% and 98.96% without a significant between-session difference. On the other hand, targets [-m], [-n], [-ŋ], and [-p˺] all started with an accuracy rate of around 90% in the first sub-test and performed a significantly increasing trend. Their perception scores stayed above 90% after Sub-test 2. It is worth noting that the figures for targets [-t˺] and [-k˺] were the lowest compared with other targets in the first four sub-tests. Precisely, these two targets both initiated their perceptual development with an accuracy score of around 70%. Their identification scores experienced a noticeable increase to around 90% in Sub-test 4 then maintained above 90% for the rest of 10 sub-tests.

5.3.2 Results of Monosyllable Repetition Test during Intervention

With random effects of Subject and Item, the mixed-effects model of the production accuracy took Target ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-ø]), Session (Sub-test 1 to Sub-test 14) and Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81) as fixed effects. Both Target [χ^2 (1) = 30.735, $p < 0.001$] and Session [χ^2 (1) = 68.713, $p < 0.001$] can effectively predict the production accuracy. There was no contribution from Type frequency of syllable [χ^2 (1) = 1.82, $p > 0.05$] to this model. See Table 10 for the final model.

(Significant codes: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$)

Fixed effects	Estimate	SE	z	p
(Intercept)	2.840	0.503	5.642	< 0.001***
Target: [-p˺]	-2.942	0.425	-6.922	< 0.001***
Target: [-t˺]	-4.200	0.426	-9.856	< 0.001***
Target: [-k˺]	-3.451	0.424	-8.145	< 0.001***
Target: [-m]	-2.321	0.431	-5.386	< 0.001***
Target: [-n]	-2.014	0.436	-4.619	< 0.001***
Target: [-ŋ]	-2.968	0.425	-6.985	< 0.001***
Session: 1	-2.101	0.372	-5.642	< 0.001***
Session: 2	-1.422	0.374	-3.805	< 0.001***

Session: 3	-1.687	0.372	-4.534	< 0.001***
Session: 4	-0.850	0.382	-2.225	< 0.05*
Session: 5	-1.422	0.374	-3.805	< 0.001***
Session: 6	-0.850	0.382	-2.225	< 0.05*
Session: 7	-1.367	0.374	-3.655	< 0.001***
Session: 8	-0.911	0.381	-2.392	< 0.05*
Session: 9	-0.725	0.385	-1.884	> 0.05
Session: 10	-0.850	0.382	-2.225	< 0.05*
Session: 11	-0.595	0.388	-1.533	> 0.05
Session: 12	-0.241	0.400	-0.602	> 0.05
Session: 13	-0.661	0.387	-1.710	> 0.05
Random effects	**Variance**	**SD**		
Subject	0.953	0.976		
Item	1.330	1.153		

Table 10. Statistical Model of Production Data in Intervention

Participants' development of seven targets in production during the intervention is illustrated in Figure 6. Fourteen sub-tests are represented by the numbers on the horizontal axis and the production accuracy in percentage is indexed on the vertical axis. Each line refers to the production rate of a target.

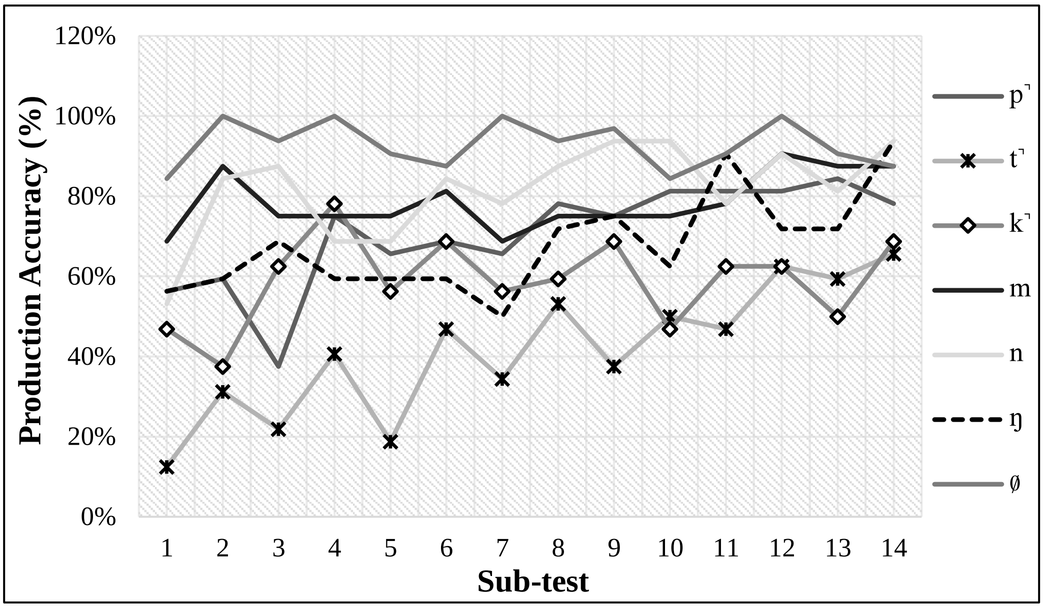

Figure 6. Development of Production Accuracy in Intervention

According to Figure 6, all the tested targets generally showed a gradual increasing tendency in their production scores. It is assumed that the learning of a target is successful if an accuracy rate of 70% is observed in two successive sessions for this target (Dinnsen & Gierut, 2008). Indicated by this standard, Mandarin participants achieved successful production learning of seven targets in different sub-tests as intervention processed.

Compared with other targets, target [-∅] represented the highest production accuracy during these fourteen sub-tests. Ranging from 84.38% to 100%, the figures for target [-∅] performed no significant between-session differences. With an accuracy rate at 84.38% and 100% observed in Sub-test 1 and 2 respectively, its production development was completed in Sub-test 2.

Afterward, the production of targets [-m] and [-n] was successfully obtained in Sub-test 3. Started with an accuracy score at 68.75% ([-m]) and 53.13% ([-n]) in

Sub-test 1, the production figures for these two targets witnessed a significant increase and maintained above the adopted criterion (an accuracy rate of 70%) in Sub-test 2 ([-m]: 87.5%; [-n]: 84.38%) and Sub-test 3 ([-m]: 75%; [-n]: 87.5%). Two more targets ([-ŋ] and [-p˺]) were developed later in Sub-test 9. Precisely, the production accuracy of target [-ŋ] fluctuated between 50% and 68.75% from Sub-test 1 to Sub-test 7. A significant increase in production score to 71.88% and 75% was obtained in Sub-test 8 and 9 for this target. Target [-p˺] experienced similar fluctuation in its production score before it realized an accuracy rate above 70% in Subtest 8 (78.13%) and Sub-test 9 (75%).

By contrast, targets [-t˺] and [-k˺] referred to the lowest score in most of the sub-tests. Although there was a significant increasing tendency in the production development of targets [-t˺] and [-k˺], these two targets were not able to maintain a production accuracy rate above 70% in two successive sub-tests. Mandarin participants failed to accomplish the production learning of targets [-t˺] and [-k˺] until the end of the intervention.

5.3.3 Results of Perception-production Correlation during Intervention

As can be seen from Table 11, raw scores of tested targets ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-ø]) in perception test and production test of each sub-test were converted into percentages for non-parametric correlation analysis.

Identification Percentages							
Sub-test	[-p˺]	[-t˺]	[-k˺]	[-m]	[-n]	[-ŋ]	[-∅]
1	86.46%	77.08%	70.83%	91.67%	88.54%	89.58%	95.83%
2	92.71%	80.21%	88.54%	91.67%	91.67%	91.67%	97.92%
3	94.79%	89.58%	85.42%	91.67%	96.88%	96.88%	95.83%
4	96.88%	88.54%	91.67%	95.83%	95.83%	100%	97.92%
5	94.79%	90.63%	92.71%	96.88%	95.83%	96.88%	97.92%
6	97.92%	93.75%	92.71%	96.88%	97.92%	100%	96.88%
7	100%	95.83%	90.63%	96.88%	96.88%	97.92%	98.96%
8	100%	93.75%	93.75%	96.88%	96.88%	97.92%	97.92%
9	98.96%	94.79%	98.96%	94.79%	96.88%	100%	98.96%
10	100%	95.83%	95.83%	95.83%	97.92%	97.92%	97.92%
11	100%	97.92%	95.83%	98.96%	98.96%	98.96%	97.92%
12	100%	96.88%	96.88%	98.96%	94.79%	98.96%	98.96%
13	100%	96.88%	96.88%	98.96%	96.88%	100%	98.96%
14	100%	98.96%	96.88%	100%	93.75%	100%	97.92%
Production Percentages							
Sub-test	[-p˺]	[-t˺]	[-k˺]	[-m]	[-n]	[-ŋ]	[-∅]
1	56.25%	12.5%	46.88%	68.75%	53.13%	56.25%	84.38%
2	59.38%	31.25%	37.5%	87.5%	84.38%	59.38%	100%
3	37.5%	21.88%	62.5%	75%	87.5%	68.75%	93.75%
4	75%	40.63%	78.13%	75%	68.75%	59.38%	100%

5	65.63%	18.75%	56.25%	75%	68.75%	59.38%	90.63%
6	68.75%	46.88%	68.75%	81.25%	84.38%	59.38%	87.5%
7	65.63%	34.38%	56.25%	68.75%	78.13%	50%	100%
8	78.13%	53.13%	59.38%	75%	87.5%	71.88%	93.75%
9	75%	37.5%	68.75%	75%	93.75%	75%	96.88%
10	81.25%	50%	46.88%	75%	93.75%	62.5%	84.38%
11	81.25%	46.88%	62.5%	78.13%	78.13%	90.63%	90.63%
12	81.25%	62.5%	62.5%	90.63%	90.63%	71.88%	100%
13	84.38%	59.38%	50%	87.5%	81.25%	71.88%	90.63%
14	78.13%	65.63%	68.75%	87.5%	93.75%	93.75%	87.5%

Table 11. Identification and Production Percentages in Intervention

According to the result of non-parametric correlation test (Spearman’s rank correlation coefficient) via cor.test (x, y, method="spearman") function in R (R Core Team, 2018), there was a moderately positive correlation ($\rho = 0.506$, $p < 0.01$) between identification percentages and production percentages for targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ]. As participants' identification accuracy of these targets increased during the fourteen sub-tests, their production accuracy went up accordingly. By contrast, no significant perception-production correlation ($\rho = 0.505$, $p > 0.05$) was detected for target [-∅].

5.4 Discussion

This chapter discusses Experiment 3 to testify the aforementioned SLM-r predictions about Mandarin speakers' learning of Cantonese syllable-final segments, the perception-production interface during the learning procedure of targets, and the effectiveness of perception-only training.

Recalling the results of Experiment 1 about the perceptual assimilation pattern of Cantonese syllable-final segments by Mandarin speakers, Cantonese [-∅] were consistently categorized as an identical target with the least learning difficulties and other six targets ([-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ]) were classified as similar sounds which would pose bigger learning challenge. It is thus hypothesized that Cantonese [-∅] would be easier and earlier acquired than [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ]. Mandarin participants' development of Cantonese syllable-final segments in Experiment 3 indicates the influence of L1 experience on L2 phonetic learning and verifies such hypotheses as predicted by Experiment 1. It is assumed that the learning of a target is successful if participants meet the basic criterion (an accuracy rate of 70%) for this target in two successive sessions (Dinnsen & Gierut, 2008). After the first session of training, identical target [-∅] was firstly learned in both perception and production in Sub-test 2. Although six similar targets ([-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ]) were perceptually acquired in Sub-test 2, their production development was completed relatively later. Two similar sounds [-m] and [-n] obtained successful production in Sub-test 3, after which similar targets [-ŋ] and [-p˺] were developed in Sub-test 9. However, Mandarin participants failed to accomplish satisfying production of similar sounds [-t˺] and [-k˺] until the end of the intervention. Generally, identical target ([-∅]) with less learning difficulties was developed earlier than those less learnable similar sounds ([-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ]) as predicted by SLM-r.

Although SLM-r generally predicts the differences in the learning difficulties between identical and similar targets of Cantonese syllable-final consonants, it fails to clarify why Cantonese [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ] are all classified as similar sounds but targets [-t˺] and [-k˺] are more difficult than other similar targets for Mandarin speakers. As observed in Experiment 3, targets [-t˺] and [-k˺] represented the lowest perception accuracy and failed to achieve production learning until the end of the intervention. Beyond the SLM-r predictions, such exception could point to potential effects other than the influence of L1 experience on L2 phonetic learning. This will be tackled in the general discussion after further results of Experiment 4 in the next chapter.

Responding to divided opinions of perception-production relationship in previous studies, the production-precedence pattern (Goto, 1971; Sheldon & Strange, 1982; McCandliss et al., 2002; Hattori & Iverson, 2009; Ingvalson et al., 2011), the perception-precedence pattern of SLM (Flege, 1995), and the perception-production co-evolve pattern of SLM-r (Flege & Bohn, 2021) are proposed. Instead of these one-size-fits-all patterns, the observed results of Experiment 3 point to a diverse perception-production relationship varying according to the learnability of different targets in L2 phonetic learning. For less learnable similar targets [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ], their perception interacted with production following a perception-precedence pattern. Successful perception (all similar targets in

Subtest 2) was achieved in advance to guide satisfying production ([-m] and [-n] in Subtest 3; [-ŋ] and [-p̚] in Sub-test 9) during the development of these targets. Indicated by a moderately positive correlation ($\rho = 0.506$, $p < 0.01$), participants' production accuracy increased along with their identification accuracy. As testified in the studies by Hao & de Jong (2016), Thorin et al. (2018), and Casillas (2020), the findings of the current study again empirically verify the Perception-precedence Hypothesis of SLM (Flege, 1995) that perception develops before production during the L2 phonetic learning.

On the other hand, the learning of identical target [-∅] was completed at the beginning of the intervention (Sub-test 2). The perception and production accuracy of this target maintained the ceilings across the intervention and performed no significant between-session differences. Therefore, no significant correlation ($\rho = 0.505$, $p > 0.05$) was found between the perception and production of the identical target [-∅]. Although it is difficult to decide which of the proposed patterns (production-precedence pattern, perception-precedence pattern or perception-production co-evolve pattern) has emerged in the development of identical target [-∅], the diverse perception-production relationship captured by Experiment 3 reveals that flexibility according to the learnability of different L2 targets should be considered when interpreting the interface between perception and production of L2 phonetic learning.

The noticeable increasing trend was realized in fourteen sub-tests of the Identification Test, which indicates the effectiveness of perception training on

Mandarin participants' perceptual development of Cantonese syllable-final segments. The effectiveness of perception-only training can be further generalized to Mandarin participants' production development. Apart from perception growth, participants' production abilities of targets also improved gradually during the intervention. These findings acknowledge the contribution of perception training on L2 learner's learning of segmental units as discussed by McCandliss et al. (2002), Handley, Sharples, & Moore (2009), Baese-Berk (2010 & 2019), Anderson (2011), Shinohara & Iverson (2018) and so on.

5.5 Summary

Chapter 5 described sixteen Mandarin participants' learning procedure of Cantonese syllable-final segments ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]) through seven sessions of perception training. Based on Cantonese phonology (Cheung, 2002), stimuli including [tʃap̚ 3], [tʃat̚ 3], [tʃak̚ 3], [tʃam 33], [tʃan 33], [tʃaŋ 33] and [tʃa 33] were utilized for intervention. Before and after each training session respectively, there was one sub-test covering the Identification Test and Monosyllable Repetition Test. Also, Discrimination Training and Identification Training with instant feedback and compulsory replay were included in training sessions. From sixteen Mandarin participants' responses to fourteen sub-tests during the intervention, the effect of L1 experience on L2 phonetic learning was first discussed. In line with the SLM-r hypotheses, identical target ([-∅]) with less learning difficulties was developed earlier than those predicted more difficult similar sounds ([-m], [-n], [-ŋ], [-p̚], [-t̚] and [-k̚]). However, SLM-r fails to clarify why Cantonese [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ] are all classified as similar sounds but [-t˺] and [-k˺] are more difficult than other similar targets for Mandarin speakers. This issue will be brought up again in the general discussion. As intervention progressed, the perception-production relationship performed a diverse pattern and varied according to different learning targets. The learning of similar targets ([-m], [-n], [-ŋ], [-p˺], [-t˺] and [-k˺]) performed a perception-precedence pattern, while no significant correlation ($\rho = 0.505$, $p > 0.05$) was found between the perception and production of identical target [-∅]. The gradual improvement in perception and production by Mandarin participants during the training reveals the effectiveness of perception-only training on facilitating L2 phonetic learning.

Chapter 6 Experiment 4: Post-test

6.1 Methodology

6.1.1 Participants

Among those thirty-two Mandarin participants in Experiment 2, sixteen participants were randomly selected as the experimental group for the intervention in Experiment 3. The rest sixteen participants were employed again as the control group and assessed together with the experimental group in Experiment 4. According to the survey of language background and language use (see Appendix 1), the experimental group (6 males, 10 females; mean age ± SD: 24.5 ± 2.21 years; age range: 20-29 years) with intervention had been exposed to Hong Kong

Cantonese for 10.03 ± 6.09 months. Other sixteen native Mandarin speakers (9 males, 7 females; mean age ± SD: 25.2 ± 3.01 years; age range: 19-30 years) were assigned to the control group without intervention. They had been in Hong Kong for 8.94 ± 7.02 months. No significant differences were identified from the language background survey between these two groups in gender ($W = 104$, $p > 0.05$), age ($W = 97$, $p > 0.05$), and the years of exposure to Hong Kong Cantonese ($W = 151$, $p > 0.05$).

6.1.2 Stimuli and Manipulation

To testify training effectiveness and compare the experimental group (with intervention) with the control group (without intervention), the design of identical Experiment 2 (served as the pre-test) and Experiment 4 (served as the post-test) was adopted. Experimental stimuli and settings in Experiment 4 were copied from Experiment 2 but presented to subjects in a new randomized order.

6.1.3 Experimental Apparatus and Procedure

Experiment 4 was administered through E-prime 2.0 (Psychology Software Tools Inc., 2016) in a soundproof phonetic lab via a high-quality headphone (Sony MDR-EX32LP). The whole procedure included the Two-alternative Forced-choice Identification Test and Monosyllable Repetition Test.

In the Two-alternative Forced-choice Identification Test, the phonological representations (IPA transcription) of tokens A and B were presented on the screen for 1500 ms in the format of "_ _ coda A (token A)" and "_ _ coda B (token B)". At the same time, the recording of token X was played. Participants were instructed to focus on the coda of experimental syllables then identify which phonological representations matched with the audio signal by pressing button A or B (stood for ‘token A’ or ‘token B’ respectively).

In Monosyllable Repetition Test, participants were instructed to repeat each of 27 target syllables at a normal speaking rate and a comfortable loudness level through a microphone.

6.1.4 Rating Scheme

The collected raw identification data of Experiment 4 were transformed into a response of either 0 (wrong) or 1 (correct) for statistical analysis. For production tests, there were a total of 432 tokens (27 targets*16 subjects) from the experimental group and 432 tokens (27 targets*16 subjects) collected in the control group. These 864 tokens were graded as target-like production (1) or non-target-like production (0) by two trained phoneticians independently. To lower raters’ bias against subjects from different groups and tests, these two trained phoneticians conducted ratings without knowing the subject and test information of each token. According to the result of inter-rater reliability measured by Cohen’s Kappa Coefficient Test via ‘irr’ R package (Gamer et al., 2007) in R (R Core Team, 2018), two raters achieved significant agreement (Cohen’s Kappa Coefficient = 0.662, $p < 0.001$) in their grading.

6.2 Data Analysis

6.2.1 Statistical Analysis of Two-alternative Forced-choice Identification Test

The identification data was analyzed via both between-group (experimental vs. control group) comparison and within-group (pre-test vs. post-test) comparison.

To compare two groups' performances in the identification test of Experiment 4, the identification data first went through a model of between-group (experimental vs. control group) comparison. The transformed scores (0 or 1) of identification data were fit in a mixed-effects model (Gries, 2009) using the 'lme4' R package (Bates et al., 2020) in R (R Core Team, 2018). This model included four factors: Target ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]; reference level = [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone; reference level = Set 1 syllable), Type frequency of syllable ([-p̚]: 31, [-t̚]: 60, [-k̚]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81; reference level = [-p̚]: 31) and Group (the experiment group and the control group; reference level = the experimental group) with random effects from Item and Subject. To testify the respective influence from the above four factors on identification accuracy, post hoc pairwise comparisons among different parameters of each fixed effect were conducted by releveling the reference level in the same model.

Identification accuracy data of Experiment 2 (as the pre-test) and Experiment 4 (as the post-test) also went through two models (1 test * 2 groups) of within-group (pre-test vs. post-test) comparison. Different performances in Experiment 2 (as the pre-test) and Experiment 4 (as the post-test) by the experimental group indicated the intervention-elicited learning, while the same comparison in the control group reflected the natural development during the same period. In these two models of identification accuracy (one for the experimental group and one for the control group separately), Subject and Item contributed as random effects. Fixed effects included Target ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]; reference level = [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone; reference level = Set 1 syllable), Type frequency of syllable ([-p̚]: 31, [-t̚]: 60, [-k̚]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-∅]: 81; reference level = [-p̚]: 31) and Session (pre-test, post-test; reference level = post-test). Follow-up pairwise comparisons in the parameters of each factor were administered.

