Englische Philologie

Acquiring an L2 sociophonological feature: The perception and production of rhoticity by Chinese learners of English

Inaugural-Dissertation

zur Erlangung des Doktorgrades Dr. phil.
der FB 08/09
der Westfälischen Wilhelms-Universität zu Münster (Westf.)

vorgelegt von
Zeyu Li

aus
China

2020

Tag der mündlichen Prüfung: 13.08.2020
Vorsitzender des Gemeinsamen beschließenden Ausschusses: Prof. Dr. Moritz Baßler
Erstgutachter: Prof. Dr. Ulrike Gut
Zweitgutachter: Prof. Dr. Barış Kabak

D6

Table of Contents

List of Figures

List of Tables

Acknowledgements

Many people have contributed to this dissertation and helped me in one way or another throughout my graduate studies in Germany, I would like to take this opportunity to express my sincere gratitude to all of them.

First and foremost, I am deeply indebted to my first supervisor, Ulrike Gut, who has guided me through this journey with her expertise, knowledge, patience, and great support, which made the past three years a particularly meaningful and delightful experience for me. Without her help and encouragement, I would not have been able to complete this dissertation. I am also thankful to Dagmar Deuber for consenting to serve as my second supervisor during the process of my doctoral study. I am especially grateful to my external advisor, Baris Kabak, under whose guidance I was fascinated by phonology and started this wonde**r**ful /**r**/ jou**r**ney in 2016. I would like to thank Baris for being a great inspiration on my path to achieving this.

I also own my deep gratitude to my colleagues at the department of English Linguistics at Münster: Romana Kopeckova, Kristy Sigmeth, Anika Gerfer, Christina Golin, Bonnie Ho, Fola Oyebola, and Philipp Meer. This team has been a source of friendship, great support, and good advice. Thanks for reading my manuscripts and listening to my talks (for many times) and the thorough comments and suggestions they provided.

My research would have been impossible without the help and support of the following people. My special thanks go to Christina Domene Moreno, who has been my Doraemon in academia and gave me invaluable help not only in the making of this dissertation but also throughout my graduate career. I am thankful to Ole Schützler for helping me with statistics and his constructive comments on my dissertation drafts. I would also like to thank Marie Himmel for sharing her knowledge on rhoticity and helping me with the acoustic analysis of the spoken data. I am grateful to Erin Lutenski and Anika Gerfer, who kindly served as second round coders for the auditory analyses. I want to thank Johanna Hartmann and Lisha Tang for generously sharing their language attitude sound samples with me. I also owe my gratitude to Rachael Rasing, Britta Padberg-Schmitt, and Kristy Sigmeth, who recorded the perceptual stimuli for my perception study and helped me find more native English speakers in Münster.

I am particularly thankful to all of my participants in China, who volunteered to take part in the rather complicated (and sometimes weird) experiments twice over a time span of six months. I want to thank Yuqin Luan, who kindly allowed me to use the phonetic lab and the phonetic classrooms at the North China Institute of Science and Technology for data collection. I owe another big thank-you to Romana Kopeckova, Rachael Rasing, and Bonnie Ho, for advertising my little experiment in their classes. I am also grateful to the twenty German students and the ten Chinese students at Münster who participated in my perception replication study and my pilot study.

I owe my deepest gratitude to my wonderful family. To my dearest parents, Li Yu and Luohu Li, thank you for your unconditional love, support, and faith in me. To my husband, Nan Gao, thanks for always standing by me no matter what, for being interested in my research, for your endless love, patience, and encouragement. Special thanks go to Garen for being funny and cute and the little one for choosing me. Words are not enough here to express how lucky I constantly feel to have all of you in my life.

1. Introduction

As research revealing variation in the speech of second language (L2) learners has been growing within the field of Second Language Acquisition (SLA) (e.g., Bayley & Preston 1996), one of the major objectives of the study of L2 phonology has become to accurately predict and comprehensively account for the variable phonological patterns of L2 learners and the factors influencing the development of their *interlanguage* (*IL*; Selinker 1972, Major 2001). The vastness of the research on L2 phonological acquisition over the past decades has suggested that the systematic variability observed in the speech of L2 learners can be subject to both linguistic and extralinguistic factors (e.g., Bayley 1996; 2005, Hansen Edwards 2008, Kautzsch 2017). The former includes the effects of the first language (L1) phonological grammar (e.g., Lado 1957, Brown 1998), speech perception processes (e.g., Flege 1995), language universals (e.g., Eckman 1977, Major 2001), the phonetic and phonological contexts of a target language (TL) feature (e.g., Colantoni & Steele 2008), and prosodic constraints (e.g., Young-Scholten 2004). Concerning extralinguistic influence, a wide variety of learner-related predictors have been proposed to shape patterns of L2 speech, including age, gender, type of instruction, form of input, proficiency level, motivation, language learning aptitude, length of residency in a foreign country, amount of L1 usage, and other relevant variables (see Piske et al. 2001, Gut 2009: 253 - 278, or Moyer 2013 for a comprehensive review).

While the effects of the linguistic and learner-related factors have been extensively examined in the field of L2 speech acquisition, the investigation of L2 phonology from a sociolinguistic perspective has been largely neglected (e.g., Bayley 2005). That is, a particular TL phonological feature may vary and change systematically with the social contexts it occurs in (see Honey 1998 for an introduction of *sociophonology* and some sociophonological features). As such, the learners' L2 speech might be "socially conditioned" (Dowd et al. 1990: 16) and to some extent contain the sociolinguistic patterns that are inherent in the nature of the TL phonological feature. However, the question arises here as to how L2 learners acquire a TL aspect that is not only phonetically and phonologically but also sociolinguistically variable across the TL varieties or between the L1 and the TL of the learners. In other words, it remains

unclear to what extent and in what forms the sociolinguistic variation of a TL or of an L1 feature can possibly modulate the speech patterns of L2 learners as well as their choice of a linguistic form.

The primary aim of the current study is thus to contribute to our understanding of L2 sociophonological acquisition and to provide a predictive account of the sociolinguistic patterns found in L2 speech. To this end, I investigate the acquisition of English rhoticity, i.e. the realisation of rhotics in the syllable coda position, by Chinese learners targeting a British English (BrE) or an American English (AmE) accent. Rhoticity constitutes a sociophonological phenomenon that robustly indexes dialectal, stylistic, and social differences in both the learners' L1 Mandarin (Zhang 2008) and their TL English (Labov 1966, Honey 1998). In English, the presence or absence of nonprevocalic /r/ is widely recognised as one of the most salient features distinguishing English varieties worldwide (Trudgill & Hannah 2013) and "the most important phonotactic difference among accents of English" (Melchers & Shaw 2003: 19). The diachronic and synchronic variability of its realisation has been shown to change systematically with both linguistic (e.g., stress and preceding vowel environments) and social factors (e.g., gender, socioeconomic status, and education) in varieties of English (e.g., Labov 1966, Chand 2010, Li & Kabak 2017). Intriguingly, Mandarin has a stylistic diminutive marker [ɚ], which is a retroflex suffix that is phonetically and articulatorily close to English nonprevocalic /r/. The variable production of this feature in Mandarin is also driven by a series of grammar-internal and grammar-external constraints, such as the phonological nature of host rhymes, lexical tone, speaking style, as well as speakers' gender, identity, and regional background (Duanmu 2007, Zhang 2008). As such, this feature offers a particularly informative test case for examining the competing transfer opportunities of phonological and sociolinguistic aspects from the L1 in shaping the perceptual and production patterns of L2 learners.

Furthermore, the study at hand makes a theoretical contribution to existing L2 speech theories and offers a theoretical model that predicts the acquisition of L2 perception and production of a sociophonological feature over time. To this end, I enquire into the acquisition of rhoticity by Chinese learners of English by focussing on three major

issues that guide research on the acquisition of L2 phonology: *Crosslinguistic Influence* (CLI) in L2 speech, the relationship between speech perception and production, and the development of L2 speech over the course of acquisition. To that end, both the perception and production of English rhoticity by Chinese learners were tested twice before and after training over a time span of six months. Accordingly, this dissertation is structured as follows.

Chapter 2 surveys existing L2 theories describing and explaining the effects of CLI in L2 phonological acquisition. First, I introduce various hypotheses and models that offer different accounts with regard to CLI and the expected areas of difficulty for L2 learners (2.1). Furthermore, two theoretical contributions predicting the time course of the impact of CLI on the development of L2 phonology are presented (2.2). Based on the review of the tenets as well as predictive scopes of these theories, research gaps and unresolved questions in the area of CLI are identified.

Chapter 3 focusses on the relationship between the perception and production of L2 speech. It first presents the Speech Learning Model (Flege 1995), which provides a theoretical foundation for the idea of perception preceding production or perception determining production (3.1). The chapter continues with empirical evidence supporting and challenging the perception-production link, demonstrating the crucial role of phonetic constraints affecting speech perception and production, respectively (3.2). Finally, a discussion of learner variability in the perception and production of L2 speech is given. This study is further placed in a variationist framework in order to examine key factors conditioning variation in the speech of L2 learners (3.3).

Chapter 4 gives a comprehensive review of the L2 sociophonological feature under scrutiny in the present study, namely rhoticity in English. After an overview of the articulatory, acoustic, and phonological accounts of nonprevocalic /r/ and its preceding vowel contexts (4.1), both grammar-internal and -external factors that were previously shown to shape patterns of rhoticity in various English varieties are compared and discussed (4.2). Finally, the potential link between patterns of rhoticity in L2 speech and in native English varieties concerning sound change is highlighted (4.3).

Chapter 5 starts with a thorough contrastive analysis of rhoticity in Mandarin and English. The phonetic, phonological, and sociolinguistic aspects of the Mandarin retroflex marker are presented and compared with nonprevocalic /r/ in English (5.1). Based on the theoretical observations and the crosslinguistic comparisons between rhoticity in Mandarin and English, research questions and predictions for the present study are formulated in the second half of this chapter (5.2).

Chapter 6 presents the methodology and data used in the present study. This includes an overview of the participants (6.1), the experimental design for the collection of perception and production data as well as a norm preference test (6.2), and a description of the grammar-internal and -external variables under scrutiny (6.3).

Chapters 7 and 8 present the results on the perception and production of rhoticity by Chinese learners of English, respectively. The structures of the two chapters are organised alike. After a description of the auditory and acoustic analyses of the data and the statistical procedures applied to them (7.1 and 8.1), the performances of the learners (7.2 and 8.2) and factors influencing the learners' perceptual and production patterns of rhoticity on the basis of multilevel analyses are presented (7.3 and 8.3). For the perception of rhoticity, an investigation of the acoustic properties of the /V(r)/ sequences is provided, indicating the relative degree of acoustic salience of different contexts (7.4). Finally, the findings are comprehensively discussed (7.5 and 8.4) and followed by a concluding summary for each chapter (7.6 and 8.5).

Chapter 9 presents the results on the relationship between the perception and production of rhoticity by the learners and the time course of its development. I provide an overview of the perception-production link for rhoticity from a bird's eye perspective (9.1), and on both individual and learner group levels (9.2 and 9.3). This chapter then zooms into the relationship between perception and production of nonprevocalic /r/ in various preceding contexts (9.4), followed by a discussion (9.5) and a chapter summary (9.6).

In chapter 10, the results of the present study are summarised and discussed in the

light of the overarching research questions and predictions formulated in chapter 5. The effects of CLI on the acquisition of L2 rhoticity (10.1), the interplay of various factors conditioning the relationship between L2 speech perception and production over time (10.2), and the predictors and patterns shared by L2 and native English speech (10.3) are discussed. Moreover, I highlight the implications of the present findings for existing L2 speech theories and propose an L2 Sociophonological Development Model that provides an explanatory and predictive account for the acquisition of an L2 sociophonological feature over the course of learning (10.4). Finally, this chapter draws an overall conclusion and gives an outlook on future research (10.5).

2. The effect of an L1 on the development of L2 phonology

The question of how previously learned languages influence the acquisition of an L2 is one of the central issues of SLA research. This chapter introduces theoretical frameworks describing and explaining the effect of CLI on the acquisition of L2 phonology. Section 2.1 demonstrates that various hypotheses and models offer different accounts in terms of CLI. Some argue that it is the differences between L1 and TL systems that lead to major difficulties for L2 learners, while others propose that the acoustic or perceived similarity between L1 and L2 structures tends to cause learning confusions, hence greater challenges to learners. Section 2.2 further presents two models on the time course of the impact of CLI on the longitudinal development of L2 phonology. One model suggests that the L2 acquisition process is conditioned by the interaction of the degree of similarity and markedness, whereas the other claims that the interplay of L1, TL, and *language universals* determines L2 phonological development overall. Issues that current frameworks fail to address, such as context-dependent CLI, possible effects of L1 dialectal variation, as well as transfer opportunities of sociolinguistic aspects that L1 and TL categories are indexical of, are also discussed in this chapter.

2.1 CLI as a main source of learning problems

One of the first attempts to describe L1 influence on L2 acquisition is Lado's (1957) *Contrastive Analysis Hypothesis (CAH).* It is based on the assumption that the areas of difficulty that a language learner will encounter can be accurately predicted by the linguistic comparison between the learner's L1 and TL. According to Lado (1957: 1), "[…] in the comparison between native and foreign language lies the key to ease or difficulty in foreign language learning". Specifically, the similarities and differences between the L1 and TL correspond to the ease and difficulty of acquisition, respectively. The CAH postulates that TL structures that exist in the L1 would be generally easy for learners to acquire as positive L1 *transfer* is more likely to occur, whereas aspects that are absent in the L1 would present major difficulties. With respect to pronunciation learning, Lado states that L2 pronunciation errors are attributable to the differences of the phonemes and their distribution between the learner's L1 and TL.

The CAH proposed by Lado (1957) is considered a "strong" version of the hypothesis, which claims that L1 knowledge may hinder the learning of a new language because it is the difference between L1 and TL that poses the greatest difficulty to language learners. A "weak" version of the CAH suggested by Newmark (1966), on the other hand, assumes that knowledge of an L1 will not inhibit but rather facilitate the acquisition of a TL. According to Newmark (1966), negative influence of L1 emerges mainly due to a lack of acquisition, i.e., the fact that language learners are not fully aware of the TL grammar. In other words, it is not the detrimental effect of the L1 that interferes with the acquisition of an L2, but rather the learner's strategy of falling back on L1 grammar when new knowledge of the L2 has not been fully mastered. However, Eckman (1977: 316) points out that this weak version of the CAH "is not falsifiable" as it makes no claims as to where errors are more likely to occur. It is worth noting that both versions argue that it is the difference between the native and foreign language that presents major learning difficulty for L2 learners.

Evidence contradicting the "strong" and "weak" forms of the CAH comes from Oller and Ziahosseiny (1970), who propose a "moderate" version of the CAH and predict that similar TL structures are more likely to cause learning confusions and consequently greater difficulty for L2 learners. New TL aspects, on the other hand, will not add to the general confusion and difficulty of language learning. They state that "wherever patterns are minimally distinct in form or meaning in one or more systems, confusion may result. Conversely, where patterns are functionally or perceptually equivalent in a system or systems correct generalization may occur" (Oller & Ziahosseiny 1970: 186). This version further posits that one of the most difficult aspects for L2 learners is the acquisition of sounds, since "the most subtle distinctions are required either between the target and native language, or within the target language" (186). However, no empirical evidence regarding the acquisition of L2 phonology has been provided to support this version of the CAH. Oller and Ziahosseiny (1970) have only reported an empirical study on the acquisition of English spelling, where they found that it is more difficult for learners of an L1 employing a Roman alphabet to learn L2 English spelling than for learners of a non-Roman L1 system, suggesting that a similar spelling system is more challenging

to be acquired than dissimilar spelling systems.

Although the CAH was an important landmark in the fields of L2 acquisition at that time until the 1970s, certain drawbacks of the three versions of the CAH have been noted (e.g., Kramsch 2007). To begin with, as the hypotheses are formulated mostly based on observations in the practice of English teaching, there is a severe lack of empirical support in terms of the acquisition of L2 phonology. Further, while Lado (1957) makes general predictions about L2 pronunciation training, the scope of his hypothesis is restricted to phonemes and their distribution without taking the influence of context-dependent variation or sociolinguistic variation into consideration. In particular, as the key of the CAH is to define whether a TL structure is different from the L1, a related question that has to be addressed involves the classification and comparison of TL structures in relation to the phonology of L1. However, no specific criteria have been identified as to how to define a TL aspect as the same or different category. For example, if a TL sound exists in the L1 of learners, should it be identified as belonging to the identical category if the phonetic contexts it occurs in the L1 differ greatly from its context in the TL? Or should it be classified into a rather dissimilar category? A detailed instruction of categorization should therefore be provided in order to generate more reliable and consistent assumptions, especially considering that the distinction between an L1 and a TL sound and their surrounding phonetic environments is not always clear-cut. Also, no predictions as to possible developmental processes or ultimate competence for L2 learners can be made based on the CAH. Finally, a number of studies have noted that some observed L2 acquisition errors appear to be rather systematic and consistent across language learners, which are not attributable to either their L1 or TL but language *universal principles*, which refer to linguistic generalisations and shared properties across all languages in the world (e.g., Major 2001).

So, while the CAH (Lado 1957) predicts the difficulties of L2 acquisition exclusively on the basis of the distinctions between a learner's L1 and TL, Eckman (1977) revises it by integrating the notion of *implicational markedness* as a predictor of relative difficulty. He defines the concept of markedness as follows: "[A] phenomenon *A* in some language is more marked than *B* if the presence of *A* in a language implies the

presence of *B*; but the presence of *B* does not imply the presence of *A*" (320). Eckman points out that the CAH indicates possible areas of difficulty for L2 learners but fails to anticipate the "directionality of difficulty" (315). As such, in order to predict the relative degree of difficulty in L2 acquisition which the CAH alone is inadequate to account for, Eckman (1977) proposes the *Markedness Differential Hypothesis (MDH)*, which postulates that TL structures that both differ from the L1 and are implicationally more marked will be difficult for learners to acquire. The TL aspects that are absent in the L1 but are not more marked, on the other hand, will not present much difficulty. This means that the relative degree of learning difficulty is closely associated with the relative degree of markedness of the TL aspects. In particular, the more marked a TL structure, the more difficult it is for L2 learners to acquire (see Eckman 1996, Eckman 2004 for further discussion).

Eckman (1977) argues that the acquisition of German and English obstruents provides an illustrative example of the improvements of the MDH relative to the CAH. To illustrate, in English, both voiced and voiceless obstruents exist in word-initial, -medial, and -final positions, while German only allows voiceless but not voiced obstruents to occur in word-final position. According to the CAH (Lado 1957), this difference would pose an equal amount of difficulty for both German learners of English and English learners of German. However, Moulton (1962) found that German learners have more difficulties in acquiring the new contrast than English learners, who only have to learn to suppress such contrasts, suggesting "a directionality of difficulty" related to the acquisition of this structure (Eckman 1977: 317). As the distribution of the obstruents has been found to be more marked in English than in German (Dinnsen & Eckman 1975), Eckman's MDH (1977) accurately predicts the relative difficulty concerning its acquisition, i.e. that it is more difficult for German-speaking learners to acquire this more marked structure in English.

Although the MDH (Eckman 1977) supplements the CAH (Lado 1957) in accurately predicting the directionality and relative degree of difficulty in L2 speech learning by incorporating implicational markedness as a measure, and while it has been tested empirically in areas beyond phonology, it falls short on several grounds like the CAH.

For example, its scope is still restricted to the segmental level in terms of phonological acquisition and it is only feasible to be applied to features that are associated with markedness such as syllable structure. In other words, it would be challenging for researchers to make assumptions based on the MDH if a TL feature has not been specified in terms of markedness, such as rhoticity. Also, it is unclear how to define a TL sound as belonging to a similar or a different category, especially when the sound exists in different surrounding phonetic environments between the L1 and the TL. Furthermore, possible transfer opportunities of sociolinguistic aspects from L1 has not been taken into account. In addition, it fails to generate predictions about possible developmental stages or ultimate attainment for L2 learners. Major and Kim (1996) further pointed out that the MDH does not specify the relative difficulty that L2 learners might encounter when L1 and L2 structures do not differ.

While the MDH (Eckman 1977) is based on the strong version of the CAH (Lado 1957), which predicts that the differences between the L1 and L2 pose a major difficulty for language learners, Major and Kim (1996) formulated the *Similarity Differential Rate Hypothesis (SDRH)*, which builds on the moderate version of the CAH (Oller & Ziahosseiny 1970) stating that similar L2 phenomena are more difficult for learners to acquire than dissimilar structures. According to Major and Kim (1996: 467), "gross differences are more often noticed, due to their perceptual saliency, whereas minimal differences are more like to be overlooked and to result in confusion or nonlearning". The SDRH further postulates that dissimilar TL categories will be acquired at a faster rate and are more likely to be fully acquired by idealised L2 learners in comparison with similar categories.

As illustrated earlier, the MDH (Eckman 1977) revisits the CAH (Lado 1957) and employs the notion of typological markedness in order to predict the relative difficulty of L2 acquisition. Major and Kim (1996) have further noted the potential conflict between the concepts of dissimilarity and markedness in the MDH. For example, considering the acquisition of a TL contrast of *A* and *B*, if *feature A* is more different from learners' L1 and less marked than *feature B*, while *feature B* is more marked and less different from the L1 than *feature A*, it would be impossible to anticipate which one would be more challenging for language learners to acquire. According to the

MDH, TL structures that both differ from the L1 and are implicationally more marked will be difficult. From the perspective of markedness, *feature B* should be harder to acquire. However, *feature A* as the more dissimilar feature should pose greater difficulty from the dissimilarity point of view. The MDH is thus inadequate to make accurate predictions of relative learning difficulty in such a conflict. The SDRH reinterprets the MDH in terms of rate in that dissimilar L2 aspects are acquired faster than similar ones. As Major and Kim state, this basic assumption is not affected by markedness, which only functions as a "mediating factor" (1996: 475) in the SDRH and indicates that relatively more marked phenomena exhibit a slower acquisition rate than less marked ones.

Another theory that recognises L1-TL similarity as the major source of L2 phonological errors involves the *Speech Learning Model* (*SLM* Flege 1995). In contrast to the CAH (Lado 1955) and the MDH (Eckman 1975), which argue that differences between pre-existing L1 and TL structures have a negative effect on acquiring a TL structure, the SLM predicts that it is the perceived crosslinguistic similarity between learners' L1 and TL inventories that leads to major learning problems in the ultimate attainment of L2 phonology. According to Flege (1995), L2 learners perceptually relate TL sounds to pre-existing L1 categories. If two contrasting TL sounds (or L1-TL contrasts) are not sufficiently different from an existing L1 category, they will be mapped into one single category, which results in incorrect perception and production of the TL sounds. Crucially, a high degree of perceived crosslinguistic dissimilarity promotes the discovery and establishment of new L2 categories, hence the acquisition of the new sounds. Similar TL sounds, on the other hand, are the most difficult for L2 learners to acquire because they tend to be equivalent to an existing L1 category. It should be noted that the degree of crosslinguistic dissimilarity between L1 and TL sounds cannot be acoustically measured. Instead, it is testable and should be established based on empirical measurements of individual L2 learners' perceptual ability (Flege 2016).

The four CLI-related theoretical frameworks reviewed here offer different accounts as regards the main source of L2 acquisition errors. Specifically, the CAH and the MDH claim that L1-L2 differences tend to lead to L2 learning problems, while the SDRH

recognises L1-L2 similarity as a main source of learning confusions and difficulties in comparison with dissimilar TL categories. Moreover, the SLM posits that L2 phonological acquisition is driven by learners' perception and proposes that the degree of perceived crosslinguistic dissimilarity between learners' L1 and TL affects their acquisition of L2 speech.

One important aspect of CLI that has been largely absent from the existing L2 theories is the potential effect of L1 dialectal variation, i.e. different dialects or varieties of the native language spoken by learners. O'Brien and Smith (2010:297), for instance, noted that "a methodological shortcoming in previous L2 acquisition studies has been that researchers have assumed an overly homogenous L1 ignoring dialect differences". However, none of the models reviewed before predict or specify the underlying influence of the L1 dialectal background of L2 learners on their acquisition of a foreign language.

Recent studies have shown that L1 dialectal or geographical variation can modify the perception and production of non-native sounds by L2 learners and emphasized the necessity of carefully controlling for the specific dialectal background of learners. In terms of perception, Chládková and Podlipský (2011), for example, examined the perceptual assimilation of Dutch high front vowels by Bohemian Czech and Moravian Czech learners and found that the two groups of listeners followed different pathways in classifying L2 vowels to their respective L1 dialectal categories. Similarly, Escudero and Williams (2012) investigated the discrimination of Dutch vowels by Peruvian Spanish and Iberian Spanish learners and showed that the listeners of the two Spanish varieties perceived the Dutch vowels differently, owing to the acoustic differences in the native vowel production between the two dialects. Besides, Escudero et al. (2012) examined the perception of English front vowels by native speakers of Dutch spoken in the Netherlands and in Belgium and revealed that the two different listener groups differed in their perception of the English /ɛ/ - /æ/ contrast due to their L1 dialectal variation. It is worth noting that these studies have focussed exclusively on the perception of vowels by L2 listeners of different L1 dialectal background. It remains unclear whether L1 dialectal variation has similar effects on the perception of other L2 phonological aspects, such as an L2 sociophonological

feature. The present study thus attempts to fill this gap by examining the perception of English rhoticity by Mandarin learners of English from different dialect regions of China.

As regards the effect of L1 dialectal variation on L2 speech production, O'Brien and Smith (2010), for example, investigated the production of L2 German vowels by learners from three distinct North American English dialectal regions. Their findings showed that L2 vowel formant values vary distinctly by the speakers from different dialectal backgrounds, although the three L1 dialects only differ subtly in terms of the acoustic properties of vowels. Simon et al. (2015) further reported on the effect of L1 regional variation on both the perception and production of L2 English vowels by native Dutch learners from East Flemish and Brabantine. In their study, main effects of L1 dialects were observed in conditioning the learners' perceptual assimilation pattern of the L2 vowels, but not in their production, indicating a perception-production mismatch under the impact of different L1 dialect regions. A more recent investigation beyond vowels was conducted by Kautzsch (2017), who explored the influence of L1 dialectal variation by testing the production of English rhoticity, linking /r/, /t/-voicing, BATH, and LOT/THOUGHT by advanced learners from two distinct German L1 accent regions – North-Rhine Westphalia and Bavaria. The results of his study demonstrated significant differences between the two accent groups in their realisation of English rhoticity and linking /r/, but not for the remainder of the features, suggesting that the impact of L1 dialect regions can operate on L2 production in a feature-dependent manner. As such, it is important to take into account the possible influence of L1 dialectal variation in the investigation of CLI on L2 speech production. The present study will examine Mandarin learners of English from all over China in order to verify its underlying effects on the production of rhoticity.

2.2 CLI during the development of L2 phonology

While different theories have highlighted the role of CLI on L2 phonology, only a few of them pay particular attention to a longitudinal perspective on the effect of CLI over the time course of L2 phonological development. The SDRH proposed by Major and Kim (1996) describes the rate, order, and stages of L2 phonological acquisition over time on the basis of the interaction between similarity and markedness. It posits that

dissimilar TL structures are acquired at faster rates than similar ones over time, with markedness operating as a mediating factor that slows acquisition rate. As Major and Kim (1996) state, the key term “areas of difficulty” in the MDH (Eckman 1977: 321) is “a problematic and vague term, may be interpreted in terms of stages or rate” (Major & Kim 1996: 162). In other words, it is usually difficult to define whether an L2 aspect is completely acquired or not, since the criterion level that determines successful attainment of a TL structure by learners is obscure. On the other hand, the rate of acquisition is rather easier to measure and better comparable. According to Major and Kim, it is therefore more reasonable and convincing to elucidate that the rate of acquisition of a structure is faster than the other over the course of learning. As such, in the SDRH, rate of acquisition is considered the key criterion to describe the degree and development of L2 phonological acquisition.

The SDRH (Major & Kim 1996) further explains the relationship between stages and rates of L2 acquisition from a longitudinal perspective. More specifically, at early stages of acquisition, dissimilar contexts might be produced at a lower level of accuracy due to a large amount of negative transfer. Similar contexts, on the other hand, are realized more accurately because of the similar components existing in the L1. However, the SDRH predicts that, at later stages of acquisition, dissimilar contexts will be acquired at faster rates than similar phenomena, since the distinction between dissimilar features is more likely to be noticed and acquired by L2 learners over time. According to the SDRH, it is much slower for learners to master similar TL structures, despite the fact that they are produced more accurately initially. The acquisition of dissimilar TL structures will gradually surpass their similar counterparts while passing through intermediate stages and progressively approach native competence, whereas the similar structures might even exhibit a negative acquisition trajectory over the course of language learning. At the same time, the mediating factor markedness does not affect the general rate of acquisition determined by similarity. Instead, the more marked feature only slows the rate relative to the less marked phenomenon. Major and Kim further suggest that since both dissimilarity and markedness operate on the rate of acquisition, the ideal case to test the SDRH would be to hold one factor constant in order to investigate the impact of the other. For example, an adequate test of the effects of similarity should control for the degree of

markedness of the TL structures in order to compare the acquisition rate between a similar and a dissimilar phenomenon over time.

The SDRH enjoys ample empirical support from L2 phonological research that is not specifically designed to test the theory but involves longitudinal stages or various levels of acquisition (e.g., Major 1986, Bohn & Flege 1992). However, similar to the CAH and the MDH, the SDRH does not provide a clear system for the identification of the degree of L1-TL similarity, which would, however, be necessary to make valid assumptions about the acquisition rate of different TL features. Although it takes into account the longitudinal perspective of L2 phonological learning in terms of the rate of acquisition, sociolinguistic aspects that can possibly affect phonological CLI over the course of acquisition have not been considered and controlled for. In addition, like the CAH and the MDH, the scope of the SDRH is mostly restricted to phonemic level and does not investigate claims about context-dependent or prosodic variation.

Another framework that has emphasized the longitudinal development of L2 phonology is the *Ontogeny and Phylogeny Model* (*OPM* Major 2001). It formalizes the interaction between L1, TL, and universal processes over the course of language acquisition. As indicated by its name, the OPM involves two main perspectives: ontogeny and phylogeny. According to Major (2001: 81), *ontogeny* refers to "the life cycle of an individual's language", while *phylogeny* represents "the life cycle of whole languages and language types, including historical change, dialect variation, language loss, and language contact phenomena". This model assumes that every language learner speaks an interlanguage (Selinker 1972), which is composed of aspects of L1, TL, and language universals. Importantly, an IL is not viewed as "a deficient version of the L2" (Major 2001: 81). Instead, it is "a complete system in its own right" that documents the interrelationship of the above-mentioned three components (Major 2001: 81).

According to the OPM, the basic pattern of IL development is characterized by an increase of TL influence, a decrease of L1 influence, with the influence of language universals first increasing and then decreasing over time. It is important to emphasize that language universals in the OPM framework refer to "the universals of language

that are not already part of the L1 or L2 system" (Major 2001: 81). As Major suggests, an IL is composed of L1 aspects exclusively in the very initial state of acquisition, which then passes through stages with increasing amounts of TL input and progressively approaches target-like patterns. An idealized learner is expected to master the TL system completely at the end state. In other words, the model predicts dominant L1 influence at earlier stages of language acquisition and possibilities of target-like behaviour at the end. Major (2001) argued that it is also possible for L2 learners to bypass the development of linguistic universals completely. That is, a learner might simply acquire the target production without being affected by any language universals over the course of learning.

In order to comprehensively account for the interaction of L1, TL, and language universals in the IL, the OPM postulates four specific corollaries considering possible chronological development, stylistic variation, as well as the potential impact of similarity and markedness. Table 2.1 illustrates the conditions and the interplay of the three components for each corollary.

Table 2.1 Four corollaries proposed in the OPM (Major 2001)

no.	Corollaries	conditions	L1	L2	language universals
1	Chronological	over time	decrease	increase	increase then decrease
2	Stylistic	with increased stylistic formality	decrease	increase	increase then decrease
3	Similarity	in similar phenomenon	decrease slowly	increase slowly	increase slowly then decrease slowly
4	Markedness	in marked phenomenon	decrease slowly	increase slowly	increase rapidly then decrease slowly

The *Chronological Corollary of the OPM* is concerned with the time course of the development of learners' IL. As shown in the table, the following pattern of IL is predicted: influence from L1 decreases, TL increases, language universals increase and then decrease. This developmental trajectory forms the fundamental pattern of the OPM. Regarding the *Stylistic Corollary of the OPM*, a similar process is shown with speaking style becoming more formal, suggesting that the IL tends to be more target-like in formal speech or in self-monitored situations relative to informal

contexts. In the *Similarity Corollary of the OPM*, the acquisition of a similar TL structure slows down the developmental rate of the IL. In particular, L1 seems to override language universals and exerts greater impact on the development of the IL, which develops with L1 influence decreasing and L2 increasing slowly over time. Likewise, the impact of language universals also changes at a relatively slow pace. The *Markedness Corollary of the OPM* describes the chronological development of IL in variably marked phenomena. Specifically, acquiring a more marked structure is characterized by greater difficulties than acquiring a less marked structure, which results in a relatively steady decrease of L1 impact and increase of L2 impact. Moreover, language universals are more influential than L1 when a more marked TL feature is acquired. In the Markedness Corollary, the effect of L1 first decreases and then decreases slowly, L2 increases slowly, while language universals increase rapidly and then decrease slowly.

In general, the OPM provides a more explicit and comprehensive account of the development and the interplay of L1, TL, and language universals on learners' speech over the course of phonological acquisition than the SDRH. However, it has been noted that there is a lack of clear definitions of language universals, which should be explicitly stated for researchers to formulate universals-related assumptions (e.g., Gut 2009: 27). Davidson (2011) also mentioned that the distinction between an L1 transfer and a developmental error due to language universals is still unclear in the OPM, the confirmation of which would "benefit from the type of articulatory and acoustic data" (131).

To summarise, the existing L2 models on CLI have pointed out the potential source of L2 acquisition difficulties and the developmental patterns of learners' L2 speech in different stages of learning. However, the theories have mostly focussed on the acquisition of context-free phonemes and their distribution yet overlooked other phonological and phonetic contexts that L1 and TL sounds occur in. Given that the surrounding environments of a sound may differ considerably between languages, possible context-dependent CLI beyond allophonic distribution has to be considered to explain and predict L2 phonological patterns. This study attempts to offer further evidence to contribute to our understanding of L1-based CLI that result in major

learning difficulties in the development of L2 speech over time. Apart from accounting for context-dependent acquisition, the present study also explores CLI on L2 phonology from a longitudinal perspective to provide empirical evidence to support or contradict the OPM and the SDRH.

A final point that should be noted is that little research has explored the acquisition of phonological variables that are indexical of different sociolinguistic aspects between L1 and TL yet. In other words, a TL phonological variable may vary distinctly in terms of the social, dialectal, and stylistic contexts it occurs in, which, could also have different patterns in the L1 of learners along similar parameters. Rhoticity, for instance, is well known as a sociophonetic and sociophonological feature in English varieties (e.g., Schützler 2010), the realisation of which was shown to be determined by different sociolinguistic constraints such as speech style, gender, education, socioeconomic status, topic of conversation, and language contact (e.g., Labov 1966, Becker 2014). For Chinese learners acquiring English rhoticity, a critical issue needs to be considered is that the L1 Mandarin of learners is also variably rhotic in different dialect regions, which is subject to similar factors as for English rhoticity yet exhibits divergent patterns. As concerns stylistic variation, for example, rhoticity in Mandarin is known as a stylistic characteristic of colloquial language (e.g., Chen 1999, Tian 2007), which was however shown to be favoured by formal speech styles in English varieties that are gaining rhoticity such as in New York City English (Labov 1994). It is unclear how this crosslinguistic contradiction would interact and shape the realisation of rhoticity in the English speech of Chinese learners. None of the existing L2 theories can provide any insights into the possible transfer opportunities of sociolinguistic aspects to L2 speech patterns. It also remains unknown how the social meanings that an L1 or a TL sound incorporates would possibly transfer between languages over the course of speech learning. As such, it can be particularly informative to investigate the development of the acquisition of a sociophonetic or sociophonological structure by L2 learners. The aim of the present study is to fill this research gap by exploring the acquisition of English nonprevocalic /r/ by Mandarin-speaking learners. Issues regarding the underlying impact of L1 dialectal variation, the transfer of sociolinguistic aspects, as well as context-dependent CLI will be examined longitudinally in the study at hand.

3. The relationship between perception and production

Another area of longstanding interest in L2 phonological acquisition is the intuitive connection between speech perception and production, which is also known as the *perception-production link* (e.g., Isbell 2016). This link has been well recognized in children's development of L1 speech in that accurate perception is a vital and natural requisite for accurate production of sounds (e.g., Smolensky 1996, Werker & Tees 2002). In the field of L2 phonology, however, the nature of the relationship between perception and production remains unclear (Lindblom 1996, Fowler 1996). Some literature suggests that a language learner must first be able to accurately perceive a sound in order to produce it (e.g., Flege 1995), while a number of studies cite counterevidence to the intuitive link, indicating that the two domains are instead controlled by different mechanisms (e.g., Allport et al. 1987, Baker & Trofimovich 2006). This chapter reviews two theoretical frameworks and empirical evidence supporting and challenging the link between speech perception and production. Section 3.1 introduces one of the most influential theories on L2 phonology which supports the view of perception preceding production on the basis of perceived crosslinguistic dissimilarity between L1 and TL sounds. Section 3.2 emphasizes the significant role played by some of the phonetic factors conditioning speech perception and production, respectively. Section 3.3 concludes this chapter by discussing learner variation in the development of L2 speech perception and production.

3.1 Speech Learning Model

Flege's (1995) *Speech Learning Model (SLM)* is one of the most influential models focussing on L2 phonological acquisition and offering a hypothesis on the perception-production link. It proposes that many difficulties in L2 production stem from the inaccurate perception of L2 sounds. In other words, it supports the view of perception determining production or perception preceding production. As discussed in section 2.1, in the SLM, CLI plays a crucial role in shaping the perception and production performance of L2 sounds by language learners. That is, similar TL sounds are more difficult for learners to acquire relative to dissimilar sounds. The SLM recognizes this phenomenon and assumes that there is a perceptual basis regarding the phonetic representations of the sounds.

According to the SLM, L2 learners "perceptually relate positional allophones in the L2 to the closest positionally defined allophone (or 'sounds') in the L1" (Flege 1995: 238). As such, L2 sounds can be perceived as *new*, *similar*, or *identical* by learners, the process of which has been termed "*equivalence classification*" (Flege 1995: 239). That is, "a single phonetic category will be used to process perceptually linked L1 and L2 sounds" (Flege 1995: 239), if the phonetic differences between the L1 and L2 sounds are too similar to be perceived by L2 learners. Predictions of segmental perception and production can be further made on the basis of L2 phonetic category formation. To be more specific, the SLM suggests that sounds perceived to be *identical* already exist in the L1 sound system of learners and thus do not demand any further learning. *New* sounds are also relatively easy for learners to acquire since they do not belong to the L1 sound inventory and new categories will be automatically established. *Similar* sounds, on the other hand, tend to pose major problems as the phonetic differences between the similar L1 and L2 sounds are difficult for learners to perceive. As such, the establishment of a new category for a similar L2 sound is impeded due to inaccurate perception, which is likely to cause inaccurate speech production. Flege (1995: 239) hypothesizes that "the greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned". In terms of L2 speech production, the SLM suggests that "the production of a sound eventually corresponds to the properties represented in its phonetic category representation" (Flege 1995: 239).

The SLM is based on a number of empirical studies conducted by Flege and his colleagues (e.g., Flege et al. 1999, Guion et al. 2000, Flege & MacKay 2004), which take extralinguistic effects such as input, age of acquisition, and amount of L2 experience into consideration and attempt to account for language learners' changes across the life span. However, many studies have provided conflicting evidence that poses challenges to the SLM. Chan (2014), for example, investigated the perception and production of English speech sounds (eight vowels and eight consonants) by Cantonese learners of English in Hong Kong in an attempt to examine the validity of the claims predicted by the SLM (Flege 1995). A total of 40 university students

majoring in English participated in a series of speech perception and production tasks. Regarding perception experiments, two L1-L2 perception tasks aiming at detecting the participants' perceived similarity between English and Cantonese and two L2 identification tasks testing their perceptual ability of the target L2 sounds were carried out. The English stimuli were produced in English and the Chinese ones were spoken in Cantonese by a female Chinese phonologist with near-native English proficiency. The production tasks involved recording the participants reading aloud a series of isolated words, minimal word pairs, and passages. With regard to the discrimination ability of the learners, results demonstrated that the perceived crosslinguistic similarity between Cantonese and English sounds exerts adverse effects on accurate perception. Specifically, similar TL structures, such as English tense and lax vowel pairs and corresponding Cantonese lax vowels at a position-dependent level, tended to be more challenging for the Cantonese learners of English to perceive than dissimilar structures, as predicted by the SLM. Concerning the relationship between perception and production, however, it was found that the link is not always straightforward. That is, accurate perception of a target sound does not necessarily lead to accurate production. For example, the Cantonese learners of English generally failed to produce the interdental fricative /ð/, in spite of their accurate perception of the sound. Chan (2014) argued that perceived cross-language similarity is not an absolute predictor of L2 speech production and other factors such as articulatory challenges often override the effects of perception on production and on the ultimate attainment of L2 sounds.

Apart from Chan (2014), Derwing and Munro (2015) also reported that the persistent articulatory difficulty that L2 learners encounter limits the strength of the perception-production link. They found that it was relatively easy for English learners of Spanish to discriminate the distinction between a trill /r/ and a tap /ɾ/ in Spanish, but the learners still encountered major difficulties when producing the sounds, which cites counterevidence for the SLM. Similarly, Kim and Park (1995) found that it was particularly challenging for English learners of Korean to accurately produce the /l/ - /ɾ/ contrasts although they can perceive the distinction successfully. As such, the perception-production link seems to be controlled by feature-specific aspects. For example, TL features involving greater articulatory difficulties are more likely to pose

significant challenges to the link as successful production of those features tends to be hindered by articulatory constraints. L2 learners' production of a sound is therefore not only conditioned by perceptual accuracy but also by articulatory factors. This view is also supported by a more recent work on the L3 phonological acquisition of Spanish rhotics by Chinese learners of L2 English, which argues that perception may affect the course of phonological development together with articulatory routine and orthography (Patience 2018).

Certain limitations of the SLM (Flege 1995) should be noted. With its focus on the perception and production of phoneme-sized units of speech by bilinguals (or non-beginners), its scope is restricted to phonemes and allophones. The question arises here as to how to evaluate perceived cross-language dissimilarity for TL structures that are more complex and beyond segmental representations. In other words, equivalence classification involving context-dependent and prosodic levels has not been specified in the SLM. For example, Cardoso's (2011) study on the acquisition of English codas by Brazilian Portuguese learners tried to test predictions of the SLM, but found it difficult to determine whether the target English CVC structure, which is always followed by a vowel in Brazilian Portuguese, should be classified as a similar or dissimilar structure for the learners. It is also unclear how to categorize L2 sounds that have identical phonetic properties but differ from L1 sounds in terms of the phonetic and phonological contexts they exist in. For instance, a particular L2 sound may exist in the L1 system in terms of its acoustic properties, which, according to Flege's (1995) equivalence classification of sounds, can be categorized into the identical category. However, the sounds may still differ concerning surrounding contexts and social meaning in different languages, which are also assumed to affect the acquisition of the new sounds. In particular, possible CLI is not only restricted to phonetic properties but social aspects as well, provided that different social meanings are indexed in the sounds between native and target language. Flege does not make an explicit mention on these sociolinguistic aspects of crosslinguistic perception of L2 learners. Besides, the SLM does not offer any insights into the longitudinal development of L1 speech perception and production. In addition, many studies have shown that perception might not be the only factor conditioning production, which can also be affected by other factors such as articulatory difficulty,

aptitude, and language experiences among others. The present study therefore attempts to offer empirical evidence to the observed gaps by testing linguistic and extralinguistic factors conditioning the relationship between the perception and production of an L2 sociophonological characteristic by learners over time.

3.2 Phonetic factors affecting speech perception and production.

A number of recent studies have recognised a combination of phonological and phonetic factors conditioning the perception-production link such as phonetic salience, prosodic domains, and articulatory arrangements (e.g., Broselow & Kang 2013, Davidson 2011), which are vitally important to understand the nature of the relationship between speech perception and production. The theories and studies that are discussed in this section focus on the phonetic considerations that guide L2 speech perception and production. Section 3.2.1 explores how articulatory constraints interact with phonological influences on L2 speech production and presents a theoretical framework proposed by Colantoni and Steele (2008), who examined the acquisition of rhotics by L2 learners and incorporated articulatory phonetics into existing L2 speech models such as the MDH (Eckman 1975) and the SLM (Flege 1995). Section 3.2.2 further discusses the role of the prominence of phonetic cues on speech perception by L2 learners.

3.2.1 Articulatory constraints on L2 speech production

As discussed in Section 3.1, accurate production of L2 sounds by foreign language learners can be conditioned by the degree of the articulatory difficulty involved in the articulation of certain L2 structures. Davidson (2011), for example, argued that both articulatory correlates and aerodynamic pressures are important phonetic factors that interact with phonological impacts on the production of L2 speech. Broselow and Kang (2013: 3) also claimed that various L2 speech patterns "may take place not in phonology but rather at the level of phonetics, where speakers may simply fall short in mastering the correct articulation". Colantoni and Steele (2008) proposed a *Perception-Production Model of L2 Segmental Acquisition (MSA)*, which emphasizes the significant role played by articulatory constraints including aerodynamic and elasto-inertial effects on the production of L2 rhotics. It investigates the acquisition of French /ʁ/ and Spanish /ɾ/ by English-speaking learners and sets out to test the predictive acounts of L2 phonological models including the MDH (Eckman 1977) and

the SLM (Flege 1995). In their study, speech production data were collected from 20 English learners of L2 French and 19 English learners of L2 Spanish performing a series of sentence-reading and passage-reading tasks. Results revealed that the learners did not master the target rhotics in all positions simultaneously. In particular, the English learners of Spanish successfully produced /ɾ/ in intervocalic position, while the French learners only acquired frication in word-initial position. As a consequence, neither of the frameworks was successful in predicting these observed patterns, because they do not consider possible context-dependent or position-based acquisition patterns or the complex interaction of different phonetic parameters (e.g., manner versus voicing for English learners acquiring the voiced dorsal fricative French /ʁ/). Specifically, the assumption of the MDH concerning the accurate production of intervocalic Spanish /ɾ/ was partially confirmed by the findings, which, however, failed to predict the acquisition of frication before voicing and the onset-coda asymmetry in both L2s. Similarly, while the SLM accurately predicted the onset-coda asymmetry of rhotic acquisition in L2 Spanish, it failed to account for the developmental sequence of manner preceding voicing for the English learners of French.

Colantoni and Steele (2008) argued that the observed acquisition patterns that the MDH and the SLM failed to predict can be best explained by phonetic principles. According to Colantoni and Steele, learners tend to first target more salient phonetic properties in more salient positions when learning a new segment. This argument is supported by Ohala and Kawasaki (1984) and Wright (2004), who showed that sounds are not equally perceived in various contexts. As such, Colantoni and Steele suggested that additional phonetic factors of speech production should be integrated in to L2 phonological models. They developed a perception-production model taking phonetic principles including aerodynamics of voicing, Lindblom's (1990) hyper- and hypoarticulation theory, and Honikman's (1964) articulatory settings theory into consideration (see Figure 3.1).

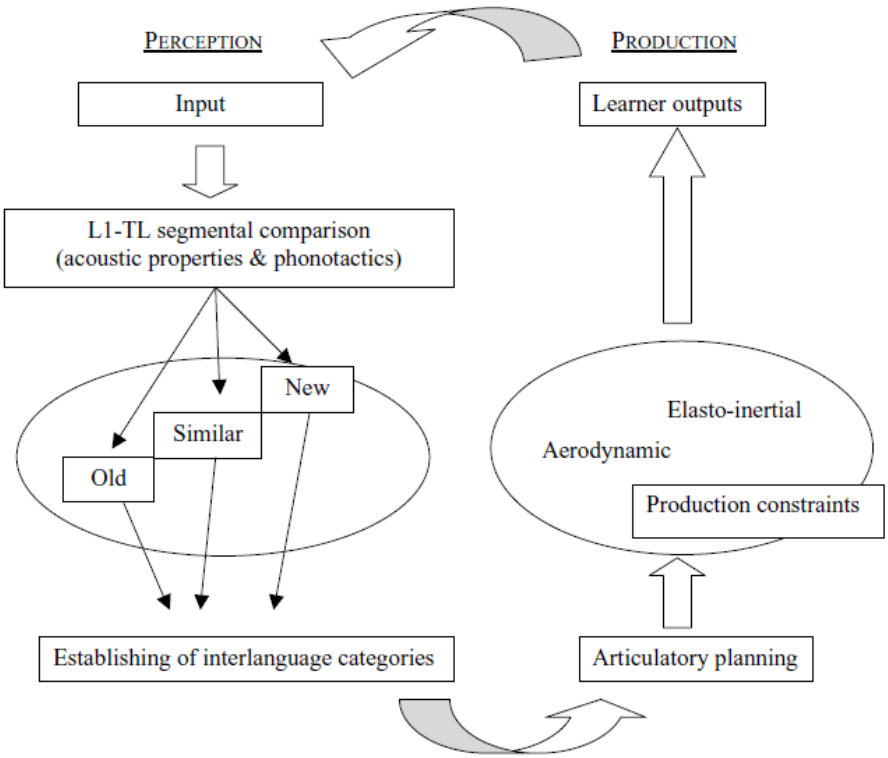


Figure 3.1 A representation of a perception-production model of L2 segmental acquisition (adapted from Colantoni and Steele 2008: 522)

Figure 3.1 shows that acquiring an L2 segment involves the interaction of perception and production. In terms of perception, information involving acoustic properties and phonotactics that is indexed in the input is first analysed by learners into segmental classifications. Colantoni and Steele (2008) noted that differences between less and more proficient learners are expected here. Specifically, less proficient learners tend to rely on acoustic properties, while more proficient learners are more likely to depend on distributional information, the mastery of which requires more linguistic experience. As a next step, the input is classified into old, similar, and new IL categories according to equivalence classification as proposed by the SLM (Flege 1995). Furthermore, the production domain starts necessary articulatory planning for the production of the IL categories established in the previous step. Crucially, it should be mentioned that accurate perception and accurate articulatory planning do not necessarily guarantee successful or target-like production. As indicated in figure 3.1, various production constraints such as aerodynamic and elasto-inertial effects can possibly condition production performance. The production is then compared with the

perceived input by L2 learners. As Colantoni and Steele (2008) state, if no distinctions are detected between the perceptual input and production output by learners or if there is a lack of articulatory control, TL categories will remain unchanged over the course of acquisition.

3.2.2 Robustness of phonetic cues on L2 speech perception

The acoustic salience of L2 structures or the perceptual needs of listeners to accurately perceive a sound are other influential phonetic factors conditioning L2 phonological acquisition. Archibald (2009:232), for example, states that "[…] phonological features are units of mental representation whose presence is cued by phonetic features […] The degree of learning difficulty may also be influenced by the robustness of the cue to the feature not just by the differences at the phonological level". In other words, the acquisition of an L2 phonology is not only concerned with the attainment of the L2 phonological grammar by learners but is also affected by the salience of the phonetic cues of specific L2 sounds. Wright (2004: 35) also points out that "phonological processes such as positional neutralization, gemination, and assimilation are motivated by interactions between the strength of perceptual cues and some notion of articulatory ease". According to Wright, *cue* means "information in the acoustic signal that allows the listener to apprehend the existence of a phonological contrast" (2004: 36). While much recent work in phonological theory such as the SLM (Flege 1995) has highlighted the influence of perception on production in L2 phonological acquisition, the potential impact of acoustic salience on speech perception is worth investigating. It should be mentioned that the SLM does not consider the influence of acoustic salience on the accurate perception of L2 sounds by learners. Instead, it argues that whether an L2 sound can be discriminated or not is determined by the extent of perceived crosslinguistic dissimilarity of the sound to individual L2 learners.

A number of empirical studies on the acquisition of the English liquid contrast by native Japanese learners have offered insights into the role played by acoustic salience on the perception-production link (see Bradlow 2008 for a comprehensive review). In general, it was found that Japanese learners encountered major difficulties acquiring English /l/ and /ɹ/ in both perception and production. However, Larson-Hall (2004) demonstrated that they were able to successfully acquire the Russian /r/. It was then

speculated that the trilled nature of the Russian /r/, which involves a more robust phonetic cue than the approximant /ɹ/ in English, is linked with its acquisition success. These findings revealed that the more robust the phonetic properties of a feature, the more likely it will be acquired by L2 learners, regardless of the impact of CLI or perceived crosslinguistic dissimilarity as predicted by the SLM. Davidson (2011) further argued that the persistent difficulties in acquiring English /ɹ/ by Japanese learners are attributable to both perceptual and articulatory factors. In terms of perception, the distinction between /l/ and /ɹ/ involves acoustic cues characterized by frequency differences in F3 and the intensity of the formant, which do not exist in L1 Japanese and are therefore difficult for Japanese learners of English to distinguish. Regarding production, English /ɹ/ is an articulatorily complicated structure that involves "a pharyngeal constriction, a coronal constriction or retroflexion, and lip rounding" (Davidson 2011: 131), which poses great challenges for its successful production by Japanese learners.

The significance of acoustic salience on L2 speech acquisition has also been shown by Colantoni and Steele (2007), who explored the role of basic phonetic principles in determining position-dependent variation in the acquisition of L2 French /ʁ/ and found acquisition asymmetries due to phonetic factors. In their study, a total of 20 intermediate and advanced-proficiency level English learners of L2 French were examined learning the French voiced uvular fricative /ʁ/. The phonetic differences between English and French rhotics differ distinctly concerning both place of articulation (coronal in English and dorsal in French) and manner (approximant in English and fricative in French). Therefore, in order to acquire /ʁ/, English learners of French must master both voicing and frication in the segment at once. Results showed that the English learners of French acquired the parameter of frication before voicing, which, according to the authors, indicates that the high degree of acoustic salience of the fricated noise of French /ʁ/ allows the accurate perception and production of the sound by the learners. This also confirms the predictions of the SLM (Flege 1995) in that the French voiced dorsal fricative as a new category for English learners is more readily acquired by the learners. Moreover, the relatively successful acquisition of frication was found to be position-based: it was first realised in the most salient onset position rather than in syllable codas.

Colantoni and Steele (2007) further argued that the observed acquisition asymmetries seem to be a consequence of basic phonetic principles rather than of language universals such as markedness. According to Colantoni and Steele, L2 aspects that are acoustically salient and articulatorily easier will be acquired first regardless of the relative degree of markedness of the structures. As the authors pointed out, basic phonetic theories could possibly make important theoretical contributions to the frameworks of L2 phonological acquisition. However, it is noteworthy that the impact of acoustic salience on the L2 speech perception is still severely understudied. Current L2 phonological models like the SLM (Flege 1995) mainly draw on the degree of perceived crosslinguistic dissimilarity between L1 and L2 sounds to account for perceptual difficulty. I would argue that the perception of a sound by L2 listeners is not only determined by CLI but also the salience of the phonetic cues. More specifically, an L2 sound might be more accurately perceived by a learner not due to its perceived crosslinguistic dissimilarity from the learner's L1 category, but its acoustic salience given its long duration or dramatic formant modulation. In my perception study, I will try to explain L2 learners' perceptual performances from both perspectives.

3.3 Variability in learners' L2 patterns

There is a growing body of literature that recognizes inter- and intra-learner variation in the relationship between speech perception and production over the course of L2 phonological acquisition. Baker and Trofimovich (2006), for instance, examined English vowel perception and production by 40 Korean learners of English and reported that the relationship between their performances in the two domains varied between learners, which was mainly determined by their age of arrival in an English-speaking country. A more recent work by Yu (2019) investigated individual variability in the perception and production of English sibilant-vowel coarticulation by native American English speakers and found "a positive correlation between how much an individual attends to context-specific variation in perception and how the sibilant contrast is realized in specific vocalic contexts" (1), suggesting a context-dependent individual variation in the perception-production link.

Cardoso (2007, 2011) explored the acquisition of English word-final stops by

Brazilian Portuguese learners of English and found variation in their perception and production patterns affected by both linguistic and extralinguistic factors. Cardoso's studies followed a sociolinguistic variationist approach for data collection and statistical analysis. In the acquisition of English, Brazilian Portuguese learners are expected to produce English codas as onsets of epenthetic vowels to various extents since codas are illicit in the L1 Brazilian Portuguese. In the production study (Cardoso 2007), six English learners on three proficiency levels (beginners, intermediate, and advanced learners) participated in three production tasks: wordlist, passage reading, and interview. A total of 1,859 tokens were collected and annotated into binary realizations: target-like production of word-final stops and L1-oriented i-epenthesis. In addition, each token was coded in terms of five linguistic variables (word type, place of articulation, word size, and stress) and three extralinguistic factors (proficiency, style, and participants).

Results suggested that the production pattern of word-final stops by the six Brazilian Portuguese learners is conditioned by both linguistic and extralinguistic factors. More specifically, place of articulation, word size, proficiency level and style were found to trigger variation in their production patterns. Target-like production was found to favor more formal speech style, in polysyllabic words, and when the coda is a coronal. Advanced learners were shown to produce a larger amount of target-like realizations than beginners and intermediate learners. Cardoso (2007) concluded that the findings supported the OPM (Major 2001), which posits dominant L1 influence at earlier stages of acquisition and the possibility of target-like behavior at later stages for language learners over the course of L2 phonological acquisition.

Following Cardoso's (2007) work on coda production, Cardoso (2011) carried out a perceptual study on the acquisition of the perception of English word-final stops by Brazilian Portuguese learners and further explored the relationship between coda perception and production. It was investigated whether the factors conditioning coda production in Cardoso (2007) also affect coda perception. A total of 51 subjects on four proficiency levels (no English, beginner, intermediate, and advanced learners) participated in a forced-choice phone identification task and were asked to discriminate whether the English pseudowords presented to them ended in a

consonant or in a vowel. The perceptual stimuli were controlled for word size, place of articulation, preceding vowel, stress, and segmental template, which were taken as potential linguistic factors predicting perceptual variation.

The findings indicated that proficiency, place of articulation, and preceding vowel are main effects conditioning coda perception. Specifically, codas are more likely to be accurately perceived by the learners of more advanced level of proficiency, in the context of segments that belong to the class of coronals and labials, and when the coda consonant is preceded by a lax vowel. Cardoso (2011) expanded the scope of the OPM (Major 2001) to the domain of developmental patterns in L2 speech perception and argued that the positive correlation between perception accuracy and proficiency level observed in his study is in line with predictions of the OPM. Based on the developmental patterns observed in the two domains, Cardoso further claimed that "to some extent, the results obtained show a correlation between speech perception and production" (2011: 433). This conclusion is based on the general trends observed that the Brazilian Portuguese learners' performance in both coda perception and production improves with the increase of their proficiency levels.

Cardoso (2011), moreover, compared the main effects predicting the observed perception and production patterns. It was hypothesized that similar main predictors would be shared in the perception and production of English coda stops due to the assumption that "production is a manifestation of perception" (Cardoso 2011: 447). However, the results provided contradicting evidence. First, word size showed significant effects on coda production but not perception, indicating that this factor could be production-specific only, which is considered to be associated with articulatory complexity (e.g., Shariatmadari 2006). Also, preceding vowel environments were found to condition perception but not production. Specifically, the lax vowels promoted accurate coda perception, while tense vowels behaved in an opposite direction. Moreover, both perception and production were found to be constrained by places of articulation. The predictive patterns of different articulatory places were however not entirely consistent between perception and production. To be more specific, while coronals constituted the only favouring phonological context of the production of word-final stops, their accurate discrimination was promoted by

both coronals and labial consonants. According to Cardoso, the pattern observed for places of articulation can possibly be accounted for by a variety of factors including markedness, acoustic salience, and input frequency. It was hypothesized that codas that are more marked, less frequent, and acoustically less salient will be harder to acquire than those that are less marked, more frequent, and more salient. Cardoso (2011) claimed that none of the previously established hierarchy of codas concerning the above-mentioned three perspectives alone can satisfactorily account for the observed perception-production pattern of coda stops. Thus, it is necessary to explain the acquisition of L2 speech within a multifaceted perspective and consider all possible influencing factors to provide a more comprehensive account for his findings.

It should be mentioned that the production data in Cardoso (2007) were collected from six Brazilian Portuguese speakers, with only two participants in each of the three proficiency groups. The small sample size should make for highly-cautious interpretations of findings, and the observed developmental patterns concerning production should thus be treated with caution. Furthermore, two different sets of participants were recruited for the perception and production tasks, respectively, which made the investigation of individual perception-production links impossible. In other words, since the production measures were not drawn from the same individuals who participated in the perceptual task, it is unclear if the link was evident at individual levels. It is also questionable as to whether comparing perception and production performance by different groups of subjects is legitimate. Last but not least, Cardoso (2011) reported a positive perception-production correlation only based on the proficiency of his participants. The more proficient the subjects, the more accurate they were found in both perception and production. I would argue that these findings suggest a developmental pattern of the acquisition of English codas by L2 learners in both domains, which, however, need not be viewed together as a perception-production correlation.

In spite of the limitations, Cardoso (2007, 2011) proposed an inspiring multifaceted approach which set out to comprehensively account for variation in speech perception and production of L2 learners. The multidisciplinary approach it employed

incorporated theories and methodologies from various fields including sociolinguistics, SLA, and formal phonology, and took both linguistic and extralinguistic factors into consideration. In this dissertation, I will also apply the multifaceted approach on the longitudinal acquisition of an L2 sociophonological variable, i.e., English rhoticity, by Mandarin-speaking learners. While the small sample size and incomparability between perception and production data present threats to validity in Cardoso's (2007, 2011) studies, I will recruit a large sample of learners for both perception and production measures over a time span of six months to generate more robust conclusions regarding the long-term relationship between perception and production at individual levels. Moreover, as rhoticity is known to be subject to both phonological and sociolinguistic factors which may also condition its perception and production by L2 learners, the present study will follow a sociolinguistic variationist approach for data collection and analysis to take into account all possible constraints on the acquisition and development of L2 rhoticity.

4. Rhoticity in English

Rhoticity, i.e., the presence of *nonprevocalic /r/*[1], is known as one of the most salient phonological features that index dialectal and sociolinguistic differences in English varieties (Labov 1966, Trudgill & Hannah 2013). This feature has been extensively examined in both diachronic and synchronic research in the fields of sociolinguistics, world Englishes, and sound variation and change, showing that both grammar-internal and -external factors are influential in modulating the gradience and variability of English rhoticity. This chapter aims at providing a foundation that will help account for inter- and intra-learner variation in the perception and production of English rhoticity by the L2 learners in the present study. This is based on the assumption that sound patterns of language variation and change in native speech might to some extent inform patterns of leaners' L2 speech (e.g., Colantoni & Steele 2007, Kabak 2019). As such, I would argue that predictors conditioning the realisation of rhoticity in native English may also affect its acquisition by L2 learners.

The following chapter is concerned with English nonprevocalic /r/. After a brief review of the phonetic and phonological aspects of the sound in section 4.1, factors that have been found to condition its presence or absence in English are presented in section 4.2. By drawing from previous work on rhoticity in world Englishes, effects of both grammar-internal and -external constraints are compared and discussed. The chapter closes with a brief discussion of the underlying relationship between native and non-native speech (section 4.3).

4.1 Phonetic and phonological aspects of rhoticity

Rhoticity refers to the production of the consonantal variants of /r/ in syllable coda position. It is phonetically and phonologically variable across different English varieties, not only in their actual phonetic realisations but also in their distribution within different phonological environments. Broadly speaking, accents of English that produce /r/ in syllable codas are considered to be *rhotic*, while those that do not realise the sound are *nonrhotic*. Regarding its various phonetic realisations, /r/ can, for

[1] nonprevocalic /r/, postvocalic /r/, and coda /r/ are often used interchangeably in previous research on rhoticity to refer to the consonantal variants of /r/ in syllable codas. In this dissertation, the term nonprevocalic /r/ will be used throughout the discussion because the realisation of linking /r/, i.e., postvocalic /r/ or coda /r/ followed by a vowel, is not examined in the present study.

example, be produced as retroflex or bunched approximants in American English, or as tap/trills in Scottish English. It should be noted that this study focusses explicitly on the phonological presence or absence of approximant /r/ in the nonprevocalic position, without taking into account different phonetic types and *linking /r/*, i.e., coda /r/ followed by vowels. The following subsections present the articulatory, acoustic, and perceptual properties of the /r/ sound as well as the connection between nonprevocalic /r/ and its preceding contexts.

4.1.1 Articulatory, acoustic, and perceptual properties of nonprevocalic /r/

Concerning the articulation of the approximant [ɹ], a number of variable gestures have been recognized. According to Espy-Wilson et al. (2000), the articulatory configurations for [ɹ] mainly involve constrictions "in the pharynx, along the palatal vault, and at the lips" (344). It usually involves "a small retraction of the root of the tongue, 4-6 centimetres above the larynx" (Lindau 1978: 554). Two contrasting types are typically identified: *retroflex* and *bunched,* which are also referred to as *apical* and *molar* (e.g., Delattre & Freeman 1968). In the bunched category, the tongue dorsum is raised toward the palate and the tongue tip is lowered, while the tongue tip curls up and points toward the hard palate in the retroflex /r/ realisation. Delattre and Freeman's (1968) seminal work on the X-Ray motion pictures of American /r/ further divides the binary classification of /r/ into eight different types of articulation by examining native male English speakers with different regional dialects, showing a wide variability of possible articulatory properties of /r/.

The various articulatory configurations of /r/ interestingly lead to the same acoustic properties. The acoustic properties of nonprevocalic /r/ in American English are usually characterised by a sharp lowering of the third and fourth formants (F3 and F4) (e.g., Stevens & Blumstein 1975), a simultaneous rise of the second formant (F2) and drop of F3 (e.g., Stevens 1999), or a low-frequency F3 with the first formant (F1) and F2 values similar to those found in a canonical central vowel (e.g., Espy-Wilson 1992). According to earlier empirical work, the actual formant values of [ɹ] range from approximately 250 Hz to 550 Hz for F1, 900 Hz to 1500 Hz for F2, and 1300 Hz to 1950 Hz for F3 in American English across genders and various phonological conditions (Delattre & Freeman 1968, Espy-Wilson 1992, Espy-Wilson et al. 2000). The relative consistency of the acoustic properties of approximant [ɹ] across different

articulation types might be attributed to the retraction of tongue root that results in lowered F3 and F4 regardless of the position of the tongue tip (Lindau 1978: 556). Delattre and Freeman (1968), on the other hand, claimed that there is no correlation between tongue shapes and formant values. The spectrograms of bunched and retroflex articulation in their investigation exhibited no significant distinction in formant frequency. Delattre and Freeman (1968: 67) concluded that the articulatory correlates of acoustic differences lie in "constriction and syllabic position" of approximant [ɹ]. More specifically, the narrowing of the palatovelar constriction lowers the F3 and the degree of narrowness correlates with the degree of F2 - F3 proximity.

The view that a low-frequency F3 contributes crucially to the perception of rhoticity has been challenged by several empirical studies on the perception of rhoticity. Heselwood (2009) and Heselwood and Plug (2011), for example, carried out a series of perceptual experiments on the relative contributions of F2 and F3 to the perception of rhoticity. In Heselwood (2009)'s study, forty phonetically trained participants listened to a series of rhotic and nonrhotic English tokens in unfiltered and filtered conditions. In the filtered condition, all acoustic formants above F2 during the vocalic portion were removed by lowpass filtering, while all spectral components up to 5.5 kHz were presented in the unfiltered version. An unexpected result showed that a clear majority of the listeners reported hearing stronger rhoticity in the filtered condition of rhotic tokens and even regarded the filtered nonrhotic tokens without F3 as rhotic. Heselwood (2009) therefore claimed that instead of inducing the perception of rhoticity, F3 is shown to restrict its perception. Heselwood and Plug (2011) further corroborated Heselwood's (2009) results and observed that the removal of F3 strengthens the perception of rhoticity by their participants. It has been concluded by Heselwood and Plug (2011) that a low or falling frequency of F3 is not crucial for rhoticity perception but rather the proximity of F3 to F2 values.

Evidence contradicting Heselwood (2009) and Heselwood and Plug (2011) comes from a study by Nagy (2008), who examined the relationship between the rate of perceived occurrence of rhoticity (which Nagy referred to as "perceptual frequency") and acoustic formant values (referred to as "formant frequency"). Results showed that

patterns of perceptual frequency are supported by formant frequency. To be more specific, younger speakers that were perceived to be more rhotic also had significantly lower F3 relative to old speakers. Similarly, college-educated speakers that were perceived to be the least rhotic compared to their uneducated counterparts were found to have the highest F3 values. As such, Nagy (2008) argued that there is a strong correlation between the perception of rhoticity and the F3 values of the nonprevocalic /r/ variants and "contexts where [ɹ] appears more frequently have acoustically distinct forms" across different speakers and linguistic environments (101).