6.2.2 Statistical Analysis of Monosyllable Repetition Test during Intervention

The production data were analyzed via both between-group (experimental vs. control group) comparison and within-group (pre-test vs. post-test) comparison.

Firstly, two groups' production performances in the Monosyllable Repetition Test of Experiment 4 were assessed via a model of between-group (experimental vs. control group) comparison. With random effects from Item and Subject, the mixed-effects model for the production data took Target ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]; reference level = [-∅]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone; reference level = Set 1 syllable), Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81; reference level = [-p˺]: 31), and Group (the experiment group and the control group; reference level = the experimental group) as four fixed effects. Post hoc tests were administered to compare pairwise parameters of each fixed factor separately.

Secondly, the within-group (pre-test vs. post-test) comparisons of production data from Experiment 2 (as the pre-test) and Experiment 4 (as the post-test) were administered to indicate the intervention-elicited learning by the experimental group and the natural development by the control group. With the random effects of Subject and Item, two mixed-effects models for the production data were assigned to two groups separately. The model took Target ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-ø]; reference level = [-ø]), Condition of syllable (Set 1 syllable, Set 2 syllable with different onset, Set 3 syllable with different nucleus and Set 4 syllable with different tone; reference level = Set 1 syllable), Type frequency of syllable ([-p˺]: 31, [-t˺]: 60, [-k˺]: 54, [-m]: 64, [-n]: 118, [-ŋ]: 49 and [-ø]: 81; reference level = [-p˺]: 31) and Session (pre-test, post-test; reference level = post-test) as fixed effects. Post hoc tests were administered to compare between pairwise parameters of each factor.

6.3 Results

6.3.1 Results of Two-alternative Forced-choice Identification Test

6.3.1.1 Between-group (Experimental vs. Control group) Comparison

To compare two groups' performances in the identification test of Experiment 4, the mixed-effects model for between-group (experimental vs. control group) comparison assessed four factors: Target, Condition of syllable, Type frequency of syllable and Group with random effects from Item and Subject. Significant contribution of Target [χ^2 (1) = 46.206, p < 0.001], Group [χ^2 (1) = 13.003, p < 0.01] and Target × Group interaction [χ^2 (1) = 39.108, p < 0.001] was identified in the results, whereas no noticeable effects were found from Condition of syllable [χ^2 (1) = 1.094, p > 0.05] and Type frequency of syllable [χ^2 (1) = 0.784, p > 0.05] in this model. See Table 12 for the final model.

(Significant codes: ****p* < 0.001, ***p* < 0.01, **p* < 0.05)

Fixed effects	**Estimate**	**SE**	***z***	***p***
(Intercept)	3.861	0.470	8.211	< 0.001***
Target: [-p̚]	-0.271	0.553	-0.490	> 0.05
Target: [-t̚]	0.949	0.528	-1.798	> 0.05
Target: [-k̚]	-0.356	0.547	-0.651	> 0.05
Target: [-m]	0.521	0.603	0.863	> 0.05
Target: [-n]	-0.146	0.556	-0.263	> 0.05
Target: [-ŋ]	-0.1093	0.563	-0.194	> 0.05

Group: Control	0.134	0.528	0.251	> 0.05
Target: [-p̚]×Group: Control	-0.379	0.566	-0.669	> 0.05
Target: [-t̚]×Group: Control	-1.637	0.518	-3.158	< 0.01**
Target: [-k̚]×Group: Control	-1.960	0.538	-3.647	< 0.001***
Target: [-m]×Group: Control	-0.773	0.621	-1.245	> 0.05
Target: [-n]×Group: Control	-1.727	0.549	-3.147	< 0.01**
Target: [-ŋ]×Group: Control	0.127	0.605	0.211	> 0.05
Random effects	**Variance**	**SD**		
Subject	0.521	0.722		
Item	1.201	1.096		

Table 12. Statistical Model of Two Groups' Identification Data in Post-test

The results of the post hoc test about two groups' identification performances in seven targets are drafted in Figure 7. In this figure, targeted segments are labeled on the horizontal axis and the percentage points of the vertical axis stand for the mean identification accuracy. The bar filled in light gray and dark gray refers to an average identification rate of a target by the experimental group and the control group respectively. The mean ± standard error of each object is illustrated by the error bar.

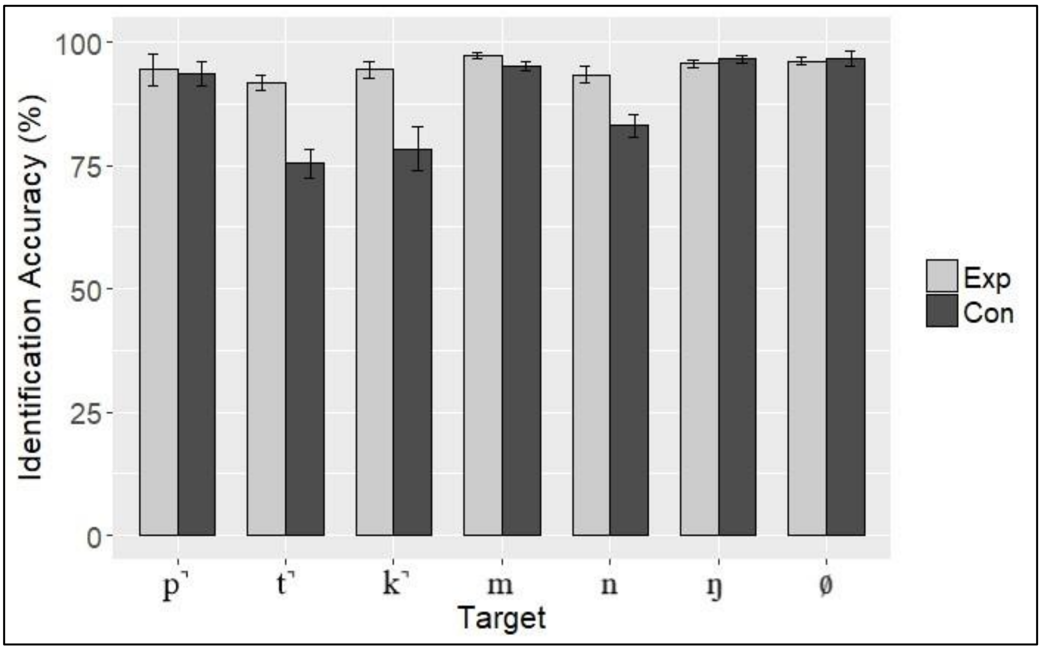

Figure 7. Two Groups' Identification Accuracy in Post-test

According to Figure 7, the experimental group's identification scores of all the targets went up to 90% in the post-test of Experiment 4 after the intervention. For the control group, the highest scores above 95% were obtained in targets [-m], [-ŋ] and [-ø] followed by the figures for targets [-p˺] (93.48%) and [-n] (82.34%). By contrast, identification rates in targets [-t˺] and [-k˺] were lowest at 75% and 77.72% respectively.

Between-group differences were found in targets [-t˺], [-k˺] and [-n]. With the training effects, participants in experimental group performed noticeably better than those in control group in above three targets. For target [-t˺], the experimental group outperformed the control group with a higher identification score [experimental-control: 91.85%-75.00%, ($p < 0.01$)]. For target [-k˺], the experimental group still surpassed control group with a score at 94.29%-77.72% ($p < 0.001$). Similar advantage of the experimental group over the control group was found in the target

[-n] [experimental-control: 94.29%-82.34%, ($p < 0.001$).

6.3.1.2 Within-group (Pre-test vs. Post-test) Comparison

To compare participants' perceptual performance in Experiment 2 (as the pre-test) and Experiment 4 (as the post-test), two models of within-group (pre-test vs. post-test) comparisons measured the identification results by the experimental group and the control group respectively. The model for the experimental group took Target, Condition of syllable, Type frequency of syllable, and Session as four fixed effects with random effects of Item and Subject. Significant contribution of Target [χ^2 (1) = 55.731, $p < 0.001$], Session [χ^2 (1) = 185.24, $p < 0.001$] and Target × Session interaction [χ^2 (1) = 28.645, $p < 0.001$] was identified in the results, while no noticeable effects were found from Condition of syllable [χ^2 (1) = 1.187, $p > 0.05$] and Type frequency of syllable [χ^2 (1) = 0.834, $p > 0.05$] in this model. See Table 13 for the final model.

(Significant codes: ****p* < 0.001, ***p* < 0.01, **p* < 0.05)

Fixed effects	**Estimate**	**SE**	***z***	***p***
(Intercept)	3.734	0.422	8.841	< 0.001***
Target: [-p˺]	-0.264	0.499	-0.529	> 0.05
Target: [-t˺]	-0.841	0.479	-1.755	> 0.05
Target: [-k˺]	-0.371	0.496	-0.749	> 0.05
Target: [-m]	0.585	0.548	1.066	> 0.05
Target: [-n]	-0.234	0.499	-0.469	> 0.05

Target: [-ŋ]	-0.003	0.511	-0.006	> 0.05
Session: Pre-test	-0.782	0.370	-2.111	< 0.05*
Target: [-p˺]×Session: Pre-test	-0.587	0.466	-1.260	> 0.05
Target: [-t˺]×Session: Pre-test	-1.159	0.438	-2.649	< 0.01**
Target: [-k˺]×Session: Pre-test	-1.309	0.457	-2.864	< 0.001***
Target: [-m]×Session: Pre-test	-0.0003	0.537	-0.001	> 0.05
Target: [-n]×Session: Pre-test	-0.633	0.464	-1.364	> 0.05
Target: [-ŋ]×Session: Pre-test	0.713	0.512	1.392	> 0.05
Random effects	**Variance**	**SD**		
Subject	0.459	0.677		
Item	0.945	0.972		

Table 13. Statistical Model of Experimental Groups' Identification Data in Pre-test & Post-test

In the model for within-group (pre-test vs. post-test) comparison of the control group, the noticeable effects of Target [χ^2 (1) = 81.71, p < 0.001] and Session [χ^2 (1) = 40.788, p < 0.001] were detected, while Condition of syllable [χ^2 (1) = 3.3656, p > 0.05] and Type frequency of syllable [χ^2 (1) = 0.1912, p > 0.05] cannot effectively predict the dependent variable (perception responses). See Table 14 for the final model.

(Significant codes: ***p < 0.001, **p < 0.01, *p < 0.05)

Fixed effects	**Estimate**	**SE**	***z***	***p***
(Intercept)	3.862	0.368	10.505	< 0.001***

Target: [-p˺]	-0.788	0.395	-1.994	< 0.05 *
Target: [-t˺]	-2.533	0.381	-6.644	< 0.001***
Target: [-k˺]	-2.332	0.382	-6.097	< 0.001***
Target: [-m]	-0.448	0.402	-1.114	> 0.05
Target: [-n]	-1.479	0.389	-3.809	< 0.001***
Target: [-ŋ]	0.054	0.415	0.131	> 0.05
Session: Pre-test	-0.586	0.091	-6.458	< 0.001***
Random effects	**Variance**	**SD**		
Subject	0.544	0.737		
Item	0.929	0.964		

Table 14. Statistical Model of Control Groups' Identification Data in Pre-test & Post-test

As summed up in Table 15, the experimental group with the intervention and the control group without the intervention obtained different identification gains. After seven sessions of intervention, participants in experimental group achieved marked improvement in identifying targets [-p˺], [-t˺], [-k˺], [-m], [-n] and [-∅]. By contrast, control group improved their identification scores in only four targets [-p˺], [-t˺], [-k˺] and [-m]. Besides, the trained participants got higher scores than their untrained counterparts in targets [-t˺], [-k˺], and [-n].

Experimental Group	Control Group

Target segment	Pre-training	Post-training	*p*	Pre-training	Post-training	*p*
[-p̚]	84.51%	94.84%	< 0.001	88.32%	93.48%	< 0.05
[-t̚]	67.93%	91.85%	< 0.001	63.59%	75.00%	< 0.001
[-k̚]	73.37%	94.29%	< 0.001	65.49%	77.72%	< 0.001
[-m]	94.57%	97.28%	< 0.05	90.22%	95.11%	< 0.01
[-n]	83.42%	94.29%	< 0.001	81.79%	82.34%	> 0.05
[-ŋ]	95.38%	95.65%	> 0.05	94.57%	96.47%	> 0.05
[-ø]	92.36%	96.18%	< 0.05	94.10%	96.53%	> 0.05

Table 15. Two Groups' Identification Accuracy in Pre-training & Post-training

6.3.2 Results of Monosyllable Repetition Test

6.3.2.1 Between-group (Experimental vs. Control group) Comparison

The mixed-effects model for the between-group (experimental vs. control group) comparison of production data in Experiment 4 took Target, Condition of syllable, Type frequency of syllable, and Group as four fixed effects. Except that no contribution was spotted in Condition of syllable [χ^2 (1) = 1.858, p > 0.05] and Type frequency of syllable [χ^2 (1) =0.153, p > 0.05], results indicated that Target [χ^2 (1) = 30.457, p < 0.001] and Group [χ^2 (1) = 20.319, p < 0.001] can significantly predict the dependent variable (production responses) with random effects of Item and Subject in the model. Factor Target did not interact with Factor Group [Target × Group: χ^2 (1) = 9.803, p > 0.05]. See Table 16 for the final model.

(Significant codes: ****p* < 0.001, ***p* < 0.01, **p* < 0.05)

Fixed effects	**Estimate**	**SE**	**z**	**p**
(Intercept)	1.955	0.363	5.389	< 0.001***
Target: [-p̚]	-0.233	0.412	-0.566	> 0.05
Target: [-t̚]	-1.658	0.410	-4.043	< 0.001***
Target: [-k̚]	-1.758	0.411	-4.281	< 0.001***
Target: [-m]	0.157	0.415	0.379	> 0.05
Target: [-n]	-0.351	0.411	-0.854	> 0.05
Target: [-ŋ]	0.219	0.417	0.526	> 0.05
Group: Control	-1.376	0.261	-5.281	< 0.001***
Random effects	**Variance**	**SD**		
Subject	0.421	0.649		
Item	0.200	0.447		

Table 16. Statistical Model of Two Groups' Production Data in Post-test

Post hoc pairwise comparison for two groups' different performances of seven targets in production was drafted in Figure 8. As can be seen from Figure 8, the horizontal axis represents the targets and the vertical axis stands for the mean production accuracy. With an error bar representing the mean ± standard error, each light gray bar and dark gray bar refers to an average production rate of a target by the experimental group and control group respectively.

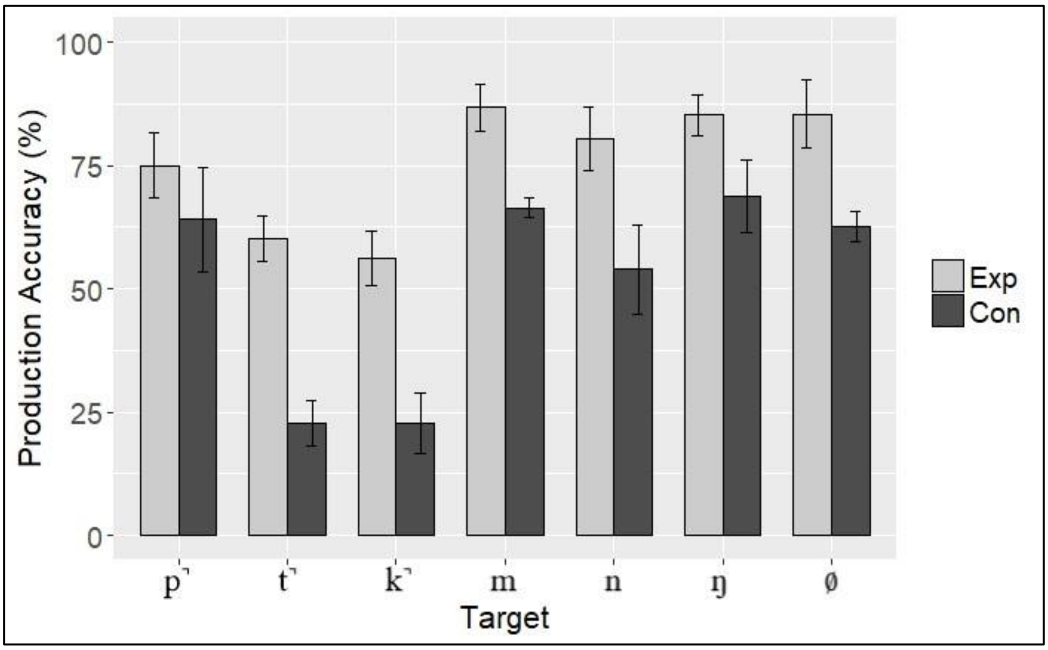

Figure 8. Two Groups' Production Accuracy in Post-test

According to Figure 8, the experimental group acquired the lowest production accuracy in targets [-t˺] and [-k˺] (60.16% and 56.25% respectively), whereas the figures for [-ø] (85.42%), [-m] (86.72%), [-n] (80.47%), [-ŋ] (85.16%) and [-p˺] (75%) were significantly higher. Similar pattern but with much lower scores were identified in the performance of the control group. The figures for targets [-t˺] (22.66%) and [-k˺] (22.66%) by the control group referred to the lowest. Participants in this group performed better in the targets [-ø], [-m], [-n], [-ŋ] and [-p˺] with a score at 62.5%, 66.41%, 53.91%, 68.75% and 64.06% respectively.

Between-group differences were found in targets [-m], [-n], [-ŋ], [-t˺], [-k] and [-ø]. With the training effects, participants in experimental group performed noticeably better than those in control group in above six targets. The figures of the experimental group were around 20% higher than those of the control group for targets [-m] [experimental-control: 86.72%-66.41%, ($p < 0.05$)], [-n] [experimental-control: 80.47%-53.91%, ($p < 0.01$)], [-ŋ] [experimental-control: 85.16%-68.75%, ($p < 0.05$)] and [-∅] [experimental-control: 85.42%-62.5%, ($p < 0.05$)]. The differences between the experimental group and control group were even bigger in targets [-t̚] and [-k̚]. For target [-t̚], experimental group's production score was almost three times higher than that by control group [experimental-control: 60.16%-22.66%, ($p < 0.001$)]. For target [-k̚], the experimental group surpassed the control group with a score at 56.25% versus 22.66% ($p < 0.05$).

6.3.2.2 Within-group (Pre-test vs. Post-test) Comparison

Two groups' production scores in Experiment 2 (as the pre-test) and Experiment 4 (as the post-test) were measured by two models of within-group comparison. These two models took Target, Condition of syllable, Frequency of syllable, and Session as four fixed effects. In the model for the experimental group, the basic model with random effects of Subject and Item was incrementally augmented after including Target [χ^2 (1) = 39.323, $p < 0.001$], Session [χ^2 (1) = 216.41, $p < 0.001$] and Target × Session [χ^2 (1) = 72.345, $p < 0.001$]. No substantial contribution was found from Condition of syllable [χ^2 (1) = 2.176, $p > 0.05$] and Type frequency of syllable [χ^2 (1) = 1.232, $p > 0.05$]. See Table 17 for the final model.

(Significant codes: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$)

Fixed effects	**Estimate**	**SE**	***z***	***p***

(Intercept)	2.032e+00	4.681e-01	4.342	1.4e-05***
Target: [-p˺]	-7.225e-01	5.282e-01	-1.368	0.17141
Target: [-t˺]	-1.51e+00	5.199e-01	-2.913	0.0036**
Target: [-k˺]	-1.70e+00	5.190e-01	-3.286	0.001**
Target: [-m]	1.083e-01	5.511e-01	0.196	0.84427
Target: [-n]	-2.190e-01	5.432e-01	-0.403	0.68686
Target: [-ŋ]	1.267e-02	5.473e-01	0.023	0.98152
Session: Pre-test	-3.196e-06	4.196e-01	0.000	0.99999
Target: [-p˺]×Session: Pre-test	-1.81e+00	5.143e-01	-3.528	0.00041***
Target: [-t˺]×Session: Pre-test	-5.15e+00	8.733e-01	-5.900	3.6e-09***
Target: [-k˺]×Session: Pre-test	-2.09e+00	5.271e-01	-3.972	7.1e-05 ***
Target: [-m]×Session: Pre-test	-1.55e+00	5.350e-01	-2.900	0.00373**
Target: [-n]×Session: Pre-test	-1.81e+00	5.305e-01	-3.424	0.00061***
Target: [-ŋ]×Session: Pre-test	-6.73e-01	5.386e-01	-1.250	0.21123
Random effects	**Variance**	**SD**		
Subject	0.754	0.869		
Item	0.245	0.495		

Table 17. Statistical Model of Experimental Group's Production Data in Pre-test & Post-test

In the model for the control group, only Target [χ^2 (1) = 34.989, p < 0.001] can significantly predict the dependent variable (production responses) with random effects of Item and Subject, whereas there were no effects of Condition of syllable

[χ^2 (1) = 1.481, p > 0.05], Type frequency of syllable [χ^2 (1) = 0.169, p > 0.05] and Session [χ^2 (1) = 1.197, p > 0.05]. See Table 18 for the final model.