Apart from F3, the extent of formant transition across /V(r)/ sequences has been considered an important phonetic cue in perceiving coda rhotic as well. Storme (2018) investigated the perception of the French coda [ʁ] by 20 native French speakers. In his study, nonce-word stimuli controlled for the presence or absence of postvocalic rhotic preceded by [i] and [a] were constructed and listeners were asked to discriminate rhotic-nonrhotic contrasts such as *amito* vs. *amirto* and *amato* vs. *amarto.* Results showed that the participants were better at perceiving rhoticity following [a] than [i]. According to Storme (2018), [ʁ] has a stronger lowering effect on the F2 of [a] than [i], which results in a greater distance of the F2 values between *amato* and *amarto* than *amito* and *amirto*, hence the greater perceptibility of the [aʁ] sequence relative to [iʁ]. This confirms findings by Gendrot (2014), who reported that both formant transition and duration are important acoustic cues in the perception of coda [ʁ] in French (cited in Storme 2018). As Gendrot suggested, the duration of [a] in French is found to be longer than that of [i], which may lead to different discrimination ability of the rhotic after [a] and [i], and further research investigating the effect of preceding vowel quality on the perception of coda /r/ is needed.

4.1.2 Phonological and empirical accounts for the existence of /Vr/ sequences

There seems to be little agreement with regard to the phonological treatment of /Vr/ sequences in rhotic English varieties (e.g., Woolley 1978: 5). Kenyon (1950) and Hultzen (1950), for instance, considered the vocalic nuclei in NURSE, NEAR, SQUARE, START, FORCE, and NORTH in American English as vowels and transcribed them with /ɝ, ɪɚ, ɛɚ, aɚ, ɔɚ, uɚ/. Thomas (1958) and Fairbanks (1960), alternatively, only treated the vocalic nucleus in NURSE as a vowel /ɝ/, while analysing the others as sequences of vowel plus glide or semivowel and transcribing them as /ɪr, ɛr, ar, ɔr, ur/

respectively. Moreover, some linguists regarded the nonprevocalic diphthongal offglide [ɚ] as an allophone of /r/ and transcribed all of the nuclei in NURSE, NEAR, SQUARE, START, FORCE, and NORTH as vowel plus /r/, namely /Vr/ sequences (e.g., Gleason 1961). According to Woolley (1978: 5), a consensus only exists regarding the phonetic facts that "[ɝ] is a mid-central monophthongal vowel and [ɪɚ, ɛɚ, aɚ, ɔɚ, uɚ] are centering diphthongs". However, it is still not clear whether all of them should be treated as /Vr/ sequences or whether a phonological treatment of the monophthong [ɝ] different from those centering diphthongs is necessary.

Woolley (1978) further provided empirical insights into the phonological treatment of English nonprevocalic /r/. By exploring the perceptual similarity across different /Vr/ sequences based on a framework of perceptual confusions (cf. Miller & Nicely 1995), he showed that the centering diphthongs NEAR, SQUARE, START, FORCE are perceptually distinct from one another, with frequent misidentifications between the monophthongal [ɝ] and each of the centering diphthongs. Woolley argued that the monophthong functions "as an intersection of confusion" and "serves as a perceptual link" for the diphthongs (1978: 21). As such, it is concluded that the analysis of the perceptual confusions among /Vr/ sequences does not support a treatment of the monophthongal [ɝ] separate from the diphthongs since the perception of the monophthong is not distinct from them. Woolley (1978) suggests that it is then legitimate to handle both monophthongal [ɝ] and the centering diphthongs as /Vr/ sequences.

A more recent study has revealed durational differences of nonprevocalic /r/ between monophthongal and diphthongal /Vr/ sequences, which may confirm the necessity of distinctive phonological treatments for the monophthongal [ɝ] after all. Kuecker et al. (2015) investigated the duration of the rhotic portion in words with /r/ in the coda position. The durations of the rhotic parts of the /r/ tokens were measured in recordings of 15 native speakers of American English. Their findings demonstrated that the NURSE vowel is usually completely rhoticized with an average rhotic portion amounting to 94% of the whole vocalic nuclei. It is followed by 76% for LETTER which is slightly less r-coloured relative to its stressed counterpart. The rhotic portion of the centering diphthongs such as FORCE, SQUARE, and NEAR is the shortest (58%)

(626). However, no significant distinction in duration of the rhotic was found concerning different syllable structures between singleton /r/ in the coda position and /r/ in consonant clusters, indicating that the duration of the rhotic was subject to the preceding vowel contexts but not the syllable structures that it exists in.

Apart from monophthongal [ɜ˞], START was always realized as a rhotacized vowel as well, the realisation of which was found to be conditioned by speaking rate, consonantal context, and speaker differences (Espy-Wilson 1992: 754). For instance, Espy-Wilson (1992) suggests that /a/ and /r/ are completely assimilated and coarticulated in the production of *cartwheel* by American English speakers and that "no discernible acoustic cue points to separate /a/ and /r/ segments" (754). The /a/ and /r/ are therefore realized as one segment acoustically with consistent formant values and a relatively steady F3 throughout the vocalic /ar/ sequence. Likewise, Ulbrich and Ordin (2014: 31) indicated that "in sequences of /ar/ the central vowel quality of a vocalised /r/ cannot be distinguished from the preceding /a/".

Lockenvitz et al. (2015) also proposed that START is a rhotacized vowel instead of a vowel followed by a consonantal /r/. They described the binary distinction between *consonantal* and *vocalic* realisations for English /r/ from a speech-language pathology point of view. According to Lockenvitz et al. (2015), *vocalic /r/* refers to a rhotic vowel or an *r*-colored vowel, which has "a rhotic tongue gesture coterminous with the duration of the vocalic segment" (614). Items such as NURSE, LETTER and START belong to this category, the production output of which is entirely rhotacized. A vowel followed by a rhotic consonant, on the other hand, "has the tongue gesture for the vowel followed by a change in tongue shape" to produce the separate /r/ segment (614). Thus, the /Vr/ sequences in NEAR, SQUARE and FORCE are typically produced with the rhotic segment surfacing only at the end of the vowel.

The distinction between rhotacized vowels and vowels followed by a rhotic consonant is also indicated by the vocalisation of nonprevocalic /r/ in nonrhotic varieties of English. It is well-known that nonprevocalic /r/ has undergone reduction in England in the mid-18th century. According to Heffner (1950), "in Southern British speech this *r* simply disappears from words like *far, farm, fur, first*, but is replaced by [schwa] in

words like *fear, fair, fire, floor, boor*, or in *fierce, fairs, fires, floors, boors.*" (149-150). In other words, the nonprevocalic /r/ is deleted in the rhotacized vowels (i.e., NURSE, LETTER and START) but is replaced by a schwa in vowels followed by a rhotic consonant (i.e., NEAR, SQUARE and FORCE). Nonprevocalic /r/ has been observed to be realised as schwa in nonrhotic varieties of English. Gick (2002) provided articulatory evidence for the connection between schwa and nonprevocalic /r/ in that both of them involve a significant retraction of the tongue root, which could possibly account for the replacement of nonprevocalic /r/ following a non-low vowel (e.g., in NEAR, SQUARE and FORCE) in nonrhotic English varieties.

Hickey (2014) further pointed out the direct connection between nonprevocalic /r/ and its preceding contexts from a diachronic perspective by exploring the effect of nonprevocalic /r/ on its preceding vowels in English. Two major types of merger were described: NURSE merger and HORSE-HOARSE/POOR-POUR mergers. The NURSE merger involves the collapse of a front/back distinction in *fir, fur*, and *fern*, which is attributed to the rhotacization of the entire syllable nucleus. Lawson et al. (2013) have further proposed that the NURSE merger in Scottish English is associated with the coarticulation of the vowels and the following /r/, and that the bunched /r/ has "a stronger global coarticulatory force over preceding vowel tongue configurations" than a retroflex /r/ does (198). Similarly, Hickey (2014: 13) indicated that "the present development is part of a long-term tendency in the phonetic typology of English for material in syllable codas to spread back to the nucleus or for the articulatory difference between nucleus and coda to be reduced". Lehiste (1962: 64) also proposed that nonprevocalic /r/ "appears to depend to a considerable extent upon the vowels associated with them".

To summarise, the earlier work reviewed in this section has demonstrated that the realisation of nonprevocalic /r/ in English varieties is closely associated with the preceding vowel contexts that it occurs in. The studies provided different phonological and empirical accounts for the existence of the /Vr/ sequences. In general, NURSE and LETTER are monophthongs and need to be viewed differently and investigated separately from the centering diphthongs NEAR, SQUARE, START, and FORCE considering their distinct durations (Kuecker et al. 2015) and the vocalisation

processes of the sequences from a diachronic perspective (Heffner 1950, Hickey 2014). The monophthongs are typically realised as rhoticised vowels, whereas the diphthongs are produced as a vowel followed by a consonantal /r/ (Lockenvitz et al. 2015). Among the diphthongs, the realisation of START approximates to that of the monophthongs as a highly rhoticised vowel.

The clear qualitative and quantitative differences in the vowels within such sequences are associated with various degrees of acoustic salience of them. It is well acknowledged that the salience of segmental sequences can be predicted by the strength of different acoustic cues (e.g., Wright 2004). According to Wright (2004), the term *cue* refers to "information in the acoustic signal that allows the listener to apprehend the existence of a phonological contrast" (36). Earlier work has shown that inherent phonetic properties could possibly contribute to the perception of prominence in English. Fry (1955, 1958), for instance, found that duration, fundamental frequency, and amplitude can affect the relative acoustic salience of a syllable to listeners, with duration being the strongest cue among the three. Kochanski et al. (2005) also found that acoustic salience in English is primarily characterized by duration and loudness, while fundamental frequency only plays a minor role. Therefore, the acoustic salience of the /Vr/ sequences is expected to exert significant impact on the acquisition of the perception of rhoticity by L2 learners. More specifically, acoustically more salient sequences are predicted to be easier for learners to perceive and acquire, while those that are of less salient cues are relatively difficult for accurate perception.

In section 7.4, the salience of the acoustic cues (duration and formant frequency (F2 and F3 values)) of the /V(r)/ sequences and the acoustic properties of the preceding vowels (F1 and F2 values) of the stimuli employed in the perception experiment of the present study will be measured and compared in order to predict and explain the accuracy orders in the perception of rhoticity by L2 learners. It is one of the aims of this study to find out to what extent the connection between nonprevocalic /r/ and its preceding environments can possibly affect the development of the perception of English rhoticity by learners over time. Furthermore, it will be explored whether universal patterns regarding the /Vr/ sequences can be observed between the speech of native English speakers and L2 learners.

4.2 Factors influencing the production of rhoticity in world Englishes

This section is concerned with grammar-internal and -external factors that can account for the variable realisation of rhoticity in world Englishes. Researchers have sought to understand how various constraints influence the phonological realisation of nonprevocalic /r/ in English worldwide since Labov's (1966) seminal work on the social stratification of New York city English, which focussed on rhoticity as a gradient sociolinguistic phenomenon. Previous studies investigating the variation and change of rhoticity in different English varieties typically employed multivariate sociolinguistic approaches to examine potential constraints on rhoticity. It is the goal of the present study to examine whether the same factors conditioning L1 rhoticity operate on L2 acquisition of nonprevocalic /r/ as well. In the following subsections, the effects of both grammar-internal and -external variables investigated in a multitude of English varieties by earlier work are compared and discussed.

4.2.1 Grammar-internal variables

This section reviews and discusses the four most frequently investigated grammar-internal variables conditioning the presence or absence of nonprevocalic /r/ across world Englishes. The factors are as varied as preceding vowel, stress, word context, and lexical frequency.

Preceding vowel

Preceding vowel environment has been shown to be one of the major factors influencing the realisation of nonprevocalic /r/ across English varieties worldwide. In Nagy and Irwin's (2010) study on rhoticity in Boston English, the authors demonstrated that the stressed central vowel NURSE is the most favouring context for the presence of /r/, whereas the unstressed central vowel LETTER is the least. Their results showed that the factor weights of the full vowels ranked in the following order of START > SQUARE > CURE > NEAR > FORCE, which, according to Nagy and Irwin (2010), did not coincide with any phonological feature of vowels. A striking difference in the effect of the preceding vowel was further shown between age groups. In the younger group, preceding START and CURE vowels outranked NURSE in favouring rhoticity, while a preceding NURSE vowel was the most favouring of the /r/ variant in the older group. Piercy (2012) investigated /r/ loss in Dorset English, a representative English dialect in southwest England. The results of her study showed

that a preceding NURSE vowel is the most promoting of nonprevocalic /r/, while the back vowels in FORCE strongly disfavour rhoticity. The effects of preceding vowels ranked in the following order: NURSE > NEAR > START > LETTER > CURE > SQUARE > FORCE. Piercy (2012) compared her findings with Nagy and Irwin's (2010) and concluded that a preceding NURSE vowel was always the most favouring context of rhoticity, with the back vowels being disfavouring of rhoticity independent of whether a variety is gaining or losing rhoticity. The other preceding vowels had rather mixed effects. Dickson and Hall-Lew (2017) also demonstrated that the back vowels in NORTH, FORCE, and START disfavour rhoticity compared to the front vowel environments in Edinburgh English. Similarly, Becker (2014)'s study on rhoticity in New York City English, a dialect that is gaining rhoticity like Boston English, showed that rhoticity was strongly favoured in central vowel NURSE. The other vowels disfavouring rhoticity in Becker (2014) were ranked from the highest to the lowest factor weight in the following order: NEAR > START > SQUARE > FORCE. In comparison with previous studies on rhoticity in American English dialects, Nagy and Irwin (2010) further concluded that the lack of universality in the hierarchy of various preceding vowel effects "may fruitfully be used to test for influence of one dialect on another" (267).

In *English as a Second Language (ESL)* varieties, the preceding vowel environments have been shown to influence the presence of nonprevocalic /r/ as well. However, distinctive or even contrasting patterns relative to the *English as a Native Language (ENL)* varieties were found. Rosenfelder (2009), for instance, explored rhoticity in educated Jamaican English based on speech samples from the Jamaican component of *International Corpus of English (ICE)*. In her study, the highest percentages of /r/ realisation were found following NEAR (66.7 %) and FORCE (63.9 %) vowels, while only 25.2 % of the instances of NURSE were realised rhotic. The effects of preceding vowel in Maltese English (Bonnici 2010) also diverge from earlier work in the ENL varieties as NURSE together with LETTER vowels were found to favour the nonrhotic variant. All other preceding vowels slightly favour the overt rhotic realisation including the back vowel FORCE, which usually disfavours rhoticity in native English varieties. Moreover, in Chand's (2010) study on rhoticity in urban Indian English, a preceding LETTER vowel was shown to be the least promoting context of the presence

of nonprevocalic /r/.

The picture of the effects of preceding context on rhoticity in *English as a Foreign Language (EFL)* is even more complex due to the stronger impact of different substrate languages relative to ENL and ESL varieties. In Kang's (2013) study on Korean English, for example, it was found that the preceding vowel showed only a statistically weak effect as compared to ENL varieties investigated by Piercy (2012) and Nagy and Irwin (2010). The effects of the preceding context differed from those in native varieties insofar as START was shown to be the most promoting of nonprevocalic /r/ in Korean English, with SQUARE being the least favourable context for the rhotic. By contrast, the back vowel FORCE that disfavours rhoticity in ENL varieties is found to promote the presence of /r/ in Korean English. Li and Kabak's (2017) study on rhoticity in Chinese English investigated speech of 13 Mandarin-speaking college English teachers. In their study, preceding vowel is reported as the only significant phonological factor affecting the realisation of rhoticity in Chinese English. The ordering of factor weights was parallel with those of both Piercy (2012) and Nagy and Irwin (2010) in that the preceding NURSE vowel favoured the overt /r/ variant and the back vowels in FORCE disfavoured it. One distinction between Chinese English and ENL varieties concerns the effect of LETTER, which attracted the rhotic variant in Chinese English but operated differently elsewhere. Li and Kabak (2017) argued that the perceptual and morphophonological resemblances between English rhotic schwa and the Chinese retroflex suffix could possibly account for this distinction as the rhotic character in Chinese is realized as [əɚ] and is thus perceptually similar to English rhotic schwa, it occurs only at the end of a word, and mostly functions as a diminutive suffix. In the present study, preceding vowel context is expected to affect the production of rhoticity by L2 learners. Further, I will explore whether commonalities between the patterns of rhoticity regarding preceding environments in native and non-native speech will be observed. As earlier work highlighted the conflicting patterns among world Englishes, and ESL and EFL varieties are likely to be affected by their respective substrate languages, the non-native pattern by Chinese learners in the present study will be primarily compared with the ENL varieties. Possible effects of preceding vowel on the perception of rhoticity by L2 listeners remain unknown and will be investigated in the

present study as well.

Stress

The effects of word stress on the realization of rhoticity are relatively consistent across English varieties: stressed syllables favour the presence of nonprevocalic /r/ while unstressed syllables disfavour it. This is the case for Dorset English (Piercy 2012), Boston English (Irwin & Nagy 2007), English in Boston and two New Hampshire towns (Nagy & Irwin 2010), and Scottish English (Dickson & Hall-Lew 2017: 241, Schützler 2010: 17) among ENL varieties as well as for educated Jamaican English (Rosenfelder 2009) as an ESL variety. As Nagy and Irwin (2010) suggested, the universal effect of stressed syllables favouring rhoticity can be explained by the greater articulatory effort and duration in stressed syllables compared to unstressed ones.

However, Chand (2010: 20) found that word stress constitutes a weak but significant effect on rhoticity in urban Indian English and the /r/ variant in the stressed syllable is more likely to be deleted compared to unstressed syllables. Another exception concerns Maltese English (Bonnici 2010), in which both stressed and unstressed central vowels were found to disfavour rhoticity, with the unstressed central vowel more strongly favouring the presence of rhoticity than stressed central vowel. Li and Kabak (2017), on the other hand, reported that both stressed central vowel NURSE and unstressed central vowel LETTER strongly favour rhoticity in Chinese English, thus no direct influence of word stress was observed. In their study, word stress did not come out as a main effect, which is not surprising given that stress was highly correlated with vowel quality and that all unstressed test tokens had the LETTER vowel. The effect of LETTER favouring rhoticity in Chinese English may be explained by the perceptual and morphological similarities between the Chinese retroflex suffix and English schwa. Additionally, the fact that Mandarin Chinese is a syllable-timed language in which stress does not operate on the syllabic or prosodic level might also account for the distinction between Chinese English and other English varieties in terms of the effect of stress on rhoticity. The present study will also test the potential impact of phonological stress on the L2 acquisition of the production of rhoticity and compare its findings with rhoticity in Chinese English (Li & Kabak 2017) to see

whether the same patterns can be found between L2 learners in the study at hand and the proficient Chinese English speakers.

Word context

Previous studies on rhoticity in world Englishes have yielded conflicting evidence for the effects of word context, i.e., position of nonprevocalic /r/ in word and syllable structures that /r/ exists in. For ENL varieties in Boston and New Hampshire towns, Nagy and Irwin (2010) reported that the environments that favour rhoticity most are the linking (*winter is*) and the word-final prepausal (*winter_*) contexts. By contrast, both morpheme-final and morpheme-internal /r/ in consonant clusters disfavour rhoticity. This is in line with findings for Dorset English (Piercy 2012), where the word-final context favours rhoticity more than /r/ in consonant clusters. Becker (2014) also describes that linking *r* is the most favouring context for the overt /r/ variant, which differs from Piercy (2012) and Nagy and Irwin (2010) in that morpheme-internal /r/ followed by a consonant was shown to be a favouring context of rhoticity in New York city English.

Similar patterns of rhoticity regarding word stress have been shown in ESL varieties. The effects of word context were examined by Chand (2010) for Indian English, Bonnici (2010) for Maltese English, and Rosenfelder (2009) for educated Jamaican English. Rosenfelder (2009) found that rhoticity favours singleton /r/ rather than preconsonantal contexts, namely in consonant clusters. Moreover, the /r/ realisation is more likely to occur when followed by a pause instead of by another word beginning with a consonant. Bonnici (2010), likewise, found that rhoticity is favoured when followed by a pause and disfavoured when followed by a consonant.

Regarding EFL varieties, Kang (2013) reported that nonprevocalic /r/ was produced most frequently in the linking environment, followed by /r/ in word-final and word-medial coda position in Korean English. Consonant clusters, by contrast, disfavoured rhoticity. Kang noted that the dissimilarity between the effects of word context in Korean English and in Dorset and Boston English concerning word-medial single /r/ might be attributable to phonological grammars of different substrate languages. The possible effect of word context on the acquisition of rhoticity by L2

learners will also be examined in the present study. It is expected that singleton is a favouring context of the presence of nonprevocalic /r/ relative to consonant clusters for Chinese learners.

Lexical frequency

Lexical frequency is another factor conditioning the realisation of rhoticity that has been repeatedly examined across world Englishes. Nagy and Irwin (2010) found that lexical frequency only played a statistically weak role, with more frequent words disfavouring rhoticity and rarer words favouring rhoticity. It seems that rarely used words were leading the sound change in Boston English to rhoticity. Piercy (2012) also indicated that rare words in Dorset English tended to be rhotic. Contrary to Nagy and Irwin, Piercy concluded that it was the frequent words that took the lead in rendering Dorset English nonrhotic. In Scottish English (Dickson and Hall-Lew 2017), by contrast, the effect of lexical frequency did not reach statistical significance.

For ESL varieties, Rosenfelder (2009) demonstrated that the percentage of /r/ realisation in Jamaican English increases with decreasing lexical frequency, which confirmed the findings for Boston and Dorset English dialects that rare words favour rhoticity more than frequent words. According to Rosenfelder (2009), unfamiliar words tend to evoke more monitored speech, where speakers are more likely to follow the prestigious rhotic Jamaican English pronunciation norms and even produce spelling pronunciations due to orthography. By contrast, speakers are at ease and tend to follow the nonrhotic Jamaican Creole norms when producing a more familiar word.

Kang (2013) found that more frequent words favour the realisation of rhoticity in Korean English while nonprevocalic /r/ is more likely to be deleted in rare words. This stands in contrast to results for ENL varieties where more frequent words tend to be nonrhotic (Nagy & Irwin 2010, Piercy 2012). Kang (2013) explained that the production of rhoticity increases with more familiar and easily pronounceable words because Korean English speakers possibly pay more attention to the rhotic production, given that American English is the target norm for Korean English speakers.

4.2.2 Grammar-external variables

This section focuses on grammar-external constraints influencing the presence or

absence of the /r/ variants shown in various English varieties. The potential effects examined in earlier work are as varied as conversation subject (Love & Walker 2012), speech style (Kang 2013, Sundkvist & Gao 2015, Bonnici 2010), ethnicity (Thomas 2007), education (Tan 2012, Bonnici 2010), socioeconomic status (Labov 1966, Dickson & Hall-Lew 2017), language contact (Chand 2010, Kang 2013), linguistic background (Li & Kabak 2017, Chand 2010, Bonnici 2010), gender (Tan & Gupta 1992, Hartmann & Zerbian 2010, Dickson & Hall-Lew 2017, Bonnici 2010), occupation (Chand 2010), age (Chand 2010, Bonnici 2010, Rosenfelder 2009), proficiency level (Kang 2013), attitude (Bonnici 2010), and target norm (Li & Kabak 2017). This section mainly discusses the grammar-external factors that are potentially associated with the acquisition of rhoticity by L2 learners, namely speech style, gender, norm orientation, and linguistic background, which will be investigated in the present study for Chinese learners.

Speech style

In terms of rhoticity in ENL varieties, Labov (1966: 234) first demonstrated that English in New York City remained mostly nonrhotic for working-class speakers in casual speech despite its overall tendency towards rhoticity. Rhoticity is therefore characterized as a feature restricted to formal speech in New York City English (Labov 1994). Conversely, a more recent work by Becker (2014) challenged the notion of rhoticity as a formal feature and noted that "rhoticity predominates in interview speech across age and class backgrounds and so can no longer be considered superposed or formal" (147).

As concerns ESL varieties, Chand (2010) found that informal contexts promote the production of rhoticity in urban Indian English since less attention to speech is generally demanded. Provided that the nonrhotic RP is the target norm of Indian English, Chand's finding is in line with Labov's (1966), who suggested that speakers' use of standard variants increases with more formal speech style. Accordingly, urban Indian English speakers are expected to be more nonrhotic and behave closer to their RP target norm in medium and high formality contexts.

The degree of rhoticity has been shown to vary distinctly with formality levels in EFL

varieties as well. Sundkvist and Gao (2015), for example, compared the degree of rhoticity in the speech of eight Yunnan English speakers from China across tasks of three formality levels and demonstrated that the percentage of rhoticity increases with increasing formality of the speaking styles. Sundkvist and Gao (2015) interpreted their findings as being related to the target norm of the Yunnan English speakers and the level of attention paid to speech. As rhoticity is assumed to be the overt norm for Yunnan English speakers, "a closer approximation to the norm is observed as the level of attention to speech, or 'formality', increases" (14). The study by Li and Kabak (2017), by contrast, has suggested that spontaneous speech motivates more rhotic realisations than highly formal contexts in Chinese English. This result is further explained by possible stylistic transfer from the speakers' substrate language Mandarin, in which the retroflex suffix is a characteristic of colloquial language for rhotic Mandarin speakers. It is therefore concluded that their stylistic choices concerning the L1 rhotic suffix exert an impact on the presence of rhoticity in Chinese English.

To sum up, previous literature has shown that the presence of rhoticity varies distinctly across different speech styles. Most studies demonstrated that the production rate of the target /r/ variant in a particular variety increases with higher formality levels, while Li and Kabak found crosslinguistic stylistic influence on the production of rhoticity in Chinese English. The present study will explore the impact of speech style on the acquisition of rhoticty by L2 learners and attempt to offer further empirical evidence to our understanding of stylistic effects in the acquisition of an L2 sociophonological variable over time.

Gender

Gender has also been suggested to influence the realization of rhoticity in English. In this respect, Hartmann and Zerbian (2009) found that rhoticity in Black South African English is predominantly a characteristic of female speech, with affluent female speakers displaying considerably higher /r/ production than their male counterparts. Since rhoticity is deemed more prestigious in the younger Black South African community, younger women are more inclined to use the rhotic variant. Male speakers, on the other hand, tend to speak nonrhotic English and use their African

languages more frequently relative to female speakers. Similarly, Tan and Gupta (1992) reported that female speakers tend to produce more rhoticity in Singapore English, because women are more status-conscious and tend to use the prestigious speech forms they perceive. Another possible reason is that female speakers seem to be generally more confident speaking English than the males so that they can produce more target rhotic variants than males in general. (Tan & Gupta 1992). In Chand (2010: 21-22), gender was shown to be the most significant factor predicting rhoticity in urban Indian English, with women being more nonrhotic than males. The female Indian speakers are apt to produce less rhoticity because nonrhotic realisation is the prestigious form in urban Indian English.

For EFL varieties, Kang (2013) showed that female students produced nonprevocalic /r/ at a higher rate than male students in South Korea. This supports previous studies that found that female speakers are more sensitive to the prestigious sociolinguistic variables and tend to lead linguistic changes considering that Korean English "is linguistically closer to American English than British English" (Kang 2013: 1).

The previous studies reviewed in this section have demonstrated that there is a close relationship between gender and target norm for the production of English rhoticity. That is, female speakers are more prestige-conscious and target-oriented than male speakers and are more successful in following their respective target norms when producing rhoticity. As such, in the present study, it is expected that female L2 learners will be more target-oriented than their male counterparts in attaining their target /r/ variant when producing English speech.

Norm orientation

Although the target or prestigious (non)rhotic norm in a particular English variety is not always examined in previous studies on rhoticity, its main effect on the production of nonprevocalic /r/ and interaction with speakers' gender, class, and the formality level of speech have been observed in previous work such as Chand (2010) and Kang (2013). The production of the prestigious /r/ variant is typically favoured by female speakers relative to male speakers and in formal speech contexts relative to informal situations. As such, I would argue that the overall degree of rhoticity in a variety tends

to align with the prestige associated with the /r/ variant.

However, it should be noted that the target or prestigious norm in previous studies is often assumed to vary with certain group of speakers on the basis of their age, gender, social class, and education rather than individual speakers. In other words, the target norm of individual speakers has been largely overlooked in earlier work. For example, Kang's (2013) study on rhoticity assumed rhotic American English to be the target norm for all the Korean participants, because those of higher proficiency level were shown to be more rhotic than less competent participants. Similarly, Chand (2010) supposed that nonrhoticity is the overt prestigious norm for urban Indian English speakers, which is taken to account for the variation in the degree of rhoticity associated with gender and speech style. However, as previous research has shown, the diverging realisation of rhoticity is characterized by significant inter- and intra-speaker variability so that the individual target norm chosen by a particular English speaker or learner should be controlled for.

An empirical study that takes into account individual target norm variation is Kautzsch (2017), who investigated the acquisition of rhoticity by advanced German learners of English. It was found that German learners targeting American English were significantly more rhotic than those targeting British English. Li and Kabak (2017)'s study on rhoticity in Chinese English also takes the individual speaker's target norm into consideration. They collected the data involving speakers' individual target norms via a questionnaire and included it in the multivariate analysis as a fixed effect. It was predicted that Chinese English speakers targeting American English would produce more nonprevocalic /r/ in English than those targeting British English. As expected, the results revealed that the factor individual target norm exerts statistically significant impact on the production of rhoticity in Chinese English with American English oriented speakers being more rhotic than their British-oriented counterparts. The present study will follow Kautzsch (2017) and Li and Kabak (2017) and investigate the impact of individual target norm choice on the acquisition of English rhoticity by L2 learners over time.

Linguistic background

CLI from the first language has also been expected to influence the production of nonprevocalic /r/ in ESL and EFL varieties. As Melchers and Shaw (2003: 131) suggested, CLI plays a significant role in conditioning rhoticity in postcolonial Englishes. To be more specific, if the substrate language of a particular English variety allows structures such as *VrC* or *Vr#* as for example in Indian languages and Malay, rhoticity will be favoured in the second language English. Conversely, if such sequences do not exist in the substrate language, rhotic realisations will be disfavoured. Li and Kabak (2017) explored the role of L1 Mandarin in the production of rhoticity in Chinese English, as Mandarin is a variably rhotic language depending on the Chinese dialects spoken by speakers. However, no direct correlation with respect to the speakers' L1 background was observed and it was concluded that the target norm exerts a greater influence on the production of /r/ in college English teachers' speech thus obscuring any effect of CLI. As English teachers have years of experience learning and teaching the target language, it is not surprising that their speech patterns display more convergence with the phonological patterns of their target norms instead of L1 relative to EFL learners. I would argue that CLI will exert a greater influence on the acquisition of rhoticity by L2 learners. Specifically, learners who speak a rhotic L1 will produce more rhotic words in their L2 English than those speaking a nonrhotic L1 regardless of their target norm, particularly at initial stages of acquisition. In the present study, learners' degree of rhoticity in their L1 Mandarin will be assessed and its effects will be examined from a longitudinal perspective.

4.3 Chapter summary

In summary, rhoticity is a phonetically and phonologically variable phenomenon across individual speakers and varieties of English. To begin with, there is a wide variability of possible articulatory configurations of nonprevocalic approximant /r/, which tend to lead to the same acoustic properties — a low-frequency F3 value relative to other segments. However, studies on the perception of rhoticity have yielded conflicting evidence for the role of F3 and little is known about the acoustic cues employed in the perception of English rhoticity by native and non-native listeners so far. Earlier work on the perception of the French coda rhotic revealed that the duration and formant transition of the vocalic nuclei contribute to the discriminability of the coda rhotic, which might have some implications on the perception of rhoticity in English as well, although French rhotic is acoustically and

articulatorily different from the English one. Empirical work along this line of research is therefore needed.

Moreover, the phonological account of nonprevocalic /r/ in English has caused a long-standing debate as well in terms of the transcription and the treatment of monophthongal [ɜ˞] and centering diphthongs. Previous research has observed the similarity between schwa and nonprevocalic /r/ in nonrhotic English varieties, the close relationship between /r/ and its preceding contexts, and the difference between rhotacized vowels and vowels followed by a rhotic consonant. In the present study, the different phonetic and phonological properties of the /Vr/ sequences are expected to influence the acquisition of English rhoticity by L2 learners. In particular, I will examine the underlying phonetic factors conditioning the perception and production of rhoticity by L2 learners including the articulatory complexity and acoustic salience of the /Vr/ sequences.

In the review of factors conditioning the realisation of nonprevocalic /r/ in English varieties, I showed that the behaviour of /r/ is predictable in that it follows similar pathways along a set of structural and sociolinguistic constraints in native English. A crucial point that should be noted here is that the patterns of learners' L2 speech might to some extent mirror sound variation and change in native speech. Earlier work in different strands of linguistic research has shed light on the connection between L2 speech variation and sound change. Colantoni and Steele (2007), for example, suggested that the phonetic principles that underlie the learners' developmental patterns in acquiring an L2 are the same as those that govern sound evolution and change in the language. They found that the onset-coda asymmetry regarding the acquisition of rhotics in their study resembles the one observed in native speakers' performance and argued that "L2 learners proceed by small manipulations of the degree of constriction in the same way native speakers do when facing variation" (Colantoni and Steele 2007: 400). Himmel (2018)'s study on the L1 phonological attrition of English rhoticity for English-German bilinguals further offers insight into the connection between the typological, diachronic, and synchronic patterns of English rhoticity in native speech and bilingual speech. A more recent work by Kabak (2019) proposed a Dynamic Equational Approach to phonological patterns and argued

that the gradience and variability in sound patterns "arise by an interaction of both language-internal and language-external factors with sound structures in *all* facets and shapes of language evolution and change, which are arguably driven by similar principles" (225). According to Kabak, it is the same set of factors that operate on L2 phonological patterns as well as sound variation and change, albeit with different "weights and magnitude of influence" (249).

As an interesting tie between the variable patterns attested in L2 phonology and sound change has been revealed, I argue that the acquisition of L2 phonological aspects that index synchronic and diachronic changes across native language varieties such as rhoticity should be investigated to offer further empirical insights into the possible correlations between learners' variation and the diachronic evolution of the structure. In order to investigate whether and to what extent the variability in L1 rhoticity can inform its acquisition and developmental process in L2 phonology, both grammar-internal and -external factors that have previously been shown to modulate rhoticity in English varieties will be examined for Chinese learners as well. I expect that the patterns of L2 rhoticity observed in the present study may mimic the variable realisation of rhoticity in native English, and the development of native and non-native English rhoticity may share universal processes, predictors, and patterns.

5. The present study: Acquiring an L2 sociophonological feature

The previous chapters highlighted the major role of CLI and perception on the development of L2 phonology over time and the potential connection between the variable patterns of native and non-native speech. In the review of previous theories and empirical work that shed light on our understanding of L2 phonological acquisition, a number of research gaps and unsolved questions in relation to the L2 acquisition of a sociophonological characteristic have been identified. The study at hand offers an attempt to address these issues by exploring the development of the perception and production of English rhoticity by Mandarin-speaking learners over a time span of six months. Before discussing the research questions and predictions of this study in Section 5.2, this chapter starts with a description of the retroflex marker [ɚ] in Mandarin and a contrastive analysis of rhoticity in Mandarin and English in Section 5.1.

5.1 Comparing rhoticity in English and Mandarin

Chapter 4 demonstrated that rhoticity is a phonetically and phonologically variable feature across English varieties, whose behaviour is constrained by a multitude of grammar-internal and -external factors, leading to its diachronic and synchronic variability. The present study is concerned with the acquisition of this feature by Chinese learners, whose L1 Mandarin is sociophonologically variable in terms of rhoticity as well. The following sections compare relevant phonetic, phonological, and sociolinguistic aspects of rhoticity in Mandarin and in English, which offers a foundation to predict potential patterns of rhoticity in Chinese learners' L2 English speech.

5.1.1 Diminutive suffix [ɚ] in Mandarin

Chinese languages can be divided into seven dialect families: Northern, Wu, Yue, Min, Hakka, Xiang, and Gan, each of which in turn contains many dialects (Yuan 1989). According to Duanmu (2007), speakers from different dialect families usually have difficulty understanding each other. Despite the lack of intelligibility across various Chinese dialect families, all of them share the same morphosyntactic structures and

written language. The study at hand only focusses on Mandarin[2], which is the national lingua franca developed on the basis of the Northern dialect family. It is a standard form of spoken Chinese and the official language of the People's Republic of China. Mandarin is a typical tone language, which means that the pitch contour over a syllable can distinguish word meanings. The language does not make use of lexical stress (Hyman 1977). The majority of Chinese words are monosyllabic, and consonant clusters are not allowed in any position (see Duanmu [2007] for a comprehensive description of Mandarin phonology).

The retroflex suffix [ɚ][3] is a diminutive marker in Mandarin. [ɚ] is considered a main characteristic of the Beijing dialect and is historically related to a Chinese character '儿', which bears the meaning of 'son' or 'child'. Apart from indicating an 'endearing' and 'adorable' sense of the stem word, it sometimes serves to semantically differentiate words in rhotic Chinese dialects (Chen 1999). For instance, [ɕauʐən] refers to a nasty person while its rhotacized counterpart [ɕauʐəɚ] represents a small-sized person or a kid. Nonetheless, Li and Thompson (1989: 9) claimed that the diminutive suffix is "an articulatory feature with no semantic significance" in most cases. Duanmu (2007: 212) also argued that the suffix mostly operates as a stylistic attribute other than a morphosyntactic one. The rhotic marker is predominantly used in colloquial language and expresses "a sense of intimacy, casualness and colloquialism" (Tian 2007: 217). According to Lü (1994, cited in Zhang [2008]: 210), the process of retroflex suffixation exhibits a representative "mellow and smooth" quality of Beijing vernacular speech (149). It is also worth noting that the rhotic suffix "impressionistically has a pharyngeal retroflex quality surprisingly similar to that of the rather unusual postvocalic-*r* of English" (Gick & Wilson 2006: 648).

Regarding its geographical distribution, the rhotic marker is widely used in northern Chinese dialects albeit with considerable phonological and phonetic variation (e.g., Bao 1996, Zhang 2000). The Mandarin speech produced by speakers from the north of China has various degrees of r-colouring. Chen (1999) claimed that many Mandarin characters are habitually rhotacized in North China independent of

[2] Other terms for Mandarin are *Putonghua*, *Standard Chinese*, *Standard Mandarin*, *Beijing Mandarin*, or *Mandarin Chinese*.

[3] It should be noted that the [ɚ] suffix can also be transcribed as [r]. The present study follows Duanmu (2007) and transcribes the diminutive suffix as [ɚ].

orthographic representations, with their nonrhotic counterparts having a rather unnatural sound to rhotic Mandarin speakers. In contrast, Mandarin spoken in the south of China is predominantly nonrhotic due to the fact that the retroflex suffix does not exist in most southern Chinese dialects (e.g., Shanghai, Hokkien, and Cantonese, etc.). Chengdu, a dialect spoken in Sichuan province located in the Southwest of China, is however shown to be rhotic and therefore regarded as a Southern member of the rhotic Mandarin family (Duanmu 2007: 223). Producing the diminutive suffix or not does not pose a problem in the communication between rhotic and non-rhotic Mandarin speakers.

Another fact that is noteworthy about the Mandarin rhotic suffix involves the potential stylistic effect it achieves. According to Chen (1999), the *r*-coloured realisation in Mandarin is typically associated with colloquial language and informal situations instead of formal speech contexts. In other words, a rhotic Mandarin speaker would be more likely to produce an abundance of rhotic markers in a casual conversation but lean towards nonrhotic realisation while delivering a formal speech. Likewise, one can seldom find its occurrence in the broadcast news on the China Central Television (CCTV) channels or in relatively formal speech genres such as legal presentation or government speech.

The phenomenon of diminutive suffixation in Mandarin has been relatively extensively explored by phonologists with three main approaches: description or transcription (e.g., Chao 1965, Wang & He 1985), Optimality Theory (Feng 2001, Ma 1997; 1998, Zhang 2000, Tian 2007), and rule-based analyses (Lin 1989, Duanmu 1990; 2007, Wang 1993). This section will mainly focus on the rule-based phonological description by Duanmu (2007), who provides a relatively comprehensive account of the phonological behaviour of rhotic suffixation in Mandarin.

According to Duanmu (2007), there are 19 possible syllable rhymes in Standard Mandarin: [z, ʐ̩, i, u, y, a, ɤ, e, o, ai, əi, au, əu, in, ən, an, əŋ, aŋ, uŋ], all of which can have the rhotic suffix [ɚ] attached to them. The diminutive suffix is specified as Coronal-[+ retroflex] by Duanmu (2007: 218). Duanmu's rule-based analysis

indicates that the suffix [ɚ] merges with the host rhyme it attaches to, the process of which is conditioned by the following three steps: ADD [ɚ], RHYME-HARMONY, and MID. The first rule ADD [ɚ] means that the suffix is added onto the coda. If a sound in the host rhyme is incompatible with [+ retroflex], it will be replaced by [ɚ] and cannot be reattached. For example, if the suffix attaches to a host rhyme ended with the nasal [n], which is incompatible with the feature [+ retroflex], the [n] in the rhyme will be replaced by the marker [ɚ]. The second constraint RHYME-HARMONY requires that the rhyme sounds cannot differ in [retroflex]. To be more specific, the high front vowel [i] and [y] would be pushed into the onset by the attached suffix and change to prenucleus glides [j] and [ɥ]. The syllabic consonant [z] or [ʐ] will be pushed into the onset as well. The last requirement MID refers to the default height of the nucleus. This only applies to sounds in the nucleus that are unspecified for height. For instance, [ɚ] is unspecified for height, so the insertion of a mid-vowel [ə] will occur to indicate the default height of the nucleus. Under this circumstance, [əɚ] becomes mid and the requirement of MID is therefore fulfilled. Table 5.1 below shows the surface forms of five rhymes [z, ʐ, i, in, y] illustrating the three rules described above. The sounds pushed into the onset are represented in parentheses.

Table 5.1 [ɚ] suffixation on [z, ʐ, i, in, y] following ADD [ɚ], RHYME-HARMONY, and MID (adapted from Duanmu 2007: 219)

Unsuffixed	**ADD [ɚ]**	**RHYME-HARMONY**	**MID**	**Surface form**
z	zɚ	(z)ɚ	(z)əɚ	(z)əɚ
ʐ	ʐɚ	*(ʐ)ɚ*	(ʐ)əɚ	(ʐ)əɚ
i	iɚ	(j)ɚ	(j)əɚ	(j)əɚ
in	iɚ	(j)ɚ	(j)əɚ	(j)əɚ
y	yɚ	(ɥ)ɚ	(ɥ)əɚ	(ɥ)əɚ

Thus, the 19 rhymes in Mandarin are reduced to 11 rhymes after being [ɚ]-suffixed, which is summarized in Table 5.2. [ʳ] refers to the feature [+ retroflex] on the sound it attaches to. The [ʳ] is attached to diphthongs instead of [ɚ] because the suffixed forms must agree with the maximal size of a syllable in Mandarin, i.e. CGVX[4] (Duanmu

[4] According to Duamnu (2007: 71), in the CGVV, C stands for a consonant, G a glide, V a vowel, and X either a vowel or a consonant.

2007: 71). The rhyme VX only allows up to two timing slots, which can be a long vowel, a diphthong, a vowel plus a nasal, or in the r-suffixed form a vowel plus [ɚ] (Duanmu 2007: 82).

Table 5.2 19 unsuffixed rhymes and 11 [ɚ]-suffixed rhymes in Mandarin (adapted from Duanmu 2007: 219 - 220)

Unsuffixed	**Suffixed**
z, z̩, əi, ən, i, in, y	əɚ
u	uɚ
o	oɚ
e	eɚ
ɤ	ɤɚ
a, ai, an	aɚ
au	au^{r}
əu	əu^{r}
əŋ	əŋ^{r}
aŋ	aŋ^{r}
uŋ	uŋ^{r}

Apart from proposing this rule-based phonological description, Duanmu also reported findings by Chao (1968: 51), indicating that this suffix interacts with syllabic tones for particular rhymes in Mandarin. Specifically, the suffixed surface forms of [i] and [ie] are dissimilar under the first (level) and second (rising) tones. For instance, [i] becomes [iəɚ] and [ie] becomes [ieɚ]. With the third (falling-rising) and fourth (falling) tones, however, the suffixed forms are not distinct any more, and both of them would become [iəɚ]. The same prosodic effect has been observed by Wang and He (1985). However, there is no valid explanation accounting for the interaction between tone and diminutive suffixation so far.

A series of studies on rhoticity in Mandarin have been conducted in order to provide empirical evidence for the previous phonological accounts. Lee (2005), for example, carried out an acoustic and articulatory investigation of [ɚ]-suffixed rhymes in Mandarin. Acoustic findings showed that all r-coloured rhymes have a low F3 value

relative to nonrhotic rhymes as an indication of rhotacization. The quality of the preceding sound in the host rhyme was found to be further modified by the [ɚ] suffix. To the best of my knowledge, Lee (2005) is the first study that explores the phonetic aspects of diminutive suffixation in Mandarin Chinese, which has added to the limited empirical work along this line of research. However, it should be noted that the study only reported data collected from one Beijing Mandarin speaker. Any conclusion deduced from such a limited data set should be treated with caution. Besides, the Mandarin syllable structure template used in this study is questionable especially considering the description of triphthongs. According to Duanmu (2007), for instance, Mandarin only allows up to two timing slots in the rhyme, which means that triphthongs do not exist. This inconsistency makes it difficult to use Lee's (2005) phonetic findings to complement previous rule-based phonological analyses in Duanmu (2007).

Furthermore, an articulatory examination carried out by Gick and Wilson (2006) provided empirical evidence to Duanmu's (2007) MID rule, which requires the insertion of a schwa when the sound in the nucleus is unspecified for height. The authors investigated the phonetic explanation for the epenthesis of a perceived schwa between the high tense vowel + liquid sequence (e.g., *hee[ə]l, hai[ə]l, hi[ə]re*) across different languages, among which Beijing Mandarin [i] + [ɚ] rhotacized structure serves as an example. The authors argued that the phenomenon of epenthesis is "a phonetic by-product of one strategy for reconciling an intrinsic conflict between articulatory targets" (Gick & Wilson 2006: 636). In other words, the perceived schwa in English is the result of the tongue passing through a schwa-like configuration during its transition between tongue dorsum/root targets. In their study, ultrasound data collected from three adult Beijing Mandarin speakers were measured to test whether this phenomenon occurs similarly in English and Chinese. Results showed that all speakers displayed tongue root advancement for /i/ and tongue root retraction for /r/. As expected, the tongue passed through the 'schwa space', showing an apparent schwa-like configuration. Gick and Wilson's (2006) study confirmed previous rule-based phonological descriptions of [ə]-insertion between the high front vowel [i] and the rhotic suffix [ɚ] in Mandarin from an articulatory perspective. The articulatory process of [ɚ] epenthesis in Mandarin is therefore shown to be in line with English. In Mandarin, however, the [i] in the rhyme will be further pushed into

the onset and change to a prenuclear glide [j]. This leads to a fully rhotaziced [əɚ] since Mandarin only allows up to two segments in rhymes (Duanmu 2007). Thus, the surface rhymes of English NEAR and Mandarin rhotacized high front vowel [i] are phonetically and phonologically distinct, with [(j)əɚ][5] in Mandarin and [iəɚ] in English.

A more recent acoustic study by Huang (2010) revealed that the acoustic properties of /r/-suffixed monophthongs are consistent with Duanmu's (2007) phonological description. Huang (2010) conducted an acoustic analysis of seven /r/-suffixed monophthongs in Chinese. The spoken data were collected from eight adult rhotic Mandarin speakers from Beijing (4 females, 4 males). Results showed that the retroflex suffixation has a centralizing effect on all monophthongs alongside the lowering of F3. It was found that all /r/-suffixed monophthongs in Mandarin are highly rhotacized and centralized, with F1 being raised in high vowels and lowered in the low vowel /a/, and F2 lowered in front vowels and raised in the back vowel /ɤ/. Moreover, Huang found that there is a statistically significant interaction between the rhotic suffix and preceding vowel quality. To conclude, this study showed that all of the Mandarin rhotic monophthongs were significantly rhotacized with a low F3 value and centralized according to F1 and F2 except for /u/, whose F1 and F2 are unaffected by the suffix.

Following Huang (2010), who examined the acoustic consequences of the rhotic suffix on monophthongs, Liu (2017) reported the acoustic properties of all rhymes comprising monophthongs, diphthongs and nasals. The author measured the first three formants of speech produced by 13 rhotic Mandarin speakers (4 males, 9 females) from Shenyang city and Hebei province in North China. Liu found that the rhotic suffix centralizes its preceding vowels and lowers their F3 significantly in general. To be more specific, F1 increased significantly for the non-low vowels and decreased for the low vowel. Further, F2 became higher for the back vowels but lower for the front vowels. As expected, F3 values decreased significantly for all the monophthongs. In terms of nasal codas, the alveolar nasal coda /n/ was replaced by the suffix, with the preceding /a/ changed just like the monophthong /a/. The velar nasal /ŋ/ was retained

[5] [j] is put in brackets because it becomes part of the onset instead of syllable rhyme after [ɚ]-suffixation.

with the F3 of the preceding vowel being lowered. As for diphthongs, the lowering of F3 was displayed from early on, while their F1 and F2 were not changed significantly. Like Huang (2010), the results in Liu (2017) provided acoustic evidence for the previous rule-based phonological accounts as well.

While the phonetic and phonological aspects of rhoticity in Mandarin have been relatively well studied, little empirical work has been carried out so far on the social meaning and sociocultural associations of this feature. Zhang (2008) explored the sociolinguistics of rhoticity in Beijing dialect and claimed that "rhotacization takes on semiotic saliency through co-occurrence with key Beijing cultural terms and frequent use in written representations of authentic Beijing-ness" (201). In addition, it was demonstrated that rhoticity is associated with a salient male local character type and is characterised by "a constellation of social meanings based in ideological construals of Beijing and its people" (216). However, the sociolinguistic status of rhoticity in other regions of China is severely understudied. It is still not clear how patterns of rhoticity vary and change across different Mandarin dialects in China. Future empirical work on the sociolingusitics of the diminutive suffix in different regions of China is therefore necessary to contribute to our understanding of the social meaning and status of the diminutive suffix in Mandarin.

To conclude, the diminutive suffix [ɚ] is a feature that is widely used by rhotic Mandarin speakers from North China in casual speech. Phonologically, the suffix is specified only for the feature Coronal-[+retroflex]. It attaches to the end of a stem and the sounds already there might be replaced, pushed into the onset, or fully rhotacized depending on their distinctive phonological features. Phonetic studies showed that the [ɚ] suffix has a low F3 and a centralizing effect on preceding vowels. Articulatory examinations revealed that the tongue configurations of [ɚ] in Mandarin resemble those of English. Moreover, rhoticity is recognised as a sociolinguistic feature in Mandarin, the status of which however remains unclear in regions outside Beijing. While the [ɚ] suffix itself is assumed to be impressionistically and phonetically similar to the nonprevocalic /r/ in English, they differ considerably with regard to a number of phonological and sociolinguistic contexts. A contrastive analysis between the [ɚ] suffix in Mandarin and nonprevocalic /r/ in English and possible types of CLI

are discussed in detail in the following section.

5.1.2 Contrastive analysis

The review of the diminutive suffix [ɚ] in (rhotic) Mandarin dialects and nonprevocalic /r/ in (American) English demonstrated that there are both similarities and dissimilarities between rhoticity in the two languages. This section gives an overview of the contrastive analysis of this feature in Mandarin and English, which is summarised in Table 5.3.

Table 5.3 Contrastive analysis of rhoticity in Mandarin and English

		(rhotic) Mandarin	**(American) English**
phonetic	realisation	approximant	
	articulation	mainly retroflex and bunched	
	acoustics	low F3	
phonological	preceding environment	all 19 rhymes, highly centralised/rhotacized vowels like English NURSE and LETTER	mainly START, FORCE, SQUARE, NURSE, LETTER, and NEAR
	syllable structure	singleton	singleton and consonant clusters
	suprasegmental-level interaction	lexical tone	word stress
orthographic	orthography	not necessary, occurs habitually	only when orthographically present
stylistic	speech style	colloquial language, informal speech	more monitored speech
social	gender	feminine characteristics	females more target-oriented
	socioeconomic status	prestigious e.g., in Beijing dialect	middle class, educated, prestigious

Concerning the phonetic properties of nonprevocalic /r/, Mandarin and English bear remarkable resemblance. Both are realised as retroflex approximants[6] with relatively low F3 values. Moreover, the same tongue configurations are recognised for the articulation of nonprevocalic /r/ in both languages: retroflex and bunched. Thus, no phonetic CLI is expected for Mandarin learners acquiring English nonprevocalic /r/. However, when the phonological and phonetic contexts that nonprevocalic /r/ occurs in are taken into account, CLI is expected at a context-dependent level, since the phonological environments that nonprevocalic /r/ occurs in differ distinctly between Mandarin and English. It is therefore informative to further compare the phonological contexts of rhoticity in the two languages.

With respect to preceding vowel environments, English nonprevocalic /r/ occurs mainly following START, FORCE, SQUARE, NURSE, LETTER, and NEAR[7], while the Mandarin suffix can be attached to all 19 rhymes. It should be noted that NURSE, LETTER, NEAR, and START exist in both Mandarin and English, whereas SQUARE and FORCE do not belong to the Mandarin vowel inventory. The process of rhotacization in English is relatively straightforward, with rhotic speakers producing the retroflex sound. Specifically, vowels in NURSE, LETTER and START can be highly rhotacized, whereas NEAR, SQUARE and FORCE are mostly produced as a vowel followed by a consonantal /r/ (e.g., Lockenvitz et al. 2015). When attaching the retroflex suffix to a host rhyme in Mandarin, however, more complicated rules must be followed as part of the Mandarin phonological grammar. As introduced in section 5.1.1, the process of diminutive suffixation is conditioned by three phonological rules: ADD [ɚ], RHYME-HARMONY, and MID (Duanmu 2007), through which the 19 rhymes in Mandarin are reduced to 11 rhymes after being suffixed. In particular, NEAR is incompatible with [+retroflex] and will be pushed into the onset, with the insertion of a mid-vowel to indicate the default height of the nucleus. Thus, rhotic NEAR in Mandarin ([(j)əɚ]) is highly rhotacized and perceptually distinct from English NEAR. As such, considering the preceding contexts of nonprevocalic /r/, the retroflex suffix in Mandarin is phonetically closest to English NURSE and LETTER, which are preceded

6 General American is the rhotic target norm for the Chinese learners of English in the present study, where nonprevocalic /r/ is categorically realised as a retroflex approximant, other phonetic types of /r/ such as tap or trill in Scottish English will not be discussed here.

7 In American English, nonprevocalic /r/ is also realised following diphthongs i.e., HOUR and REQUIRE. They are however relatively infrequent and thus not tested in the present study.

by central vowels. They are followed by START, which was shown to be highly rhotacized in English and also exists in Mandarin. SQUARE and FORCE vowels that do not exist in Mandarin are therefore considered dissimilar contexts. The crosslinguistic similarity hierarchies of the six preceding contexts of nonprevocalic /r/ in English for Mandarin learners can be stated as follows: NURSE/LETTER > START > NEAR > SQUARE/FORCE.

As for syllable structure, English nonprevocalic /r/ occurs both as a singleton and in consonant clusters. Mandarin, however, does not allow consonant clusters. Therefore, the syllable coda is assumed to be the same context for Mandarin learners to produce rhoticity in English, while consonant clusters are rather dissimilar contexts. In terms of prosodic-level interaction, the presence of English rhoticity was shown to vary with word stress, but only interacts with syllabic tone in Mandarin. It is not yet clear whether Chinese learners would handle the rhoticity-stress interaction in English like native English speakers, and this will be investigated in the present study.

Rhoticity also differs between Mandarin and English in terms of orthography. In Mandarin, the orthographic presence of the retroflex suffix is not necessary. Rhotic Mandarin speakers use the suffix habitually, which differs distinctly across individual speakers and speech styles. In English, however, it is closely associated with spelling. The grapheme <r> provides an unambiguous cue for the realisation of nonprevocalic /r/ in rhotic English varieties. Spelling, on the other hand, is rather a misleading representation in nonrhotic English accents such as RP, where /r/ is categorically deleted in nonprevocalic position despite its orthographic presence. The distinction of orthography between Mandarin and English may cause confusions for Chinese learners. Specifically, learners may overlook orthographic indications and produce nonprevocalic /r/ habitually as in their L1 Mandarin. After acquiring the rule of orthography of English rhoticity, it is possible for learners to produce it more often in a reading task where spelling is provided relative to in unscripted speech.

Speech style is also expected to operate on the acquisition of rhoticity by Chinese learners of English. In Mandarin, rhoticity is a stylistic feature of colloquial language and is found to be used more frequently in informal situation. In English, variation in

speech styles is also shown in variably rhotic varieties, where the production of the prestigious variants is more pronounced in highly monitored speech styles (Chand 2010, Sundkvist & Gao 2015).

Possible CLI of sociolinguistic aspects that rhoticity is indexical of in Mandarin and English will also be investigated in the present study. In terms of gender, the diminutive suffix in Mandarin carries a sense of "endearing". Its usage adds a "mellow and smooth" (Lü [1994]: 149, cited in Zhang [2008]: 210) sense to the speech, which can be assumed to be more closely associated with feminine characteristics. Therefore, female Chinese learners may be more likely to produce rhoticity in Mandarin concerning its meaning. In variably rhotic English varieties, the production of rhoticity was found to be more prominent in female speech when it is deemed to be a prestigious feature, like in Black South African English (Hartmann & Zerbian 2009), because female speakers are more norm-sensitive and tend to lead the process of sound change. As for the possible social status English rhoticity indicates, it is typically associated with middle-class speech, more educated speakers and is regarded as a prestigious feature such as in New York City English (Labov 1966) and Singapore English (Tan 2012). In Mandarin, on the other hand, the indexical nature of rhoticity is severely understudied. Earlier work by Zhang (2008) focussed exclusively on rhoticity in the Beijing dialect and reported that rhoticity is a prestigious feature for local Beijing residents. In particular, Zhang associated rhoticity with masculinity in the Beijing dialect despite that the feature itself is considered to be feminine. Male Mandarin speakers in Beijing were shown to use this feature more frequently than females to reinforce their local and traditional identity as a "Beijinger". However, the status of rhoticity in other regions of China remains largely unknown. It is yet not clear whether there is variability in the sociolinguistic patterns of rhoticity across different Mandarin dialects in China.

5.2 Research questions and predictions

In this section, three research questions that the present study aims to address, and relevant predictions based on the theoretical and empirical literature reviewed previously are presented and discussed.

RQ1. How does the effect of an L1 operate on the acquisition of English rhoticity by L2 learners over time?

In the review of the theoretical frameworks on L2 phonological acquisition in Chapter 2, I demonstrated that CLI plays a significant role in affecting the acquisition and development of L2 phonology. However, a number of gaps and unsolved questions in terms of CLI have been identified. Specifically, different hypotheses and models disagree on the main sources of L2 learning problems. That is, it remains unclear whether it is the dissimilarities or similarities between the L1 and TL systems that lead to major difficulties for L2 learners. Furthermore, the scope of these theories is largely restricted to the acquisition of phonemes and their distribution without offering specific accounts for possible context-sensitive CLI beyond the segmental level, the effects of L1 dialectal background, or L1 transfer of sociolinguistic aspects during the course of L2 sociophonological acquisition. It is one of the objectives of the present study to fill these research gaps to fully capture the mechanisms of CLI on L2 phonology over time. The sociophonological feature rhoticity, whose surface patterns are conditioned by a multitude of structural and sociolinguistic factors in both the L1 Mandarin and the TL English, offers an instructive case to explore the impact of context-dependent CLI and L1 sociolinguistic transfer on the development of L2 phonology. The following CLI-related predictions are made based on the theoretical models reviewed in Chapter 2:

CAH/MDH. The CAH and the MDH propose that the dissimilarity between language systems tends to pose major challenges for L2 learners. Accordingly, the acquisition of English rhoticity is expected to be relatively easy for rhotic Mandarin speakers yet particularly difficult for nonrhotic Mandarin learners of English. Furthermore, based on the crosslinguistic similarity hierarchy of the English /V(r)/ sequences for Chinese learners proposed in section 5.1.2, the following acquisition order from ease to difficulty is expected: NURSE/LETTER > START > NEAR > SQUARE/FORCE.

SLM. The SLM postulates that similar target language structures are difficult for learners to perceive and produce, while dissimilar structures are relatively easy.

According to the SLM, the following acquisition order of the contexts of nonprevocalic /r/ from ease to difficulty is expected: SQUARE/FORCE > NEAR > START > NURSE/LETTER.

SDRH. The SDRH predicts the rate of acquisition based on CLI and suggests that dissimilar TL structures are acquired at a faster rate than similar sounds. Based on the SDRH, the acquisition rate of English /V(r)/ sequences from fast to slow is expected as follows: SQUARE/FORCE > NEAR > START > NURSE/LETTER.

OPM. The OPM predicts the impact of CLI on the development of learners' interlanguage from a longitudinal perspective. The development of L2 rhoticity acquisition may be best described in terms of the OPM's Chronological Corollary. It is therefore expected that L1 would play a dominant role in conditioning patterns of L2 rhoticity at earlier stages of acquisition. The impact of TL-related aspects would gradually override L1 and exert greater influence on learners' IL over time.

Moreover, the contrastive analysis of rhoticity showed that Mandarin and English differ distinctly in terms of the phonological environments that nonprevocalic /r/ occurs in, as well as orthography, stylistic and social factors. Transfer opportunities in these aspects that previous theories did not specify are also expected to surface in the acquisition of English rhoticity by Chinese learners. The following predictions are made on the basis of the contrastive analysis in Section 5.1.2:

syllable structure Chinese learners are expected to encounter greater difficulty when producing nonprevocalic /r/ in consonant clusters than in singleton position, since consonant clusters are not allowed in Mandarin phonology.

stress It is assumed to be difficult for Chinese learners to acquire the interaction between rhoticity and stress in English, as the latter does not form part of Mandarin phonology.

speech style Chinese learners are expected to produce nonprevocalic /r/ more frequently in relatively informal tasks than in more controlled speech.

orthography Some Chinese learners may produce intrusive /r/ in the nonprevocalic

position when <r> is not orthographically indicated.

gender Chinese learners are expected to associate nonprevocalic /r/ with feminine characteristics in English speech. Consequently, female learners may produce nonprevocalic /r/ more frequently than male speakers.[8]

In particular, the five predictions above are expected to vary over the course of L2 phonological acquisition according to the OPM (Major 2001). The Chinese learners in the present study will be tested twice within a time span of six months in order to capture the development of their performance over time. As predicted by the OPM, the impact of CLI from Mandarin on English is assumed to be more dominant at initial states of acquisition only, which will gradually decrease over time.

RQ2. What is the relationship between the perception and production of rhoticity by L2 learners? How does this change over time? Which factors condition the perception-production link?

The second research question focusses on the relationship between speech perception and production by L2 learners over the course of phonological acquisition. While L2 speech theories such as the SLM offer a solid theoretical foundation to the idea of perception leading production, a number of empirical studies have provided counterevidence challenging the link, indicating that perception and production might be separate domains conditioned by the interaction of different phonetic variables such as articulatory difficulty and acoustic salience. Variability between learners was also shown to determine their relationship over time. In particular, it is not clear whether perception and production performances by learners are constrained by the same factors. The study at hand sets out to explore the interplay of speech perception and production as well as the underlying forces conditioning both domains. The following theory-based predictions are made for the perception and production of English rhoticity by Chinese learners:

[8] It should be noted that Zhang (2008) claimed that rhoticity is associated with masculinity but only in Beijing. The status of rhoticity is unique in the Beijing dialect because it is deemed a prestigious feature by local Beijingers, and male speakers tend to use the feature frequently to emphasize their local identity. The sociolinguistic status of this feature in other Chinese dialects remains unknown. Since no male learners in the present study are from Beijing, I argue that the connection between rhoticity and masculinity for Beijing speakers will not affect the learners' acquisition of rhoticity in this study.

SLM. As the SLM suggests, accurate perception of a TL structure can predict its accurate production by the learner. Accordingly, Chinese learners' perceptual performances in discriminating English rhoticity in different vowel environments and syllable structures are expected to predict acquisition difficulty. More specifically, /V(r)/ structures that are more accurately perceived by a learner are assumed to be more accurately produced by the learner as well.

MSA. The MSA further takes into account the role of articulatory constraints, indicating that both articulation and perception would contribute to the successful acquisition of an L2 sound. As such, a Chinese learner is expected to first acquire rhoticity in a context where it can be accurately perceived by the learner and does not involve major articulatory difficulties.

Another aspect that previous theories fail to account for concerns the development of the relationship between L2 speech perception and production over time. In the present study, it is predicted that both the perception and production performances of Chinese learners will improve over the course of acquisition, albeit with inter- and intra-speaker variability. In particular, it is predicted that structures where perceptual confusion is more likely to surface would be acquired slowly, while perceptually easier structures are acquired at a relatively fast rate.

RQ3. Which factors previously attested to affect rhoticity in sound variation and change guide its development in L2 speech? Do the patterns of rhoticity in L2 phonology bear resemblance to its patterns in native English speech?

The third research question sets out to explore the connection between patterns of rhoticity in native and non-native English speech. It was argued that the development of L2 rhoticity for Chinese learners may to some extent mirror the diachronic and synchronic variability of rhoticity in world Englishes. As such, universal processes and constraints characterizing its variation and gradience might be shared by native English speakers and Mandarin learners. In Chapter 4, grammar-internal and -external factors predicting patterns of rhoticity in different English varieties were compared

and discussed, which are assumed to operate on the L2 acquisition of English rhoticity in a similar manner as in native English varieties. In the present study, the following predictions regarding the production patterns of rhoticity by Chinese learners concerning these factors are made:

preceding vowel Front vowel environments are expected to favour rhoticity more than back vowels.

syllabic stress Nonprevocalic /r/ in stressed syllable favours the presence of rhoticity more than unstressed ones.

syllable structure Singleton /r/ is more promoting of rhoticity relative to consonant clusters.

word type Content words favour the presence of rhoticity more than function words

speech style Learners tend to produce their target nonprevocalic /r/ variant more frequently in more monitored speech than in less controlled task.

gender Female learners are expected to be more target-oriented than their male counterparts when producing English rhoticity.

norm orientation Chinese learners targeting BrE are assumed to be less rhotic than those targeting AmE.

linguistic background Rhotic Mandarin learners from the North of China are expected to produce more rhoticity in English than nonrhotic Mandarin speakers.

Crucially, the factors guiding the variable and changing patterns of English rhoticity largely overlap with the CLI-related predictions under RQ1. To be more specific, similar factors were expected to influence patterns of rhoticity in L1 Mandarin and L2 English, the directions of patterns predicted by which however do not always align (see Table 5.3). Competing hypotheses regarding preceding vowels, syllabic stress, speech style, and gender are observed, which are summarized in Table 5.4. The acquisition of English rhoticity by Mandarin learners thus offers an excellent test case to gain insights into the competing roles of L1- and TL-related structural and sociolinguistic aspects conditioning the patterns of rhoticity in L2 English speech.

Table 5.4 Competing predictions between the effects of Mandarin and English on non-native speech

	Mandarin-based pattern	**English-based pattern**
preceding vowel	NURSE/LETTER > START > NEAR > SQUARE/FORCE (CAH/MDH) SQUARE/FORCE > NEAR > START > NURSE/LETTER (SLM/SDRH)	NURSE > front vowels > back vowels
syllabic stress	insensitive to stress	stressed > unstressed
speech style	informal > formal	More target-like in formal speech
gender	female > male	female more target-oriented

6. Methodology and materials

This chapter provides details about method and data for the present study. Following an overview of the participants in Section 6.1, Section 6.2 presents the experimental tasks used to collect the speech perception and production data as well as various sociolinguistic characteristics of the learners. Descriptions of the grammar-internal and -external variables investigated in the statistical models are presented in Section 6.3.

6.1 Participants

48 native Mandarin speakers who are first-year university students majoring in English at the North China Institute of Science and Technology (NCIST) participated in the present study twice in October 2017 (T1) and April/May 2018 (T2), respectively. They include 16 males and 32 females, aged between 17 and 20. They are from 24 different provinces in mainland China and thus represent a broad spectrum of the country dialectally and geographically. All reported to be native Mandarin speakers with no knowledge of other foreign languages except English and without any experiences in an English-speaking country for more than two weeks before the first time of data collection. The learners have had at least six years of experience learning English before entering university, since English is one of the key subjects in school and part of the National College Entrance Examination (NCEE) in China. As the L1 dialectal background and target norm variation of learners are generally considered to be key factors conditioning the acquisition of L2 speech, the participants were controlled for their self-stated target norms in the questionnaires and places of origin in China to obtain symmetrically filled cells with an equal number of participants. As one female learner from South China targeting AmE did not participate at T2, data of 47 subjects were eventually transcribed for both data points and coded for further analyses. Table 6.1 below shows the number of participants in each subgroup, providing us with a relatively balanced set as testing the contribution of L1 dialect regions and target norm variation to L2 rhoticity perception and production. The background statistics of the informant set is summarized in Table 6.2.