(Significant codes: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$)

Fixed effects	**Estimate**	**SE**	***z***	***p***
(Intercept)	0.656	0.336	1.952	0.051
Target: [-p˺]	-0.165	-2.031	-0.385	0.7006
Target: [-t˺]	-2.017	0.435	-4.634	3.59e-06***
Target: [-k˺]	-2.031	0.435	-4.668	3.04e-06***
Target: [-m]	0.153	0.428	0.357	0.721
Target: [-n]	-0.262	0.427	-0.613	0.539
Target: [-ŋ]	0.424	0.432	0.981	0.327
Random effects	**Variance**	**SD**		
Subject	0.138	0.371		
Item	0.241	0.491		

Table 18. Statistical Model of Control Group's Production Data in Pre-test & Post-test

According to the results of within-group (pre-test vs. post-test) comparisons, two groups' production gains in different targets from Experiment 2 (as the pre-test) to Experiment 4 (as the post-test) were summed up below in Table 19.

	Experimental Group	**Control Group**

Target segment	Pre-training	Post-training	*p*	Pre-training	Post-training	*p*
[-p̚]	39.06%	75%	< 0.001	57.81%	64.06%	> 0.05
[-t̚]	1.56%	60.16%	< 0.001	21.09%	22.66%	> 0.05
[-k̚]	17.97%	56.25%	< 0.001	20.31%	22.66%	> 0.05
[-m]	61.72%	86.72%	< 0.001	70.31%	66.41%	> 0.05
[-n]	49.22%	80.47%	< 0.001	64.06%	53.91%	> 0.05
[-ŋ]	75.78%	85.16%	< 0.05	76.56%	68.75%	> 0.05
[-∅]	85.42%	85.45%	> 0.05	67.71%	62.5%	> 0.05

Table 19. Two Group's Production Accuracy in Pre-training & Post-training

Except for no valid change in the score of the target [-∅], the experimental group showed significant improvement in production scores of targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] after the intervention. The biggest increase was achieved in targets [-t̚] (from 1.56% to 60.16%) and [-k̚] (from 17.97% to 56.25%). Both started at around 40% in the pre-test of Experiment 2, targets [-p̚] and [-n] increased around 30% in their accuracy scores before reaching 75% and 80.47% respectively in the post-test of Experiment 4. Targets [-m] and [-ŋ] also increased noticeably to 86.72% and 85.16% compared with their figures in the pre-test. Opposite to the situation of the experimental group, the production scores of the control group did not realize noticeable improvement in any targets from pre-test to post-test.

6.4 Discussion

With the illustrated results of Experiment 4, the learning outcomes of Cantonese syllable-final segments by Mandarin speakers predicted according to SLM-r and the effectiveness of perception-only training are brought up again in this chapter.

The learning outcomes of the experimental group (by training) and the control group (by natural development) after three weeks both point to the influence of L1 experience on L2 learning. Turning back to perceptual assimilation pattern of Cantonese syllable-final segments by Mandarin speakers in Experiment 1, Cantonese [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ] were categorized as similar sounds with bigger learning difficulties than learnable identical target [-∅]. Although two Mandarin groups in Experiment 4 obtained learning gains for six similar targets to different extents, they were struggling to achieve satisfying attainment for similar sounds [-t˺] and [-k˺]. In perception, the control group's perception scores for similar targets [-t˺] and [-k˺] fluctuated around the basic criteria (70%) and referred to the lowest. In production, the accuracy scores of similar targets [-t˺] and [-k˺] by two groups were the lowest. By contrast, identical target [-∅] referred to the best performances by two groups in most cases. These findings keep in line with the above SLM-r predictions and verify that similar targets are more difficult to be learned than identical categories.

However, these SLM-r predictions are not able to specify why the learning outcomes of targets [-t˺] and [-k˺] are less salient than those of other similar sounds

([-p˺], [-m], [-n] and [-ŋ]). Similar deficiency of such SLM-r predictions emerged in the discussion about the deferred development of targets [-t˺] and [-k˺] after the successful learning of other similar sounds in Experiment 3. Therefore, it is again suggested that potential effects other than the influence of L1 experience on the developmental procedure and learning outcomes of L2 phonetic learning should be considered. Alternative accounts and heuristic amendments of different L2 learning models will be put forward in the general discussion to uncover this issue.

The advantage in performance by the experimental group over the control group suggests the effectiveness of the perception-only intervention. After three weeks, both the experimental group and the control group achieved an identification score above the basic criterion (70%) for all the targets. However, it should be noted that the identification scores by the experimental group are significantly higher than those by the control group. In production, the experimental group realized substantial production gains in almost all the targets with the help of intervention. By contrast, the control group failed to make valid improvements for any targets by natural development and their scores in these targets were all below 70%.

6.5 Summary

The results of Experiment 4 (Post-test) were discussed in Chapter 6. Except for those sixteen participants selected as the experimental group for the intervention in Experiment 3, the rest of the sixteen participants were employed again as the control group and assessed together with the experimental group in Experiment 4. By adopting an identical set of tests, Experiment 2 and Experiment 4 can serve as the pre-test and post-test. Thus, the experimental group went through a procedure of pre-test, 7 sessions of perception training (lasted for three weeks), and post-test, whereas the control group was assessed via the pre-test, a three-week-interval, and the post-test. The findings of Experiment 4 are twofold. Firstly, the predicted effects of L1 experience on the L2 phonetic development by SLM-r hypotheses are generally verified through better learning outcomes of the learnable identical target [-ø] than those of similar targets [-t̚] and [-k̚]. However, the SLM-r predictions of the current study are not able to specify why the learning outcomes of targets [-t̚] and [-k̚] are less salient than those of other similar sounds ([-p̚], [-m], [-n] and [-ŋ]), which will be tackled in the general discussion. Secondly, the effectiveness of perception training is supported by experimental groups' advantages in perception and production of targets over the control group.

Chapter 7 General Discussion

Chapter 7 goes through the results of the four experiments and outlines the developmental patterns of Cantonese syllable-final segments ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-ø]) by Mandarin speakers. The implications of the results will be discussed. In Section 7.1, the research questions and hypotheses in this study are tackled with empirical results. The current observations are compared with previous findings to fill the gap in the application of SLM-r in understating Mandarin speakers' learning of Cantonese syllable-final segments. New evidence is elaborated to enrich the discussion of the relationship between L2 perception and production. Section 7.2 provides alternative hypotheses by PAM-L2 and L2LP to supplement the SLM-r predictions of the current study. In Section 7.3, heuristic amendments will be put forward to advance the application of the aforementioned L2 models. Section 7.4 turns to the effect of phonetic nature and co-articulation features of Cantonese syllable-final segments on Mandarin and Cantonese speakers' performances of target sounds. In Section 7.5, the valuable role of phonological contrast in L2 speech learning is supplemented. Section 7.6 ends up with a discussion of the ultimate attainment in L2 phonetic learning.

7.1 The Development of Cantonese Syllable-final Segments by Mandarin Speakers

The current study investigates the learning of Cantonese syllable-final segments by Mandarin speakers from the perspectives of perception and production. Four experiments were conducted centering around the cross-L1-L2 similarity and the perception-production interface.

7.1.1 The Effect of L1 Experience in SLM-r

The **Research Question 1** of the current study is to measure the cross-language similarity of syllable-final segments between Mandarin and Cantonese by Mandarin speakers. As agreed by PAM-L2, SLM/SLM-r, and L2LP, L1 experience plays an important role in molding L2 phonetic learning and the cross-L1-L2 similarity is the crux predicting the learning difficulties and learning outcomes of different L2 sounds. After a glimpse of four measurement methods for cross-L1-L2 similarity (the comparison of ① IPA symbols, ② acoustic similarity, ③ articulatory similarity, and ④ perceptual similarity) in the extant literature, the Perceptual Assimilation Test was adopted in the current study.

In the Perceptual Assimilation Test (Experiment 1), two tasks were incorporated including a Four-alternative Forced-choice Categorization Test and a Goodness-of-fit Rating Test. The categorization percentage, similarity rating score, and fit index were adopted to quantify the perceptual similarity of targets. Decided by the cross-L1-L2 perceptual similarity, L2 sounds can be categorized as identical, similar, and new sounds in SLM-r (Flege & Bohn, 2021). These three categories are different in the learning difficulties and learning outcomes of their developmental procedure. The identical L2 sounds refer to the least learning difficulties and L2 learners would find it relatively easy to develop identical L2 categories with the help of positive L1-L2 transfer. By contrast, the learning of similar sounds is the most challenging (Flege, 1995; Flege & Bohn, 2021). It is difficult for L2 learners to extract the phoneme correctly from the assimilated L1-L2 category (equivalence classification), which prevents the successful formation of similar categories (Flege, 1995; Flege & Bohn, 2021). If L2 tokens are regarded as new sounds, the formation of new L2 categories is processive with cumulating L2 experience and the learning of such sounds will be ultimately successful (Flege, 1995; Flege & Bohn, 2021). To interpret the results of the Perceptual Assimilation Test (Experiment 1) with SLM-r, none of the seven Cantonese syllable-final segments were perceived as a dissimilar ‘new’ sound to Mandarin speakers. Cantonese [-∅] was classified as an identical category to Mandarin [-∅] with the least learning difficulties. Other six targets were categorized as similar sounds (Cantonese [-p˺], [-t˺], [-k˺] to Mandarin [-∅]; Cantonese [-m], [-ŋ] to Mandarin [-ŋ]; Cantonese [-n] to Mandarin [-n]) and would pose more difficulties in Mandarin speakers' learning.

Research Question 2 attempts to verify the effects of L1 experience on L2 phonetic learning in the SLM-r framework. To attest the above SLM-r hypotheses deduced by cross-L1-L2 similarity, the developmental patterns of Cantonese syllable-final segments by an experimental group of Mandarin speakers through pre-test (Experiment 2), seven sessions of perception training (Experiment 3), and post-test (Experiment 4) were depicted. The findings of the pre-test (Experiment 2) revealed that identical target /-∅/ was better perceived and produced by Mandarin speakers, while the performances of similar targets [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ] in both perception and production were less satisfying. During the perception training in Experiment 3, the development of identical target [-∅] was firstly completed in both perception and production in Sub-test 2. Although the successful perceptual learning for all similar targets was achieved in Sub-test 2, their production development was processed much later (similar sounds [-m] and [-n] in Sub-test 3; [-ŋ] and [-p˺] in Sub-test 9). Until the post-test (Experiment 4), Mandarin speakers' production abilities of similar targets [-t˺] and [-k˺] were still developing. Collectively, learnable identical target ([-∅]) was developed earlier than those more difficult similar sounds ([-m], [-n], [-ŋ], [-p˺], [-t˺] and [-k˺]) by Mandarin speakers. These findings acknowledge the effect of the previously existing experience of

Mandarin syllable-final segments on Mandarin speakers' learning of Cantonese syllable-final segments and manifest the predicted learning difficulties and learning outcomes as SLM-r hypothesized.

7.1.2 The L2 Perception-production Interface

The L2 perception-production relationship, as the **Research Question 3** of the current study, was mainly discussed via the interface between perception and production during the intervention (Experiment 3).

The observed findings reveal that diverse patterns of perception-production interface as a function of the learnability of different targets could manifest during L2 phonetic learning. In the case of less learnable similar targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ], perception and production were moderately positively correlated ($\rho = 0.506$, $p < 0.01$) and participants' production accuracy increased along with their identification accuracy during the intervention. The perception of all the similar targets was achieved in Sub-test 2, which is earlier than the development of their productions (similar sounds [-m] and [-n] in Sub-test 3; [-ŋ] and [-p̚] in Sub-test 9). This points to a perception-precedence pattern. On the other hand, the perception and production of the easily learnable identical target [-∅] were settled after Sub-test 2 and maintained an accuracy score of around 90% throughout the intervention. No significant correlation ($\rho = 0.505$, $p > 0.05$) was found between the perception and production for this target.

Concerning diverse opinions about the perception-production interaction in the L2 phonetic learning, the production-precedence pattern (Goto, 1971; Sheldon &

Strange, 1982; McCandliss et al., 2002; Hattori & Iverson, 2009; Ingvalson et al., 2011), the perception-precedence pattern of SLM (Flege, 1995), and the perception-production co-evolve pattern of SLM-r (Flege & Bohn, 2021) are proposed. The case of similar targets in the current study agreed with the SLM hypotheses that successful perception would develop before satisfying production in L2 learning and the perception accuracy would place an ceiling on the production accuracy of L2 sounds (Flege, 1995; Flege et al., 1995; Flege et al, 2003). The case of identical target [-∅] failed to provide overt support to any of the above three patterns. As the perception and production of this target were completed with a ceiling accuracy score when the experiments were initiated, it is difficult to decide which of the production-precedence, perception-precedence, or perception-production co-evolve strategies was utilized by Mandarin speakers to achieve the successful learning of identical target [-∅]. The observed diverse patterns here negate those one-size-fits-all patterns proposed by previous studies. In line with Cheng & Zhang (2009), Nagle (2018a), and Baese-Berk (2019), it is suggested that the perception-production relationship in L2 phonetic learning is relatively plastic and it could vary according to different learning targets. The current study further specifies that the learnability of different learning targets could be an important indicator for the diverse patterns of the L2 perception-production interface. If this is the case, the learning of less learnable targets (e.g., similar targets in SLM/SLM-r) tends to rely on the perception-precedence pattern, while it is likely for easily learnable targets (e.g., identical targets in SLM/SLM-r) to achieve satisfying perception and production simultaneously.

The perception-production interface has been a hotspot in the research of L2 phonetic learning. Some recently updated studies intend to find out other possible factors that could modulate the L2 perception-production interface. Jia, Strange, Collado, & Guan (2006), Fabra & Romero (2012)，Saito, Kazuya, & van Poeteren (2018), and Melnik-Leroy et al. (2021) propose that the L2 perception-production relationship could vary according to different developmental stages in the L2 phonetic learning. Generally, the L2 perception-production link is relatively stronger at the initial stage and the development of production is mainly induced by the successful perception. In the late learning phase, the relationship between these two modalities becomes weaker. The dependence of production on perception has been weakened and the achievement of a good production is affected by other factors (e.g., age of learning, learning aptitude, speech training experience, and so on). Other studies suggest that individual differences should be considered, as individual differences in cognitive skills, language use, language aptitude and so on would modulate how the L2 perception-production interface manifests (Cerviño-Povedano & Mora, 2010; Darcy, Mora, & Daidone, 2016; Nagle, 2018b; Nagle & Baese-Berk, 2021). By assessing Japanese speakers' performances of English contrast /ɹ/-/l/ and English speakers' performances of Japanese singleton-geminate consonants, Kato & Baese-Berk (2020) points out that both input types and target contrasts can influence the L2 perception-production relationship. It is thus suggested that the combinations of different perspectives rather than a single factor should be considered when addressing the L2 perception-production interface. The present study is in line with the above research and intends to deepen the understanding of the perception-production relationship in L2 speech. However, it remains unknown whether the influence of the learnability of different learning targets (as proposed in the current study) would take effect with the aforementioned factors (as suggested in the previous studies) when modulating the L2 perception-production interface. This leaves room for future discussion.

7.1.3 The Effectiveness of Perception Training

Research Question 4 intends to testify the effectiveness of perception training on promoting L2 phonetic learning with the findings of Experiment 3 (the perception training) and Experiment 4 (the post-test).

In the seven sessions of perception training, the perception and production of Cantonese syllable-final segments by one experimental group of Mandarin speakers performed a gradual improvement. Seven targets soon achieved successful perceptual learning in Sub-test 2 of the intervention. Except for targets [-t˺] and [-k˺], other targets also completed production learning successively.

Afterward, the achievements of this experimental group were compared with those of another untrained control group in the post-test. The identification scores of targets [-p˺], [-t˺], [-k˺], [-m], [-n] and [-∅] by the experimental group increased significantly to 90%, whereas the control group improved their identification scores in only four targets [-p˺], [-t˺], [-k˺] and [-m] by naturalistic development. With the help of perception training, the experimental group got noticeably higher identification scores than the control group in targets [-t̚], [-k̚] and [-n]. In production, the experimental group showed significant improvement in the scores of six syllable-final consonants ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ]) and satisfied the adopted criterion at 70% for targets [-p̚], [-m], [-n], [-ŋ], and [-∅] from pre-test to post-test. Oppositely, the untrained control group was unable to realize improvement in the production of any targets and the production scores of all the targets by this group were below 70% until post-test. With the training effects, the experimental group performed noticeably better than the control group in the production of targets [-m], [-n], [-ŋ], [-t̚], [-k], and [-∅] in post-test.

In line with preceding research (McCandliss et al., 2002; Handley et al., 2009; Baese-Berk, 2010 & 2019; Anderson, 2011; Shinohara & Iverson, 2018), the above-obtained results acknowledge the efficiency of perception training on the development of L2 perception and production. The perception training for beginning L2 learners has not necessarily to be HVPT (Cebrian et al., 2019). Instead, low variability phonetic training (one set of stimuli embedded in the same phonetic context by one talker in the current study) is also effective in promoting the L2 phonetic learning of inexperienced and weak-ability learners. According to SLM-r (Flege & Bohn, 2021), the quantity and quality of received L2 input in the ambient language environment by L2 learners is one important prerequisite for the detection of cross-L1-L2 similarity and the category formation of L2 sounds. However, the amount of L2 input could be insufficient during naturalistic learning (Sakai & Moorman, 2018; Cebrian et al., 2019) and exposure to accented target language can lower the quality of L2 input (Flege & Eefting, 1987; Flege, 1991; Flege & Bohn, 2021). Laboratory phonetic training is a reliable source of intensive and qualified L2 input. This answers why the experimental group with intervention obtained better learning gains of Cantonese syllable-final segments than the control group by naturalistic learning in this study.

On the other hand, the effectiveness of perception training could be related to the L2 perception-production interface as observed in the current study. Previous studies about the L2 perception-production relationship propose the production-precedence pattern (Goto, 1971; Sheldon & Strange, 1982; McCandliss et al., 2002; Hattori & Iverson, 2009; Ingvalson et al., 2011), the perception-precedence pattern (Flege, 1995), and the perception-production co-evolve pattern (Flege & Bohn, 2021). As discussed in Section 7.1.2, the learning of six Cantonese similar targets ([-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ]) performed a perception-precedence pattern and tended to rely more on correct perception before successful pronunciation. Therefore, these targets benefited directly from the perception training and achieved significant perception gains firstly. The perception gains then guided the production learning and were generalized to the production improvement after intervention. To optimize the effectiveness of L2 phonetic training and better facilitate L2 learners' learning, it is suggested that training methodologies should be adjusted according to the perception-production patterns represented by different learning targets.

7.2 Alternative Hypotheses of PAM-L2 and L2LP

The SLM-r framework generally predicted the learning of Cantonese syllable-final segments by Mandarin speakers as discussed in Section 7.1.1, but it failed to clarify why targets [-t˺] and [-k˺] were more difficult than other similar targets ([-m], [-n], [-ŋ] and [-p˺]) for Mandarin speakers. Different tenets employed by PAM-L2 and L2LP could lead to different predictions supporting or negating the SLM-r hypotheses. This section attempts to seek alternative accounts from PAM-L2 and L2LP and supplement the SLM-r hypotheses on this issue.

PAM-L2 (Best & Tyler, 2007) focus on the perception of L2 contrasts by inexperienced learners. In the framework of PAM-L2 (Best & Tyler, 2007), Category-goodness Difference Assimilation (with deviant goodness fit) or Single-category Assimilation (with equal goodness fit) will occur if two L2 categories are assimilated to one L1 category. The assimilation pattern of Cantonese [-t˺] and [-k˺] to Mandarin [-∅] was in accord with these two cases. To further decide whether Cantonese contrast [-t˺]-[-k˺] belong to Category-goodness Difference Assimilation or Single-category Assimilation, post hoc comparisons of similarity rating scores between Cantonese [-t˺]-Mandarin [-∅] and Cantonese [-k˺]-Mandarin [-∅] was conducted. The result indicated that no significant difference was found (3.925-3.975: $W = 3180$, $p > 0.05$). Accordingly, Cantonese contrast [-t˺]-[-k˺] belongs to Single-category Assimilation which refers to the highest learning difficulties in PAM-L2. As predicted by PAM-L2, it is assumed that Mandarin speakers would find it rather challenging to differentiate Cantonese [-t˺] and [-k˺] at the beginning as these two tokens are perceived as the equivalence to Mandarin [-∅]. Sufficient exposure to the differences between [-t˺]-[-k˺] is necessary to promote perception learning. Mandarin speakers would develop a new phonetic category for at least one of the contrasting phones before the formation of the other new category.

L2LP attempts to model the entire developmental process from the initial to the end stage of L2 speech perception of L2 contrasts by both inexperienced and experienced learners (Escudero, 2005). In the three learning scenarios of L2LP, the NEW scenario refers to the case that two L2 sounds are mapped to one L1 category, which is similar to the Single-category Assimilation or Category-goodness Difference Assimilation in PAM-L2 (Escudero, 2005). As Cantonese [-t˺] and [-k˺] were mapped to Mandarin [-∅], the learning of contrast [-t˺]-[-k˺] is the case of the NEW scenario. To achieve successful learning in this scenario, Mandarin speakers should create new Cantonese categories or split previously existing Mandarin categories to accommodate the contrasting relationship of these two sounds. Deduced from the L2LP hypothesis, the learning of the NEW scenario is the most challenging for L2 learners (Escudero, 2005).