Table 6.1 Distribution of speakers based on the variables TARGET NORM and PLACE OF ORIGIN

	South China	**North China**
American English	12	11
British English	12	12

Table 6.2 Gender, target norm, age, age of acquisition, origin, and Chinese dialects spoken by the participants

subject	**gender**	**target**	**age**	**AoA**	**Origin**	**Chinese dialects**
4131	M	BrE	18	7	S	Hainan/Cantonese
4132	F	AmE	18	8	N	Tianjin
4133	M	BrE	19	9	N	Shandong
4134	F	BrE	19	8	N	Shandong
4135	F	AmE	18	7	N	Shandong
4141	M	BrE	19	6	N	Shaanxi
4144	F	BrE	20	10	S	Guangxi/Cantonese
4146	F	AmE	20	9	S	standard Mandarin
4148	F	BrE	18	8	N	Shaanxi
4149	F	AmE	18	8	S	Zhejiang
4150	F	BrE	19	7	S	Yunnan
4151	F	BrE	17	9	N	Dongbei
4154	F	BrE	19	8	S	Min
4155	M	BrE	19	6	N	Shanxi
4157	F	AmE	20	6	N	Dalian
4159	F	AmE	18	8	S	Sichuan
4160	M	BrE	19	9	S	Sichuan
4301	M	BrE	18	9	S	Hainan
4302	F	AmE	18	8	S	Hainan/Cantonese
4303	F	AmE	19	8	N	standard Mandarin
4306	F	BrE	17	4	N	Hebei
4310	F	AmE	19	5	S	Hunan
4311	F	BrE	19	9	N	Shaanxi
4313	M	BrE	18	9	S	Anhui
4314	M	AmE	18	9	S	Guangxi
4319	M	AmE	18	7	N	Gansu
4323	F	AmE	18	10	S	Sichuan
4325	M	AmE	20	9	N	Shanxi
4326	F	AmE	19	6	N	standard Mandarin
4329	M	AmE	18	9	S	Hubei

4401	F	BrE	18	4	N	Jiangsu
4403	F	AmE	18	10	N	standard Mandarin
4404	F	BrE	17	9	S	standard Mandarin
4406	F	AmE	18	8	N	Hebei
4408	F	BrE	19	10	N	Hebei
4409	M	AmE	18	6	S	Hunan
4411	F	AmE	17	7	N	Shaanxi
4412	M	AmE	19	10	N	Ningxia
4413	M	BrE	18	12	S	Anhui
4414	F	BrE	19	9	S	Zhuang
4417	M	AmE	19	12	S	Guizhou
4418	F	BrE	17	8	N	Gansu
4419	F	BrE	18	5	S	Ningbo
4420	F	AmE	18	6	S	Yunnan
4423	F	AmE	18	9	S	Sichuan
4424	F	AmE	18	10	S	Sichuan
4428	M	BrE	20	10	N	Shanxi/Cantonese

AoA: age of acquisition; F: female; M: male; AmE: American English; BrE: British English; S: South China; N: North China

The participants attended a compulsory English pronunciation course between the two times of data collection, in which the features of various standard English norms, namely AmE and BrE, and how to correctly produce English sounds on both segmental and suprasegmental levels were taught. Besides, after being introduced the distinctions between various norms, they were required to select a target between AmE and BrE at the beginning of the course and acquire respective features under the instruction of a female Chinese lecturer, who has had more than twenty-year teaching experience at the NCIST. Participants' performance before training and after six months of acquisition will be examined.

6.2 Experimental design

Data collection consisted of three main parts, which included one perception and three production tasks as well as a norm preference survey. As a first step, students filled in a sociolinguistic questionnaire eliciting personal biographical information including gender, age, age of acquisition, place of origin in China, dialects and other foreign languages spoken, media exposure, target norm, and the extent to which they were aware of the difference between American and British accents (see Appendix A). Questionnaires in both Mandarin and English were provided but no students availed

themselves of the English option.

6.2.1 Oddity discrimination task

The purpose of this experiment was to investigate the perception of English rhoticity by Mandarin-speaking learners of English. As little is known about the perceptual sensitivity to rhoticity by L2 learners of English, an oddity discrimination task was designed to examine whether learners were able to perceive the presence or absence of nonprevocalic /r/ in English. In the oddity experiment, listeners were asked to decide whether the three sound items presented in a trial were similar or different. They needed to further indicate the odd member out if they heard one or respond that all items were identical.

The oddity task was chosen over the more traditional AX and ABX (or AXB/XAB) tasks mainly due to its "greatest memory load and stimulus uncertainty" (Strange & Shafer 2008: 161) among all the trail structures. Rhoticity is recognized as a salient feature in English, the perception of which is assumed to be relatively easy (e.g., Flege 2016). Also, the participants in the present study have had years of English learning experience. Therefore, a more complex task type was preferred to tap listeners' phonetic or phonological rather than acoustic sound discrimination abilities. An AX or ABX discrimination task would not be appropriate due to its simplicity and low cognitive load (see a detailed comparison of tasks in e.g., Colantoni et al. 2015: 94-98 or Strange & Shafer 2008: 159-162).

6.2.1.1 Stimuli

18 disyllabic English words were selected, which contain nonprevocalic /r/ varying by six preceding vowel categories (NURSE, START, NEAR, SQUARE, LETTER, and FORCE) and three syllable structures (/r/ in consonant clusters, in word-final coda position, and in word-medial coda position) (see Table 6.3). The words were embedded in a carrier sentence ("I say ____ to you"). The full set of perceptual stimuli are listed in Appendix B1. The stimulus materials were naturally produced by eight adult native English speakers of rhotic and nonrhotic accents living in Germany. Four American English speakers (three females, one male) were recorded reading the 18 sentences with the presence of nonprevocalic /r/, while the four British English speakers (three

females, one male) produced the stimuli in nonrhotic BrE accent. The eight speakers produced the stimulus sentences twice with similar intonation and speed to ensure the comparability of the stimuli across speakers. Each speaker was recorded individually in a quiet room using a ZOOM H2n recorder. The recordings were digitized (23kHz).

Table 6.3 Test stimuli by preceding vowel and syllable structure

Preceding vowel	consonant clusters	word-medial coda	word-final coda
START	*remark*	*garden*	*guitar*
NURSE	*concern*	*perfect*	*prefer*
LETTER	*southern*	*perhaps*	*louder*
SQUARE	*affairs*	*careful*	*prepare*
NEAR	*pioneers*	*weirdo*	*severe*
FORCE	*newborn*	*morning*	*before*

6.2.1.2 Procedure

All participants were tested in individual sessions in a soundproofed phonetic lab at the NCIST using a notebook computer (ThinkPad 470), with stimuli presented at a self-selected comfortable volume level via a HD-201 Sennheiser headset. The experimental procedure was developed in E-prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). Responses were collected using a Chronos button box.

Written instructions in English were provided on the computer screen that presented the task, which was accompanied by oral clarifications by the examiner. The same amount of instruction was given to all participants. They were asked to listen to each trial and to respond by pressing the corresponding buttons on the Chronos button box. Before the experiment began, a practice session comprising four non-test trials was provided to familiarize participants with the task. They were told to focus on the test items embedded in the carrier sentences and were encouraged to respond as quickly as they could since their reaction time would be measured as well. Each trial was only played once and the responses of the participants could not be changed once given. Feedback was provided during the practice session but not during the experiment. The formal experiment only started when the participants believed that they have fully understood the instructions and were ready to proceed.

In the experimental session, 36 test trials and 10 distractor trials were presented in random order. Each trial consisted of sentences produced by three different speakers with both female and male voices. A total of 36 experimental trials tested each of the 18 rhotic/nonrhotic contrasts, half of which were *change trials* containing an odd item out (e.g., AAB, ABA, BAA, ABB, BAB, BBA). The odd one appeared with equal frequency in all three possible serial positions. The remaining *no-change trials* consisted of three physically distinct stimulus sentences produced by different native speakers of a single rhotic or nonrhotic category (e.g., AAA, BBB). The participants indicated whether the odd stimulus on each trial was the first, the second, or the third by pressing a button marked "1", "2", or "3" on the Chronos button box if they heard one. If participants thought the three different instances in a trial are from the same category, they were instructed to click a fourth button marked "4". In addition, ten distractor trials (*alive-arrive; began-begin; began-begun; begin-begun; breathe-breeze; came-game; crumble-grumble; label-table; sudden-southern; 'newborn-new'born*) produced by the same speakers were included alongside the target trials in this experiment so that listeners were less likely to deduce the research focus.

The interstimulus interval (ISI) between the three stimulus sentences in all trials was 500 ms, with the intertrial interval (ITI) being 3 seconds long at most. The experiment proceeded to the next trial immediately after the participants responded, or after 3 seconds automatically if no response was detected. The whole experiment lasted for around 25 minutes with two breaks inserted in-between to ensure that participants could focus on the task without being affected by fatigue or boredom. Result files were automatically generated by E-prime 3.0 after each participant had completed the task.

6.2.2 Production data

For the production data, all participants were recorded in a soundproof acoustic laboratory at the NCIST individually using a high-quality microphone (NEUMANN U87). All recordings were digitalized and transferred onto a computer (Apple Mac Pro MC561CH/A) with 48kHz and 32 bit. Data collection included three different tasks, which will be discussed in the order in which they were administered: (i)

English picture naming (ii) English passage reading and (iii) Mandarin passage reading. The materials were given to the participants prior to the recording session. The participants were first given five minutes to familiarize themselves with the materials and then they were requested to perform the tasks in this particular order one by one without interruption after they had settled comfortably in the sound-attenuated phonetic lab.

6.2.2.1 Picture naming task

In this task, participants were asked to name twelve pictures presented on a piece of paper in English (see Appendix B2). The picture-naming was administered as the first task for the following reasons. To begin with, it is a relatively simple and informal test which only elicited single words and could therefore help subjects to relax in the highly-observed recording condition. Further, this task was chosen over reading a word list since no orthography was involved in the written form so that participants were less likely to detect the variables of interest. The picture names were controlled for six preceding vowel environments with n = 2 in each cell. Table 6.4 below shows the test items by preceding vowel contexts in the picture-naming task.

Table 6.4 Test items in the picture-naming task

preceding vowels	**lexical sets**	**test items**
stressed central vowels	NURSE	*nurse, earth*
unstressed central vowels	LETTER	*flower, letter*
low vowels	START	*star, car*
non-low front vowels	SQUARE	*chair, bear*
	NEAR	*beer, ear*
non-low back vowels	FORCE	*four, north*

6.2.2.2 Reading passage task

In this task, participants were asked to read aloud an English passage that consisted of a 390-word story—"*The boy who cried wolf",* which is an adapted version of Aesop's fable by Deterding (2006) as below:

There was once a poor shepherd boy who watched his flocks in the fields next to a dark forest near the foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran down to the village shouting "Wolf, wolf." As soon as they heard him, the villagers all rushed from their homes, full of concern for his safety, and two of them stayed with him for a while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful. However, not long after, a wolf was looking for a change in its usual diet of chicken and duck, so it actually did come out from the forest and began to threaten the sheep. Racing down to the village, the boy of course cried out even louder than before, but as all the villagers were convinced that he was trying to fool them a third time, nobody bothered to come and help him. And so the wolf had a feast.

This passage was designed to elicit phonetic data in English that contains clear tokens of all English sounds (Sharbawi and Deterding 2010). It has been used in previous studies on rhoticity in Yunnan English (Sundkvist & Gao 2010), Chinese English (Li & Kabak 2017), and Brunei English (Sharbawi & Deterding 2010). Altogether this passage yielded 24 test tokens containing potential nonprevocalic /r/ in six different vowel categories, which are summarized in table 6.5.

Table 6.5 Test items in the passage reading task

preceding vowels	**lexical sets**	**test items**[9]
stressed central vowels	NURSE	*heard, concern, were, third*
unstressed central vowels	LETTER	*shepherd, afternoon, villagers, pleasure, later, however, after, louder, bothered*
low vowels	START	*dark*
non-low front vowels	NEAR	*near*
	SQUARE	*their, air, there*
non-low back vowels	FORCE	*poor, for, more, course, before*

6.2.2.3 Mandarin passage reading task

For this task, participants read aloud a Mandarin reading passage as the last element

[9] It should be mentioned that the distribution of test items is not optimal, as the preceding NEAR and START contexts only have one item each.

of the recording session. In the text, 18 Chinese characters that are typically rendered with retroflex suffix in the rhotic Mandarin dialects were included (see Appendix B4). The suffix *-er* (儿) in Mandarin was not presented in the written form in order to avoid potential impact of orthography. Rhotic Mandarin speakers were expected to produce the retroflex suffix habitually seeing its orthographic form or not. The aim of this task was to examine the degree of rhoticity in speakers' L1 Mandarin which was predicted to play a role on the production of rhoticity in their L2 English.

6.2.3 Norm preference survey

This task aims at measuring Chinese learners' norm preference towards different English varieties, namely the nonrhotic BrE or the rhotic AmE. In spite of the fact that the participants already reported their target norms in the questionnaire, it remains unclear whether they were truly aware of the distinction between different norms at the point of data collection. Some of them may have not decided on their target yet as well, which would potentially affect the effect of norm orientation in the present study.

A verbal-guise experiment was therefore designed to serve as an indirect measure to counterbalance participants' self-reported explicit target norm collected in the questionnaire. The verbal-guise technique has been widely employed in the study of language attitudes (e.g., Chiba et al. 1995, Mckenzie 2008). It is a modified version of the classic matched-guise test (Lambert et al. 1960). In the matched-guise test, a single speaker was used to produce different languages or varieties. This was however criticized for being unnatural and artificial, which failed to reflect the authentic situation of language usage (Hiraga 2005). The verbal-guise test, on the other hand, involves multiple speakers who provide authentic English accents. Although the verbal-guise technique risks introducing potentially confounding paralinguistic variables such as different voice qualities (see Kerswill 2002, Buchstaller 2006), it is favoured when it is difficult to find balanced bilinguals or multilinguals in two or more languages or varieties (e.g., EI-Dash & Tucker 1975).

In the present study, four authentic news broadcast clips produced by four different female native English and Chinese news reporters were carefully selected (Hartmann 2020): a native BrE speaker, a native AmE speaker, and two highly proficient Chinese

news reporters with American- and British-oriented English accents respectively.

The voices were presented in the following order: American English, Chinese British English, Chinese American English, and British English. After playing each recording, participants were requested to evaluate the speaker based on five statements: 1) This person is easy to understand; 2) I am used to this kind of pronunciation; 3) I like her pronunciation very much; 4) I want to speak English like this person; and 5) I want her to be my English teacher. A six-level Likert scale[10] anchored at 1 (not at all) to 6 (very much) was employed for the participants to rate each speaker. They were asked to focus on the speakers' pronunciation instead of the contents of the news or voice qualities to minimise potential influence of speaker and news differences. Each recording lasted 30 seconds. Participants had 30 seconds after each recording to rate each accent.

This experiment was administered in a computer-equipped phonetic classroom at NCIST. Every participant worked with a networked laptop computer managed by a server controlled by the experimenter. The audio recordings were presented to each participant via Salar E28 headsets connected to the computer's headphone ports.

It is worth noting that the present survey differs from other large-scale language attitude research as it only concerns participant's individual norm preference towards standard English norms. Previous studies on language attitudes towards varieties of English usually involve judging various traits of the speakers such as power, status, competence and solidarity (e.g., Bayard & Green 2003, Hartmann 2020). This was, however, beyond the scope of the present study and therefore not examined in detail.

The average rating of different voices (ranging between 1 to 6) by each participant was calculated and then examined in the mixed-effects regression analyses as four independent continuous predictors within the NORM PREFERENCE factor group.

[10] A five- or seven-level Likert scale was avoided on purpose based on results of a pilot study conducted by the author before the data collection. Participants of the pilot study were found to favour the neutral option without showing any preferences. Therefore, a six-level Likert scale was used in the official data collection so that the participants are forced to indicate their preferences towards different accents.

6.3 Independent variables

The perception and production data collected were coded for independent variables that had previously been shown or expected to exert potential impact on the acquisition of rhoticity by Chinese learners of English. Both grammar-internal and -external constraints were grouped and analysed statistically. Table 6.6 gives a brief overview of the external factors and the corresponding tasks in which they were examined. In the following, details involving each factor group will be illustrated.

Table 6.6 Grammar-external factors collected in the different tasks

Task	**Factor Group**
Questionnaire	GENDER AGE OF ACQUISITION PLACE OF ORIGIN TARGET NORM MEDIA EXPOSURE ACTUAL ENGLISH USAGE
Verbal-guise test	NORM PREFERENCE
Mandarin production	DEGREE OF RHOTICITY IN MANDARIN
English production	SPEECH STYLE
Oddity discrimination	REACTION TIME

6.3.1 Grammar-internal variables

Three phonological variables were considered: PRECEDING VOWEL, SYLLABLE STRUCTURE, and WORD TYPE. The perceptual stimuli were coded for PRECEDING VOWEL and SYLLABLE STRUCTURE, while the production data collected were coded for all three variables. Concerning WORD TYPE, nouns, verbs, adjectives and adverbs were regarded as content words, with the others being coded as function words. Table 6.7 below lists the internal factor groups with an example for each variable.

Table 6.7 Summary of grammar-internal factors

Factor group	Factors	Example
PRECEDING VOWEL	START	*garden*
	NURSE	*prefer*
	LETTER	*louder*
	SQUARE	*careful*
	NEAR	*severe*
	NORTH	*morning*
SYLLABLE STRUCTURE	consonant clusters	*concern*
	word-medial coda	*perhaps*
	word-final coda	*severe*
WORD TYPE	content words	*dark*
	function words	*for*

6.3.2 Grammar-external variables

Among the external factor groups, eight demographic and speaker-specific variables were identified: GENDER, AGE OF ACQUISITION, PLACE OF ORIGIN, TARGET NORM, MEDIA EXPOSURE, ENGLISH USAGE, NORM PREFERENCE, and DEGREE OF RHOTICITY IN MANDARIN. They correspond to the level of speakers, which is the highest level in a multilevel analysis. SPEECH STYLE as a stylistic factor was classified to the level of text units and varies with production tasks. In addition, individual responses in the oddity discrimination task were coded for the REACTION TIME measured by the Chronos button box. Table 6.8 summarizes the distribution and types of the external variable groups in the present study.

Table 6.8 Summary of grammar-external factors

Factor Group	Type	Factors/Criteria	No. of learners
GENDER	Categorical	female	31/47
		male	16/47
AGE OF ACQUISITION	Continuous	ranging from 4 to 12	NA
PLACE OF ORIGIN	Categorical	North	23/47
		South	24/47
TARGET NORM	Categorical	American English	23/47
		British English	24/47
MEDIA EXPOSURE	Categorical	rhotic English varieties	26/47
		nonrhotic English varieties	3/47
		both	18/47
ENGLISH USAGE	Continuous	ranging from 1 to 20 hours/week	NA
NORM PREFERENCE	Continuous (ranging from 1-6)	native American English	NA
		native British English	
		Chinese American English	
		Chinese British English	
DEGREE OF RHOTICITY IN MANDARIN	Continuous	ranging from 2 to 15 instances	NA
SPEECH STYLE	Categorical	picture naming	All speakers
		passage reading	All speakers
REACTION TIME	Continuous	ranging from 13 to 2955 ms	NA

7. Perception of rhoticity by Chinese learners of English

This chapter is concerned with the development of perception of English nonprevocalic /r/ by Chinese learners. Section 7.1 briefly describes the preparation of the data and the statistical procedures applied to the results. After general descriptive statistics of the perceptual results based on *A-prime (A')* scores and accuracy rates attained by the learners in section 7.2, section 7.3 presents multilevel analyses of potential factors conditioning the learners' perceptual patterns of rhoticity at both data collection times. In section 7.4, acoustic properties of the perceptual stimuli are presented and the relative salience of different phonetic contexts of nonprevocalic /r/ is investigated. The results are comprehensively discussed in section 7.5, followed by a concluding summary in section 7.6.

7.1 Data preparation and statistical analysis

A total of 1632 and 1601 perceptual responses for the 47 participants were elicited from the oddity discrimination experiment at T1 and T2, respectively, which did not include cases where no responses were detected by the Chronos button box. Alongside the accuracy and reaction time, each response was also coded for the stimuli and the two phonological variables PRECEDING VOWEL and SYLLABLE STRUCTURE. Furthermore, grammar-external constraints that are of potential interest were also coded and analysed, including listeners' place of origin in China, target norm, norm preference, media exposure, gender, and reaction time. This set of data was further analysed using mixed-effects logistic regression models in R (R Core Team 2013) using the Lme4 package (Bates et al. 2014) in order to test the statistical significances of different factors predicting the perceptual performances of the listeners. Post-hoc pairwise comparisons with Bonferroni correction to adjust p-values for multiple comparisons were performed using the lsmeans package (Lenth 2016) and the emmeans package (Lenth 2018).

A' scores[11] were calculated for each learner and each preceding vowel environment to take possible response bias into account. The measure of discrimination sensitivity

[11] Here, I follow Flege and Mackay (2004) and Guion et al. (2000) and calculated A-prime scores for the perceptual data. A-prime is used instead of D-prime because the oddity discrimination task in the present study is a four alternative forced choice (4AFC) test. D-prime, on the other hand, is typically calculated for 2AFC test like the AX and ABX tasks, where listeners are only given two answer options.

was based on the proportion of *hits (H)* and *false alarms (FA)*. *H* refers to the correct selection of the odd item out in change trials, while *FA* indicates the incorrect selection of an odd member in no-change trials. An *A'* score of 1.0 represents perfect discrimination, whereas 0.5 or below indicates a lack of sensitivity to a contrast. The *A'* scores were calculated using the formula below provided by Snodgrass, Levy-Berger, and Haydon (1985: 451):

a) If $H > FA$ then $A' = 0.5 + [(H-FA)\,(1 + H - FA)]/[4H\,(1 - FA)]$;
b) If $H = FA$ then $A' = 0.5$;
c) If $H < FA$ then $A' = 0.5 - [(FA - H)\,(1 + FA - H)]/[4FA\,(1 - H)]$

7.2 Descriptive statistics

7.2.1 Discrimination ability

Beginning with the overall perceptual sensitivity to English rhoticity, Figure 7.1 illustrates the mean *A'* scores obtained by the 47 Chinese learners at T1 and T2 respectively. It shows that the mean *A'* scores amounted to 0.38 at T1, which increased to 0.47 at T2, suggesting an overall lack of sensitivity to the English rhotic-nonrhotic contrast by the Chinese learners.

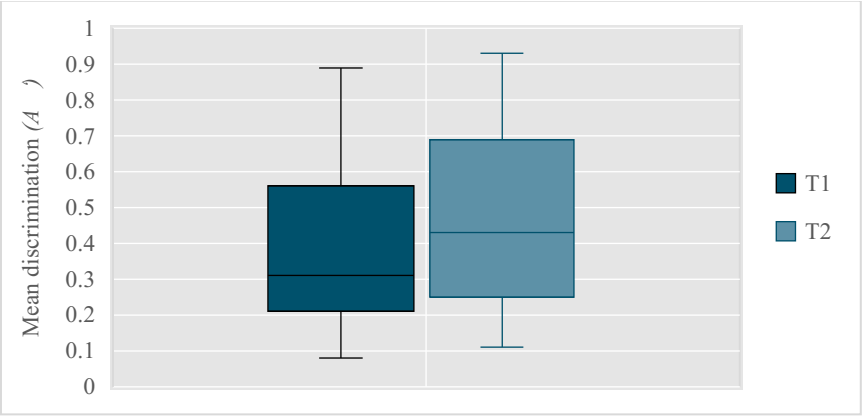


Figure 7.1 Boxplots of mean discrimination scores obtained for all learners at T1 and T2

In order to ascertain whether and to what extent individual learners' perceptual ability in discriminating nonprevocalic /r/ in English changed over time, a paired sample *t*-test was performed to compare their perceptual performance before and after the six months of phonological training. Results revealed that, on average, the Chinese

learners had significantly higher discrimination scores at T2 compared with T1 (t (46) = -3.1, p = 0.003).

The discrimination scores received by individual participants at T1 and T2 are shown in Figure 7.2, indicating a great degree of inter- and intra-learner variation in the perception of rhoticity by the Chinese learners over time. The individual *A'* scores varied widely across participants from 0.08 to 0.89 at T1 and from 0.11 to 0.92 at T2.

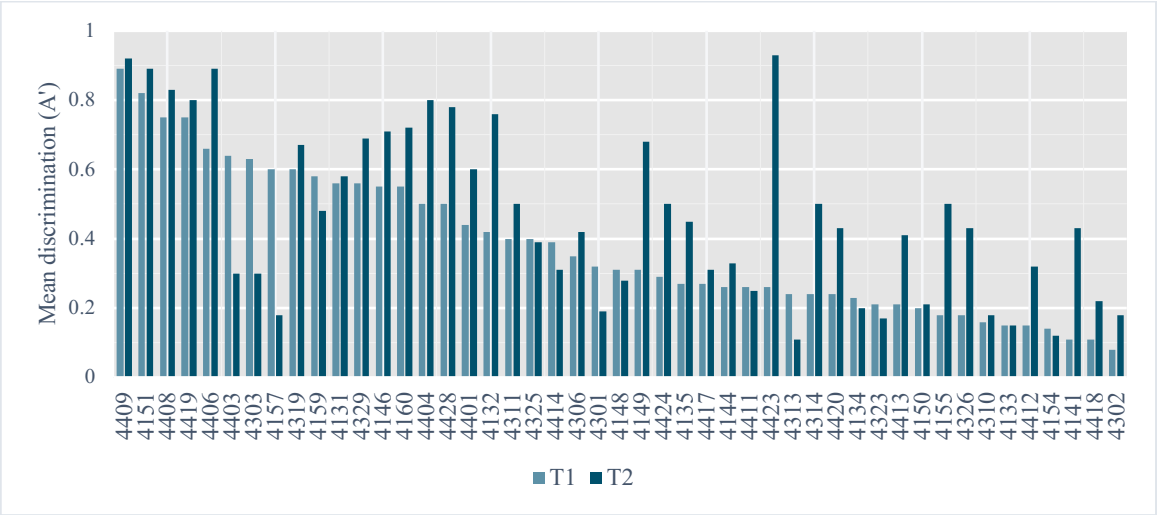


Figure 7.2 Mean discrimination scores obtained by the 47 Chinese learners at T1 and T2. Learners are arranged along the x-axis according to their performance at T1.

Learners 4409 and 4151, for instance, were highly sensitive in discriminating English rhoticity, with the *A'* scores being higher than 0.8 at both data collection times. Learners 4133 and 4154, in contrast, displayed extreme insensitivity towards rhoticity, with the discrimination scores amounting to below 0.2 at both T1 and T2. At T1, 16 participants were found to be perceptually sensitive to English nonprevocalic /r/, with the mean *A'* scores amounting to or being higher than 0.5. This number of learners further increased to 20 at T2. Furthermore, the learners differed in their own performance between the two data collection times. About 72% (34 out of 47) of the learners showed various degrees of improvement from T1 to T2, ranging from moderate improvements such as for learner 4319 (increased from 0.6 to 0.67), to learner 4423, who showed a dramatic increase in her discrimination scores from 0.26 at T1 to near-perfect discrimination (0.93) at T2. Nine participants (19%) exhibited a

decrease of perceptual sensitivity to rhoticity such as learners 4303 and 4157, while the *A'* scores received by four of the learners (9%) remained stable. Learners 4325 and 4411, for example, did not differ in their perceptual performance between T1 and T2.

Figure 7.3 shows the mean *A'* scores for the six preceding vowel environments at T1 and T2. In general, all of the preceding contexts received higher *A'* scores at T2 relative to T1. The learners were only perceptually sensitive to /r/ preceded by a SQUARE vowel at T1, while being neutral to FORCE and insensitive to the preceding NEAR, START, NURSE, and LETTER contexts. After six months of training, the learners further gained perceptual sensitivity towards the preceding FORCE, NEAR and START environments, while still showing a lack of sensitivity in discriminating rhoticity in the stressed central vowel NURSE and the unstressed central vowel LETTER, albeit with slight improvements. These results suggest that it is particularly difficult for the Chinese learners to perceive rhoticity preceded by central vowels.

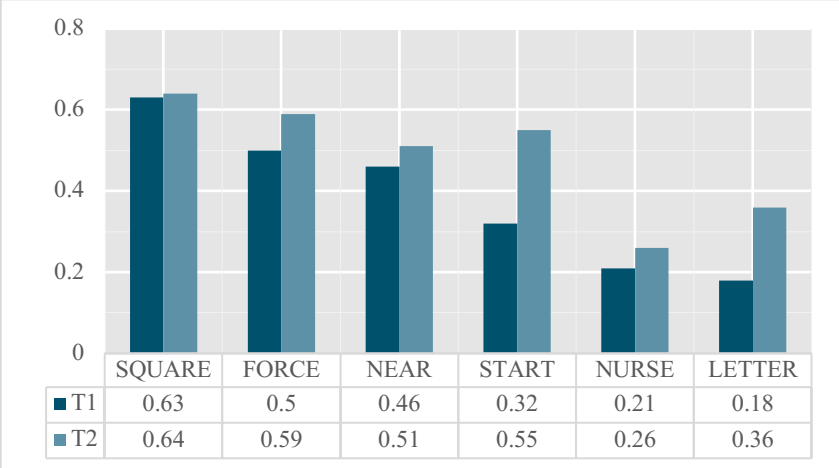


Figure 7.3 Mean A' scores obtained for six preceding contexts at T1 and T2

7.2.2 Proportions of accurate responses

Table 7.1 illustrates the rate of accurate responses by the learners in perceiving rhoticity in six different vowel environments at T1 and T2. Nonprevocalic /r/ preceded by a SQUARE vowel was found to be the easiest context for the learners to discriminate, with the percentages of accuracy amounting to 56.9% at T1 and 57.8%

at T2. Similarly, more than half of the FORCE contrasts was perceived accurately (50.2% at T1 and 54.8% at T2). 47.9% of nonprevocalic /r/s with a preceding NEAR vowel and 39.6% of those with a preceding START context were accurately distinguished by the learners at T1, which increased to 50.4% and 52.3% respectively at T2. Nonprevocalic /r/ with a preceding NURSE vowel was relatively difficult for the learners to distinguish, with only 29.3% and 34.7% being correctly perceived at T1 and T2 respectively. Similarly, the learners discriminated merely 29% of the LETTER contrasts successfully at T1, which increased to 42.8% at T2.

Table 7.1 Proportions of accurate responses on the basis of PRECEDING VOWEL

Vowel	SQUARE		FORCE		NEAR		START		NURSE		LETTER	
	T1	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
n total	274	263	273	272	280	272	270	262	266	268	269	264
n accurate	156	152	137	149	134	137	107	137	78	93	78	113
Percentage	**56.9**	**57.8**	**50.2**	**54.8**	**47.9**	**50.4**	**39.6**	**52.3**	**29.3**	**34.7**	**29**	**42.8**

Table 7.2 shows the percentages of accurate responses by the Chinese learners in discriminating nonprevocalic /r/ in different syllable structures. The accuracy rates of all three syllable contexts increased from T1 to T2. Singleton /r/ in word-medial position was the most difficult context for the learners to perceive at T1 (37.6%). Word-final single /r/, by contrast, was found to be the most favourable context for the learners in terms of perception (46.6% at T1 and 51.8% at T2). Nonprevocalic /r/ in consonant clusters was less difficult relative to word-medial single /r/ (42.5% at T1 and 48.1% at T2).

Table 7.2 Proportions of accurate responses on the basis of SYLLABLE STRUCTURE

Syllable structure	Consonant clusters		Word-final singleton		Word-medial singleton	
	T1	**T2**	**T1**	**T2**	**T1**	**T2**
n total	543	536	552	448	537	617
n accurate	231	258	257	232	202	291
Percentage	**42.5**	**48.1**	**46.6**	**51.8**	**37.6**	**47.2**

Table 7.3 presents the perceptual performance by the Chinese learners from different L1 backgrounds in China. As shown in the table, whether the learners originated in the rhotic North or the nonrhotic South of China does not seem to affect their

perception of nonprevocalic /r/ in English. Similar percentages of accurate responses were achieved by the learners from the North and the South, with 42.9% and 41.7% of the test trials being accurately identified at T1, which further increased to 49.1% and 48.5% respectively at T2.

Table 7.3 Proportions of accurate responses on the basis of PLACE OF ORIGIN

L1 region	Rhotic North		Nonrhotic South	
	T1	**T2**	**T1**	**T2**
n total	799	797	833	804
n accurate	343	391	347	390
Percentage	**42.9**	**49.1**	**41.7**	**48.5**

Table 7.4 shows the perceptual accuracy percentages by the Chinese learners according to their self-stated individual target norms. The learners aiming at AmE seemed to be slightly more accurate in perceiving rhoticity than those targeting BrE. The AmE-oriented learners accurately perceived 43.2% of the rhotic-nonrhotic contrasts at T1 and 50.7% at T2, whereas those targeting nonrhotic BrE only correctly responded to 41.3% and 46.8% of the trials at T1 and T2 respectively.

Table 7.4 Proportions of accurate responses on the basis of TARGET NORM

Target norm	American English		British English	
	T1	**T2**	**T1**	**T2**
n total	843	827	789	774
n accurate	364	419	326	362
Percentage	**43.2**	**50.7**	**41.3**	**46.8**

The rates of accurate responses by the learners divided by gender are illustrated in Table 7.5. Female and male learners did not differ in terms of their perceptual performance in discriminating English rhoticity. Both of the groups accurately perceived around 42% of the test contrasts at T1, which further increased to approximately 49% after six months.

Table 7.5 Proportions of accurate responses on the basis of GENDER

Gender	female		male	
	T1	**T2**	**T1**	**T2**
n total	1073	1045	559	556
n accurate	445	511	235	270
Percentage	**42.4**	**48.9**	**42.0**	**48.6**

7.2.3 Reaction time

Table 7.6 illustrates the mean reaction time of the Chinese learners when perceiving rhoticity in different vowel contexts. As concerns T1, it took only 441 ms on average for the learners to respond to a SQUARE contrast. Conversely, the reaction time of learners in responding to START and NURSE contexts was much longer, amounting to 562 ms and 568 ms, respectively. Similarly, the listeners had an average reaction time of more than 600 ms when perceiving rhoticity in the contexts of LETTER, NURSE and START at T2. By contrast, the mean reaction time of the learners in responding to a SQUARE vowel only amounts to 361 ms at T2. In general, nonprevocalic /r/ in START, NURSE, and LETTER contexts had relatively long reaction times in the experiment, which was followed by FORCE and NEAR, while listeners responded most quickly to SQUARE.

Table 7.6 Mean reaction time (ms) by learners in each preceding vowel context

	FORCE		LETTER		NEAR		NURSE		SQUARE		START	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
n total	273	272	269	264	280	272	266	268	274	263	270	262
Mean (ms)	**528**	**546**	**525**	**600**	**495**	**581**	**568**	**601**	**441**	**361**	**562**	**612**
SD (ms)	430	518	431	546	385	505	501	598	375	567	525	598

The mean reaction time for learners to distinguish the presence or absence of nonprevocalic /r/ in various syllable structures is shown in Table 7.7. The subjects were fairly quick in responding to rhoticity in word-final single position (501 ms at T1 and 558 ms at T2), while being slightly slower in discriminating nonprevocalic /r/ in consonant clusters (528 ms at T1 and 576 ms at T2). Results revealed that participants took the longest reaction time to respond to word-medial single /r/ contrasts (529 ms at T1 and 608 ms at T2).

Table 7.7 Mean reaction time (ms) by learners in each syllable structure

	consonant clusters		word-final singleton		word-medial singleton	
	T1	T2	T1	T2	T1	T2
n total	543	536	552	448	537	617
Mean (ms)	**528**	**576**	**501**	**558**	**529**	**608**
SD (ms)	450	527	440	544	446	587

Table 7.8 shows the mean reaction time of the Chinese learners for accurate and

inaccurate responses respectively. With respect to T1, the mean reaction time did not differ considerably between accurate and inaccurate responses. Results showed that, on average, it took the listeners 500 ms to respond to a contrast accurately and 533 ms in giving a wrong answer. In contrast to T1, the reaction time between accurate and inaccurate responses differs significantly at T2 (t (1599) = 3.462, $p < 0.001$). As shown in the table, the learners had a much longer reaction time for inaccurate responses (630 ms) compared to that of the accurate ones (534 ms) at T2.

Table 7.8 Mean reaction time (ms) in accurate and inaccurate responses

	Accurate response		Inaccurate response	
	T1	T2	T1	T2
n total	690	781	942	820
Mean (ms)	**500**	**534**	**533**	**630**
SD (ms)	438	499	451	601

It has to be pointed out that the reaction time showed in Tables 7.6 - 7.8 are with relatively high standard deviation, suggesting that the mean values are less informative given the huge variability across listeners and contexts. Therefore, section 7.2.3 only provided us with a rough idea of the distribution of reaction time. The statistical effects of reaction time will be examined as a continuous variable in the following multilevel analyses which take into account the inbalance of data.

7.3 Multilevel analyses

The perception data were further analysed via a mixed effects logistic regression model in R (R Core Team 2013) using the Lme4 package (Bates et al. 2014) in order to identify significant factors that account for the inter- and intra-learner variability in perceiving English rhoticity by the Chinese learners. Table 7.9 illustrates the results of the multivariate analysis of the T1 data with the perceptual accuracy of each response as dependent variable, and participants and stimulus as random intercepts.

*Table 7.9 Best-fit mixed effects regression model of PERCEPTUAL ACCURACY at **T1** as a binary dependent factor, with PARTICIPANT and STIMULUS as Random Intercepts*

Fixed effects	**Levels**	**Estimate**	**Std. error**	***z*-value**	***p* <**	
		-0.002	0.268	-0.010	0.992	
PRECEDING VOWEL	LETTER	-0.975	0.368	-2.648	0.008	**
	NEAR	-0.094	0.363	-0.259	0.780	
	NURSE	-0.959	0.368	-2.605	0.009	**
	SQUARE	0.303	0.364	0.833	0.405	
	START	-0.470	0.365	-1.289	0.198	
Random effects	**Type**	**Variance**	**Std. Dev**			
PARTICIPANT	(Intercept)	0.286	0.535			
STIMULUS	(Intercept)	0.150	0.388			
		Min.	**Median**	**Max.**		
Scaled residuals		-2.003	-0.517	2.356		

As concerns T1, PRECEDING VOWEL was found to be the only significant phonological factor conditioning the perception of rhoticity by the Chinese learners. As shown in Figure 7.4, nonprevocalic /r/ preceded by a SQUARE vowel is the most promoting phonological environment of accurate responses. It is followed by the contexts of FORCE, NEAR and START, while the learners were found to be insensitive to preceding central vowels NURSE (p = 0.009) and LETTER (p = 0.008). Post-hoc pairwise comparisons with Bonferroni correction showed that the learners' performance differed significantly between LETTER and SQUARE (p = .008) as well as NURSE and SQUARE (p < .01). Their discrimination ability between the other contexts did not reach statistical significance. The best-fit model eliminated SYLLABLE STRUCTURE, REACTION TIME and the other external constraints including GENDER, TARGET NORM, and PLACE OF ORIGIN for the perception of nonprevocalic /r/ by the learners at T1.

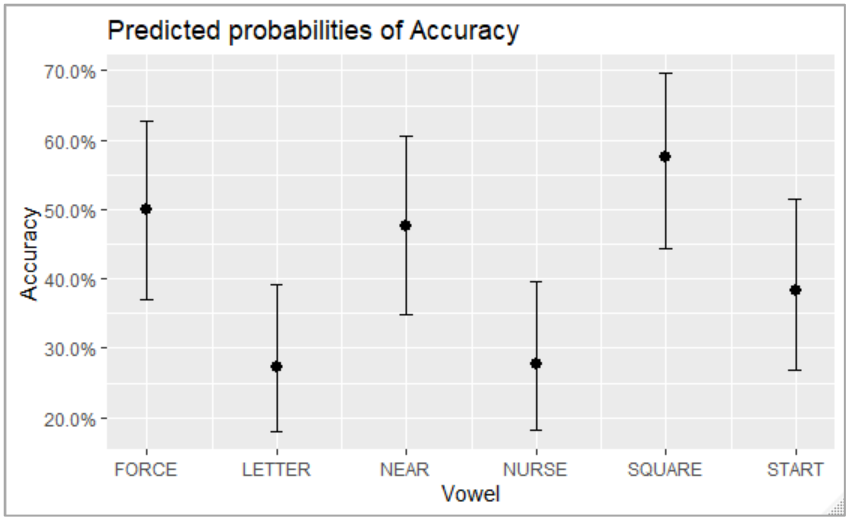


Figure 7.4 Main effects of PRECEDING VOWEL on perceptual accuracy at T1

A mixed effects logistic regression model was fit for the learners' perceptual performance at T2 as well, the results of which are shown in Table 7.10. At T2, PRECEDING VOWEL exerts a significant effect in predicting the perception of rhoticity by the learners in the similar hierarchy as at T1: SQUARE > FORCE > NEAR > START > LETTER > NURSE (see Figure 7.5). The participants tended to respond to /r/ following SQUARE, FORCE, NEAR and START vowels more accurately than following LETTER and NURSE. It should be noted that the order of LETTER and NURSE differs from T1 in that a preceding stressed central vowel NURSE is shown to be more difficult for learners to perceive (p = 0.002) than its unstressed counterpart at T2. However, pairwise comparisons with Bonferroni correction for PRECEDING VOWEL revealed that the difference between NURSE and LETTER was not statistically significant. At T2, significant differences in the learners' perceptual performance were only found between the contexts of FORCE and NURSE (p = .033) as well as SQUARE and NURSE (p = .006).

*Table 7.10 Best-fit mixed effect regression model of PERCEPTUAL ACCURACY at **T2** as a binary dependent factor, with PARTICIPANT and STIMULUS as Random Intercepts*

Fixed effects	**Levels**	**Estimate**	**Std. error**	***z*-value**	***p* <**	
(intercept)		0.392	0.236	1.657	0.097	
PRECEDING VOWEL	LETTER	-0.524	0.294	-1.781	0.075	
	NEAR	-0.193	0.293	-0.659	0.510	
	NURSE	-0.907	0.296	-3.060	0.002	**
	SQUARE	0.144	0.294	0.489	0.625	
	START	-0.209	0.319	-0.654	0.513	
REACTION TIME	continuous	-0.000	0.000	-2.982	0.003	**
Random effects	**Type**	**Variance**	**Std. Dev**			
PARTICIPANT	(Intercept)	0.456	0.676			
STIMULUS	(Intercept)	0.079	0.281			
		Min.	**Median**	**Max.**		
Scaled residuals		-2.329	-0.457	2.481		

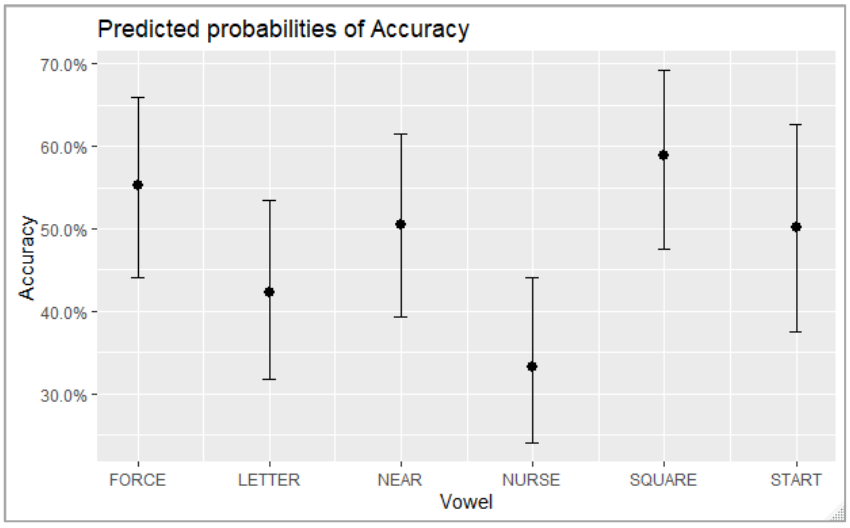


Figure 7.5 Main effects of PRECEDING VOWEL on perceptual accuracy at T2

Furthermore, the best-fit model at T2 revealed significant effects exerted by REACTION TIME on the discrimination of the rhotic-nonrhotic contrasts (p = 0.003). As shown in Figure 7.6, the reaction time taken by the learners to accurately respond to a test trial is much shorter than when giving a wrong answer. In other words, the quicker a learner responded to a trial, the more likely the response would be accurate. Other potentially influential internal and external factors on the perception of rhoticity by

the learners were found to be statistically insignificant and were therefore eliminated by the best-fit model for T2.

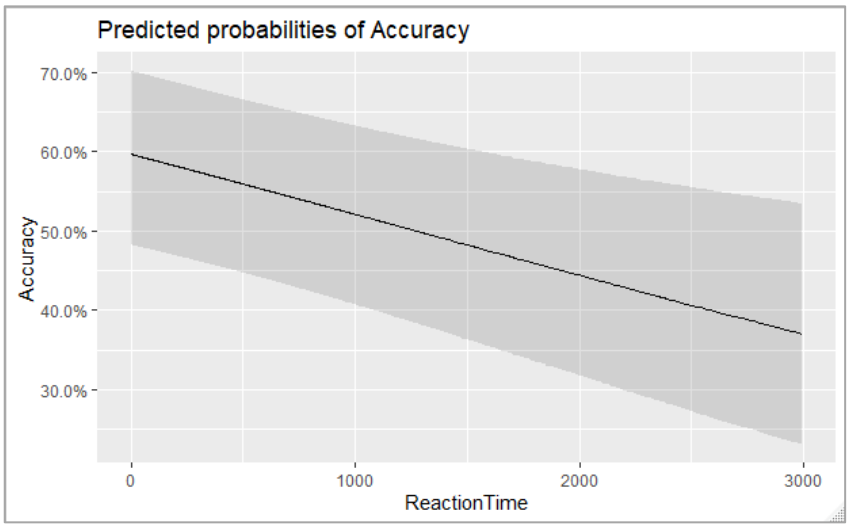


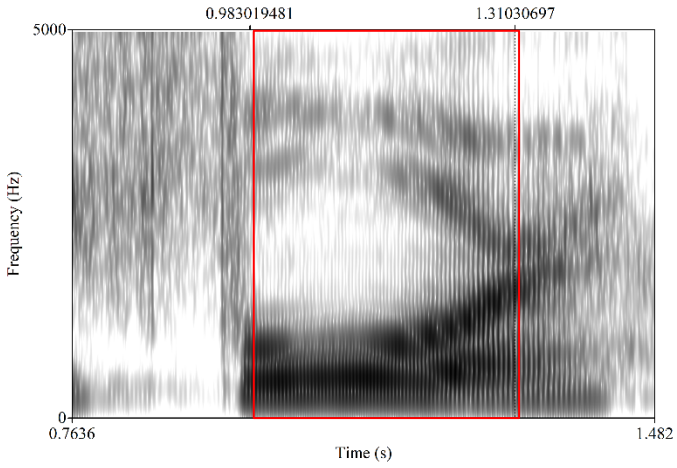
Figure 7.6 Main effects of REACTION TIME on perceptual accuracy at T2

7.4 Acoustic properties of perceptual stimuli

This section reports on the acoustic properties of the perceptual stimuli in the oddity discrimination task in order to identify the degree of acoustic salience of various phonetic contexts and the potentially different preceding vowel realisations between rhotic and nonrhotic stimuli. It was predicted that the perceptual accuracy of rhoticity is associated with the acoustic salience of the context that nonprevocalic /r/ occurs in and their preceding vowel qualities. Thus, the phonetic properties of the /V(r)/ sequences (duration, F2, and F3) and the preceding vowels (F1 and F2) in the oddity discrimination task in the present study were measured and presented. More specifically, the rhotic stimuli produced by three female native American English speakers and the nonrhotic ones by three female native British English speakers were analysed acoustically in Praat (Boersma 2001). The reason for measuring speech produced by female speakers exclusively was to avoid potential physiological differences between genders and therefore facilitate comparison. Since all of the stimuli were well controlled with respect to the duration (2 seconds per stimulus), loudness, pitch, and intonation of the whole utterance, no normalisation procedure was undertaken. The acoustic correlates of rhoticity were measured together with their

preceding vowels, given that the preceding contexts exert significant influence on predicting the Chinese learners' perceptual patterns of rhoticity.

The boundaries for the /V(r)/ sequences of rhotic and nonrhotic tokens were identified on the basis of visual inspection of the spectrogram by the author. As illustrated in Figure 7.7 the beginning of the sequences was marked at the onset of the preceding vowel, avoiding possible formant transitions from the preceding consonants at the same time. The end boundaries were marked at the lowest state of F3 values for rhotic productions and at the end of the stable formant structure for nonrhotic realisations. F1, F2, and F3 values at five measurement points (10%, 30%, 50%, 70%, and 90%) and their duration in milliseconds were extracted automatically using a Praat script. The formant contours were drawn using the programme VisibleVowel.org (Heeringa & Van de Velde 2018).



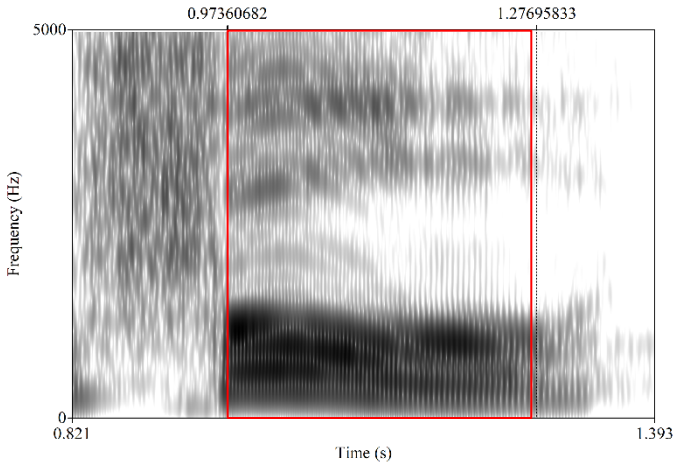


Figure 7.7 Examples of rhotic and nonrhotic realisations of the /V(r)/ sequences in FORCE

7.4.1 Duration

The mean duration of the /V(r)/ sequences in the six vowel contexts was first measured in rhotic and nonrhotic contexts respectively. According to Figure 7.8, the rhotic realisation of the contexts of FORCE, LETTER, NEAR, SQUARE and START is slightly longer than their nonrhotic counterparts. The context of NURSE yields similar duration independent of the presence or absence of nonprevocalic /r/. Since no dramatic difference was observed in terms of the mean duration of /V(r)/ sequences between rhotic and nonrhotic conditions, they were lumped together and compared as shown in Figure 7.9.

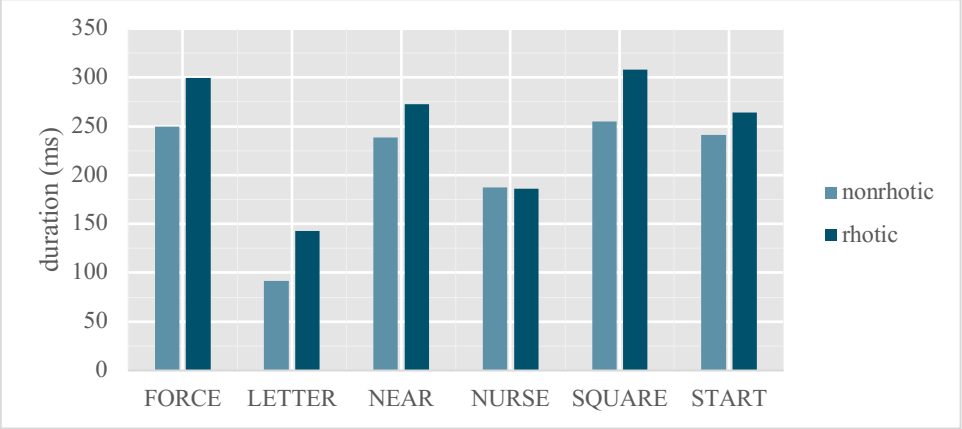


Figure 7.8 Duration (ms) of /V(r)/ sequences in the six phonetic contexts in rhotic and nonrhotic realisations

Figure 7.9 further illustrates the duration of the /V(r)/ sequences for each preceding vowel context. It shows that the sequences of START, FORCE, SQUARE, and NEAR are much longer than the monophthongs NURSE and LETTER. More specifically, SQUARE has the longest mean duration (285 ms), which is followed by NEAR (257 ms), FORCE (254 ms) and START (254 ms). NURSE and LETTER, by contrast, are strikingly shorter than the other contexts, with the mean duration amounting to 186 ms for stressed central vowel NURSE and merely 119 ms for unstressed LETTER.

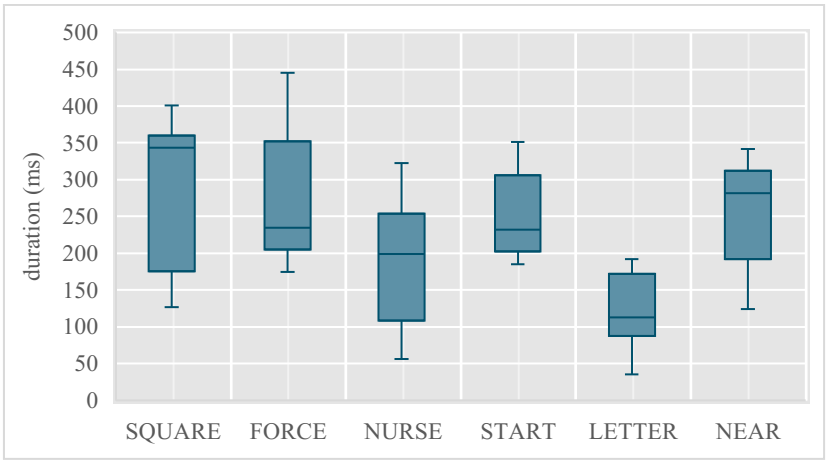


Figure 7.9 Duration (ms) of /V(r)/ sequences in six preceding vowel environments

Figure 7.10 shows the duration of the /V(r)/ sequences in different syllable structures, namely nonprevocalic /r/ in word-final position, in word-medial position, and in consonant clusters. Word-final single /r/ sequences have the longest mean duration, which amounts to about 294 ms. It is followed by the sequences in consonant clusters (248 ms). The word-medial singletons are considerably shorter than the other contexts, with the mean duration amounting to only 158 ms.

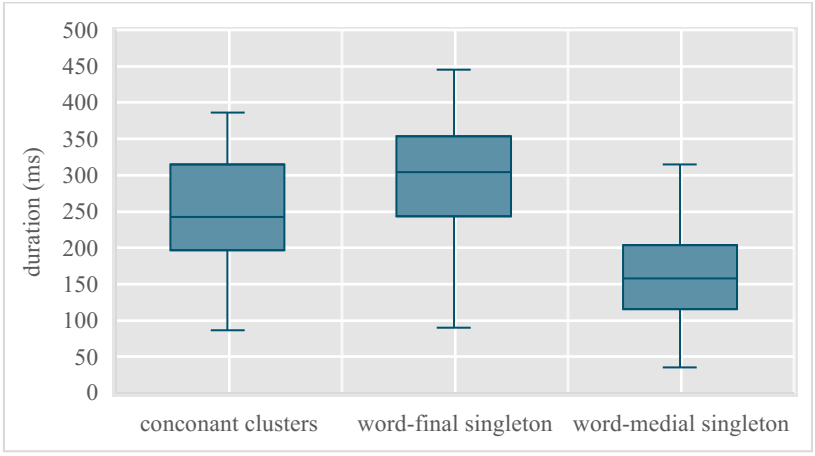


Figure 7.10 Duration (ms) of /V(r)/ sequences in three syllable structures

7.4.2 Formant movements

In Figure 7.11, F2 movements of the six phonetic contexts are shown for rhotic and nonrhotic realisations respectively. As concerns the nonrhotic condition, the F2 values

of all the /V(r)/ sequences are relatively stable. In NEAR context, F2 decreases from 2500 Hz to 2000 Hz, which overlaps with that of the SQUARE sequences with mean F2 lying at around 2000 Hz. In the contexts of LETTER and NURSE, F2 values amount to approximately 1750 Hz. In START sequences, F2 remains stable at a slightly higher level than 1250 Hz, while the context of FORCE demonstrates the lowest F2 values at around 1000 Hz.

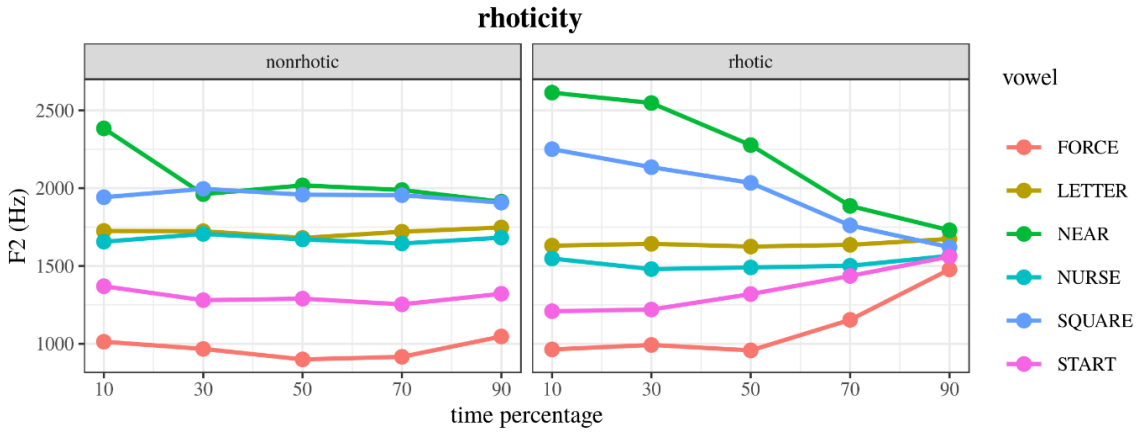


Figure 7.11 F2 values (Hz) of the /V(r)/ sequences in the six phonetic contexts in rhotic and nonrhotic realisations

As concerns the rhotic realisations, substantial formant changes were found in all phonetic sequences except NURSE and LETTER. As shown in Figure 7.11, the mean F2 values of the different vowel contexts start with similar values to their respective nonrhotic counterparts and converge at between 1500 Hz and 1750 Hz at the end of each sequence. To begin with, in the NEAR context, F2 sets out from the highest point among all at above 2500 Hz and decreases dramatically to around 1750 Hz at the end of the sequence. For SQUARE, F2 starts with 2250 Hz and decreases to around 1625 Hz. In the contexts of FORCE and START, F2 begins from slightly below 1000 Hz and 1250 Hz respectively and then rises moderately to approximately 1500 Hz. The mean F2 values of NURSE and LETTER, by contrast, remain at a steady level between 1500 Hz and 1750 Hz across measurement times and are slightly lower than for their nonrhotic counterparts. All in all, Figure 7.11 shows a rise of F2 for FORCE and START, a decrease of F2 for NEAR and SQUARE, and a steady state for LETTER and NURSE. The absolute mean values at the end points for all contexts are, however, similar.

The substantial rise of F2 for the rhotic realisations of FORCE and START is due to the production of nonprevocalic /r/, which causes the modification of articulatory parameters for the change of vowel quality from the back vowels to a central vowel. Similarly, the dramatic decrease of F2 for NEAR and SQUARE is subject to the transformation of vowel quality from the front vowels to a central one. In the context of the central vowels LETTER and NURSE, by contrast, no articulatory modifications are necessary for any changes of vowel quality. Therefore, F2 remains stable for the central vowels in both rhotic and nonrhotic realisations.

Figure 7.12 shows the mean F3 values of the /V(r)/ sequences in different preceding environments in rhotic and nonrhotic conditions respectively. In terms of the nonrhotic sequences, F3 of all phonetic contexts remains relatively high, setting out together from around 2800 Hz. The end points of most contexts were found to be above 2800 Hz except in FORCE, the formant trajectory of which ends slightly lower than the others at approximately 2600 Hz.

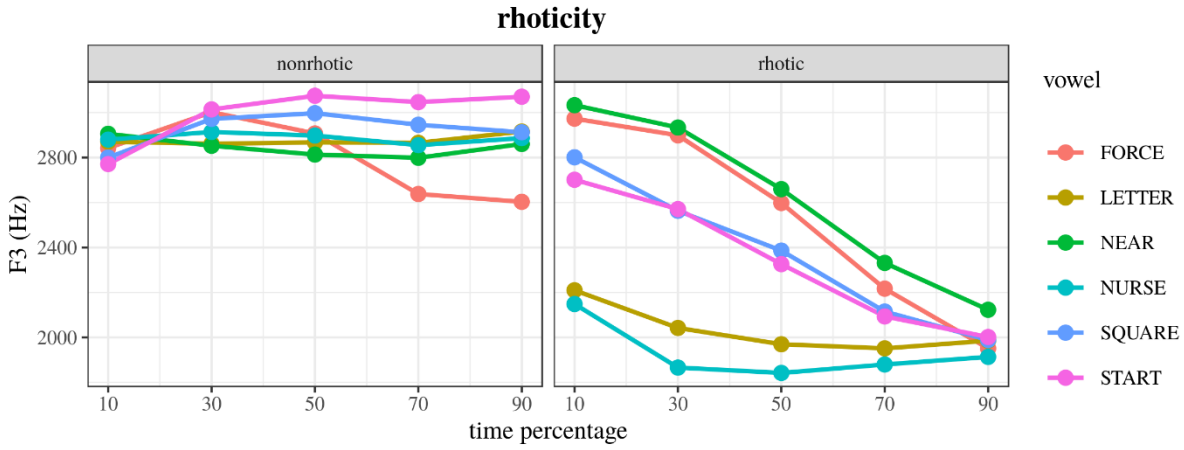


Figure 7.12 F3 values (Hz) of the /V(r)/ sequences in the six phonetic contexts in rhotic and nonrhotic realisations

In contrast to the nonrhotic conditions, F3 values of the r-coloured sequences change considerably across measurement time. The formant transitions of NEAR and FORCE were found to be the most dramatic, starting from around 3000 Hz and dropping substantially to around 2000 Hz. In the contexts of SQUARE and START, F3 sets out from slightly lower than NEAR and FORCE at approximately 2800 Hz and reaches the end state at around 2000 Hz as well. The mean F3 values of LETTER and NURSE, by

contrast, are rather steady and decline only slightly from 2200 Hz to 2000 Hz. It is noteworthy that F3 in various phonetic environments converge to around 2000 Hz at the end points. Similar to F2 in the contexts of NURSE and LETTER, no dramatic transitions were found for the mean F3 values of the central vowels, which differs greatly from the other four vowel contexts.

7.4.3 Acoustic salience of rhoticity

Since duration operates as an important acoustic correlate of prominence in English, the duration of nonprevocalic /r/ together with its preceding phonetic contexts were examined in the present study. In terms of the difference between monophthongal and diphthongal realisations, it is expected that diphthongs are significantly longer than monophthongs. The diphthongs START, FORCE, SQUARE, and NEAR were therefore assumed to be perceptually more prominent to listeners than the monophthongs LETTER and NURSE. Accordingly, rhoticity in the longer diphthongal contexts were assumed to be relatively easier for listeners to perceive than the shorter monophthongal productions. Zooming into the duration of each phonetic context, SQUARE was identified as the most salient context with the longest duration, immediately followed by the other diphthongs NEAR, FORCE, and START. The contexts of LETTER and NURSE, on the other hand, are much shorter and thus predicted to be less prominent than the other environments. The prominence hierarchy of rhoticity concerning the duration of the six preceding phonetic contexts follows the order of SQUARE > FORCE, NEAR, and START > NURSE > LETTER.

As concerns rhoticity in different syllable structures, /V(r)/ sequences in syllable-final word-medial position were found to be the shortest and therefore the least salient among all. Considering the other two contexts, the mean duration of /V(r)/ sequences in word-final position is slightly longer than that in consonant clusters and comes to be the most perceptible context. It is then assumed that listeners would perform the best at discriminating the rhotic-nonrhotic contrasts in word-final position. Nonprevocalic /r/ in word-medial position, on the other hand, would constitute the most difficult perceptual context. As such, the prominence of rhoticity in various syllable structures follows the order of word-final singleton > consonant clusters > word-medial singleton.

Another factor that can possibly affect the degree of acoustic salience of a syllable concerns acoustic modulation along different acoustic dimensions such as frequency and amplitude (Wright 2004). According to Kawasaki (1982) and Ohala (1992), the more changes involved in an acoustic dimension of a speech signal, the greater the acoustic salience. Wright (2004) further proposed that the amount of the increase of salience is associated with the direction of change. Since the direction and the degree of formant transitions can possibly account for the acoustic salience of a syllable, both F2 and F3 movements of the /V(r)/ sequences were investigated in the present study. As concerns F2 in the rhotic realisations, both NEAR and SQUARE show a decrease of F2, with the change of values being more dramatic for NEAR (800 Hz) than for SQUARE (500 Hz). The F2 values of FORCE, on the other hand, increase strikingly (500 Hz). The rhotic realization of START involves a rather moderate rise of F2 (300 Hz). In the contexts of NURSE and LETTER, F2 remains stable throughout the articulation of the sequences. According to the amount of F2 change, the prominence ordering of the six phonetic contexts is NEAR > SQUARE and FORCE > START > NURSE and LETTER.

In terms of F3 values, the rhotic realisations of FORCE, NEAR, SQUARE, and START display various degrees of decrease. Formant transitions are the most dramatic for FORCE and NEAR (around 1000 Hz), which is followed by a relatively moderate drop in SQUARE and START (around 800 Hz). The F3 values for the contexts of NURSE and LETTER are rather stable and decrease slightly by around 200 Hz. The relative prominence of the six phonetic contexts based on F3 shows the following order: FORCE and NEAR > SQUARE and START > NURSE and LETTER.

Taking the modulation of both F2 and F3 movements into consideration, the prominence hierarchy of rhoticity follows NEAR > FORCE > SQUARE > START > NURSE and LETTER. However, it should be noted that this order is only based on the amount of formant frequency changes of various contexts without taking the direction of modulation into account. Concerning the inclination of changes, NEAR and SQUARE are featured by a drop of both F2 and F3 values, whereas the contexts of START and FORCE involve a rise in F2 and a drop in F3 values. It remains unclear whether or which direction of movements promotes acoustic salience of rhoticity.

If both duration and formant transitions are considered, LETTER is to be always expected as the least salient phonetic context, followed by its stressed counterpart NURSE. The other diphthongal realisations are more salient, with longer durations and more dramatic formant changes relative to the central vowels. However, the exact prominence of NEAR, FORCE, SQUARE, START cannot be determined yet. It can only be concluded that SQUARE is the context with the longest duration while NEAR is shown to yield the most dramatic formant transitions.

7.4.4 Realisation of preceding vowels

Figure 7.13 presents the acoustic properties of the rhotic and nonrhotic realisations of the six preceding vowels, respectively. The F1 and F2 values of the vowels at an early time point (namely 10%) were shown in order to avoid possible impact of the following nonprevocalic /r/ on the vowels to the largest extent. As Figure 7.13 illustrates, the rhotic realisations of SQUARE are markedly higher and slightly more front than its nonrhotic counterparts, showing the most distinct vowel qualities among the six vowels. Furthermore, the rhotic realisations of NURSE seem to be slightly higher than the nonrhotic ones (not entirely separate though). On the other hand, the vowel qualities of FORCE, NEAR, LETTER, and START produced by the rhotic and nonrhotic speakers are largely overlapped.

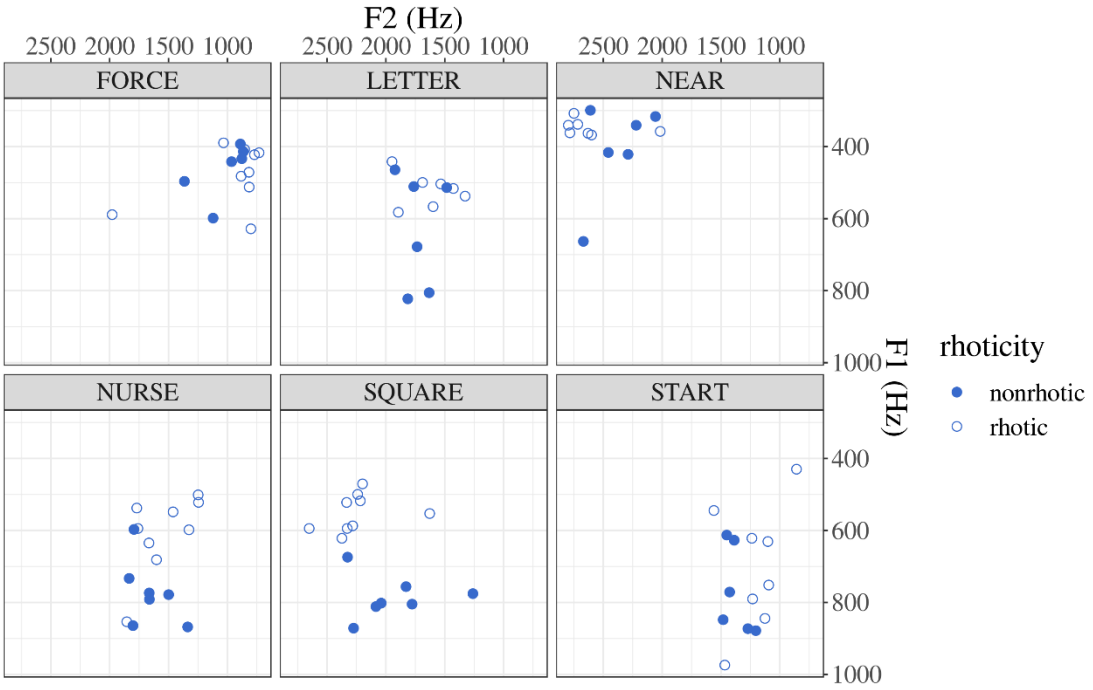


Figure 7.13 Position of the six preceding vowels in the acoustic vowel space

7.5 Discussion of results

7.5.1 Development of perception

As predicted, the perceptual performance of the learners improved significantly from T1 to T2. The learners became more sensitive in distinguishing English rhoticity after the six-month training and learning. It should be noted that the generally low accuracy rates and *A'* scores could be attributed to the design of the oddity task. To begin with, the present experiment employed the most complex trial structures among all discrimination task types. Specifically, three physically distinct stimuli naturally produced by both male and female native English speakers were incorporated in each trial. Moreover, as both change trial and no-change trial were included, there were 8 possible combinations of stimuli (AAB, ABA, BAA, BBA, BAB, ABB, AAA, and BBB) throughout 36 test trials. Instead of the AX and ABX discrimination tasks that only allow for two answer categories, the oddity task provides the listeners with four answer options, with three for them to indicate the odd item out in change trials, and an extra one for them to respond to no-change trials. In addition, distractor trials were

presented to the learners in random order alongside the experimental trials. It is then not surprising that the discrimination scores attained by the learners were relatively low considering the complex and difficult nature of the oddity discrimination task. However, it is yet not entirely clear whether the potential task effect is the only reason for the learners' poor perceptual performance. Ideally, native English speakers of a rhotic variety or L2 learners from a rhotic L1 background should be considered as control groups for the perceptual experiment in the future.