As observed in the current study, the performances and learning outcomes of targets [-t˺] and [-k˺] by Mandarin speakers referred to the least satisfying throughout the experimental procedure. The relatively higher learning difficulty of Cantonese contrast [-t˺]-[-k˺] suggested by PAM-L2 and L2LP hypotheses provides a possible account for the obtained results. From this perspective, PAM-L2 and L2LP hypotheses supplement SLM-r predictions on the developmental pattern of Cantonese targets [-t˺] and [-k˺].

7.3 The Effect of Distribution Patterns of L2 Targets on L2 Phonetic Learning

Previous L2 studies collectively value the important role of L2 learners' knowledge about previously existing L1 categories in modulating L2 performances but not too many of them consider such issue in the context of sound change in distribution patterns of target segments. It is possible that sound changes in either L1 or L2 categories can influence how would the target sounds be mapped between L1 and L2, which can further pose an effect on L2 learners' performances. Extant studies about the L2 phonetic learning models overwhelmingly target on contrastively distributed L1 and L2 sounds with little attention paid to a situation where target segments might be experiencing a sound merger. After seeking inspiration from a few studies about the modulation of merging L1 categories on L2 performances, the current study primarily focuses on how would the suspension of contrastive distribution of L2 categories influence the L2 learning of such sounds. The coronal-dorsal merging tendency of Cantonese syllable-final segments observed in this study opens up a chance to unveil this issue, based on which, this section intends to suggest related heuristic amendments to the application of theoretical models in L2 phonetic learning.

Only a few studies touch upon the influence of L1 sound merger on L2 performances. In Soo, Johnson, & Babel (2021), Cantonese-English bilinguals' productions of /n/ and /l/ in both Cantonese and English are measured via mid-frequency spectral tilt (H4-2KHz) and F2-F1 spacing under the influence of

Cantonese /n/-/l/ merger. According to the results of comparisons at group-level in this study, /n/-/l/ contrast is poorly differentiated in Cantonese but is well-maintained in English. This indicates that segments /n/ and /l/ behave separately in Cantonese and English. The /n/-/l/ merger in L1 Cantonese transfers no effects to Cantonese-English bilinguals' differentiation of corresponding segments in L2 English. The results of comparisons at the individual level in the same study, however, point to other possibilities. Some of the participants transfer the /n/-/l/ merger from Cantonese to English and produce neutralized /n/ and /l/ in both languages. Thus, the suspension of contrastive distribution of target segments in L1 would pose an influence on L2 performances. Similarly, Kim (2012) explores the effects of the L1 sound merger on L2 performances by examining adult Korean English learners' production of corresponding stops in both Korean (aspirated vs. lax stops) and English (voiced vs. voiceless stops) via VOT and f0. These two acoustic parameters react differently to the sound merger. The VOT merger between aspirated and lax stops is surfaced in participants' L1 Korean productions while the VOT contrast between voiced and voiceless stops are well-preserved in their L2 English. By contrast, the f0 merger in Korean is transferred to the productions of corresponding stops in English. These results point to an effect of the L1 sound merger on L2 production. Furthermore, it is suggested that such an effect may function variously in different acoustic parameters of target segments.

The observed coronal-dorsal merger of Cantonese syllable-final segments in this study provides a chance to identify the effect of changes in distribution patterns of L2 categories on L2 phonetic learning. In Experiment 2 (the pre-test), Cantonese participants satisfied the adopted criteria (70% of accuracy) in the perception of all the targets but with a significantly lower score in targets [-t̚] and [-k̚]. As for production, only dorsal targets [-ŋ] and [-k̚] achieved an accuracy score below 70%. These participants showed a noticeable preference for coronal targets ([-n] and [-t̚]) instead of the dorsal ones ([-ŋ] and [-k̚]) in their production. These results generally support the coronal-dorsal merger and the alveolarization merging tendency as observed by previous studies (see Bauer, 1979; Yeung, 1981; Chen, 1999; Zee, 1999a & 1999b; Law et al., 2001; Wong, 2005; Ding, 2010; Bauer & Benedict, 2011; To et al., 2013; To et al., 2015).

According to the perceptual pattern of Cantonese syllable-final segments by Mandarin speakers, Cantonese [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] were all categorized as similar targets in SLM-r and should refer to comparative learning difficulties. However, targets [-t̚] and [-k̚] represented the lowest learnability compared with other similar targets, which exceeds the prediction of SLM-r. During the intervention of Experiment 3, Mandarin speakers' production scores of targets [-t̚] and [-k̚] were below 70% most of the time and these two targets failed to obtain successful production learning until the end of the training. In the comparison of the experimental group (with intervention) and control group (without intervention) in Experiment 4, targets [-t̚] and [-k̚] again referred to the least satisfying performances by Mandarin speakers of both groups. The relatively lower learnability of these two targets can result from a potential influence of the coronal-dorsal merger of syllable-final segments in contemporary Hong Kong Cantonese. According to Chen (1999) and Wong (2005), it is more often to observe the coronal-dorsal merger in the younger generation of Cantonese native speakers. The employed Mandarin participants of these two groups all were university students. Before the experiment, they had an exposure to Cantonese for about 9.48 ± 6.7 months, during which they could be immersed in the phonetic variants of syllable-final coronal and dorsal targets in the daily contact with their Cantonese schoolmates. Obviously, neither by intervention nor by naturalistic development during the period of the current study is enough for these participants to master the variants of [-t˺]-[-k˺] merger or to differentiate between targets [-t˺] and [-k˺]. It is worth noting that the relatively lower learnability of the coronal-dorsal merger of syllable-final segments is also observed in Cantonese children's L1 acquisition. In To et al. (2013), the phonetic variants of syllable-final consonants (/-n/-/-ŋ/ and /-t˺/-/-k˺/) are recorded in Cantonese children's cross-sectional output at different ages. If the realizations of such phonetic variants of the coronal-dorsal merger are strictly considered to be incorrect, both coronal and dorsal targets refer to the relatively lower learnability and are developed later than other syllable-final consonants. Precisely, Cantonese /-p˺/ and /-m/ are developed by age 4;6, which happens earlier than the acquisition of /-k˺/ (acquired by age 5;0), /-n/, /-ŋ/ and /-t˺/ (not acquired even until 11 years old). Therefore, the changes in distribution patterns of learning targets affect not only the L1 acquisition but also the L2 learning of Cantonese syllable-final consonants.

As reviewed at the beginning of this section, the effects of the L1 sound merger on L2 performances could function variously at different comparison levels of participants (Soo et al., 2021) and in different acoustic parameters of target segments (Kim, 2012). Different modulation patterns about the influence of L2 merger on L2 phonetic learning are also obtained in the current study. Mandarin speakers experienced fewer difficulties in overcoming Cantonese [-n]-[-ŋ] merger than [-t̚]-[-k̚] merger, which can be attributed to Mandarin speakers' previously existing knowledge of the contrastive distribution of Mandarin syllable-final segments [-n] and [-ŋ]. Although the syllable-initial segments [t] and [k] are contrastively distributed in participants' L1 Mandarin grammar, the cross-syllable-position mapping is not necessarily transferred from syllable-initial to syllable-final positions at the level of position-sensitive allophones (Flege & Davidian, 1984; Flege et al., 1987; Flege, 1988b; Flege, 1992). Mandarin speakers were not able to take advantage of their L1 knowledge about Mandarin syllable-initial [t]-[k] contrast when learning Cantonese syllable-final segments [-t̚] and [-k̚]. Therefore, Mandarin speakers managed to develop Cantonese [-n] and [-ŋ] as two independent segments after the intervention, while targets [-t̚] and [-k̚] referred to the least learnability among all the similar targets.

PAM-L2 and L2LP hypotheses provide an alternative account in Section 7.2 and supplement the deficiency of SLM-r predictions in the clarification of the lowest learnability of targets [-t̚] and [-k̚]. The current study instead proposes that the relatively higher learning difficulties of these two targets can be attributed to their merging tendency in Cantonese. The above frameworks regrettably neglect the potential effects of distribution patterns of learning targets on L2 phonetic development. Heuristic amendments to the application of the aforementioned L2 models are put forward here and it is suggested that the distribution patterns of L2 targets should be considered together with the perceived cross-L1-L2 similarity when predicting the development of different L2 sounds. The suspension of contrastive distribution of L2 categories can pose higher learning difficulties for L2 learners as outlined in the current study. It remains unclear whether Mandarin speakers would manage a native-like coronal-dorsal merger or develop the merging targets as two independent ones with accumulating immersion in daily Cantonese. This leaves room for future research to investigate the learning of merging categories in a target L2.

7.4 The Effect of Inherent Phonetic Nature of Syllable-final Segments on L1 and L2 Speech Processing

As proposed in Section 7.3, the relatively lower learnability of Cantonese coronal-dorsal merging targets could be influenced by the suspension of coronal-dorsal contrasts in contemporary Hong Kong Cantonese. However, the inherent nature of Cantonese syllable-final segments in this language-specific merging pattern should not be ignored. As naturally-produced voice segments of a human language, Cantonese syllable-final segments have certain intrinsic acoustic/articulatory features and inevitably interact with other sounds in different phonetic contexts of Cantonese through co-articulation. This section endeavors to explore how would such features of Cantonese syllable-final segments constrain Mandarin and Cantonese speakers' perception and production of such targets.

The acoustic/articulatory features of different syllable-final segments govern their perceptibility levels in the manner and place of articulation, which would further influence listeners' performances towards corresponding targets. Different levels of acoustic salience are manifested in the articulation manner of different syllable-final segments. Carlisle (2001) and Tropf (2019) delineate how L2 learners would respond to coda segments with different levels of sonority. It is suggested that more sonorant coda segments (e.g., nasal codas) are perceptually more salient and would be better preserved than those less sonorant ones (e.g. plosive codas) by L2 learners in the L2 speech learning. This explains why the employed Mandarin participants performed better in more sonorant syllable-final nasals ([-n] and [-ŋ]) than less salient syllable-final stops ([-t̚] and [-k̚]) even though these two pairs of targets are undergoing the same merging process in the place of articulation. A similar pattern is observed in one study about Cantonese native speakers' perception and production of Cantonese consonant endings by Law et al. (2001). As nasal endings carry more acoustic information than stop endings, fewer errors are detected in the perception and production of more salient nasal endings [n] and [ŋ] than stop endings [t] and [k]. In the current study, more salient nasal targets [-n] and [-ŋ] were satisfactorily perceived by the Cantonese participants, while their perception of stop targets [-t̚] and [-k̚] failed to meet the adopted criteria (70% of accuracy rate).

On the other hand, syllable-final segments with different articulation place differ in their perceptibility. This affects how would listeners respond to the place contrast of syllable-final segments. Precisely, the articulators of syllable-final coronals and dorsals are intraoral tongue tip and tongue dorsum which cannot be seen by perceivers. By contrast, the labial targets are normally pronounced with lip constriction, which can provide robust visual cues about the place of articulation for listeners (Winters, 2000; Hume & Tserdanelis, 2002; Khouw & Ciocca, 2007). From this perspective, labial segments are assumed to be more salient and would be better perceived than the coronal and dorsal ones. The performances of Cantonese syllable-final stops ([-p˺], [-t˺] and [-k˺]) by both Mandarin and Cantonese participants in the current study mainly benefit from the perceptual salience of labial segments. For Mandarin speakers, target [-p˺] was better performed and developed earlier than targets [-t˺] and [-k˺]. For Cantonese speakers, target [-p˺] was satisfactorily perceived and produced while the performances of targets [-t˺] and [-k˺] were less satisfying.

The co-articulation features between the preceding vowels and syllable-final stops in VC collocations could impact perceivers' performances of syllable-final targets. Earlier studies depict that both release burst (the intraoral venting at the release of articulator constriction) and formant transitions (formant changes in the coarticulation of VC collocations) can provide informative cues of place contrast of stop consonants in the syllable-final position (Mal´ecot, 1958; Winitz, Scheib, & Reeds, 1972; Manuel, 1991; Byrd, 1992; Jun, 2004; Wright, 2004). The world languages deploy different perceptual cues to indicate the place contrast of syllable-final stops. The phonological system of English comprises both released and unreleased allophones of syllable-final stops (Henderson & Repp, 1982; Lisker, 1999; Zsiga, 2000; Kochetov & So, 2007). However, English native speakers tend to acknowledge the released form as the norm (Sumner & Samuel, 2005) and prefer released burst over formant transitions for the place cues of syllable-final stops (Chang, 2016). When the released burst is absent, the place cues of syllable-final stops are only provided by formant transition (Jun, 2004). In Thai and Korean, syllable-final stops are compulsory unreleased and transition cues, therefore, play a predominant role in listeners' perception of place contrasts of syllable-final stops (Abramson & Tingsabadh, 1999; Tsukada, 2006; Tsukada & Roengpitya, 2008; Tsukada, Nguyen, Roengpitya, & Ishihara, 2007; Chang & Mishler, 2012; Chang, 2016).

Similar to the cases in Thai and Korean, perceptual cues by release bursts are not available for the place contrast of Cantonese syllable-final stops (Cheung, 1986). Instead, previous studies turn to the post-vocalic coarticulation features of Cantonese unreleased syllable-final stops for their place cues (Ciocca, Wong, & So, 1994; Khouw & Ciocca, 2007; Yiu, 2016; Yiu, Archangeli, & Yip, 2021). Ciocca, Wong, & So (1994) observes three-way different F2 values at the vowel offset during VC transitions of Cantonese syllable-final stops (/p/, /t/, and /k/). It is therefore suggested that the F2 frequency can signal the place contrast of the above targets. In Khouw & Ciocca (2007), labial stop /p/ presents noticeably lower F2 values than the other two stops (/t/ and /k/) in the coarticulation of preceding vowel /a/. However, no significant F2 differences are found between targets /t/ and /k/. Khouw & Ciocca (2007) surmises that the employed Cantonese adolescent speakers are influenced by the coronal-dorsal merging tendency of syllable-final stops in contemporary Hong Kong Cantonese and therefore fail to produce different F2 frequencies for sequences /at/ and /ak/. The stimuli of the current study also include the collocation of central-low vowel /a/ and unreleased syllable-final stops [-p˺], [-t˺], and [-k˺]. According to the results of Khouw & Ciocca (2007), the place cue of the target [-p˺] is relatively more salient as signaled by a significantly lower F2 value than those of targets [-t˺] and [-k˺] in such collocations. If so, it makes sense that target [-p˺] was better performed and earlier developed than targets [-t˺] and [-k˺] by Mandarin speakers in the present study.

Language-specific patterns of L1 or L2 cannot account for all the perceptual biases manifested during L2 speech learning (Flege & Bohn, 2021). The present study attributes the lower learnability of Cantonese coronal and dorsal targets partly to the Cantonese-specific ongoing coronal-dorsal merger. This, however, does not mean that the aforementioned phonetic nature and co-articulation features of Cantonese syllable-final segments do not play a role in both L1 and L2 phonological processes by Cantonese and Mandarin speakers. It is encouraged that future research is developed about how the inherent features of target sounds would take effect during the language-specific speech process in L1 and L2.

7.5 The Phonological Contrast in L2 Phonetic Learning

The development of a speech sound includes not only the learning of phonetic features represented by the sound but also the understanding of the contrastive relationship between this sound and other corresponding sounds in the phonological system (Ohala, 2005; Al-Hindawi, Al-Hassnawi & Al-Ebadi, 2018). The reviewed models also acknowledge that accumulating input of the phonological contrast of the targeted sounds can facilitate the development of these sounds in L2 phonetic learning. SLM-r speculates that L2 learners' growth of L2 lexicon strengthens their awareness of the lexical functional differences between the targeted segments and other corresponding sounds, which is one important stage in the formation process of an L2 phonetic category (Bundgaard-Nielsen et al., 2011; Flege & Bohn, 2021). In PAM-L2 (Best & Tyler, 2007), the category formation of Category-goodness Difference Assimilation type and Single-category Assimilation type especially relies on L2 learners' cumulative exposure to the contrasting words of the two phones of an L2 contrast. L2LP (Escudero, 2005) contends that L2 perceptual development is a gradual meaning-driven procedure. L2 learners will adjust their L2 perception grammar according to the meaning differences in the contrast of the perceived forms and speakers' intended forms (Escudero, 2005; van Leussen & Escudero, 2015). This section discusses the important role of phonological contrast of Cantonese syllable-final segments in Mandarin speakers' learning of these sounds.

In Chang (2016), category 'zero' (refer to the absence of a final stop) is employed with /p/, /t/, and /k/ to assess the perception of unreleased final stops by Korean and English speakers. In accord with this preceding work, the absence of a final consonant is also considered in the present study. Albeit [-∅] represents not a concrete sound, the introduction of this category in the current study truly reflects the phonological relationship of syllable-final segments in Cantonese. In the phonological system of Cantonese, the absence of syllable-final consonants is permissible and can be symbolized as category [-∅]. This category and other corresponding segments are contrastively distributed and function collaboratively to distinguish meanings. Such contrastive relationship was realized through carrier syllables [CVp̚], [CVt̚], [CVk̚], [CVm], [CVn], [CVŋ] and [CV∅] in the present study. Before the experiment, participants were familiarized with the IPA symbol [-∅] and instructed to refer this category to the absence of syllable-final consonants in syllable structure [CV∅].

The error pattern of category [-∅] reflects Mandarin speakers' awareness of the contrasting relationship between [-∅] and other syllable-final consonants in Cantonese. As can be seen from Table 20 and Table 21, Mandarin speakers sometimes mistakenly realized category [-∅] as other Cantonese syllable-final consonants. They mainly mis-identified/mis-produced syllable [CV∅] as syllables [CVp̚], [CVt̚], and [CVk̚]. In only a few cases, these participants confused [CV∅] with [CVm], [CVn], and [CVŋ]. However, these error types only account for a small portion. Mandarin speakers maintained an accuracy score above the adopted criterion ((70% of accuracy rate) for Cantonese category [-∅] from pre-test to post-test. This suggests that participants can grasp the concept of [-∅] for [CV∅] syllables and successfully distinguish this category from other syllable-final consonants ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ]) in Cantonese.

Test	**Error Type**					
	[-p˺]	[-t˺]	[-k˺]	[-m]	[-n]	[-ŋ]
Pre-test	1.04%	1.04%	2.08%	1.04%	1.04%	1.39%
Sub-test 1	1.04%	1.04%	2.08%	0.00%	0.00%	0.00%
Sub-test 2	0.00%	1.04%	0.00%	0.00%	1.04%	0.00%
Sub-test 3	0.00%	2.08%	2.08%	0.00%	0.00%	0.00%
Sub-test 4	0.00%	1.04%	1.04%	0.00%	0.00%	0.00%
Sub-test 5	0.00%	1.04%	1.04%	0.00%	0.00%	0.00%
Sub-test 6	1.04%	2.08%	0.00%	0.00%	0.00%	0.00%
Sub-test 7	0.00%	1.04%	0.00%	0.00%	0.00%	0.00%
Sub-test 8	0.00%	0.00%	2.08%	0.00%	0.00%	0.00%
Sub-test 9	0.00%	1.04%	0.00%	0.00%	0.00%	0.00%
Sub-test 10	1.04%	1.04%	0.00%	0.00%	0.00%	0.00%
Sub-test 11	0.00%	1.04%	1.04%	0.00%	0.00%	0.00%
Sub-test 12	0.00%	0.00%	1.04%	0.00%	0.00%	0.00%
Sub-test 13	0.00%	0.00%	1.04%	0.00%	0.00%	0.00%
Sub-test 14	0.00%	1.04%	1.04%	0.00%	0.00%	0.00%
Post-test	0.00%	1.74%	1.74%	0.00%	0.00%	0.35%

Table 20. Experimental Group’s Error Pattern of Category [-∅] in Identification

Test	**Error Type**					
	[-p̚]	[-t̚]	[-k̚]	[-m]	[-n]	[-ŋ]
Pre-test	6.25%	3.12%	3.12%	0.00%	2.10%	0.00%
Sub-test 1	6.25%	6.25%	3.12%	0.00%	0.00%	0.00%
Sub-test 2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sub-test 3	6.25%	0.00%	0.00%	0.00%	0.00%	0.00%
Sub-test 4	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sub-test 5	3.12%	0.00%	6.25%	0.00%	0.00%	0.00%
Sub-test 6	6.25%	6.25%	0.00%	0.00%	0.00%	0.00%
Sub-test 7	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sub-test 8	0.00%	0.00%	6.25%	0.00%	0.00%	0.00%
Sub-test 9	3.12%	0.00%	0.00%	0.00%	0.00%	0.00%
Sub-test 10	3.12%	6.25%	6.25%	0.00%	0.00%	0.00%
Sub-test 11	3.12%	6.25%	0.00%	0.00%	0.00%	0.00%
Sub-test 12	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Sub-test 13	6.25%	3.12%	0.00%	0.00%	0.00%	0.00%
Sub-test 14	6.25%	6.25%	0.00%	0.00%	0.00%	0.00%
Post-test	6.25%	4.17%	4.17%	0.00%	0.00%	0.00%

Table 21. Experimental Group’s Error Pattern of Category [-∅] in Production

As discussed before, the phonetic learning of novel L2 categories can be progressed with exposure to the phonological contrast of these sounds (Escudero, 2005; Best & Tyler, 2007; Flege & Bohn, 2021). The introduction of category [-∅] helps to present the holistic view about the phonological relationship of all the acceptable segments in the syllable-final position of Cantonese for Mandarin speakers. During the perception training of the present study, the audio stimuli of the carrier syllables [CVp̚], [CVt̚], [CVk̚], [CVm], [CVn], [CVŋ] and [CV∅] reinforce the phonetic features of Cantonese syllable-final segments. Meanwhile, the pairwise presented stimuli strengthen participants' awareness about the contrastive relationship among these targets. The obtained results indicate that Mandarin participants in the experimental group benefit from such design and progress eventually in the learning of Cantonese syllable-final segments ([-p̚], [-t̚], [-k̚], [-m], [-n], [-ŋ] and [-∅]). However, it is difficult to decide whether the different learning outcomes between [-∅] and the other segments are attributed to the fact that category [-∅] represents not a concrete sound. Therefore, it is encouraged to conduct future studies and figure out an alternative experimental method when comparing these Cantonese syllable-final segments.