Despite this challenge, some of the learners still achieved ceiling performance with particularly high *A'* scores. A significant overall improvement of their perception of rhoticity was found over the course of time, albeit with inter- and intra-individual variability. In terms of the variation between learners, the participants' performance ranged from complete insensitivity towards rhoticity to perfect discrimination ability. Moreover, considerable variability within individual learners between two data collection times and across different phonetic contexts was also found. A large proportion of participants showed various degrees of improvement, while the others were either stable or became even less sensitive in discriminating English rhoticity. On the other hand, the learners' perceptual performance also varied according to phonetic dimensions, showing a higher level of sensitivity to certain environments yet being insensitive to others.

As demonstrated by the mean *A'* scores obtained for different phonetic contexts at T1 and T2, the Chinese learners seemed to acquire perceptual sensitivity towards rhoticity context by context. The results showed that they were first only sensitive to preceding SQUARE vowels and subsequently gained sensitivity towards FORCE, NEAR, and START at T2. With preceding central vowels NURSE and LETTER, however, the learners still failed in discriminating rhoticity after the six months of further acquisition. Different levels of difficulty are assumed to be inherent in the phonetic contexts that nonprevocalic /r/s occur in. The Chinese learners in the present study were found to be perceptually more sensitive to acoustically salient contexts: at T1, they showed sensitivity only to the context of SQUARE, which represents the longest phonetic context among those tested. By T2, the learners were able to discriminate rhoticity in all of the four diphthongal realisations, which, according to the acoustic

measurements of the perceptual stimuli, have significantly longer duration and more dramatic formant transitions relative to the monophthongal realisations. NURSE and LETTER, on the other hand, are acoustically less salient in terms of both duration and formant modulation and also turned out to be the most difficult contexts for the learners to perceive rhoticity in at both T1 and T2.

7.5.2 Factors influencing perception

PRECEDING VOWEL

The multilevel analyses revealed that the perception of rhoticity by the Chinese learners is significantly influenced by the preceding phonetic contexts. It is not surprising that the development of rhoticity perception is controlled by the preceding environments considering the close relationship between nonprevocalic /r/ and the preceding vowels in both Mandarin and English. Regarding Mandarin, earlier work shows that the diminutive suffix centralizes the preceding vowels (Huang 2010). Specifically, sounds that are compatible with the feature [+ retroflex] will be completely rhotacized, while those that are incompatible with the feature will be pushed into the onset (Duanmu 2000). As a result, the 19 rhymes in Mandarin are reduced to 11 when a rhotic suffix is attached as described in section 5.1. Concerning English, nonprevocalic /r/ is also not independent from its preceding vowel contexts, and speakers tend to reduce the phonetic and articulatory differences between syllable coda /r/ and nucleus (e.g., Lehiste 1962, Hickey 2014). Lawson et al. (2013) also found that nonprevocalic /r/ has a coarticulatory effect on its preceding vowels in Scottish English, which possibly accounts for the merger of the historically-distinct NURSE vowels. Moreover, Storme (2018) and Gendrot (2014) reported that the perception of French coda rhotic is affected by its preceding vowel context, showing a universal impact of preceding vowel on the perception of rhoticity crosslinguistically. Besides, preceding vowel has also been found to be a significant phonological factor conditioning the production of rhoticity in world Englishes, with front vowel environments being more favouring of rhoticity than back vowels (e.g., Nagy & Irwin 2007).

The preceding vowels predict the likelihood of correct perception of rhoticity by the Chinese listeners in the following order: SQUARE > NEAR > FORCE > START > NURSE >

LETTER. This primarily correlates with the degrees of acoustic salience of the six phonetic contexts measured and discussed in section 7.4. To begin with, SQUARE had the longest duration and was shown to be the easiest context for the learners to discriminate rhoticity in. NEAR as the most salient context because of its most dramatic formant transition became the second promoting context of accurate responses. The contexts of FORCE and START are diphthongal realisations that also involve formant movements (simultaneous F2 rise and F3 drop), the degrees of which are however less remarkable than NEAR. The central vowels LETTER and NURSE, on the other hand, are acoustically the least salient contexts due to their short duration and steady formant movements and were found to be particularly difficult for the learners to distinguish rhoticity in as well. As the rhotic realisations of NURSE and LETTER are identified by the steady state of both *F2* and *F3*, it could be deduced that it is the formant movements from the preceding vowel to the nonprevocalic /r/ that possibly affect the perception of rhoticity other than the absolute frequency value at a specific measurement point. The Chinese learners were more accurately perceiving nonprevocalic /r/ in the contexts with long duration and sharp formant transitions rather than those that were short and with steady formant movements. All in all, the accuracy of rhoticity perception is closely associated with the preceding vowel. The more salient a /V(r)/ sequence, the greater chance for the learners to perceive rhoticity.

Furthermore, accurate perception might as well be due to the use of rather distinct vowels in the /V(r)/ sequences between RP and AE. The preceding SQUARE vowel, being the only context showing completely distinct vowel properties between rhotic and nonrhotic realisations among the six vowels as discussed in section 7.4.4, was also found to be the only context that the listeners can successfully perceive at T1. It seems that the differences in vowel qualities may have contributed to the differential discriminability of the learners. However, it is impossible to tease apart the effects of vowel qualities and acoustic salience on the perception of rhoticity in SQUARE with the current methodology, since SQUARE is arguably also the most salient contexts among the six. By contrast, NURSE as the other vowel showing different rhotic and nonrhotic realisations did not promote the perceptual sensitivity of the learners. They failed to discriminate rhoticity in NURSE at both data points, which is not surprising as

NURSE is an acoustically less salient sequence considering its duration and frequency movements. It is then reasonable to consider preceding vowel difference as one of the factors conditioning the perceptual performance of the learners, the effects of which have been overridden by the acoustic salience of rhoticity.

Interestingly, the acquisition pattern of perception as regards various preceding vowel environments is also in line with the prediction of the SLM (Flege 1995), which proposes that dissimilar TL sounds are more easily perceived and produced by L2 learners as opposed to similar TL sounds. As illustrated in section 5.2, the SLM predicts the following acquisition order of rhoticity for Mandarin learners from ease to difficulty: SQUARE/FORCE > NEAR > START > NURSE/LETTER, which almost perfectly mirrors the perceptual pattern by the learners in the present study. Since both the crosslinguistic dissimilarity of rhoticity between English and Mandarin and the degree of acoustic salience of the /V(r)/ sequences predict the same acquisition pattern for Chinese learners, a critical issue that should be immediately addressed is to distinguish the underlying force determining the observed perceptual pattern.

In order to investigate the potential factors conditioning the perceptual accuracy of rhoticity, a replication perception study was conducted by testing 20 German learners of English. To ensure the comparability between the two sets of listeners, first-year undergraduate students majoring in English at a German university were recruited to participate in the same oddity discrimination task with the same procedure as illustrated in section 6.2.1. Among them there were 16 females and 4 males, aged between 18 and 20. All the participants were native German speakers with no knowledge of Mandarin Chinese, which allows the elimination of possible effects of CLI from Mandarin. The German and the Chinese listeners were expected to display similar perceptual patterns if the acoustic salience of the perceptual stimuli was the key factor. Otherwise the German listeners would show a considerably different pattern from the Chinese learners considering the language distance between the two L1s, which is most likely to be conditioned by their L1 German as predicted by the SLM (Flege 1995) and other foreign languages spoken by them.

A total of 709 responses were collected from the 20 German listeners, which were

further analysed using a mixed-effects logistic regression model in R (R Core Team 2013) using the Lme4 package (Bates et al. 2014) with PERCEPTUAL ACCURACY as a binary dependent variable, and PARTICIPANT and STIMULUS as random effects. The results in Table 7.11 show that PRECEDING VOWEL is a significant factor conditioning the perceptual accuracy of English rhoticity by the German learners of English. As shown in Figure 7.14, a preceding SQUARE vowel was found to be the most promoting context of accurate responses, with more than half (58.5%) of the cases being discriminated accurately by the German learners (p = 0.33), which is followed by preceding NEAR (48.3%). The effects of the other /V(r)/ sequences are however not significantly different from each other according to the mixed-effects regression model. Post-hoc t-tests with Bonferroni corrections were further conducted and no statistical significance between the six preceding contexts was found.

Table 7.11 Best-fit mixed effect regression model of PERCEPTUAL ACCURACY for German learners of English

Fixed effects	**Levels**	**Estimate**	**Std. error**	***z*-value**	***p* <**	
		-0.302	0.250	-1,209	0.227	
PRECEDING VOWEL	LETTER	-0.113	0.318	-0,355	0.723	
	NEAR	0.236	0.316	0.747	0.455	
	NURSE	-0.244	0.322	-0.758	0.449	
	SQUARE	0.681	0.319	2.138	0.033	*
	START	-0.189	0.320	-0.589	0.556	
Random effects	**Type**	**Variance**	**Std. Dev**			
PARTICIPANT	(Intercept)	0.245	0.495			
STIMULUS	(Intercept)	0.041	0.203			
		Min.	**Median**	**Max.**		
Scaled residuals		-1.825	-0.639	1.699		

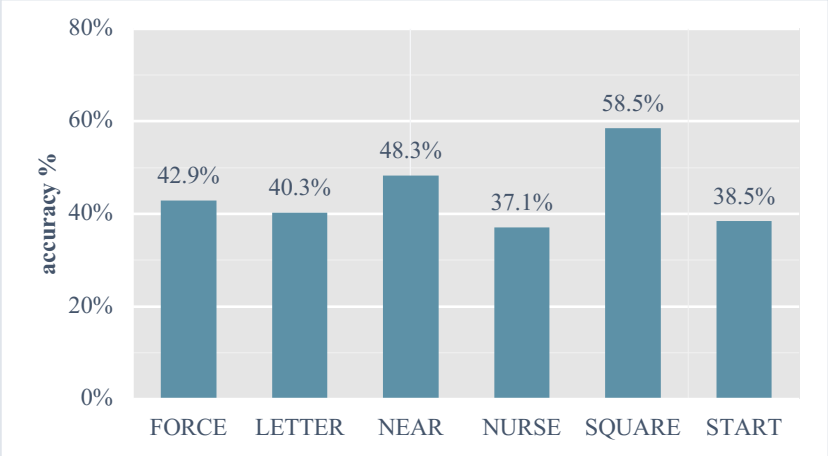


Figure 7.14 Accuracy of perceived nonprevocalic /r/ for German learners by PRECEDING VOWEL

The observed perceptual pattern of the German learners largely parallels the pattern of the Chinese learners as regards the contexts of SQUARE and NEAR. At the first time of data collection, the Chinese listeners displayed perceptual sensitivity only when rhoticity occurred in a preceding SQUARE context. In the acoustic measurement of the perceptual stimuli, SQUARE was found to be an acoustically salient context concerning its longest duration among the six /V(r)/ sequences, which can possibly account for its promoting sensitivity for the learners with different L1s. A preceding NEAR vowel

context, which was found to involve the most salient phonetic cues concerning formant modulation and was the second favouring context for the Chinese learners, was also relatively easy for the German learners to discriminate. It could therefore be concluded that the perception of English rhoticity in the contexts of SQUARE and NEAR by the learners in the study at hand is more determined by the relative degree of acoustic salience of the phonetic contexts it occurs in than the perceived crosslinguistic dissimilarity between L1 and TL as predicted by the SLM (Flege 1995).

While the contexts of NURSE and LETTER were shown to be particularly challenging for the Chinese listeners to discriminate, possibly due to their remarkable resemblance to the phonetic contexts of rhoticity in Mandarin or the fact that they are the acoustically less salient /V(r)/ sequences, they did not pose greater challenges for the German listeners. According to the mixed effects model in Table 7.11, the German learners' perceptual sensitivity to NURSE and LETTER was not significantly different from the contexts of START and FORCE. The different perceptual patterns (regarding NURSE and LETTER) between the German and the Chinese learners are likely to be attributable to the different L1s spoken by them, which is in line with the predictions of the SLM (Flege 1995). As such, both the degree of acoustic salience and the crosslinguistic dissimilarity between the TL and the L1 of learners seem to account for the context-dependent perception of English rhoticity by L2 learners. It is however not clear how these two influencing factors interact with each other in conditioning the perception of a particular L2 feature or context so far. It seems that for the Chinese learners in the present study, the impact of acoustic salience overrode L1 and exerted more influence on the perception of English rhoticity, but the effect of L1 cannot be completely ruled out either due to the slightly different perceptual pattern (regarding NURSE and LETTER) displayed by the German learners. Therefore, research on this line is necessary in the future to further shed light on our understanding of L2 speech perception.

REACTION TIME

The best-fit mixed model of the Chinese learners' perceptual performance at T2 revealed that reaction time plays a significant role in predicting the accurate

perception of rhoticity. The longer the reaction time, the more likely a contrast is perceived inaccurately by the learners. The perceptual accuracy of the learners increases with a decreasing amount of reaction time. This can be explained by the degree of acoustic salience of various phonetic contexts as well. As described in section 1.3, the acoustically more salient contexts generally take shorter reaction times and are more likely to yield accurate responses relative to the acoustically less salient ones. Taking SQUARE as an example, the learners in the present study spent only 368 ms on average to respond to a SQUARE vowel, which is the most salient context based on its long duration and dramatic decrease of both *F2* and *F3*. Accordingly, the rate of accuracy in perceiving SQUARE is also the highest among all learners. The context of LETTER, on the other hand, is the least salient in terms of its short duration and steady formant state, which is particularly difficult for the learners to perceive and requires a much longer reaction time (600 ms at T2).

Other external factors

In this study, none of the external factors were found to exert significant impact on the perception of nonprevocalic /r/ by the Chinese learners of English. Participants with different gender, place of origin and target norm did not vary in their perceptual patterns of rhoticity in English. As concerns L1 regions, although the retroflex suffix does not exist in most Mandarin dialects in the South of China, the learners there could still distinguish rhoticity with similar accuracy levels as those from the North. One possible reason for this is that the learners from the South may have regular exposure to rhotic Mandarin and English speech via various sources such as media or personal contact. According to the information regarding the learners' media exposure collected in the questionnaires, none of the learners reported to only have exposure to nonrhotic English media. In terms of target orientation, the results showed that targeting a rhotic or nonrhotic variety did not affect learners' discrimination sensitivity. As for gender, although female speakers were shown to be more target-conscious and tended to produce more target variants compared to males (e.g., Chand 2010), they did not differ in terms of their performance in distinguishing English nonprevocalic /r/.

7.6 Chapter summary

This chapter was concerned with the perception of English rhoticity by the Chinese

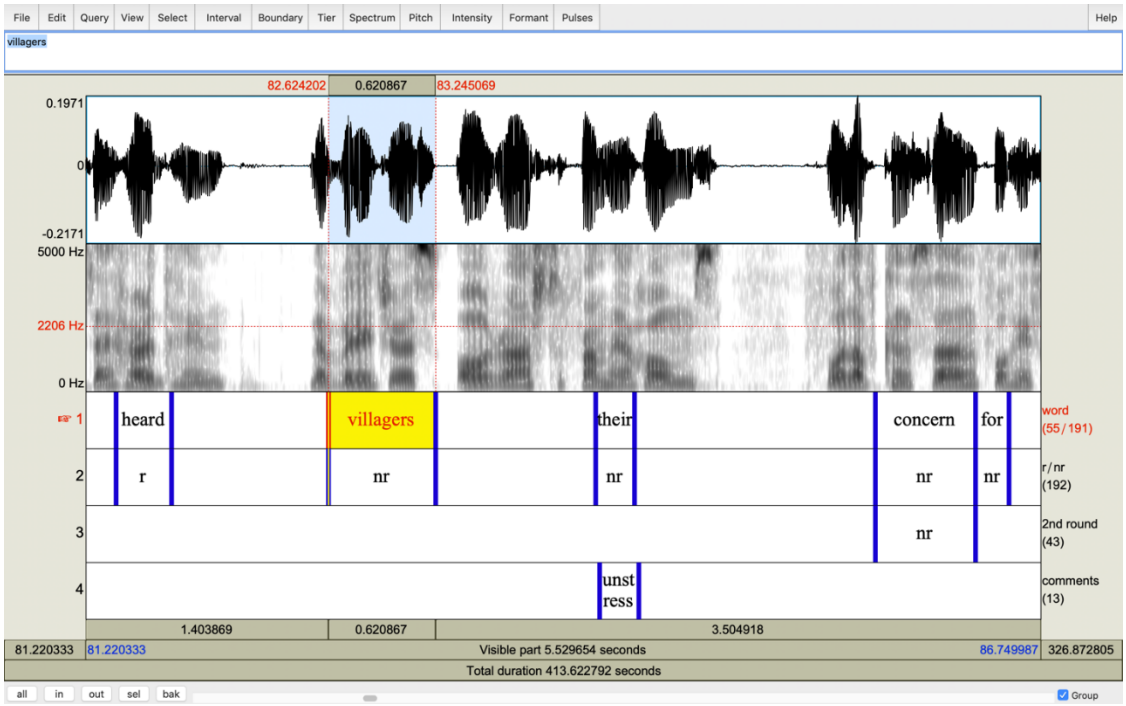
learners over time. In general, the perceptual sensitivity of the learners improved significantly after the training and the six months of learning, albeit with considerable variability between and within individual learners. The perceptual patterns of English rhoticity of the learners could be accounted for by both the acoustic salience of the phonetic contexts that nonprevocalic /r/ exists in and the degree of the crosslinguistic dissimilarity between the TL and the L1 of the learners. The acoustically more salient and the perceptually dissimilar stimuli are relatively more perceptible for the learners, while those acoustically less salient and perceptually more similar such as NURSE and LETTER are especially difficult for them to discriminate. Specifically, the acoustic salience of various phonetic contexts was determined by the duration as well as the amount and direction of formant modulation of the /V(r)/ sequences, which further supports the findings of Gendrot (2014) for the perceptibility of French coda rhotics. Importantly, the learners were found to acquire the perception of rhoticity context by context over time, from the acoustically more salient and the perceptually dissimilar ones to the less salient and similar ones. Furthermore, PRECEDING VOWEL was found to be the only main effect conditioning the perceptual patterns of English rhoticity by the learners. The other phonological factor SYLLABLE STRUCTURE and all the grammar-external predictors did not reach statistical significance.

8. Production of rhoticity by Chinese learners of English

This chapter investigates Chinese learners' acquisition of the production of English nonprevocalic /r/. After a brief introduction of the coding and analyses of the spoken data in section 8.1, section 8.2 presents general descriptive statistics of the realisation of rhoticity by the learners in different contexts over time. Section 8.3 demonstrates main effects conditioning their production based on mixed effects logistic regression models for T1 and T2 respectively. The results are further discussed in section 8.4, in which the development of the learners (section 8.4.1) and the significant internal and external factors predicting the patterns of rhoticity (section 8.4.2 and 8.4.3) are explored in detail.

8.1 Data preparation and statistical analysis

A total of 3,353 tokens for the 47 Chinese speakers were elicited from the two English tasks at both data points. They were transcribed auditorily in Praat (Boersma 2001; Version 5.3.87). The boundaries for each word with nonprevocalic /r/ were manually marked on an interval tier named "word". A binary choice between the presence and absence of nonprevocalic /r/ was made perceptually by the author on the second tier named "r/nr" (see Figure 8.1). The transcription of words and auditory coding of rhoticity were extracted automatically using a Praat script. A subsample of 799 tokens (about 20% of the production data, i.e. every fifth word in the dataset) were further examined by a phonetically-trained native American listener. A high degree of intercoder reliability amounting to 94% agreement was attained for the auditory analysis.



nr: nonrhotic; r: rhotic

Figure 8.1 Transcription and coding of speech data

Alongside the auditory analysis of the presence or absence of nonprevocalic /r/, each token was also coded for the independent variables of interest as described in section 6.3. Both grammar-internal and -external constraints were expected to exert influence on the production of rhoticity by Chinese learners of English. The data were further analysed by mixed-effects logistic regression models in R (R Core Team 2013) using the Lme4 package (Bates et al. 2014). Post-hoc pairwise comparisons with Bonferroni corrections were conducted in the lsmeans package (Lenth 2016) and its successor, the emmeans package (Lenth 2018).

8.2 Descriptive statistics

This section starts with the overall distribution of rhoticity as well as the overall proportions of target production by the 47 Chinese speakers at both data collection times (Section 8.2.1). Section 8.2.2 then looks into the individual variability in the realisation of nonprevocalic /r/. Section 8.2.3 further zooms in to show the frequencies of rhoticity by the learners with respect to different grammar-internal and -external variables.

8.2.1 Realisation of nonprevocalic /r/

Beginning with the overall distribution of the realisation of nonprevocalic /r/ in all cases by the 47 Chinese learners of English, Table 8.1 illustrates that slightly more than half of the tokens were produced with the presence of nonprevocalic /r/ at the two times of data collection. At T1, 57.6% of the tokens were realised with rhoticity, which slightly declined to 53.4% after six months of acquisition. The overall rhotic realisations by the Chinese learners amounted to 55.5% for both T1 and T2, suggesting that the speech produced by the learners is marginally rhotic.

Table 8.1 Overall realisation of rhoticity by Chinese learners at T1 and T2

	rhotic	nonrhotic	total	rhotic%
T1	965	710	1675	57.6
T2	891	778	1669	53.4
T1 and T2	1856	1488	3344	55.5

As the learners in the present study had two different target norms, non-rhotic BrE and rhotic AmE, which was expected to influence the realisation of rhoticity in their speech, the percentage of the target realisations was calculated for individual learners on the basis of their self-reported target norms in the questionnaire. To be more specific, the rate of rhotic production was calculated for the learners targeting AmE, while the rate of nonrhotic realisation was calculated for the learners aiming at BrE. As shown in Figure 8.2, the percentage of target production by the Chinese learners ranged from 13% to 100% for T1, and from 0% to 100% for T2. The mean target-like production amounted to 50% at T1, which further increased to 60% after training and six months of acquisition. However, a statistically significant difference in their production performance between T1 and T2 was not reached according to a paired *t*-test, as the variation within the groups was so high.

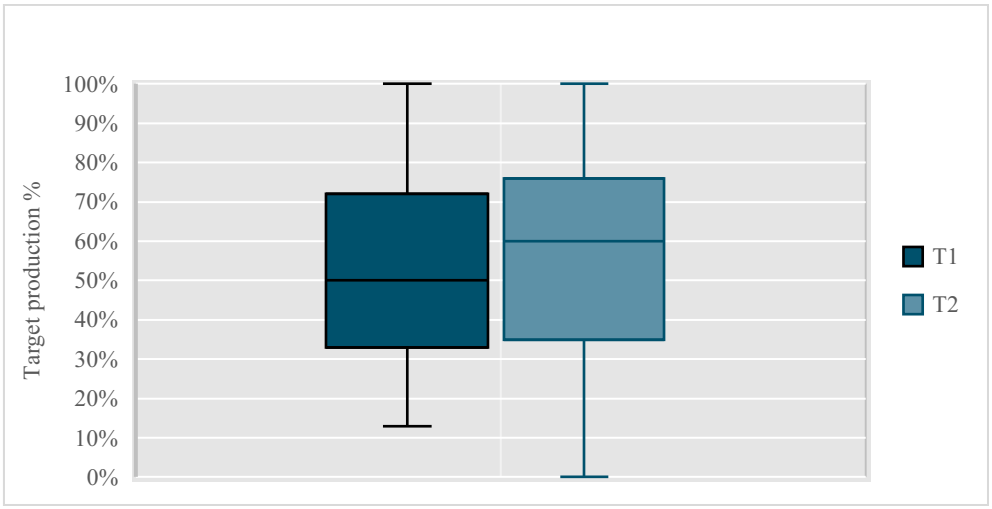


Figure 8.2 Boxplots of proportions of target production for all learners at T1 and T2

8.2.2 Inter- and intra-learner variability

This section shows the individual distribution of target production by the Chinese learners of English over time. As expected, considerable inter- and intra-learner variability was observed. The percentages of target production by each learner at both T1 and T2 are illustrated in Figure 8.3. On the one hand, remarkable differences existed between speakers, ranging from highly target-oriented production performance at both data collection times such as speakers 4131 and 4413, to speaker 4409, who completely failed to follow his target norm at T2.

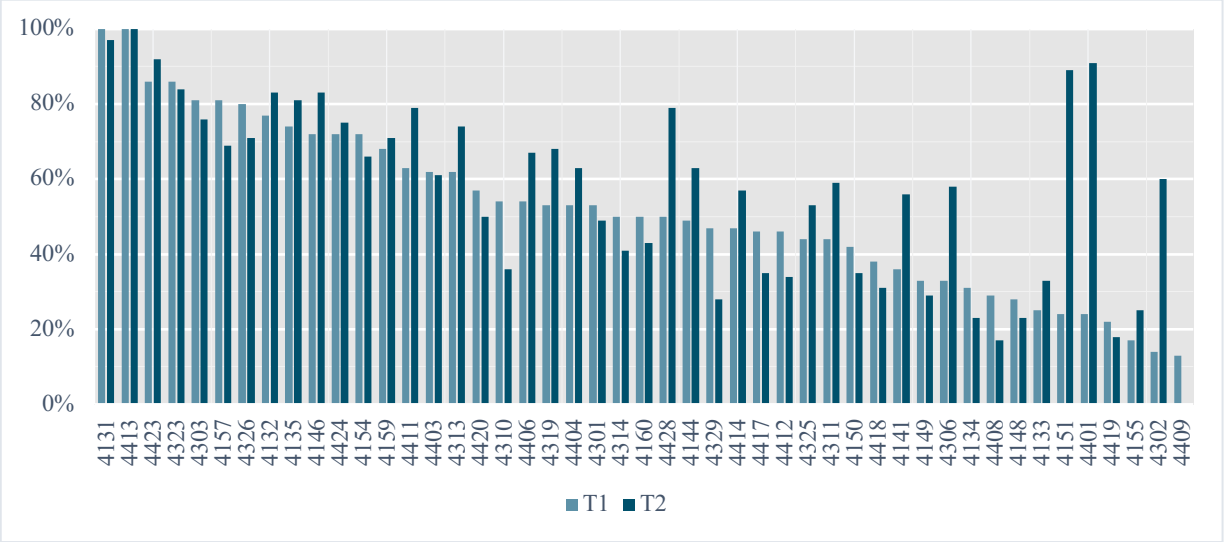


Figure 8.3 Individual distribution of target production by the 47 Chinese learners at T1 and T2. Learners are arranged along the x-axis according to their performance at T1.

On the other hand, learners differed in their own performances between the two times of data collection. Zooming into the performance of individual learners, half of them (24 out of 47) performed more target-oriented at T2 relative to T1. Speakers 4151 and 4401 only produced around 24% of the tokens target-like at T1, which improved dramatically to around 90% after six months of further learning. In contrast, some learners became less target-oriented in their production of English rhoticity such as learners 4310 and 4329. In general, no learners were found to be categorically rhotic. Ceiling performance was achieved only by learners from the South of China targeting BrE, such as learner 4413, who was categorically nonrhotic at both T1 and T2. Learner 4131 also produced all of the test tokens without nonprevocalic /r/ at T1 but became slightly rhotic at T2. Learner 4409 targeting AmE was categorically nonrhotic and produced none of the test tokens target-like at T2. It seemed particularly difficult for the learners targeting AmE to pronounce all of the test tokens with the a realised nonprevocalic /r/.

8.2.3 Quantitative distribution of rhoticity

This section investigates whether the grammar-external variables TARGET NORM, PLACE OF ORIGIN, GENDER, SPEECH STYLE (TASK), NORM PREFERENCE, as well as the grammar-internal variables PRECEDING VOWEL, SYLLABLE STRUCTURE, WORD TYPE

have an effect on the rhoticity rates of the Chinese learners.

Table 8.2 illustrates the proportion of the presence of rhoticity by the Chinese learners with different target norms at T1 and T2 respectively. In general, the learners targeting AmE produced a slightly higher amount of nonprevocalic /r/ (59.7%) than those targeting the nonrhotic BrE (51.7 %). At T1, the distinction between the speakers aiming at different norms amounted to 4%, which increased to 12.1% after the six months of acquisition, suggesting that the Chinese learners in the present study gradually converged towards their target norms with respect to the production of English rhoticity over time.

Table 8.2 Proportions of rhotic realisations by TARGET NORM

Target norm	AmE		BrE	
	T1	**T2**	**T1**	**T2**
n total	894	853	458	816
n rhotic	537	506	817	385
%	**60.1**	**59.3**	**56.1**	**47.2**

Table 8.3 demonstrates the amount of rhoticity produced by the Chinese learners coming from different L1 dialectal regions at T1 and T2. As can be seen, the learners from the North of China where the Mandarin dialects spoken are characterized by rhoticity produced a higher amount of rhotic tokens in English (63.35%) than those who originated in the South of China (48.4%). At T1, the learners from the North pronounced 67.5% of the tokens with nonprevocalic /r/, which however decreased to 59.9% at T2. The proportion of rhotic production for the learners from the South was rather consistent before and after the training and the six months of learning.

Table 8.3 Proportions of rhotic realisations by PLACE OF ORIGIN

L1 region	Rhotic North		Nonrhotic South	
	T1	**T2**	**T1**	**T2**
n total	856	818	855	851
n rhotic	578	484	417	407
%	**67.5**	**59.2**	**48.8**	**47.8**

Table 8.4 shows the percentage of rhotic realisations by female and male learners respectively. The production of rhoticity by the 47 learners varied distinctly with gender. It is clear that the female speakers favoured the presence of nonprevocalic /r/

in English more strongly than their male counterparts at both data collection times. The difference of the percentages of rhotic production between the male and the female learners amounted to 19.7% at T1, which further increased to 27% at T2.

Table 8.4 Proportions of rhotic realisations by GENDER

Gender	female		male	
	T1	**T2**	**T1**	**T2**
n total	1142	1105	569	564
n rhotic	739	672	256	219
%	**64.7**	**60.8**	**45.0**	**38.8**

Table 8.5 shows the distribution of rhotic productions in tasks with different levels of formality at T1 and T2. As shown in the table, the participants produced higher numbers of nonprevocalic /r/ in the relatively easier and more informal picture naming task (68.4% at T1 and 61.8% at T2) than in the relatively more difficult and controlled passage reading task (53.4% at T1 and 49.4% at T2). The difference in the realisation of rhoticity between the two tasks seemed to be slightly more remarkable at T1 (15%) than T2 (12.4%).

Table 8.5 Proportions of rhotic realisations by SPEECH STYLE

Task	Passage reading		Picture naming	
	T1	**T2**	**T1**	**T2**
n total	1167	1130	544	539
n rhotic	623	558	372	333
%	**53.4**	**49.4**	**68.4**	**61.8**

The language preference of the 47 learners towards four different English accents was investigated using a verbal-guise survey. The learners were asked to listen to and evaluate four English news clips produced by two native English speakers and two highly proficient Chinese English news reporters with rhotic and nonrhotic accents respectively. Table 8.6 illustrates the ratings given by the learners with different target norms.

Table 8.6 Ratings of different accents by learners targeting AmE and BrE

	American English		British-oriented Chinese English		American-oriented Chinese English		British English	
target	AmE	BrE	AmE	BrE	AmE	BrE	AmE	BrE
Mean	**3.25**	**3.04**	**3.87**	**4.44**	**4.13**	**3.91**	**4.13**	**4.96**
SD	1.73	1.46	1.48	1.20	1.48	1.56	1.54	1.07

Note: The ratings range from 1 (strongly disfavour) to 6 (strongly favour).

As shown in the table, native British English received the highest scores, which amounted to 4.13 for the learners targeting AmE and 4.96 for those targeting BrE. Native American English, on the contrary, was rated the lowest among the four accents by the learners targeting both norms. On the other hand, American and British English accents produced by highly proficient Chinese English speakers were rated much higher than native American English. As expected, the Chinese learners targeting BrE displayed more positive attitudes towards British-oriented Chinese English (4.44) than those targeting AmE (3.87). Likewise, the American English news report produced by a highly proficient Chinese English speaker was rated higher by the learners targeting AmE (4.13) than by those targeting BrE (3.91). It is worth noting that the self-reported target norms of the learners showed a closer correlation with their ratings of the Chinese English accents than the native English accents.

The preceding phonetic environment has been repeatedly shown to exert significant impact on the production of rhoticity in previous studies. Table 8.7 illustrates the distribution of rhotic realisations on the basis of the six preceding contexts at two times of data collection. The Chinese learners in the present study produced the largest amount of rhoticity when the nonprevocalic /r/ was preceded by a NURSE vowel, the percentages of which amounted to 84.4% at T1 and 79.8% at T2. A preceding NEAR vowel was the second favouring context of rhoticity, with 73% and 63% tokens being realised at T1 and T2 respectively. The learners also produced more than half of the test tokens with the presence of nonprevocalic /r/ when it was preceded by LETTER, SQUARE, and START contexts at both data points. A preceding FORCE vowel, conversely, was shown to be the least promoting of rhoticity. The learners produced only 30.9% and 31.6% of the cases preceded by a FORCE vowel at T1 and T2

respectively.