7.6 The Ultimate Attainment in L2 Phonetic Learning

The L2 learning models including SLM/SLM-r and L2LP build up their frameworks based on a premise that L2 learners maintain the learning ability to form new phonetic categories when being exposed to a foreign language. These models, however, hold different opinions about the ultimate attainment of L2 phonetic learning. According to SLM (Flege, 1995) and SLM-r (Flege & Bohn, 2021), there is a bi-directional interaction of the phonetic categories between L1 and L2 subsystems. L2 learning is inevitably affected by L2 learners' previously existed L1 knowledge. On the other hand, the phonetic input for the successful development of new L2 sounds cannot be exactly the same as the perceived input by native speakers of L2 (Flege, 1995; Flege & Bohn, 2021). Thus, L2 learners are unable to obtain equally matching performances like those by native speakers of a target L2. L2LP (Escudero, 2005) holds a contrary opinion and suggests that L2 learners can attain native-like L2 perception with an intact L1 perception grammar since L1 and L2 perception function separately in two systems.

The present study investigates Mandarin speakers' ultimate attainment in their learning of Cantonese syllable-final segment by adopting the accuracy rate of targets as an important indicator. If indicated only by the accuracy rate, the perception accuracy of the experimental group observed in Experiment 4 (post-test) agrees with the L2LP hypothesis of native-like ultimate attainment of L2 phonetic learning. This group achieved a native-like perception score above 90% in all Cantonese syllable-final segments after the intervention. Due to the lack of acoustic analysis in the current study, future research is necessary to decide whether the acoustic cues adopted in the perception of Cantonese syllable-final segments by Mandarin speakers are as native as those by Cantonese speakers. Their production abilities of targets, however, were still developing and only the production scores of targets [-p̚], [-m], [-n], [-ŋ], [-∅] satisfied the criterion at 70%. In SLM (Flege, 1995) and SLM-r (Flege & Bohn, 2021), the ultimate attainment of L2 phonetic learning depends importantly on the perceived cross-L1-L2 assimilation patterns of target sounds. SLM-r (Flege & Bohn, 2021) further specifies that the necessary amount of L2 exposure to realize adequate perceived cross-L1-L2 similarity would be speculated by the complexity of L2 learning targets which is codetermined by the universal occurrence frequency and the monolingual acquisition time of such sounds. In this case, it is pending whether Mandarin speakers can realize satisfactory native-like production accuracy of Cantonese syllable-final segments based on the observed result of the current study, awaiting further exploration by future research.

Chapter 8 Conclusion

8.1 Major Findings of the Current Study

Four experiments were incorporated in this dissertation to tackle four research questions centering around ① the cross-linguistic similarity of syllable-final segments between Mandarin and Cantonese, ② the developmental stages of Cantonese syllable-final segments by Mandarin speakers within the SLM-r framework, ③ the perception-production interface during the gradual development of targets, and ④ the effectiveness of perception training on promoting Mandarin speakers' learning of targets. The Perceptual Assimilation Test was administered to gauge the cross-Mandarin-Cantonese similarity of syllable-final segments by Mandarin speakers in Experiment 1. SLM-r (Flege & Bohn, 2021) was deployed to make predictions on how Cantonese syllable-final segments would be developed by an experimental group of Mandarin speakers through pre-test (Experiment 2), seven sessions of perception training (Experiment 3), and post-test

(Experiment 4). During the perception training in Experiment 3, the perception-production relationship was discussed. The effectiveness of perception training was substantiated through the gradual learning of targets by the experimental group in Experiment 3 and the advantage of this group over the untrained control group in Experiment 4. Major findings and implications are concluded in this section.

According to the results of the Perceptual Assimilation test, Cantonese [-∅] represented a reasonably good exemplar to Mandarin [-∅]. On the other hand, moderately fit exemplars were represented by Cantonese [-p̚], [-t̚] and [-k̚] to Mandarin [-∅], Cantonese [-m] to Mandarin [-ŋ], Cantonese [-n] to Mandarin [-n] and Cantonese [-ŋ] to Mandarin [-ŋ]. To interpret the results with SLM-r, Cantonese [-∅] was classified as an identical category with the least learning difficulties and the other six targets were categorized as less learnable similar sounds to their Mandarin counterparts.

These SLM-r hypotheses prepare the theoretical ground for the discussion of the development of Cantonese syllable-final segments by Mandarin speakers. In pre-test, identical target [-∅] generally outperformed similar targets [-p̚], [-t̚], [-k̚], [-m], [-n] and [-ŋ] in both perception and production. As intervention progressed, learnable identical target ([-∅]) was developed earlier than those more difficult similar sounds ([-m], [-n], [-ŋ], [-p̚], [-t̚] and [-k̚]) by Mandarin speakers. Mandarin speakers' production abilities of similar targets [-t̚] and [-k̚] were still unsatisfying until the post-test.

Diverse patterns of perception-production interface as a function of the learnability of different targets in the L2 phonetic learning are pointed out in the current study instead of those one-size-fits-all patterns proposed by previous studies. In the case of less learnable similar targets [-p˺], [-t˺], [-k˺], [-m], [-n] and [-ŋ], perception and production were correlated and performed a perception-precedence pattern as proposed by SLM (Flege, 1995). On the other hand, the perception and production of the identical target [-∅] were completed with a ceiling accuracy score when the experiments were initiated and showed no correlation during the developmental procedure.

Perception training is empirically testified to be effective in promoting Mandarin speakers' learning of targets. The perception and production of Cantonese syllable-final segments by Mandarin speakers performed a gradual improvement during the intervention. This trained group benefited directly from the perception training and then outperformed the untrained group in the post-test after the intervention.

As observed in the present study, the suspension of contrastive distribution of coronal and dorsal targets in Cantonese could pose higher learning difficulties in Mandarin speakers' development of targets [-t˺] and [-k˺]. To this end, heuristic amendments to the application of theoretical models in L2 phonetic learning are advanced in this dissertation. It is suggested that a more comprehensive prediction of L2 phonetic learning should integrate the perceived cross-L1-L2 similarity with the distribution patterns of L2 targets.

8.2 Significance of the Current Study

This section outlines the empirical and theoretical significance of the current study.

With the development of Guangdong-Hong Kong-Macao Greater Bay Area, the cross-border contact between Hong Kong residents and mainland Chinese has become closer. It is increasingly popular for Mandarin native speakers to learn Hong Kong Cantonese. In Cantonese phonological system, syllable-final segments ([-p˺], [-t˺], [-k˺], [-m], [-n], [-ŋ] and [-∅]) which are undergoing a coronal-dorsal merging tendency ([-t˺]-[-k˺] and [-n]-[-ŋ]) could be novel to Mandarin speakers. However, little is known about Mandarin speakers' learning of Cantonese phonetics with even less attention paid to Mandarin speakers' perception and production development of seven Cantonese syllable-final segments. Empirically, this study supplements these insufficiently discussed topics with valuable experimental evidence.

On the other hand, the theoretical significance of the current dissertation is threefold as below.

As verified in previous studies, the reviewed L2 phonetic models successfully predict the L2 speech learning results according to the cross-L1-L2 similarity in most cases (e.g., Sheldon & Strange, 1982; Best et al., 2003; Francis et al., 2008; Mayr & Escudero, 2010; Peperkamp & Bouchon, 2011; Darcy & Krüger, 2012; Pilus, 2016; Yazawa et al., 2017; Chao et al., 2019; Chappell, 2019; Luo et al., 2020). Taking a rarely discussed L1-L2 pair (Mandarin speakers' learning of

Cantonese syllable-final segments) as a starting point, the current study generally testifies the validity of the SLM-r framework (see Section 4.4, 5.4, 6.4, and 7.1.1 for detailed discussion). Meanwhile, PAM-L2 and L2LP hypotheses provide an alternative account for the lowest learnability of Cantonese merging coronal-dorsal targets. This is in line with the preceding research and replenishes the comparison in the applicability of different L2 phonetic models. The previous studies, however, seldom reflect on the latent drawbacks of the extant frameworks. Therefore, another more important intention of the current study is to uncover the inadequacies of extant L2 phonetic models and provide plausible heuristic amendments for the better application of these models (see detailed illustration in Section 7.3). With the detected lowest learnability of Cantonese merging coronal-dorsal targets, the current study puts forward that the aforementioned L2 phonetic theories overwhelmingly target on contrastively distributed L2 sounds but pay little attention to a situation where target segments might be experiencing a sound merger. The learning difficulty and consequence of merging L2 sounds cannot be sufficiently predicted through only the perceptual cross-L1-L2 similarity. It is therefore suggested that sound changes in distribution patterns of L2 targets should be integrated into the hypotheses of the above models when predicting L2 phonetic learning. With this being done, the explanatory power of these L2 phonetic models can be strengthened. In the case that different L2 tokens are classified into the same categories according to the cross-L1-L2 similarity, distribution patterns of L2 targets can be used to weight such categorization results and further decide that merging L2 tokens are more difficult than those contrastively-distributed ones for L2 learners. Collectively, the results of the current study not only shed new light on the application of SLM-r, PAM-L2, and L2LP (refers to the discussion in Section 7.1.1 and 7.2) but also guide future studies to extend these frameworks and address the learning of L2 sounds which might be undergoing different phonological processes (e.g., merging, assimilation, co-articulation, etc.) in other languages.

The present dissertation secondly proposes a diverse and plastic L2 perception-production relationship varying according to the learnability of different L2 targets. In the case of the current study, the perceptual cross-L1-L2 similarity in SLM-r can be an important indicator for the learnability of different L2 targets. Identical targets with higher learnability would achieve satisfied perception and production simultaneously, while the learning of less learnable similar targets performs a perception-precedence pattern. In addition to the previously proposed production-precedence pattern (Goto, 1971; Sheldon & Strange, 1982; McCandliss et al., 2002; Hattori & Iverson, 2009; Ingvalson et al., 2011), perception-precedence pattern (Flege, 1995), and perception-production co-evolve pattern (Flege & Bohn, 2021), this study provides another possible account for a better understanding of the perception-production relationship across different L2 targets.

Lastly, the effectiveness of laboratory perception training is verified in this dissertation. Furthermore, it is newly suggested that the observed effectiveness of training could be related to the perception-production interface patterns represented by different L2 targets. As most of the learning targets tend to rely more on the perception-precedence pattern, they realize significant perception gains through a direct advantage of perception training. The success in perception then is generalized to the production improvement. These results bridge training methodologies and perception-production interface patterns in L2 phonetic learning. It provides important insights into the design of L2 laboratory training and cues future studies to adjust training arrangements according to the relationship between L2 perception and production of different targets.

8.3 Reflection of the Current Study and Suggestion for the Future Study

Reflection on methodology issues and theoretical application of the current study is discussed in this section, according to which, the suggestions for future study are proposed.

Different rating schemes are utilized in previous studies when deciding participants' accuracy of production. In these rating schemes, two phoneticians (Dankovičová, 1999; Nicolaidis, Edwards, Beckman, & Tserdanelis; 2003), three phoneticians (Yoneyama, Beckman, & Edwards, 2003) or even twenty phoneticians (Isaacs & Thomson, 2013) are recruited to complete the grading. The current study involves only two raters. It is suggested that the arbitrariness and bias of this two-rater scheme are minimized since these two raters are well-trained phoneticians who also use Cantonese as the native language and their grading scores are verified by the inter-rater reliabilities test. However, some empirical studies attend to lower bias and individual differences in the grading by employing more raters. It would be better to adopt a more objective rater scheme for future study.

It is encouraged to consider alternative presentation methods of the coronal-dorsal contrast in the experimental stimuli. As depicted in preceding studies, the coronal-dorsal merging of Cantonese syllable-final segments is still developing (Bauer, 1979; Yeung, 1981; Chen, 1999; Zee, 1999a & 1999b; Law et al., 2001; Wong, 2005; Ding, 2010; Bauer & Benedict, 2011; To et al., 2013; To et al., 2015) and is only more popular among the younger generation of Cantonese native speakers (Chen, 1999; Wong, 2005). There could be a long way ahead before such merger is completed and acknowledged by the Cantonese community. To better reflect the phonological fact of the syllable-final coronal-dorsal merger in Cantonese, the coronal-dorsal contrast of targeted segments is preserved in the current study. It remains unclear whether Mandarin speakers would manage a native-like coronal-dorsal merger or develop the merging targets as two independent ones with accumulating immersion in daily Cantonese. This leaves room for future research with different stimuli paradigm. For example, the phonetic variants of syllable-final coronal and dorsal targets can be introduced into experiments. Such a design will shed light on the investigation about the learning of merging categories in L2 speech development.

Future studies can also explore how other training paradigms would modulate L2 phonetic learning. Instead of utilizing only L2 targets as stimuli in the perceptual intervention, discrimination training of cross-L1-L2 similarity between an L2 sound and its L1 correlate is suggested by Chan (2012). L2 learners' abilities to discern the phonetic differences of L1-L2 sound pairs for category formation of new L2 sounds would be strengthened via such training. There is also extant literature attempting to investigate the mediation effects of production training on L2 phonetic learning (Akahane-Yamada, 1998; Hirata, 2004; Hattori, 2009; Linebaugh & Roche, 2015; Thomson & Derwing, 2015; Feng, 2020; Zhang & Yuan, 2020). It can help to decide a better training strategy for promoting Mandarin speakers' learning of Cantonese syllable-final segments after comparing the training effects of different procedures.

Flege (2021) focuses on the methodological issues in L2 speech research and attempts to propose appropriate methods eliciting L2 data for the interpretation of SLM/SLM-r. However, most of the methods in Flege (2021) were exemplified with Italian/Spanish as the L1 and English as the L2, which is not necessarily applicable to other L1-L2 pairs. It is therefore suggested that researchers of L2 speech can adjust experimental methods according to available resources flexibly and should not negate alternative methods which could yield empirically similar results. Future research can be conducted under the SLM/SLM-r framework with a strictly designed study following Flege (2021)'s paradigm.

In addition to the above methodology issues, this section finally touches upon the theoretical exploration of the present study. Among the commonly used measurement methods of cross-L1-L2 similarity (the comparison of ① IPA symbols, ② acoustic similarity, ③ articulatory similarity, and ④ perceptual similarity), the current study mainly focused on the comparison of IPA transcription and the perceptual similarity of syllable-final segments between Mandarin and Cantonese. The results of cross-L1-L2 similarity derived from the other two methods (acoustic similarity and articulatory similarity) may support or negate the theoretical hypotheses in the current study, awaiting testification of future research.

In Section 7.2, alternative hypotheses by PAM-L2 and L2LP provide a possible account for the relatively higher learning difficulty of Cantonese contrast [-t˺]-[-k˺], which supplement the SLM-r hypotheses in the current study on such issues. In preceding studies of PAM-L2 and L2LP, discrimination tasks, which assess non-native listeners' discriminative abilities in sound contrasts, are often used since these two models focus on the non-native contrasts rather than individual non-native sounds. Due to the lack of discrimination data in the present study, the alternative hypotheses by PAM-L2 and L2LP look forward to being corroborated by future studies. On the other hand, it is advanced that the suspension of contrastive distribution of L2 categories can pose higher learning difficulties in L2 phonetic development. It could furnish valuable information when future studies attempt to predict the development of different L2 categories based on not only the perceived cross-L1-L2 similarity but also the distribution patterns of L2 targets. This needs to be further verified with upcoming empirical data about the L2 phonetic learning of merging categories in other languages.

Appendixes

Appendix 1 Summary of Language Background Questionnaire

As summed up below, Mandarin Group 1 was only assessed in Experiment 1. In Experiment 2, the control group was served by the Cantonese Group (CG), while the experimental group consisted of Mandarin Group 2 (MG 2) and Mandarin Group 3 (MG 3). Then, Mandarin Group 2 went through the intervention of Experiment 3. This group again participated as the experimental group and was assessed together with the control group (Mandarin Group 3) in the post-test (Experiment 4).

I. Personal Information

1. Your age?

Group (Subject No.)	Age Rage (years)	Mean Age ± SD (years)
CG (16)	18 ~ 30	22.75 ± 3.54
MG 1 (20)	18 ~ 30	22.85 ± 4.09
MG 2 (16)	20 ~ 29	24.5 ± 2.21
MG 3 (16)	19 ~ 30	25.2 ± 3.01

2. Your gender?

Group (Subject No.)	Males	Females
CG (16)	6	10
MG 1 (20)	9	11
MG 2 (16)	6	10

MG 3 (16)	9	7

3. Where is your hometown?

Group (Subject No.)	Hometown
CG (16)	Hong Kong (16)
MG 1 (20)	Hubei (7), Shănxi (4), Heilongjiang (3), Henan (2), Shandong (2), Anhui (1), Shānxi (1)
MG 2 (16)	Shandong (4), Heilongjiang (3), Liaoning (2), Inner Mongolia (2), Anhui (1), Gansu (1), Hebei (1), Tianjin (1), Xinjiang (1)
MG 3 (16)	Shandong (5), Heilongjiang (3), Gansu (2), Anhui (1), Dalian (1), Hebei (1), Ningxia (1), Tianjin (1), Xinjiang (1)

4. When did you move to Hong Kong?

See Question 5 for the time of exposure to Hong Kong Cantonese.

5. How long have you been in Hong Kong?

Group (Subject No.)	Time of Exposure
CG (16)	22.75 ± 3.54 years
MG 1 (20)	12.9 ± 7.87 months
MG 2 (16)	10.03 ± 6.09 months
MG 3 (16)	8.94 ± 7.02 months

II. Language Usage

1. What is your first language (Multiple choices)?

□ Cantonese □ English □ Mandarin □ Others

Group (Subject No.)	First Language
CG (16)	Cantonese (16)
MG 1 (20)	Mandarin (20)
MG 2 (16)	Mandarin (16)
MG 3 (16)	Mandarin (16)

2. Except for Cantonese, English, and Mandarin, what other languages can you speak and indicate the level of proficiency?

Group (Subject No.)	Other Languages and Proficiency
CG (16)	No
MG 1 (20)	No
MG 2 (16)	No
MG 3 (16)	No

3. What languages do you use in your family (Multiple choices)?

□ Cantonese □ English □ Mandarin □ Others

Group (Subject No.)	Language Used in Family
CG (16)	Cantonese (16)
MG 1 (20)	Mandarin (20)
MG 2 (16)	Mandarin (16)
MG 3 (16)	Mandarin (16)

4. What languages do you use in school (Multiple choices)?

□ Cantonese □ English □ Mandarin □ Others

Group (Subject No.)	Language Used in School
CG (16)	Cantonese (3), Cantonese & English (13)
MG 1 (20)	Mandarin (13), Mandarin & English (7)
MG 2 (16)	Mandarin (8), Mandarin & English (8)
MG 3 (16)	Mandarin (6), Mandarin & English (10)

5. How often do you use Cantonese?

□ Always □ Sometimes □ Seldom/ Never

Group (Subject No.)	Frequency of use (Cantonese)
CG (16)	Always (16)
MG 1 (20)	Seldom/ Never (20)
MG 2 (16)	Seldom/ Never (16)
MG 3 (16)	Seldom/ Never (16)

6. Have you ever had Cantonese pronunciation training?

□ Yes (How long?) □ No

Group (Subject No.)	History of Cantonese Pronunciation Training
CG (16)	No (16)
MG 1 (20)	No (20)
MG 2 (16)	No (16)
MG 3 (16)	No (16)

7. Is your hearing normal?

□ Yes □ No

Group (Subject No.)	Normal Hearing
CG (16)	Yes (16)
MG 1 (20)	Yes (20)
MG 2 (16)	Yes (16)
MG 3 (16)	Yes (16)

8. Is your articulation normal?

□ Yes □ No

Group (Subject No.)	Normal Articulation
CG (16)	Yes (16)
MG 1 (20)	Yes (20)
MG 2 (16)	Yes (16)
MG 3 (16)	Yes (16)

Appendix 2 Stimuli, Trials of Experiment 1

2.1 Four-alternative Forced-choice Categorization Test

No.	Onset	Nucleus	Target	Tone	Option
1	k	a	p˺	33	[-ø]-[-n]-[-ŋ]-wu
2	k	a	t˺	33	[-ø]-[-n]-[-ŋ]-wu
3	k	a	k˺	33	[-ø]-[-n]-[-ŋ]-wu
4	k	a	m	33	[-ø]-[-n]-[-ŋ]-wu
5	k	a	n	33	[-ø]-[-n]-[-ŋ]-wu
6	k	a	ŋ	33	[-ø]-[-n]-[-ŋ]-wu
7	tʃ	a	p˺	33	[-ø]-[-n]-[-ŋ]-wu
8	tʃ	a	t˺	33	[-ø]-[-n]-[-ŋ]-wu
9	tʃ	a	k˺	33	[-ø]-[-n]-[-ŋ]-wu
10	tʃ	a	m	33	[-ø]-[-n]-[-ŋ]-wu
11	tʃ	a	n	33	[-ø]-[-n]-[-ŋ]-wu
12	tʃ	a	ŋ	33	[-ø]-[-n]-[-ŋ]-wu
13	tʃ	a	p˺	22	[-ø]-[-n]-[-ŋ]-wu
14	tʃ	a	t˺	22	[-ø]-[-n]-[-ŋ]-wu
15	tʃ	a	k˺	22	[-ø]-[-n]-[-ŋ]-wu
16	tʃ	a	m	22	[-ø]-[-n]-[-ŋ]-wu
17	tʃ	a	n	22	[-ø]-[-n]-[-ŋ]-wu
18	tʃ	a	ŋ	22	[-ø]-[-n]-[-ŋ]-wu

19	tʃ	ɐ	p̚	33	[-∅]-[-n]-[-ŋ]-wu
20	tʃ	ɐ	t̚	33	[-∅]-[-n]-[-ŋ]-wu
21	tʃ	ɐ	k̚	33	[-∅]-[-n]-[-ŋ]-wu
22	tʃ	ɐ	m	33	[-∅]-[-n]-[-ŋ]-wu
23	tʃ	ɐ	n	33	[-∅]-[-n]-[-ŋ]-wu
24	tʃ	ɐ	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
25	k	a	∅	33	[-∅]-[-n]-[-ŋ]-wu
26	tʃ	a	∅	33	[-∅]-[-n]-[-ŋ]-wu
27	tʃ	a	∅	22	[-∅]-[-n]-[-ŋ]-wu
28	k	a	p̚	33	[-∅]-[-n]-[-ŋ]-wu
29	k	a	t̚	33	[-∅]-[-n]-[-ŋ]-wu
30	k	a	k̚	33	[-∅]-[-n]-[-ŋ]-wu
31	k	a	m	33	[-∅]-[-n]-[-ŋ]-wu
32	k	a	n	33	[-∅]-[-n]-[-ŋ]-wu
33	k	a	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
34	tʃ	a	p̚	33	[-∅]-[-n]-[-ŋ]-wu
35	tʃ	a	t̚	33	[-∅]-[-n]-[-ŋ]-wu
36	tʃ	a	k̚	33	[-∅]-[-n]-[-ŋ]-wu
37	tʃ	a	m	33	[-∅]-[-n]-[-ŋ]-wu
38	tʃ	a	n	33	[-∅]-[-n]-[-ŋ]-wu
39	tʃ	a	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
40	tʃ	a	p̚	22	[-∅]-[-n]-[-ŋ]-wu

41	tʃ	a	t̚	22	[-∅]-[-n]-[-ŋ]-wu
42	tʃ	a	k̚	22	[-∅]-[-n]-[-ŋ]-wu
43	tʃ	a	m	22	[-∅]-[-n]-[-ŋ]-wu
44	tʃ	a	n	22	[-∅]-[-n]-[-ŋ]-wu
45	tʃ	a	ŋ	22	[-∅]-[-n]-[-ŋ]-wu
46	tʃ	ɐ	p̚	33	[-∅]-[-n]-[-ŋ]-wu
47	tʃ	ɐ	t̚	33	[-∅]-[-n]-[-ŋ]-wu
48	tʃ	ɐ	k̚	33	[-∅]-[-n]-[-ŋ]-wu
49	tʃ	ɐ	m	33	[-∅]-[-n]-[-ŋ]-wu
50	tʃ	ɐ	n	33	[-∅]-[-n]-[-ŋ]-wu
51	tʃ	ɐ	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
52	k	a	∅	33	[-∅]-[-n]-[-ŋ]-wu
53	tʃ	a	∅	33	[-∅]-[-n]-[-ŋ]-wu
54	tʃ	a	∅	22	[-∅]-[-n]-[-ŋ]-wu
55	k	a	p̚	33	[-∅]-[-n]-[-ŋ]-wu
56	k	a	t̚	33	[-∅]-[-n]-[-ŋ]-wu
57	k	a	k̚	33	[-∅]-[-n]-[-ŋ]-wu
58	k	a	m	33	[-∅]-[-n]-[-ŋ]-wu
59	k	a	n	33	[-∅]-[-n]-[-ŋ]-wu
60	k	a	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
61	tʃ	a	p̚	33	[-∅]-[-n]-[-ŋ]-wu
62	tʃ	a	t̚	33	[-∅]-[-n]-[-ŋ]-wu

63	tʃ	a	k˺	33	[-ø]-[-n]-[-ŋ]-wu
64	tʃ	a	m	33	[-ø]-[-n]-[-ŋ]-wu
65	tʃ	a	n	33	[-ø]-[-n]-[-ŋ]-wu
66	tʃ	a	ŋ	33	[-ø]-[-n]-[-ŋ]-wu
67	tʃ	a	p˺	22	[-ø]-[-n]-[-ŋ]-wu
68	tʃ	a	t˺	22	[-ø]-[-n]-[-ŋ]-wu
69	tʃ	a	k˺	22	[-ø]-[-n]-[-ŋ]-wu
70	tʃ	a	m	22	[-ø]-[-n]-[-ŋ]-wu
71	tʃ	a	n	22	[-ø]-[-n]-[-ŋ]-wu
72	tʃ	a	ŋ	22	[-ø]-[-n]-[-ŋ]-wu
73	tʃ	ɐ	p˺	33	[-ø]-[-n]-[-ŋ]-wu
74	tʃ	ɐ	t˺	33	[-ø]-[-n]-[-ŋ]-wu
75	tʃ	ɐ	k˺	33	[-ø]-[-n]-[-ŋ]-wu
76	tʃ	ɐ	m	33	[-ø]-[-n]-[-ŋ]-wu
77	tʃ	ɐ	n	33	[-ø]-[-n]-[-ŋ]-wu
78	tʃ	ɐ	ŋ	33	[-ø]-[-n]-[-ŋ]-wu
79	k	a	ø	33	[-ø]-[-n]-[-ŋ]-wu
80	tʃ	a	ø	33	[-ø]-[-n]-[-ŋ]-wu
81	tʃ	a	ø	22	[-ø]-[-n]-[-ŋ]-wu
82	k	a	p˺	33	[-ø]-[-n]-[-ŋ]-wu
83	k	a	t˺	33	[-ø]-[-n]-[-ŋ]-wu
84	k	a	k˺	33	[-ø]-[-n]-[-ŋ]-wu

85	k	a	m	33	[-∅]-[-n]-[-ŋ]-wu
86	k	a	n	33	[-∅]-[-n]-[-ŋ]-wu
87	k	a	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
88	tʃ	a	p˺	33	[-∅]-[-n]-[-ŋ]-wu
89	tʃ	a	t˺	33	[-∅]-[-n]-[-ŋ]-wu
90	tʃ	a	k˺	33	[-∅]-[-n]-[-ŋ]-wu
91	tʃ	a	m	33	[-∅]-[-n]-[-ŋ]-wu
92	tʃ	a	n	33	[-∅]-[-n]-[-ŋ]-wu
93	tʃ	a	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
94	tʃ	a	p˺	22	[-∅]-[-n]-[-ŋ]-wu
95	tʃ	a	t˺	22	[-∅]-[-n]-[-ŋ]-wu
96	tʃ	a	k˺	22	[-∅]-[-n]-[-ŋ]-wu
97	tʃ	a	m	22	[-∅]-[-n]-[-ŋ]-wu
98	tʃ	a	n	22	[-∅]-[-n]-[-ŋ]-wu
99	tʃ	a	ŋ	22	[-∅]-[-n]-[-ŋ]-wu
100	tʃ	ɐ	p˺	33	[-∅]-[-n]-[-ŋ]-wu
101	tʃ	ɐ	t˺	33	[-∅]-[-n]-[-ŋ]-wu
102	tʃ	ɐ	k˺	33	[-∅]-[-n]-[-ŋ]-wu
103	tʃ	ɐ	m	33	[-∅]-[-n]-[-ŋ]-wu
104	tʃ	ɐ	n	33	[-∅]-[-n]-[-ŋ]-wu
105	tʃ	ɐ	ŋ	33	[-∅]-[-n]-[-ŋ]-wu
106	k	a	∅	33	[-∅]-[-n]-[-ŋ]-wu

107	tʃ	a	∅	33	[-∅]-[-n]-[-ŋ]-wu
108	tʃ	a	∅	22	[-∅]-[-n]-[-ŋ]-wu

2.2 Goodness-of-fit Rating Test

No.	Onset	Nucleus	Audio Target	Display Target	Tone	Rating Scale
1	k	a	p˺	n	33	1 to 5 (least similar to very similar)
2	k	a	p˺	ŋ	33	1 to 5 (least similar to very similar)
3	k	a	p˺	∅	33	1 to 5 (least similar to very similar)
4	k	a	t˺	n	33	1 to 5 (least similar to very similar)
5	k	a	t˺	ŋ	33	1 to 5 (least similar to very similar)
6	k	a	t˺	∅	33	1 to 5 (least similar to very similar)
7	k	a	k˺	n	33	1 to 5 (least similar to very similar)
8	k	a	k˺	ŋ	33	1 to 5 (least similar to very similar)
9	k	a	k˺	∅	33	1 to 5 (least similar to very similar)
10	k	a	m	n	33	1 to 5 (least similar to very similar)
11	k	a	m	ŋ	33	1 to 5 (least similar to very similar)
12	k	a	m	∅	33	1 to 5 (least similar to very similar)
13	k	a	n	n	33	1 to 5 (least similar to very similar)
14	k	a	n	ŋ	33	1 to 5 (least similar to very similar)
15	k	a	n	∅	33	1 to 5 (least similar to very similar)
16	k	a	ŋ	n	33	1 to 5 (least similar to very similar)
17	k	a	ŋ	ŋ	33	1 to 5 (least similar to very similar)

18	k	a	ŋ	∅	33	1 to 5 (least similar to very similar)
19	tʃ	a	p˺	n	33	1 to 5 (least similar to very similar)
20	tʃ	a	p˺	ŋ	33	1 to 5 (least similar to very similar)
21	tʃ	a	p˺	∅	33	1 to 5 (least similar to very similar)
22	tʃ	a	t˺	n	33	1 to 5 (least similar to very similar)
23	tʃ	a	t˺	ŋ	33	1 to 5 (least similar to very similar)
24	tʃ	a	t˺	∅	33	1 to 5 (least similar to very similar)
25	tʃ	a	k˺	n	33	1 to 5 (least similar to very similar)
26	tʃ	a	k˺	ŋ	33	1 to 5 (least similar to very similar)
27	tʃ	a	k˺	∅	33	1 to 5 (least similar to very similar)
28	tʃ	a	m	n	33	1 to 5 (least similar to very similar)
29	tʃ	a	m	ŋ	33	1 to 5 (least similar to very similar)
30	tʃ	a	m	∅	33	1 to 5 (least similar to very similar)
31	tʃ	a	n	n	33	1 to 5 (least similar to very similar)
32	tʃ	a	n	ŋ	33	1 to 5 (least similar to very similar)
33	tʃ	a	n	∅	33	1 to 5 (least similar to very similar)
34	tʃ	a	ŋ	n	33	1 to 5 (least similar to very similar)
35	tʃ	a	ŋ	ŋ	33	1 to 5 (least similar to very similar)
36	tʃ	a	ŋ	∅	33	1 to 5 (least similar to very similar)
37	tʃ	a	p˺	n	22	1 to 5 (least similar to very similar)
38	tʃ	a	p˺	ŋ	22	1 to 5 (least similar to very similar)
39	tʃ	a	p˺	∅	22	1 to 5 (least similar to very similar)

40	tʃ	a	t˺	n	22	1 to 5 (least similar to very similar)
41	tʃ	a	t˺	ŋ	22	1 to 5 (least similar to very similar)
42	tʃ	a	t˺	ø	22	1 to 5 (least similar to very similar)
43	tʃ	a	k˺	n	22	1 to 5 (least similar to very similar)
44	tʃ	a	k˺	ŋ	22	1 to 5 (least similar to very similar)
45	tʃ	a	k˺	ø	22	1 to 5 (least similar to very similar)
46	tʃ	a	m	n	22	1 to 5 (least similar to very similar)
47	tʃ	a	m	ŋ	22	1 to 5 (least similar to very similar)
48	tʃ	a	m	ø	22	1 to 5 (least similar to very similar)
49	tʃ	a	n	n	22	1 to 5 (least similar to very similar)
50	tʃ	a	n	ŋ	22	1 to 5 (least similar to very similar)
51	tʃ	a	n	ø	22	1 to 5 (least similar to very similar)
52	tʃ	a	ŋ	n	22	1 to 5 (least similar to very similar)
53	tʃ	a	ŋ	ŋ	22	1 to 5 (least similar to very similar)
54	tʃ	a	ŋ	ø	22	1 to 5 (least similar to very similar)
55	tʃ	ɐ	p˺	n	33	1 to 5 (least similar to very similar)
56	tʃ	ɐ	p˺	ŋ	33	1 to 5 (least similar to very similar)
57	tʃ	ɐ	p˺	ø	33	1 to 5 (least similar to very similar)
58	tʃ	ɐ	t˺	n	33	1 to 5 (least similar to very similar)
59	tʃ	ɐ	t˺	ŋ	33	1 to 5 (least similar to very similar)
60	tʃ	ɐ	t˺	ø	33	1 to 5 (least similar to very similar)
61	tʃ	ɐ	k˺	n	33	1 to 5 (least similar to very similar)

62	tʃ	ɐ	k̚	ŋ	33	1 to 5 (least similar to very similar)
63	tʃ	ɐ	k̚	∅	33	1 to 5 (least similar to very similar)
64	tʃ	ɐ	m	n	33	1 to 5 (least similar to very similar)
65	tʃ	ɐ	m	ŋ	33	1 to 5 (least similar to very similar)
66	tʃ	ɐ	m	∅	33	1 to 5 (least similar to very similar)
67	tʃ	ɐ	n	n	33	1 to 5 (least similar to very similar)
68	tʃ	ɐ	n	ŋ	33	1 to 5 (least similar to very similar)
69	tʃ	ɐ	n	∅	33	1 to 5 (least similar to very similar)
70	tʃ	ɐ	ŋ	n	33	1 to 5 (least similar to very similar)
71	tʃ	ɐ	ŋ	ŋ	33	1 to 5 (least similar to very similar)
72	tʃ	ɐ	ŋ	∅	33	1 to 5 (least similar to very similar)
73	k	a	∅	n	33	1 to 5 (least similar to very similar)
74	k	a	∅	ŋ	33	1 to 5 (least similar to very similar)
75	k	a	∅	∅	33	1 to 5 (least similar to very similar)
76	tʃ	a	∅	n	33	1 to 5 (least similar to very similar)
77	tʃ	a	∅	ŋ	33	1 to 5 (least similar to very similar)
78	tʃ	a	∅	∅	33	1 to 5 (least similar to very similar)
79	tʃ	a	∅	n	22	1 to 5 (least similar to very similar)
80	tʃ	a	∅	ŋ	22	1 to 5 (least similar to very similar)
81	tʃ	a	∅	∅	22	1 to 5 (least similar to very similar)

Appendix 3 Stimuli, Trials of Experiment 2/Experiment 4

3.1 Two-alternative Forced-choice Identification Test

No.	Onset	Nucleus	Coda A	Coda B	Target	Tone
1	k	a	p̚	m	p̚	33
2	k	a	p̚	m	m	33
3	k	a	p̚	t̚	p̚	33
4	k	a	p̚	t̚	t̚	33
5	k	a	p̚	n	p̚	33
6	k	a	p̚	n	n	33
7	k	a	p̚	k̚	p̚	33
8	k	a	p̚	k̚	k̚	33
9	k	a	p̚	ŋ	p̚	33
10	k	a	p̚	ŋ	ŋ	33
11	k	a	m	t̚	m	33
12	k	a	m	t̚	t̚	33
13	k	a	m	n	m	33
14	k	a	m	n	n	33
15	k	a	m	k̚	m	33
16	k	a	m	k̚	k̚	33
17	k	a	m	ŋ	m	33
18	k	a	m	ŋ	ŋ	33

19	k	a	t˺	n	t˺	33
20	k	a	t˺	n	n	33
21	k	a	t˺	k˺	t˺	33
22	k	a	t˺	k˺	k˺	33
23	k	a	t˺	ŋ	t˺	33
24	k	a	t˺	ŋ	ŋ	33
25	k	a	n	k˺	n	33
26	k	a	n	k˺	k˺	33
27	k	a	n	ŋ	n	33
28	k	a	n	ŋ	ŋ	33
29	k	a	k˺	ŋ	k˺	33
30	k	a	k˺	ŋ	ŋ	33
31	tʃ	a	p˺	m	p˺	33
32	tʃ	a	p˺	m	m	33
33	tʃ	a	p˺	t˺	p˺	33
34	tʃ	a	p˺	t˺	t˺	33
35	tʃ	a	p˺	n	p˺	33
36	tʃ	a	p˺	n	n	33
37	tʃ	a	p˺	k˺	p˺	33
38	tʃ	a	p˺	k˺	k˺	33
39	tʃ	a	p˺	ŋ	p˺	33
40	tʃ	a	p˺	ŋ	ŋ	33

41	tʃ	a	m	t˺	m	33
42	tʃ	a	m	t˺	t˺	33
43	tʃ	a	m	n	m	33
44	tʃ	a	m	n	n	33
45	tʃ	a	m	k˺	m	33
46	tʃ	a	m	k˺	k˺	33
47	tʃ	a	m	ŋ	m	33
48	tʃ	a	m	ŋ	ŋ	33
49	tʃ	a	t˺	n	t˺	33
50	tʃ	a	t˺	n	n	33
51	tʃ	a	t˺	k˺	t˺	33
52	tʃ	a	t˺	k˺	k˺	33
53	tʃ	a	t˺	ŋ	t˺	33
54	tʃ	a	t˺	ŋ	ŋ	33
55	tʃ	a	n	k˺	n	33
56	tʃ	a	n	k˺	k˺	33
57	tʃ	a	n	ŋ	n	33
58	tʃ	a	n	ŋ	ŋ	33
59	tʃ	a	k˺	ŋ	k˺	33
60	tʃ	a	k˺	ŋ	ŋ	33
61	tʃ	a	p˺	m	p˺	22
62	tʃ	a	p˺	m	m	22

63	tʃ	a	p˺	t˺	p˺	22
64	tʃ	a	p˺	t˺	t˺	22
65	tʃ	a	p˺	n	p˺	22
66	tʃ	a	p˺	n	n	22
67	tʃ	a	p˺	k˺	p˺	22
68	tʃ	a	p˺	k˺	k˺	22
69	tʃ	a	p˺	ŋ	p˺	22
70	tʃ	a	p˺	ŋ	ŋ	22
71	tʃ	a	m	t˺	m	22
72	tʃ	a	m	t˺	t˺	22
73	tʃ	a	m	n	m	22
74	tʃ	a	m	n	n	22
75	tʃ	a	m	k˺	m	22
76	tʃ	a	m	k˺	k˺	22
77	tʃ	a	m	ŋ	m	22
78	tʃ	a	m	ŋ	ŋ	22
79	tʃ	a	t˺	n	t˺	22
80	tʃ	a	t˺	n	n	22
81	tʃ	a	t˺	k˺	t˺	22
82	tʃ	a	t˺	k˺	k˺	22
83	tʃ	a	t˺	ŋ	t˺	22
84	tʃ	a	t˺	ŋ	ŋ	22

85	tʃ	a	n	k˺	n	22
86	tʃ	a	n	k˺	k˺	22
87	tʃ	a	n	ŋ	n	22
88	tʃ	a	n	ŋ	ŋ	22
89	tʃ	a	k˺	ŋ	k˺	22
90	tʃ	a	k˺	ŋ	ŋ	22
91	tʃ	ɐ	p˺	m	p˺	33
92	tʃ	ɐ	p˺	m	m	33
93	tʃ	ɐ	p˺	t˺	p˺	33
94	tʃ	ɐ	p˺	t˺	t˺	33
95	tʃ	ɐ	p˺	n	p˺	33
96	tʃ	ɐ	p˺	n	n	33
97	tʃ	ɐ	p˺	k˺	p˺	33
98	tʃ	ɐ	p˺	k˺	k˺	33
99	tʃ	ɐ	p˺	ŋ	p˺	33
100	tʃ	ɐ	p˺	ŋ	ŋ	33
101	tʃ	ɐ	m	t˺	m	33
102	tʃ	ɐ	m	t˺	t˺	33
103	tʃ	ɐ	m	n	m	33
104	tʃ	ɐ	m	n	n	33
105	tʃ	ɐ	m	k˺	m	33
106	tʃ	ɐ	m	k˺	k˺	33

107	tʃ	ɐ	m	ŋ	m	33
108	tʃ	ɐ	m	ŋ	ŋ	33
109	tʃ	ɐ	t˺	n	t˺	33
110	tʃ	ɐ	t˺	n	n	33
111	tʃ	ɐ	t˺	k˺	t˺	33
112	tʃ	ɐ	t˺	k˺	k˺	33
113	tʃ	ɐ	t˺	ŋ	t˺	33
114	tʃ	ɐ	t˺	ŋ	ŋ	33
115	tʃ	ɐ	n	k˺	n	33
116	tʃ	ɐ	n	k˺	k˺	33
117	tʃ	ɐ	n	ŋ	n	33
118	tʃ	ɐ	n	ŋ	ŋ	33
119	tʃ	ɐ	k˺	ŋ	k˺	33
120	tʃ	ɐ	k˺	ŋ	ŋ	33
121	k	a	p˺	ø	p˺	33
122	k	a	p˺	ø	ø	33
123	k	a	t˺	ø	t˺	33
124	k	a	t˺	ø	ø	33
125	k	a	k˺	ø	k˺	33
126	k	a	k˺	ø	ø	33
127	k	a	m	ø	m	33
128	k	a	m	ø	ø	33

129	k	a	n	∅	n	33
130	k	a	n	∅	∅	33
131	k	a	ŋ	∅	ŋ	33
132	k	a	ŋ	∅	∅	33
133	tʃ	a	p̚	∅	p̚	33
134	tʃ	a	p̚	∅	∅	33
135	tʃ	a	t̚	∅	t̚	33
136	tʃ	a	t̚	∅	∅	33
137	tʃ	a	k̚	∅	k̚	33
138	tʃ	a	k̚	∅	∅	33
139	tʃ	a	m	∅	m	33
140	tʃ	a	m	∅	∅	33
141	tʃ	a	n	∅	n	33
142	tʃ	a	n	∅	∅	33
143	tʃ	a	ŋ	∅	ŋ	33
144	tʃ	a	ŋ	∅	∅	33
145	tʃ	a	p̚	∅	p̚	22
146	tʃ	a	p̚	∅	∅	22
147	tʃ	a	t̚	∅	t̚	22
148	tʃ	a	t̚	∅	∅	22
149	tʃ	a	k̚	∅	k̚	22
150	tʃ	a	k̚	∅	∅	22

151	tʃ	a	m	ø	m	22
152	tʃ	a	m	ø	ø	22
153	tʃ	a	n	ø	n	22
154	tʃ	a	n	ø	ø	22
155	tʃ	a	ŋ	ø	ŋ	22
156	tʃ	a	ŋ	ø	ø	22

3.2 Monosyllable Repetition Test

No.	Onset	Nucleus	Coda	Tone
1	k	a	p̚	33
2	k	a	t̚	33
3	k	a	k̚	33
4	k	a	m	33
5	k	a	n	33
6	k	a	ŋ	33
7	tʃ	a	p̚	33
8	tʃ	a	t̚	33
9	tʃ	a	k̚	33
10	tʃ	a	m	33
11	tʃ	a	n	33
12	tʃ	a	ŋ	33
13	tʃ	a	p̚	22

14	tʃ	a	t˺	22
15	tʃ	a	k˺	22
16	tʃ	a	m	22
17	tʃ	a	n	22
18	tʃ	a	ŋ	22
19	tʃ	ɐ	p˺	33
20	tʃ	ɐ	t˺	33
21	tʃ	ɐ	k˺	33
22	tʃ	ɐ	m	33
23	tʃ	ɐ	n	33
24	tʃ	ɐ	ŋ	33
25	k	a	ø	33
26	tʃ	a	ø	33
27	tʃ	a	ø	22

Appendix 4 Stimuli, Trials of Experiment 3

4.1 Two-alternative Forced-choice Identification Test

No.	Onset	Nucleus	Coda A	Coda B	Target	Tone
1	tʃ	a	p˺	m	p˺	33
2	tʃ	a	p˺	m	m	33
3	tʃ	a	p˺	t˺	p˺	33
4	tʃ	a	p˺	t˺	t˺	33
5	tʃ	a	p˺	n	p˺	33
6	tʃ	a	p˺	n	n	33
7	tʃ	a	p˺	k˺	p˺	33
8	tʃ	a	p˺	k˺	k˺	33
9	tʃ	a	p˺	ŋ	p˺	33
10	tʃ	a	p˺	ŋ	ŋ	33
11	tʃ	a	m	t˺	m	33
12	tʃ	a	m	t˺	t˺	33
13	tʃ	a	m	n	m	33
14	tʃ	a	m	n	n	33
15	tʃ	a	m	k˺	m	33
16	tʃ	a	m	k˺	k˺	33
17	tʃ	a	m	ŋ	m	33
18	tʃ	a	m	ŋ	ŋ	33

19	tʃ	a	t˺	n	t˺	33
20	tʃ	a	t˺	n	n	33
21	tʃ	a	t˺	k˺	t˺	33
22	tʃ	a	t˺	k˺	k˺	33
23	tʃ	a	t˺	ŋ	t˺	33
24	tʃ	a	t˺	ŋ	ŋ	33
25	tʃ	a	n	k˺	n	33
26	tʃ	a	n	k˺	k˺	33
27	tʃ	a	n	ŋ	n	33
28	tʃ	a	n	ŋ	ŋ	33
29	tʃ	a	k˺	ŋ	k˺	33
30	tʃ	a	k˺	ŋ	ŋ	33
31	tʃ	a	p˺	ø	p˺	33
32	tʃ	a	p˺	ø	ø	33
33	tʃ	a	t˺	ø	t˺	33
34	tʃ	a	t˺	ø	ø	33
35	tʃ	a	k˺	ø	k˺	33
36	tʃ	a	k˺	ø	ø	33
37	tʃ	a	m	ø	m	33
38	tʃ	a	m	ø	ø	33
39	tʃ	a	n	ø	n	33
40	tʃ	a	n	ø	ø	33

41	tʃ	a	ŋ	∅	ŋ	33
42	tʃ	a	ŋ	∅	∅	33

4.2 Monosyllable Repetition Test

No.	Onset	Nucleus	Coda	Tone
1	tʃ	a	p̚	33
2	tʃ	a	t̚	33
3	tʃ	a	k̚	33
4	tʃ	a	m	33
5	tʃ	a	n	33
6	tʃ	a	ŋ	33
7	tʃ	a	∅	33

4.3 Discrimination Training

No.	Onset	Nucleus	Coda A	Coda B	Tone
1	tʃ	a	p̚	m	33
2	tʃ	a	p̚	t̚	33
3	tʃ	a	p̚	n	33
4	tʃ	a	p̚	k̚	33
5	tʃ	a	p̚	ŋ	33
6	tʃ	a	m	t	33
7	tʃ	a	m	n	33
8	tʃ	a	m	k̚	33

9	tʃ	a	m	ŋ	33
10	tʃ	a	t˺	n	33
11	tʃ	a	t˺	k˺	33
12	tʃ	a	t˺	ŋ	33
13	tʃ	a	n	k˺	33
14	tʃ	a	n	ŋ	33
15	tʃ	a	k˺	ŋ	33
16	tʃ	a	p˺	ø	33
17	tʃ	a	t˺	ø	33
18	tʃ	a	k˺	ø	33
19	tʃ	a	m	ø	33
20	tʃ	a	n	ø	33
21	tʃ	a	ŋ	ø	33
22	tʃ	a	p˺	p˺	33
23	tʃ	a	t˺	t˺	33
24	tʃ	a	k˺	k˺	33
25	tʃ	a	m	m	33
26	tʃ	a	n	n	33
27	tʃ	a	ŋ	ŋ	33
28	tʃ	a	ø	ø	33
29	tʃ	a	p˺	p˺	33
30	tʃ	a	t˺	t˺	33

31	tʃ	a	k˺	k˺	33
32	tʃ	a	m	m	33
33	tʃ	a	n	n	33
34	tʃ	a	ŋ	ŋ	33
35	tʃ	a	ø	ø	33
36	tʃ	a	p˺	p˺	33
37	tʃ	a	t˺	t˺	33
38	tʃ	a	k˺	k˺	33
39	tʃ	a	m	m	33
40	tʃ	a	n	n	33
41	tʃ	a	ŋ	ŋ	33
42	tʃ	a	ø	ø	33
43	tʃ	a	p˺	m	33
44	tʃ	a	p˺	t˺	33
45	tʃ	a	p˺	n	33
46	tʃ	a	p˺	k˺	33
47	tʃ	a	p˺	ŋ	33
48	tʃ	a	m	t˺	33
49	tʃ	a	m	n	33
50	tʃ	a	m	k˺	33
51	tʃ	a	m	ŋ	33
52	tʃ	a	t˺	n	33

53	tʃ	a	t˺	k˺	33
54	tʃ	a	t˺	ŋ	33
55	tʃ	a	n	k˺	33
56	tʃ	a	n	ŋ	33
57	tʃ	a	k˺	ŋ	33
58	tʃ	a	p˺	∅	33
59	tʃ	a	t˺	∅	33
60	tʃ	a	k˺	∅	33
61	tʃ	a	m	∅	33
62	tʃ	a	n	∅	33
63	tʃ	a	ŋ	∅	33
64	tʃ	a	p˺	p˺	33
65	tʃ	a	t˺	t˺	33
66	tʃ	a	k˺	k˺	33
67	tʃ	a	m	m	33
68	tʃ	a	n	n	33
69	tʃ	a	ŋ	ŋ	33
70	tʃ	a	∅	∅	33
71	tʃ	a	p˺	p˺	33
72	tʃ	a	t˺	t˺	33
73	tʃ	a	k˺	k˺	33
74	tʃ	a	m	m	33

75	tʃ	a	n	n	33
76	tʃ	a	ŋ	ŋ	33
77	tʃ	a	∅	∅	33
78	tʃ	a	p̚	p̚	33
79	tʃ	a	t̚	t̚	33
80	tʃ	a	k̚	k̚	33
81	tʃ	a	m	m	33
82	tʃ	a	n	n	33
83	tʃ	a	ŋ	ŋ	33
84	tʃ	a	∅	∅	33

4.4 Identification Training

No.	Onset	Nucleus	Coda A	Coda B	Target	Tone
1	tʃ	a	p̚	m	p̚	33
2	tʃ	a	p̚	m	m	33
3	tʃ	a	p̚	t̚	p̚	33
4	tʃ	a	p̚	t̚	t̚	33
5	tʃ	a	p̚	n	p̚	33
6	tʃ	a	p̚	n	n	33
7	tʃ	a	p̚	k̚	p̚	33
8	tʃ	a	p̚	k̚	k̚	33
9	tʃ	a	p̚	ŋ	p̚	33

10	tʃ	a	p˺	ŋ	ŋ	33
11	tʃ	a	m	t˺	m	33
12	tʃ	a	m	t˺	t˺	33
13	tʃ	a	m	n	m	33
14	tʃ	a	m	n	n	33
15	tʃ	a	m	k˺	m	33
16	tʃ	a	m	k˺	k˺	33
17	tʃ	a	m	ŋ	m	33
18	tʃ	a	m	ŋ	ŋ	33
19	tʃ	a	t˺	n	t˺	33
20	tʃ	a	t˺	n	n	33
21	tʃ	a	t˺	k˺	t˺	33
22	tʃ	a	t˺	k˺	k˺	33
23	tʃ	a	t˺	ŋ	t˺	33
24	tʃ	a	t˺	ŋ	ŋ	33
25	tʃ	a	n	k˺	n	33
26	tʃ	a	n	k˺	k˺	33
27	tʃ	a	n	ŋ	n	33
28	tʃ	a	n	ŋ	ŋ	33
29	tʃ	a	k˺	ŋ	k˺	33
30	tʃ	a	k˺	ŋ	ŋ	33
31	tʃ	a	p˺	∅	p˺	33

32	tʃ	a	p˺	∅	∅	33
33	tʃ	a	t˺	∅	t˺	33
34	tʃ	a	t˺	∅	∅	33
35	tʃ	a	k˺	∅	k˺	33
36	tʃ	a	k˺	∅	∅	33
37	tʃ	a	m	∅	m	33
38	tʃ	a	m	∅	∅	33
39	tʃ	a	n	∅	n	33
40	tʃ	a	n	∅	∅	33
41	tʃ	a	ŋ	∅	ŋ	33
42	tʃ	a	ŋ	∅	∅	33
43	tʃ	a	p˺	m	p˺	33
44	tʃ	a	p˺	m	m	33
45	tʃ	a	p˺	t˺	p˺	33
46	tʃ	a	p˺	t˺	t˺	33
47	tʃ	a	p˺	n	p˺	33
48	tʃ	a	p˺	n	n	33
49	tʃ	a	p˺	k˺	p˺	33
50	tʃ	a	p˺	k˺	k˺	33
51	tʃ	a	p˺	ŋ	p˺	33
52	tʃ	a	p˺	ŋ	ŋ	33
53	tʃ	a	m	t˺	m	33

54	tʃ	a	m	t˺	t˺	33
55	tʃ	a	m	n	m	33
56	tʃ	a	m	n	n	33
57	tʃ	a	m	k˺	m	33
58	tʃ	a	m	k˺	k˺	33
59	tʃ	a	m	ŋ	m	33
60	tʃ	a	m	ŋ	ŋ	33
61	tʃ	a	t˺	n	t˺	33
62	tʃ	a	t˺	n	n	33
63	tʃ	a	t˺	k˺	t˺	33
64	tʃ	a	t˺	k˺	k˺	33
65	tʃ	a	t˺	ŋ	t˺	33
66	tʃ	a	t˺	ŋ	ŋ	33
67	tʃ	a	n	k˺	n	33
68	tʃ	a	n	k˺	k˺	33
69	tʃ	a	n	ŋ	n	33
70	tʃ	a	n	ŋ	ŋ	33
71	tʃ	a	k˺	ŋ	k˺	33
72	tʃ	a	k˺	ŋ	ŋ	33
73	tʃ	a	p˺	∅	p˺	33
74	tʃ	a	p˺	∅	∅	33
75	tʃ	a	t˺	∅	t˺	33

76	tʃ	a	t˺	∅	∅	33
77	tʃ	a	k˺	∅	k˺	33
78	tʃ	a	k˺	∅	∅	33
79	tʃ	a	m	∅	m	33
80	tʃ	a	m	∅	∅	33
81	tʃ	a	n	∅	n	33
82	tʃ	a	n	∅	∅	33
83	tʃ	a	ŋ	∅	ŋ	33
84	tʃ	a	ŋ	∅	∅	33

Bibliography

Akahane-Yamada, R., Adachi, T., Kawahara, H., Pruitt, J. S., & McDermott, E. (1998). Toward the optimization of computer-based second language production training. In *STiLL-Speech Technology in Language Learning,* 111-114.

Abramson, A. S., & Tingsabadh, K. (1999). Thai final stops: Cross-language perception. *Phonetica*, *56*(3-4), 111-122.

Anderson, G. G. (2011). *Non-Linear Dynamics of Adult Non-Native Phoneme Acquisition Perception & Production.* ProQuest Dissertations and Theses.

Antonioua, M., Best, C. T., Tylera, M. D., & Kroosa, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics, 39*(4), 558-570.

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of phonetics, 40*(4), 582-594.

Antoniou, M., & Wong, P. C. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America, 138*(2), 571-574.

Antoniou, M., & Wong, P. C. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America*, *139*(1), 271-278.

Alispahic, S., Mulak, K. E., & Escudero, P. (2017). Acoustic properties predict perception of unfamiliar Dutch vowels by adult Australian English and Peruvian Spanish listeners. *Frontiers in psychology*, *8*, 52.

Al-Hindawi, F., Al-Hassnawi, H. D. M., & Al-Ebadi, H. K. (2018). An Investigation of the Relationship between Phonetics and Phonology. *Adab Al-Kufa*, *1*(34), 23-32.

van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of phonetics, 32*(4), 455-491.

Bauer, R. S. (1979). Alveolarization in Cantonese: a case of lexical diffusion. *Journal of Chinese Linguistics, 7*(1), 132-141.

Byrd, D. (1992). Perception of assimilation in consonants clusters: A gestural model. *Phonetica*, *49*(1), 1-24.

Browman, C. P., & Goldstein, L. (1989). Gestural structures and phonological patterns. *Haskins Laboratories Status Report on Speech Research*, 1-23.

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An Overview. *Haskins Laboratories Status Report on Speech Research*, 23-42.

Browman, C. P., & Goldstein, L. (1993). Dynamics and Articulatory Phonology. *Haskins Laboratories Status Report on Speech Research*, 51-62.

Bohn, O.-S. (2002). On phonetic similarity. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 191-216). Trier, Germany: Wissenschaftlicher Verlag.

Best, C. T. (1994). The Emergence of Native-Language Phonological Influences in

Infants: A Perceptual Assimilation Model. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

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). Timonium, MD: York Press.

Boersma, P. (1998). *Functional phonology* (Doctoral Dissertation). The University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, P., & Weenink. D. (2018). Praat: doing phonetics by computer [Version 6.0.37]. Retrieved from http://www.praat.org/.

Baese-Berk, M. M. (2010). *An examination of the relationship between speech perception and production* (Doctoral Dissertation). Northwestern University.

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

Bauer, R. S., & Benedict, P. K. (2011). *Modern Cantonese phonology* (Trends in Linguistics. Studies and Monographs, Vol. 102). Berlin: Mouton de Gruyter.

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. *Journal of the Acoustical Society of America, 109*(2), 775-794.

Best, C. T., Halle, P., Bohn, O. S., & Faber, A. (2003). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 2889-2892). Barcelona, Spain: Causal Productions.

Best, C. T., & McRoberts, G. W. (2003). Infant Perception of Non-Native Consonant Contrasts that Adults Assimilate in Different Ways. *Language and Speech, 46*(2), 183-216.

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

Bundgaard-Nielsen, R.L., Best, C.T., & Tyler, M.D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33*(3), 433-461.

Best, C. T., Goldstein, L. M., Nam, H., & Tyler, M. D. (2016). Articulating What Infants Attune to in Native Speech. *Ecological Psychology, 28*(4), 216-261.

Bennett, R., Tang, K., & Ajsivinac Sian, J. (2018). Statistical and acoustic effects on the perception of stop consonants in Kaqchikel (Mayan). *Laboratory Phonology, 9*(1), 1-42.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r /and /l/: Long-term retention of learning in perception and production. *Perception & psychophysics, 61*(5), 977-985.

Brannen, K. (2002). The role of perception in differential substitution. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 47*(1-2), 1-46.

Bohn, O.-S. & Ellegaard, A. A. (2019). Perceptual assimilation and graded discrimination as predictors of identification accuracy for learners differing in L2 experience: The case of Danish learners' perception of English initial fricatives. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2070-2074). Melbourne, Australia.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Fox, J. (2020). *Package 'lme4'*. Retrieved from https://github.com/lme4/lme4/.

Carlisle, R. S. (2001). Syllable structure universals and second language acquisition. *International journal of English studies*, *1*(1), 1-19.

Cardoso, W. (2011). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research*, *27*(4), 433-465.

Chang, C. B. (2016). Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition*, *19*(4), 791-809.

Casillas, J. V. (2020). Phonetic category formation is perceptually driven during the early stages of adult L2 development. *Language and speech*, *63*(3), 550-581.

Ciocca, V., Wong, L., & So, L. K. H. (1994). An acoustic analysis of unreleased stop consonants in word final positions. In *Proceedings of the International*

Conference on Spoken Language Processing (pp. 1131-1134). Yokohama, Japan.

Cerviño-Povedano, E., & Mora, J. C. (2010). Investigating Catalan-Spanish bilingual EFL learners' overreliance on duration: Vowel cue weighting and phonological short-term memory. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Achievements and perspectives in the acquisition of second language speech: New sounds 2010* (pp. 53-64). Peter Lang.

Chang, C. B., & Mishler, A. (2012). Evidence for language transfer leading to a perceptual advantage for non-native listeners. *Journal of the Acoustical Society of America*, 132(4), 2700-2710.

Cebrian, J., Carlet, A., Gorba, A., & Gavaldà, N. (2019). Perceptual training affects l2 perception but not cross-linguistic similarity. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 929-933). Melbourne, Australia.

Chan, A. Y. (2012). Cantonese English as a second language learners' perceived relations between "similar" L1 and L2 speech sounds: A test of the speech learning model. *The Modern Language Journal, 96*(1), 1-19.

Chappell, W. (2019). Phonological (in) visibility: The perception of reduced Spanish vowels among L1-Spanish speakers, L2-Spanish learners, and English monolinguals. *Journal of Second Language Pronunciation, 5*(3), 435-463.

Cho, M. H. (2021). Perception and Production of American English Vowels by Korean University Students. *The Journal of the Korea Contents*

Association, *21*(5), 285-294.

Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A Generative Theory of the Syllable* (Linguistic Inquiry Monographs, No. 9). Cambridge, Mass.: MIT Press.

Chao, S-C., Ochoa, D. & Daliri, A. (2019). Production variability and categorical perception of vowels are strongly linked. *Frontiers in human neuroscience, 13*, 96.

Chen, N. L. L. (1999). *The velar coda variation in Hong Kong Cantonese* (Master's thesis). City University of Hong Kong, Hong Kong.

Cheung, K. (1986). *The present day phonology of Cantonese* (Doctoral Dissertation). The University of London.

Cheung, P. S. P. (1990). *The acquisition of Cantonese phonology in Hong Kong: A cross-sectional study* (Undergraduate thesis). University College London.

Cheung, H.-N. (2002). Cantonese of the 21st Century: Formation of a New Phonological System in the Hong Kong Language. *Journal of Jinan University, 24*(2), 25-40.

Cheng, B., & Zhang, Y. (2009). The relationship between speech perception and production in second language learners. *The Journal of the Acoustical Society of America, 125*(4), 2754.

Cheng, B., & Zhang, Y. (2015). Syllable structure universals and native language interference in second language perception and production: Positional asymmetry and perceptual links to accentedness. *Frontiers in psychology, 6*, 1801.

Christensen, R.H.B. (2019). *Ordinal-Regression Models for Ordinal Data*. Retrieved from https://CRAN.R-project.org/package=ordinal.

Chu, P. C. K. (2009). Do first-language (L1) phonemic categories play a role in the perception of second-language (L2) phonemic contrasts? A look from the perception of Cantonese codas by Mandarin speakers. *The Journal of the Acoustical Society of America, 125*(4), 2763.

Chu, J., Yang, C. S., Liu, G. F. (2019). Analysis of Second Language Acquisition (SLA) Speech Perception Model & the Perception of Second Language Prosody. *Revista de Cercetare si Interventie Sociala, 64*, 334-351.

Cichocki, W., House, A., Kinloch, A., & Lister, A. (1993). Cantonese speakers and the acquisition of French consonants. *Language learning, 43*(1), 43-68.

Dankovičová, J. (1999). Articulation rate variation within the intonation phrase in Czech and English. In *ICPhS99: Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 269-272). San Francisco.

Dinnsen, D. A., & Gierut, J. A. (2008). *Optimality theory, phonological acquisition and disorders* (Series: Advances in optimality theory). London: Equinox.

Ding, S. Z. P. (2010). Phonological change in Hong Kong Cantonese through language contact with Chinese topolects and English over the past century. In R. M. Millar (Ed.), *Marginal Dialects: Scotland, Ireland and beyond* (Vol. 1, pp. 198-218). University of Aberdeen Press, Aberdeen.

Darcy, I, & Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics, 40*(4), 568-581.

Darcy, I., Mora, J. C., & Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Language Learning*, 66, 741-773.

Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics, 73,* 128-143.

Eckman, F. R. (1981). On the naturalness of interlanguage phonological rules. *Language learning*, 31(1), 195-216.

Edge, B. A. (1991). The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese. *Studies in Second Language Acquisition, 13*(3), 377-393.

Escudero, P. (2005). *Linguistic perception and second-language acquisition: explaining the attainment of optimal phonological categorization* (Doctoral Dissertation). Utrecht University, LOT DissertationSeries113.

Escudero, P. (2006). Second Language Phonology: The Role of Perception. In M. Pennington (Ed.), *Phonology in context* (pp. 109-134). Macmillan, London.

Escudero, P. (2009). The linguistic perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in perception* (pp. 151-190). Berlin: De Gruyter Mouton.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(4), 551-585.

Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the acoustical society of America*, *130*(5), EL277-EL283.

Elvin, J. (2016). T*he role of the native language in non-native perception and spoken word recognition: English vs. Spanish learners of Portuguese* (Doctoral Dissertation). Western Sydney University.

Elvin, J., Williams, D., & Escudero, P. (2016). The relationship between perception and production of Brazilian Portuguese vowels in European Spanish monolinguals. *Loquens, 3*(2), 1-10.

Escudero, Paola, Benders, T. & Lipski, S. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics, 37* (4), 452-465.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics, 40*(2), 280-288.

Escudero, P., & Williams, D. (2012). Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *The Journal of the Acoustical Society of America, 131*(5), EL406-EL412. Retrieved from http://doi.org/10.1121/1.3701708.

Escudero, P., Sisinni, B., & Grimaldi, M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America, 135*(3), 1577-1584. Retrieved from http://doi.org/10.1121/1.4864477.

Fabra, L. R., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*(3), 491-508.

Faris, M. M., Best, C. T., & Tyler, M. D. (2018). Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics, 70*, 1-19.

Feng, Z. (2020). Effects of Identification and Pronunciation Training Methods on L2 Speech Perception and Production: Training Adult Japanese Speakers to Perceive and Produce English /r/-/l/. *Studies in Applied Linguistics and TESOL*, *20*(2), 57-83.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of phonetics*, *15*(1), 47-65.

Flege, J. E. (1988a). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders: A review* (pp. 224-401). Norwood, NJ: Ablex.

Flege, J. E. (1988b). The development of skill in producing word-final English stops: Kinematic parameters. *The Journal of the Acoustical Society of America*, 84(5), 1639-1652.

Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental*

Psychology, *43*(3), 701-731.

Flege, J. E. (1992). The intelligibility of English vowels spoken by British and Dutch talkers. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement, and management* (Vol.1, pp. 157-232). Amsterdam: John Benjamins.

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. 233-77). Timonium, MD: York Press.

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

Flege, J. E. (2005). Origins and development of the Speech Learning Model. In *1st Acoustical Society of America Workshop in L2 speech learning*. Simon Fraser University, Vancouver, BC, Canada.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & Hualde, J. (Eds.) *Laboratory Phonology 9* (pp. 353-380). Berlin: Mouton de Gruyter.

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., & Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied psycholinguistics*, 5(4), 323-347.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6*(3), 185-202.

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of Spanish and English: Evidence for phonetic category formation. *Journal of the Acoustical Society of America, 83*(2), 729-740.

Flege, J. E., & Wang, C. (1989). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of phonetics, 17*(4), 299-315.

Flege, J. E., & Liu, S. (2001). THE EFFECT OF EXPERIENCE ON ADULTS'ACQUISITION OF A SECOND LANGUAGE. *Studies in second language acquisition, 23*, 527-552.

Flege, J. E. & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*(1), 1-34.

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., McCutcheon, M. J., & Smith, S. C. (1987). The development of skill in producing word-final English stops. *The Journal of the Acoustical Society of America*, 82(2), 433-447.

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America, 92*(1), 128-143.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America, 95*(6), 3623-3641.

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

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

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

Flemming, E. S. (2016). *Auditory representations in Phonology*. Routledge.

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

Francis, A. L., Ciocca, V., & Ma, L. (2004). Effects of native language experience on perceptual learning of Cantonese lexical tones. *The Journal of the Acoustical Society of America*, *115*(5), 2544-2544.

Fuhrmeister, P., & Myers, E. B. (2017). Non-native phonetic learning is destabilized by exposure to phonological variability before and after training. *The Journal of the Acoustical Society of America, 142*(5), EL448-EL454.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in medicine*, 22(9), 1433-1446.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Retrieved from https://CRAN.R-project.org/package=stargazer.

Gao, M. (2005). Identifying native Cantonese stops: implication for native speakers' perception. *The Journal of the Acoustical Society of America, 117*(4), 2460.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds " L" and " R". *Neuropsychologia, 9*(3), 317-323.

Gries, S. T. (2009). *Statistics for linguistics with R: A practical introduction* (Vol. 208). Mouton de Gruyter.

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Frontiers in psychology*, *4,* 340.

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.

Gilichinskaya, Y. D., & Strange, W. (2010). Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the*

Acoustical Society of America, *128*(2), EL80-EL85.

Garibaldia, C. L., & Bohnb, O. S. (2015). Phonetic similarity predicts ultimate attainment quite well: The case of Danish /i, y, u/ and /d, t/ for native speakers of English and of Spanish. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Gamer, M., Lemon, J., & Fellows, I. (2007). *The irr Package*. Retrieved from http://www.r-project.org.

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

Henderson, J. B., & Repp, B. H. (1982). Is a stop consonant released when followed by another stop consonant? *Phonetica*, *39*(2-3), 71-82.

Hume, E., & Tserdanelis, G. (2002). Labial unmarkedness in Sri Lankan Portuguese Creole. *Phonology*, *19*(3), 441-458.

Handley, Z., Sharples, M., & Moore, D. (2009). Training novel phonemic contrasts: A comparison of identification and oddity discrimination training. In *Speech and Language Technology in Education* (SLaTE) 2009, 3-5th. Sep. 2009, Wroxall, England.

Hattori, K. (2009). *Perception and production of English /r/-/l/ by adult Japanese speakers*. (Doctoral thesis). University College London, United Kingdom.

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The*

Journal of the Acoustical Society of America, 125(1), 469-479.

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.

Hao, Y. C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics, 54*, 151-168.

Ingham, J. (2014). Perception, production and perceptual learning in the second language: A study of perceptual learning by L1 Bengali speakers of L2 English. In *The 9th Lancaster University postgraduate conference in linguistics and language teaching* (pp. 49-69). Lancaster: Lancaster University.

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

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), 135-159.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of phonetics, 39*(4), 571-584.

Jakobson, R. (1971). *Phonological studies*. The Hague: Mouton.

Jun, J. (2004). Place assimilation. In H. Bruce, K. Robert, & S. Donca (Eds.), *Phonetically based phonology* (pp. 58-86). Cambridge University Press.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69*(3), 505-528.

Jia, G., Strange, W., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, *119*, 1118-1130.

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

Kohler, K., (1981). Contrastive phonology and the acquisition of phonetic skills. *Phonetica, 38*(4), 213-226.

Kochetov, A. (2005). Phonetic Sources of Phonological Asymmetries: Russian Laterals and Rhotics. In *Proceedings of the 2005 annual conference of the Canadian Linguistic Association* (pp. 1-12).

Kim, M. R. (2012). The effects of first-language sound change on second-language speech production. *The Journal of the Acoustical Society of America*, *132*(3), 1938-1938.

Kitikanan, P. (2017). The effects of L2 experience and vowel context on the perceptual assimilation of English fricatives by L2 Thai learners. *English Language Teaching, 10*(12), 72-82.

Khouw, E., & Ciocca, V. (2007). An acoustic and perceptual study of initial stops produced by profoundly hearing impaired adolescents. *Clinical linguistics &*

phonetics, *21*(1), 13-27.

Kochetov, A., & So, C. K. (2007). Place assimilation and phonetic grounding: A cross-linguistic perceptual study. *Phonology*, *24*(3), 397-432.

Lisker, L. (1999). Perceiving final voiceless stops without release: Effects of preceding monophthongs versus non-monophthongs. *Phonetica*, *56*(1-2), 44-55.

Larson-Hall, J. (2004). Predicting perceptual success with segments: A test of Japanese speakers of Russian. *Second Language Research, 20*(1), 33-76.

Lakshmanan, U., & Selinker, L. (2001). Analysing interlanguage: how do we know what learners know?. *Second Language Research*, *17*(4), 393-420.

Linebaugh, G., & Roche, T. B. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language and Learning*, *9*(1), A1-A17.

Law, S.-P., Fung, R. S.-Y., & Bauer, R. S. (2001). Perception and production of Cantonese consonant endings. *Asia Pacific Journal of Speech, Language and Hearing*, *6*(3), 179-195.

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, *125*(2), 1138-1152.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36(1)*, 141-157.

Leung, M.-T., & Law, S.-P. (2001). HKCAC: The Hong Kong Cantonese adult language corpus. *International Journal of Corpus Linguistics, 6*, 305-326.

Leung, M.-T., Law, S.-P., & Fung, S.-Y. (2004). Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior research methods, instruments, & computers, 36*(3), 500-505.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*: University of California Press.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the acoustical society of America*, *94*(3), 1242-1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874-886.

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

Luo, J., Li, V. G., & Mok, P. P. K. (2020). The Perception of Cantonese Vowel Length Contrast by Mandarin Speakers. *Language and speech, 63*(3), 635-659.

van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in psychology*, *6*, 1000.

Malécot, A. (1958). The role of releases in the identification of released final stops:

a series of tape-cutting experiments. *Language*, *34*(3), 370-380.

Manuel, S. Y. (1991). Some phonetic bases for the relative malleability of syllable-final versus syllable-initial consonants. In *Proceedings of the 12th International Congress of Phonetics Sciences* (pp. 118-121). Aix-en-Provence, France.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 89-108.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(3), 279-297.

Milenova, M. (2015). The acquisition of new and similar sounds: Evidence from Bulgarian learners of modern Greek. In G. Kotzoglou, K. Nikolou, E. Karantzola, et al. (Eds.), *The 11th international conference on Greek linguistics* (pp. 1113-1121). Mytilene: University of the Aegean.

Mokari, P. G., & Werner, S. (2017). Perceptual assimilation predicts acquisition of foreign language sounds: The case of Azerbaijani learners' production and perception of Standard Southern British English vowels. *Lingua, 185,* 81-95.

Melnik-Leroy, G. A., Turnbull, R., & Peperkamp, S. (2021). On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French /u/-/y/ contrast. Second Language

Research. Retrieved from: https://doi.org/10.1177/0267658320988061.

Nagle, C. L. (2018a). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, *68*(1), 234-270.

Nagle, C. L. (2018b). Perception, production, and perception-production: Research findings and implications for language pedagogy. *Contact Magazine*.

Nagle, C. L. (2019). Perception, imitation, and production: exploring a three-way perception-production link. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 1297-1301). Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Nagle, C. L., & Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception-production research: revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, 1-26.

Nicolaidis, K., Edwards, J., Beckman, M., & Tserdanelis, G. (2003). Acquisition of obstruents in Greek. In *Proceedings of the 6th International Conference of Greek Linguistics*. Creek: Rethymno.

Ohala, J. J. (2005). The marriage of phonetics and phonology. *Acoustical science and technology*, *26*(5), 418-422.

Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish-English spontaneous code-switching. *Journal of Phonetics*, *52*, 121-137.

Pilus, Z. (2016). Perception of voicing in English word-final obstruents by Malay speakers of English: Examining the Perceptual Assimilation

Model. *Malaysian Journal of ELT Research, 1*(1), 12.

Piske, T., I. Mackay & J. Flege. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29*(2), 191-215.

Peperkamp, S., & Bouchon, C. (2011). The Relation between perception and production in L2 Phonological Processing. In *12th Annual Conference of the International Speech Communication Association*. Florence, Italy.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461-472.

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

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

Peperkamp, S., & Dupoux, E. (2003). Reinterpreting loanword adaptations: the role of perception. In *Proceedings of the 15th International congress of phonetic sciences* (pp. 367-370). Barcelona, Spain: Causal Productions.

Psychology Software Tools Inc. (2016). E-prime software [Version 2.0]. Retrieved from http://www.pstnent.com.

R Core Team. (2018). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/.

Riney, T. J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, *49*(2), 275-302.

Reiterer, S. M., Hu, X., Sumathi, T. A., Singh, N. C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in psychology*, *4*, 782.

Ripley, B. D., & Venables, W. N. (2021). *Package 'nnet'*. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4/.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10, 209-231.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*(3), 322-338.

Selinker, L., & Rutherford, W. E. (2013). *Rediscovering interlanguage*. New York: Routledge.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*(3), 243-261.

Strange, W. (1999). Levels of abstraction in characterizing cross-language phonetic similarity. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 2513-2519). University of California at Berkeley, Department of Linguistics.

Strange, W., Yamada, R., Kubo, R., Trent, S., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *Journal of the Acoustical Society of America, 109*(4), 1691-1704.

Strange, W., Bohn, O. S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America, 115*(4), 1791-1807.

Strange, W., Bohn, O. S., Nishi, K., & Trent, S. A. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America, 118*(3), 1751-1762.

Strange, W. (2007). Cross-language phonetic similarity of vowels. *Language experience in second language speech learning:* In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning, In honor of James Emil Flege* (pp. 35-55). Berlin: John Benjamins.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampin (Eds.), *Phonology and second language acquisition* (pp. 153-192). Amsterdam: John Benjamins.

So, L. K. H., & Dodd, B. J. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language, 22*(3), 473-495.

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.

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

Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l. *Journal* of Phonetics, 66, 242-251.

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

Saito, K., & van Poeteren, K. (2018). The perception–production link revisited: The case of Japanese learners' English/ɹ/performance. *International Journal of Applied Linguistics*, *28*(1), 3-17.

Soo, R., Johnson, K. A., & Babel, M. (2021). *Sound change in spontaneous bilingual speech: A corpus study on the Cantonese n-l merger in Cantonese-English bilinguals*. Retrieved from https://doi.org/10.31219/osf.io/9c3w6.

Tsukada, K. (2006). Cross-language perception of word-final stops in Thai and English. *Bilingualism: Language and Cognition*, *9*(3), 309-318.

Tarone, E. (2012). Interlanguage. *The encyclopedia of applied linguistics*, 1-7.

Tropf, H. (2019). Sonority as a variability factor in second language phonology. In A. James, & J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 173-192). Berlin, Boston: De Gruyter Mouton.

Tsukada, K., & Roengpitya, R. (2008). Discrimination of English and Thai words ending with voiceless stops by native Thai listeners differing in English experience. *Journal of the International Phonetic Association*, *38*(3), 325-347.

To, C. K., Cheung, P. S., & McLeod, S. (2013). A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones. J*ournal of Speech, Language, and Hearing Research, 56*(1), 103-122.

To, C. K. S., McLeod, S., & Cheung, P. S. P. (2015). Phonetic variations and sound changes in Hong Kong Cantonese: Diachronic review, synchronic study and implications for speech sound assessment. *Clinical linguistics & phonetics, 29*(5), 333-353.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344.

Thorin, J., Sadakata, M., Desain, P., & McQueen, J. M. (2018). Perception and production in interaction during non-native speech category learning. *The Journal of the Acoustical Society of America, 144*(1), 92-103.

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the gap. *Studies in Second Language Acquisition, 33*(3), 339-372.

Tsukada, K., Nguyen, T. T. A., Roengpitya, R., & Ishihara, S. (2007). Cross-language perception of word-final stops: Comparison of Cantonese, Japanese, Korean and Vietnamese listeners. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1781-1784). Dudweiler: Pirrot.

Tsukada, K., Birdsong, D., Mack, M., Sung, H., Bialystok, E., & Flege, J. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica, 61*(2-3), 67-83.

Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation of native and non-native fricative place contrasts. *Developmental Psychobiology*, *56*(2), 210-227.

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn* (pp. 607-630). Dept. of English, School of Communication & Culture, Aarhus University.

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.

Tuninetti, A., Whang, J., & Escudero, P. (2019). Non-native vowel perception in a 4IAX task: the effects of acoustic distance. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 240-244). Melbourne, Australia.

Venables, W. N., & Ripley, B. D. (2002) *Modern Applied Statistics with S-Plus* (Fourth edition). Springer.

Winters, S. (2000). Turning phonology inside out, or testing the relative salience of audio-visual cues for place of articulation. *Working Papers in Linguistics of Ohio State University*, *53*. 168-199.

Wright, R. (2004). A review of perceptual cues and cue robustness. In H. Bruce, K. Robert, & S. Donca (Eds.), *Phonetically based phonology* (pp. 34-57). Cambridge University Press.

Wong, S. Y. (2005). *Phonological variation: the case of the syllable-final velar nasal in Cantonese* (Masters' Thesis). The University of Hong Kong.

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36*(2), 250-267.

Winitz, H., Scheib, M. E., & Reeds, J. A. (1972). Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. *The Journal of the Acoustical Society of America*, *51*(4B), 1309-1317.

Yeung, S. W. (1981). *Some aspects of phonological variations in the Cantonese spoken in Hong Kong* (Master's thesis). The University of Hong Kong, Hong Kong.

Yeon, S. H. (2004). *Teaching English word final alveolopalatals to native speakers of Korean* (Doctoral Dissertation). The University of Florida.

Yiu, S. (2016). Co-articulation of unreleased final stops to preceding vowels in Cantonese: An ultrasound study. *The Journal of the Acoustical Society of*

America, *140*(4), 3222-3223.

Yip, M. (1997). Consonant-vowel interaction in Cantonese. *LINGUISTIC MODELS, 20*, 251-274.

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

Yazawa, K., Kondo, M., & Escudero, P. (2017). Modelling Japanese speakers' perceptual learning of English /iː/ and /ɪ/ within the L2LP framework. In *Proceedings of the Phonetic Teaching and Learning Conference 2017* (pp. 115-119). UCL, London.

Yang, Y. X. (2020). *Effects of Cross-linguistic phonetic similarity on Chinese learners' perception and production of Russian speech sounds* (Doctoral thesis). Hunan University.

Yoneyama, K., Beckman, M. E., & Edwards, J. (2003). Phoneme frequencies and acquisition of lingual stops in Japanese. Unpublished manuscript. Ohio State University.

Yiu, S., Archangeli, D., & Yip, J. (2021). Articulatory Strategies for Place Contrasts of Unreleased Final Stops on Preceding Vowels: Evidence from Ultrasound Imaging. In *Proceedings of the Annual Meetings on Phonology* (pp. 1-5). Toronto, Canada.

Zee, E. (1999a). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics, 27*(1), 120-167.

Zee, E. (1999b). Chinese (Hong Kong Cantonese). In the *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (pp. 8-60). Cambridge: Cambridge University Press.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of phonetics*, *28*(1), 69-102.

Zhang, C. W., Chen, S., & Leung, M. T. (2019). The Perception and Production of Cantonese Syllable-final Stops in Mandarin Speakers: An Analysis of Perceptibility Scale. In *Proceedings of Hanyang International Symposium on Phonetics and Cognitive Sciences of Language 2019* (pp. 145-146). Korea: Hanyang University. Retrieved from http://site.hanyang.ac.kr/web/hisphoncog/online-handbook.

Zhang, R., & Yuan, Z. M. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, *42*(4), 905-918.

Zhu, X. N., Jiao, L., Yan, Z. C., & Hong, Y. (2008). Three ways of Rusheng sound change. *Studies of the Chinese language*, 4, 324-338.