Table 8.7 Proportions of rhotic realisations by PRECEDING VOWEL

Vowel	NURSE		NEAR		LETTER		SQUARE		START		FORCE	
	T1	**T1**	**T2**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**	**T1**	**T2**
n total	179	141	139	178	633	609	233	230	140	146	385	367
n rhotic	151	103	88	142	401	353	146	119	75	73	119	116
%	**84.4**	**73.0**	**63.3**	**79.8**	**63.3**	**58.0**	**62.7**	**51.7**	**53.6**	**50.0**	**30.9**	**31.6**

As Table 8.8 presents, the proportion of rhotic realisations also varied distinctly with nonprevocalic /r/ occurring in different syllable structures. 64.7% and 64.2% of the tokens in consonant clusters were realized with rhoticity at T1 and T2. It was followed by a single /r/ in word final position, with 57.2% and 50.7% being realized by the learners. Concerning single nonprevocalic /r/ in word-medial position, only 8.2% and 14.9% of the test tokens were realized at T1 and T2 respectively. However, this discrepancy can be attributed to the unbalanced amount of data for the different syllable contexts. It is noteworthy that only one word, *aft*ernoon, was included in the word-medial singleton category. It is therefore not legitimate to draw any conclusion with respect to the production of singleton /r/ in word-medial position based on such a limited data set. Considering the remaining two syllable structures, the learners were found to favour nonprevocalic /r/ in consonant clusters slightly more than in word-final single position.

Table 8.8 Proportions of rhotic realisations by SYLLABLE STRUCTURE

Syllable structure	Consonant clusters		Word-final singleton		Word-medial singleton	
	T1	**T2**	**T1**	**T2**	**T1**	**T2**
n total	539	524	1123	1098	49	47
n rhotic	349	327	642	557	4	7
%	**64.7**	**62.4**	**57.2**	**50.7**	**8.2**	**14.9**

Table 8.9 demonstrates the production of rhoticity in different word types at both times. As shown in the table, rhoticity was pronounced more frequently in content words than in function words. The average production percentages between T1 and T2 amounted to 60.3% for content words and approximately 35% for function words.

Table 8.9 Proportions of rhotic realisations by WORD TYPE

Word type	Content word		Function word	
	T1	**T2**	**T1**	**T2**
n total	1424	1348	287	321
n rhotic	897	776	98	115
%	**63.0**	**57.6**	**34.1**	**35.8**

8.3 Multilevel analyses

The data were further analysed in a mixed-effects logistic regression model in R using the Lme4 package to investigate potential grammar-internal and -external predictors controlling the presence or absence of rhoticity in the English speech of the Chinese learners.

Table 8.10 illustrates the main effects affecting the participants' production of rhoticity at T1. PRECEDING VOWEL was found to be the only significant phonological factor conditioning the realisation of nonprevocalic /r/ by the Chinese learners. The following hierarchy was revealed as in Figure 8.4: NURSE > NEAR > LETTER > SQUARE > START > FORCE. To be more specific, the stressed central vowel NURSE was shown to be the most promoting context of the realisation of nonprevocalic /r/ ($p < .001$), which was followed by LETTER ($p < .001$), NEAR ($p = .003$), and SQUARE ($p = .006$), while START and FORCE strongly disfavoured rhoticity. Here, pairwise comparisons with Bonferroni correction were further conducted. Results showed that FORCE is significantly different from NURSE ($p < .001$), LETTER ($p = .002$), and NEAR ($p = .038$). Differences between the other vowels did not reach statistical significance.

Table 8.10 Best-fit mixed effect regression model of the presence or absence of rhoticity at T1 as a binary dependent variable, with PARTICIPANT *and* WORD *as random intercepts*

Fixed effects	**Levels**	**Estimate**	**Std. error**	***z*-value**	***p* <**	
(intercept)		-0.208	0.419	-0.496	0.620	
PRECEDING	LETTER	1.607	0.427	3.761	0.000	***
VOWEL	NEAR	1.925	0.638	3.019	0.003	**
	NURSE	2.727	0.593	4.597	0.000	***
	SQUARE	1.451	0.529	2.745	0.006	**
	START	0.989	0.625	1.584	0.113	
GENDER	male	-1.167	0.326	-3.582	0.000	***
SPEECH STYLE	informal	0.766	0.348	2.198	0.028	*
PLACE OF ORIGIN	South	-1.094	0.306	-3.572	0.000	***
Random effects	**Type**	**Variance**	**Std. Dev**			
PARTICIPANT	(Intercept)	0.9185	0.958			
WORD	(Intercept)	0.7099	0.843			
		Min.	**Median**	**Max.**		
Scaled residuals		-5.863	0.278	3.271		

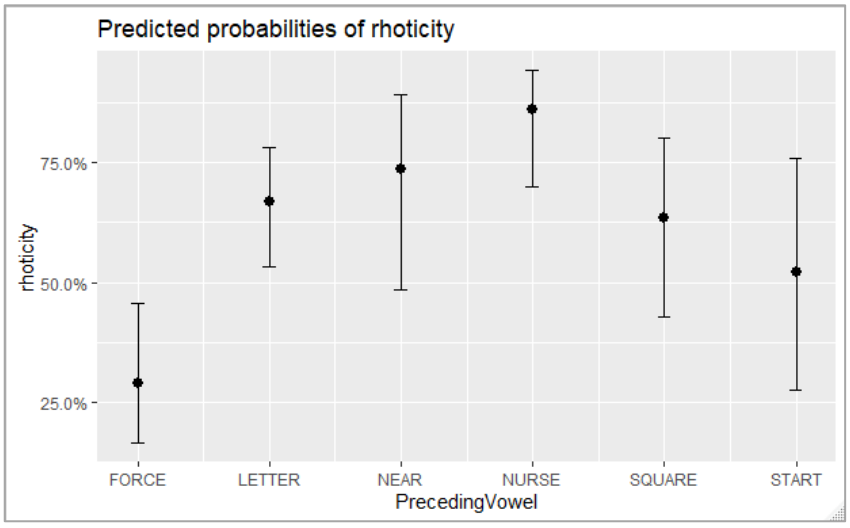
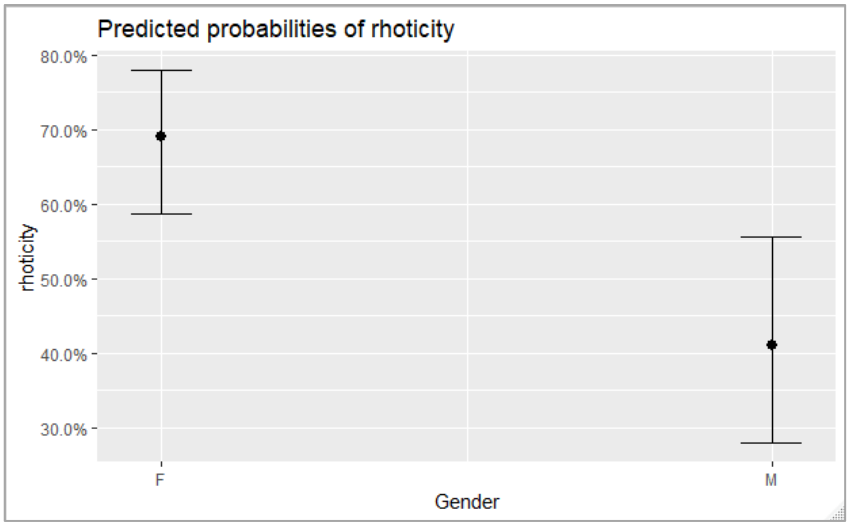


Figure 8.4 Main effects of PRECEDING VOWEL *on rhoticity production at T1*

As concerns the other phonological variables examined, the best-fit model eliminated the effects of SYLLABLE STRUCTURE, SYLLABIC STRESS and WORD TYPE. In terms of SYLLABLE STRUCTURE, single /r/ in word-medial position was disregarded possibly because only one item *afternoon* was included in this category. No significant difference was found between the conditions of non-prevocalic /r/ in consonant

clusters and in word final position. Similarly, SYLLABIC STRESS and WORD TYPE were statistically insignificant in predicting production patterns of English rhoticity by the learners at T1.

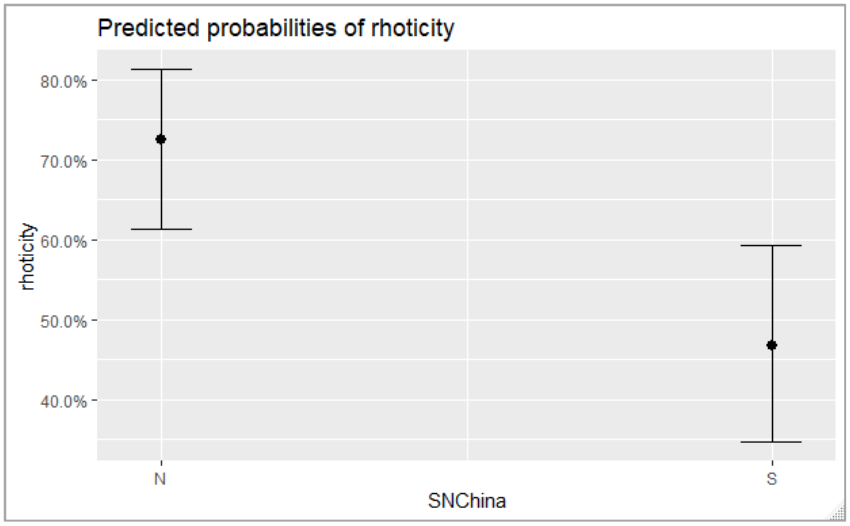
Apart from phonological factors, three external constraints have been suggested to operate on the production of rhoticity by the Chinese learners including GENDER, PLACE OF ORIGIN, and SPEECH STYLE. To begin with, Figure 8.5 illustrates that the presence of nonprevocalic /r/ varied with GENDER in that the female speakers tend to produce rhoticity more frequently than the male learners ($p < .001$).



F: female; M: male

Figure 8.5 Main effects of GENDER on rhoticity production at T1

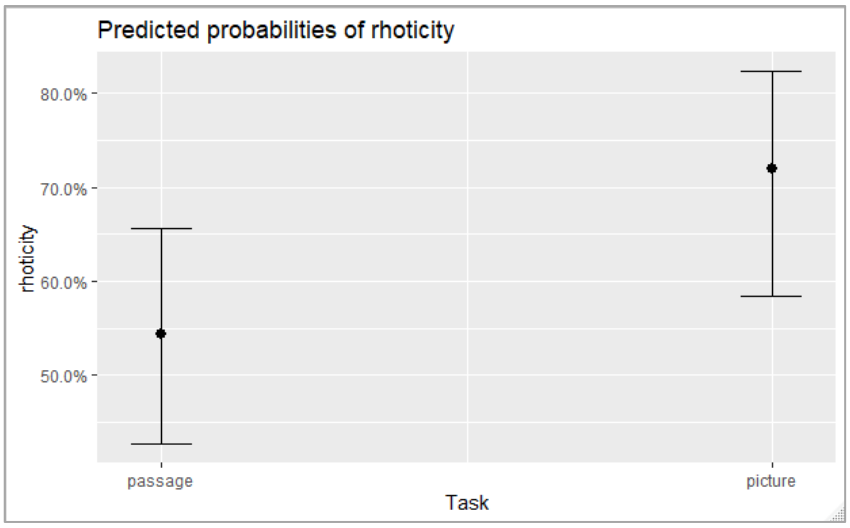
The patterns of its realisation are further conditioned by the speakers' regional backgrounds in China (see Figure 8.6). As expected, the speakers from the nonrhotic South of China strongly disfavour rhoticity compared to those coming from the North, where rhotic Mandarin dialects are spoken ($p < .001$).



N: (rhotic) North China; S: (nonrhotic) South China

Figure 8.6 Main effects of PLACE OF ORIGIN *on rhoticity production at T1*

Moreover, SPEECH STYLE was shown to exert a relatively weak but significant effect (p = .049) in predicting the production of rhoticity. As shown in Figure 8.7, the realisation of nonprevocalic /r/ is promoted by the informal context (picture-naming task) rather than the relatively more formal situation (passage-reading task). Other external factors tested including AGE OF ACQUISITION, TARGET NORM, MEDIA EXPOSURE, and NORM PREFERENCE were not found to be statistically influential in controlling the production of rhoticity by the Chinese learners of English at T1.



passage: (more controlled) passage-reading task;
picture: (less controlled) picture-naming task

Figure 8.7 Main effects of SPEECH STYLE (TASK) on rhoticity production at T1

The best-fit mixed-effects logistic regression model of the presence of rhoticity by the Chinese learners after the six months of acquisition is shown in Table 8.11. PRECEDING VOWEL still exerted a significant impact on the production of nonprevocalic /r/ by the learners. Like T1, the stressed central vowel NURSE is the most favouring environment of rhoticity among all the six preceding contexts examined ($p < .001$), with a preceding back vowel FORCE being the least promoting of the realisation of rhoticity. According to Figure 8.8, the effects of different preceding vowels at T2 ranked following the order of NURSE > NEAR > LETTER > SQUARE > START > FORCE, which was similar as at T1 except for the effects of NEAR and LETTER. At T1, a preceding unstressed central vowel LETTER slightly favoured rhoticity more than a high-front NEAR vowel. Subsequent pairwise comparisons with Bonferroni correction revealed that LETTER and NEAR are not significantly different from each other. Like T1, significant differences were only found between FORCE and NURSE ($p < .0001$), FORCE and LETTER ($p = .024$), and FORCE and NEAR ($p = .033$).

Table 8.11 Best-fit mixed effect regression model of the presence or absence of rhoticity at T2 as a binary dependent variable, with PARTICIPANT *and* WORD *as Random Intercepts*

Fixed effects	**Levels**	**Estimate**	**Std. error**	***z*-value**	***p* <**	
(intercept)		0.264	0.841	0.314	0.753	
PRECEDING	LETTER	1.310	0.415	3.159	0.002	**
VOWEL	NEAR	1.838	0.600	3.063	0.002	**
	NURSE	2.708	0.579	4.667	0.000	***
	SQUARE	0.996	0.516	1.931	0.053	
	START	0.871	0.591	1.474	0.141	
GENDER	male	-1.631	0.425	-3.841	0.000	***
WORD TYPE	function	-0.850	0.411	-2.067	0.039	*
CEA	(continuous)	0.325	0.140	2.328	0.020	*
CEB	(continuous)	-0.445	0.156	-2.850	0.004	**
Random effects	**Type**	**Variance**	**Std. Dev**			
Participant	(Intercept)	1.578	1.256			
Word	(Intercept)	0.650	0.806			
		Min.	**Median**	**Max.**		
Scaled residuals		-5.888	0.232	12.565		

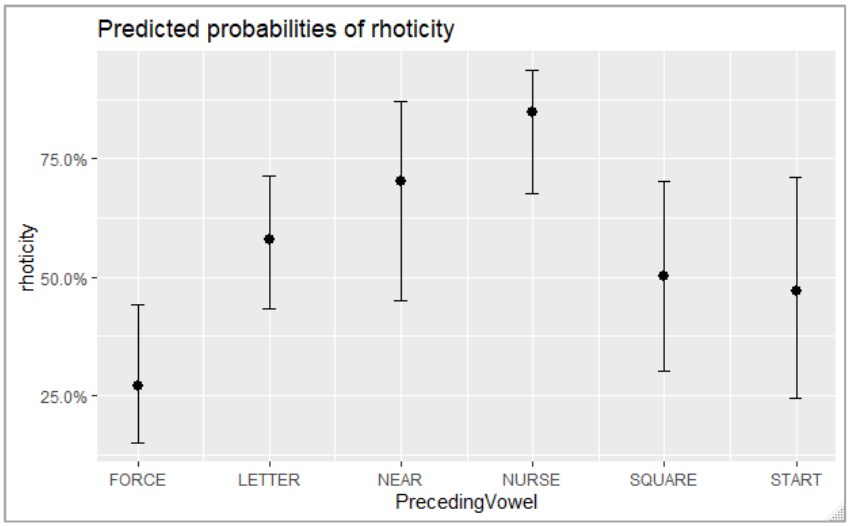
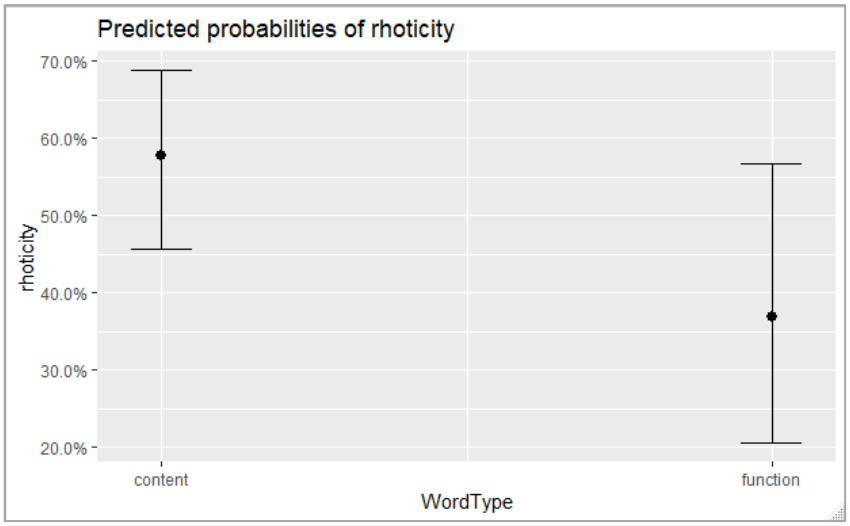


Figure 8.8 Main effects of PRECEDING VOWEL *on rhoticity production at T2*

In addition, a relatively weak but statistically significant impact was achieved by WORD TYPE. According to Figure 8.9, content words favour the presence of nonprevoclic /r/ more than function words (p = .039). Similar as T1, SYLLABLE

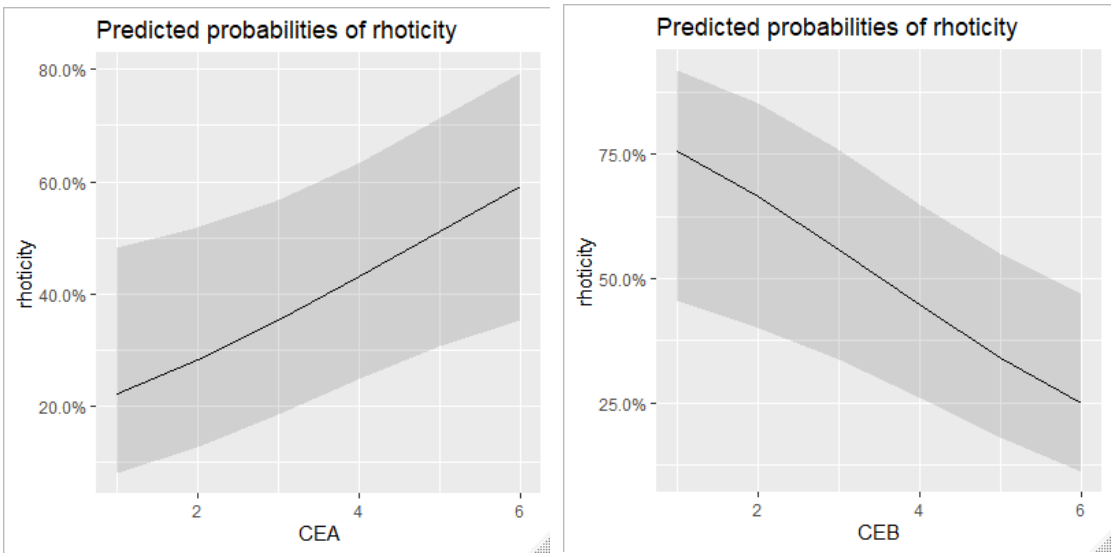
STRUCTURE and SYLLABIC STRESS were eliminated by the best-fit model for T2.



content: content words; function: function words

Figure 8.9 Main effects of WORD TYPE on rhoticity production at T2

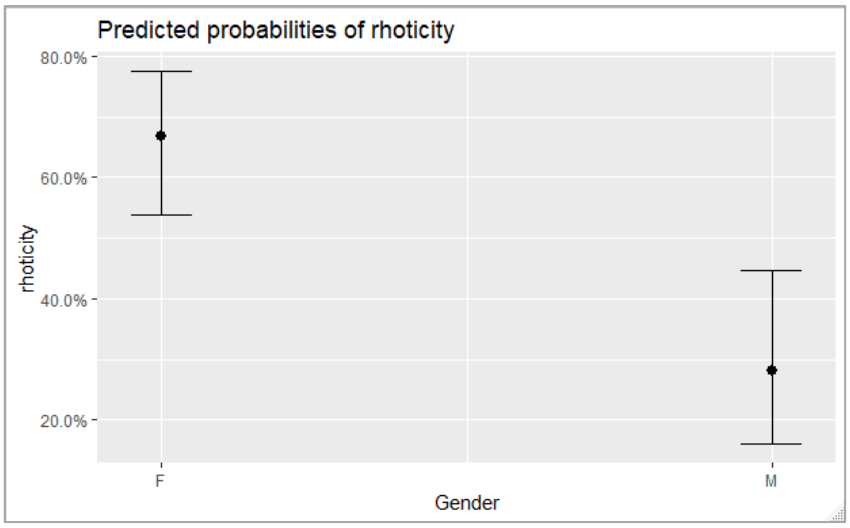
While the main effects of internal variables conditioning the production of rhoticity by the Chinese learners were generally consistent between T1 and T2, the significant external factors found at T2 differed considerably from those found at T1. In contrast to T1, the best-fit model at T2 eliminated the effects of PLACE OF ORIGIN and SPEECH STYLE and promoted speakers' NORM PREFERENCE as a significant predictor. In modelling the NORM PREFERENCE factor group, the average ratings of the four English accents (native American English, native British English, American-oriented Chinese English, British-oriented Chinese English) by individual participants were examined as four continuous variables. Results show that the Chinese learners' realization of rhoticity was significantly conditioned by their attitudes towards American- and British-oriented Chinese English accents. As shown in Figure 8.10, the learners that rated American-oriented Chinese English higher were shown to favour rhoticity (p = .02). Specifically, their amount of rhoticity production increases, as their ratings of the American-oriented accent increases. On the contrary, those who showed more positive attitudes towards the British-oriented Chinese English accent produced a significantly lower amount of nonprevocalic /r/ (p = .004). Their ratings of the native American and British English accents, however, did not reach statistical significance.



CEA: American-oriented Chinese English; CEB: British-oriented Chinese English

Figure 8.10 Main effects of NORM PREFERENCE on rhoticity production at T2

Additionally, GENDER still remains a main effect influencing the presence of rhoticity at T2, like at T1. Figure 8.11 illustrates that the female learners in the present study were shown to be more rhotic than their male counterparts ($p < .001$). Like T1, the external variables AGE OF ACQUISITION, TARGET NORM, and MEDIA EXPOSURE did not show statistically significant effects on the realisation of rhoticity by the Chinese learners at T2.



F: female; M: male

Figure 8.11 Main effects of GENDER *on rhoticity production at T2*

8.4 Discussion of results

In the following, it will be discussed in how far the individual predictions outlined in section 5.2 concerning the production of rhoticity can be confirmed or have to be rejected. Section 8.4.1 deals with the general phonological acquisition process of rhoticity by the Chinese learners of English as predicted by the theoretical frameworks. The interplay of grammar-internal and -external factors modulating the production of nonprevocalic /r/ by the Chinese learners over time is discussed in detail in sections 8.4.2 and 8.4.3.

8.4.1 Development of rhoticity

One general assumption regarding the acquisition process of the production of English rhoticity was that Chinese learners would produce a higher amount of test tokens according to their self-chosen target norms at T2 than at T1, albeit with considerable inter- and intra-learner variability. The results indicate that these expectations are partially met: on average, the learners produced a higher amount of words target-like at T2 compared to T1. However, a paired *t*-test showed that the difference between T1 and T2 did not reach statistical significance. One possible explanation concerns the considerable variability between and within the learners as observed. The striking inter- and intra-speaker variation found in the present study is in parallel with earlier work on the production of rhoticity in EFL varieties such as

Yunnan English (Sundkvist & Gao 2015), Chinese English (Li & Kabak 2017), and in other ESL varieties such as educated Jamaican English (Rosenfelder 2009). This finding also resonates with a number of studies in the field of L2 phonological acquisition, which highlight the high degree of variability between and within individual L2 learners (e.g., Bayley & Tarone 2011).

Furthermore, the lack of improvement from T1 to T2 with respect to the learners' production could be attributable to the state of their self-reported target norms. In the questionnaire, very few learners reported to know the difference between AmE and BrE. Their self-stated target norms were also largely inconsistent with their preferences shown in the norm preference survey. It is therefore argued that the learners were not fully aware of the distinction between different English accents and thus their actual language preference at T1, which is likely to result in the inconsistency between their target norms and actual realisations as observed in the present study. On the other hand, it is also possible that some learners switched their target norms over the time span of six months. Various reasons can possibly contribute to the modification of their individual choices such as media exposure, language contact, and change of language attitude.

As discussed in section 5.2, the CAH (Lado 1957) and the MDH (Eckmann 1977) predicted that it would be relatively easy for rhotic Mandarin speakers to produce rhoticity and for nonrhotic Mandarin speakers to speak nonrhotic English accents. Figure 8.12 illustrates the distribution of rhotic productions by the Chinese learners with different L1 dialectal backgrounds and target norms at both T1 and T2.

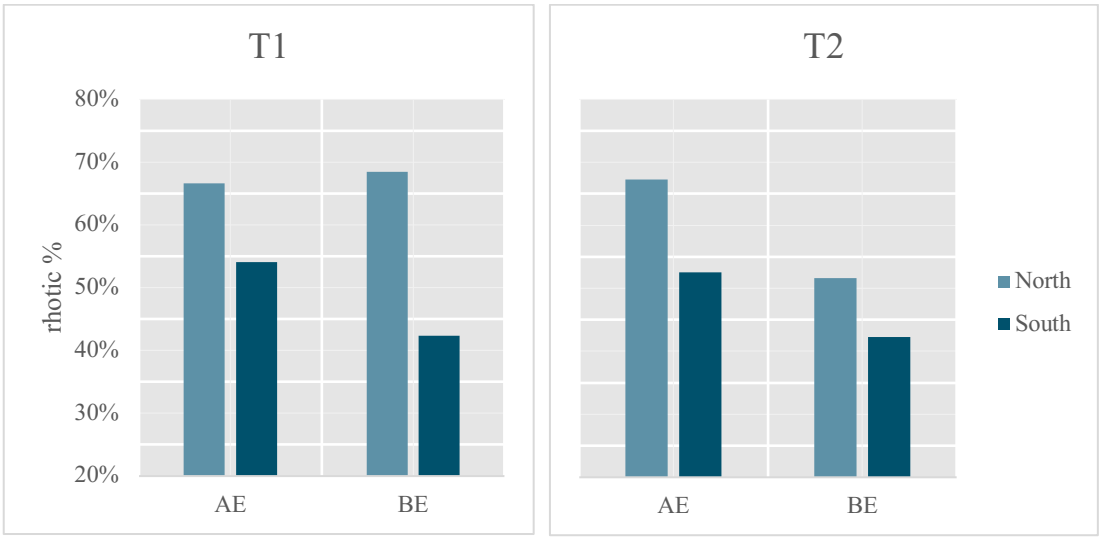


Figure 8.12 Proportions of rhotic realisations by Chinese learners on the basis of place of origin and target norm at T1 and T2

As predicted, the learners from the North of China targeting AmE were the most rhotic and produced nearly 70% of the test tokens with nonprevocalic /r/ at T1 and T2. The learners from the South targeting BrE, on the other hand, were the least rhotic over time as expected. The learners from the North aiming at BrE were shown to be highly rhotic at T1, which became much more target-oriented after six months. The improvement over time suggests that it is relatively difficult but attainable for rhotic Mandarin speakers to acquire nonrhoticity. For the learners from the South targeting AmE, however, the pattern of their production did not improve from T1 to T2, indicating the major difficulty in acquiring English rhoticity by the nonrhotic Mandarin speakers. The acquisition patterns predicted by the CAH and the MDH are therefore supported by the findings of the present study.

In the light of the OPM's (Major 2001) Chronological Corollary, it was assumed that L1 would play a dominant role in conditioning patterns of L2 rhoticity at earlier stages of acquisition. The impact of TL-related aspects would gradually override L1 and exert greater influence on learners' IL over time. This hypothesis is confirmed by the complex interplay of the underlying factors found to condition the production of rhoticity by the learners at T1 and T2. To be more specific, the multilevel analyses revealed that L1 regional background and L1 stylistic transfer were significant in predicting the learners' patterns of rhoticity at T1, the effects of which were however

eliminated at T2. The TL-related factor NORM PREFERENCE took control in affecting their realisation of rhoticity significantly after the six months of learning.

8.4.2 Grammar-internal variables

Preceding vowel

Regarding possible effects of the preceding contexts of nonprevocalic /r/, it was expected that preceding vowels would exert a significant impact on the production of rhoticity. More specifically, front vowel environments would favour rhoticity more than back vowels. This hypothesis is confirmed as the back vowel FORCE was shown to be the least favouring of nonprevocalic /r/ in the speech of the Chinese learners at both T1 and T2. This corroborates findings by Piercy (2012), Dickson and Hall-Lew (2017), and Becker (2014), who also found a preceding FORCE vowel to be the least favouring environment of the presence of rhoticity in ENL varieties including Dorset English, Scottish English and New York City English. This finding, however, diverges from the previous studies on rhoticity in educated Jamaican English (Rosenfelder 2009), Maltese English (Bonnici 2010), and urban Indian English (Chand 2010), where the back vowel FORCE exerts a rather promoting effect on the realisation of rhoticity. In comparison with earlier work on rhoticity in EFL varieties, the present findings contradict Kang's (2013) on Korean English but support those found by Li and Kabak (2017) for Chinese English.

The results further show that a stressed central vowel NURSE is the most promoting for nonprevocalic /r/ by the learners at both data points. This is in line with earlier work on rhoticity in Boston English (Nagy and Irwin 2010), Dorset English (Piercy 2012), and New York City English (Becker 2014), where a preceding NURSE vowel is also reported to be the most favouring context of rhoticity. However, previous studies in the ESL varieties yielded rather conflicting evidence concerning the effect of NURSE such as educated Jamaican English (Rosenfelder 2009), Maltese English (Bonnici 2010), and urban Indian English (Chand 2010). As such, the present pattern concerning the most and least promoting contexts of rhoticity in the speech of the Chinese learners bears more resemblance to the variably rhotic ENL varieties regardless of whether they are gaining or losing rhoticity in the course of sound variation and change. Although the preceding vowel was found to be a significant constraint on rhoticity in ESL varieties, no universal agreement on the ordering of

individual vowel effects can be found between those varieties and the present study. The divergence between ENL and ESL varieties could be possibly explained by the strong effects exerted by various substrate languages and prestigious norms on the production of nonprevocalic /r/ in ESL varieties (e.g., Melchers and Shaw 2003). Its realisation in the ENL varieties, on the other hand, can possibly be interpreted as having a strong articulatory basis, while being less subject to language or dialect contact. This fairly universal pattern concerning preceding contexts found between ENL varieties and the speech of the learners in this study suggests that the production of rhoticity by the Chinese learners is primarily conditioned by the degree of articulatory difficulty.

The main effects of the unstressed central vowel LETTER in predicting the production of rhoticity in the speech of the Chinese learners is also noteworthy. In the present study, it operated as the second favouring context of rhoticity at T1 and ranked at the third at T2. A preceding LETTER vowel was almost never found to be a promoting environment of rhoticity in other varieties of English. Nagy and Irwin (2010), for instance, reported that the LETTER vowel favours rhoticity the least in Boston English. Bonnici (2010) and Chand (2010) also found LETTER to strongly disfavour rhoticity relative to other preceding vowels in Maltese English and urban Indian English. An exception is Li and Kabak (2017), who reported that LETTER attracts the rhotic variant in Chinese English due to the transfer from substrate language Mandarin. As discussed in section 5.1, the diminutive suffix in Mandarin is perceptually and morphophonologically similar to LETTER in English. It is therefore not surprising that the unstressed LETTER vowel favours rhoticity in the speech of the Chinese learners, which can be attributed to possible structural CLI from L1 Mandarin.

The effects of the other preceding vowels on rhoticity rank in the following order: NEAR > SQUARE > START. The high front vowel favours rhoticity more than relatively low and back contexts. This also confirms that a preceding front vowel is more promoting of nonprevocalic /r/ than back vowel environments. For the Chinese learners of English, it seems that high vowels favour rhoticity more than low vowels as well. This pattern, however, is not in parallel with any other English varieties. As proposed by Nagy and Irwin (2010), the effects of preceding contexts on rhoticity

usually do not coincide with any phonological feature of vowels. I argue that the production of rhoticity by L2 learners is mostly constrained by articulatory difficulty. Taking the realisation of START as an example, its articulation of a low vowel plus nonprevocalic /r/ sequence involves greater tongue configurations than for preceding mid or high vowels (ie. SQUARE and NEAR), which could possibly lead to the higher degree of articulatory effort and therefore a greater probability of reduction in the speech of the Chinese learners.

In section 5.2, predictions regarding the effects of preceding vowel were made according to the CAH, the MDH, the SDRH, and the SLM based on the crosslinguistic similarity hierarchy of the English /V(r)/ sequences for Chinese learners. In general, the present findings confirm the predictions of the CAH and the MDH, which propose that TL structures that already exist in learners' L1, namely NURSE and LETTER, would be easier for Chinese learners to produce. Accordingly, the predictions of the SLM and the SDRH, which claim that similar structures are more difficult for Chinese learners to acquire than dissimilar ones, have to be rejected.

Syllable structure

As regards syllable structure, it was assumed that the word-final single coda position favours rhoticity more than it in consonant clusters considering the fact that Mandarin does not allow structure like consonant clusters at all. However, this hypothesis has to be rejected as the learners in the present study did not differ very much in producing rhoticity in the two syllable types. Nonprevocalic /r/ occurred slightly more often in consonant clusters (63.6%) than in word-final single coda position (54%). This stands in contrast to rhoticity in world Englishes as well such as in Dorset English (Piercy 2012), Boston English (Nagy and Irwin 2010), educated Jamaican English (Rosenfelder 2009) and Korean English (Kang 2013), where nonprevocalic /r/ was found to favour word-final coda positions rather than pre-consonantal contexts.

Meanwhile, the learners were shown to strongly disfavour single /r/ in word-medial position relative to word-final position. This parallels with Scobbie (2006), who found that word-final nonprevocalic /r/ is realised more often than word-internal /r/ in Scottish English. Nevertheless, future research is needed to investigate possible

effects of word-medial single /r/ on the presence of rhoticity, concerning that *afternoon* is the only test word in this group in the present study. The nonprevocalic /r/ in *afternoon* is often deleted by the learners possibly because of the high lexical frequency of this word. As proposed by Piercy (2012), Nagy and Irwin (2010), and Rosenfelder (2009), rare words favour rhoticity more strongly than frequent words in Boston English, Dorset English, and educated Jamaican English. It is thus unclear whether the high proportion of deletion of the nonprevocalic /r/ in *afternoon* is due to its lexical frequency or syllable structure.

Word type

The prediction regarding word type stated that content words favour the presence of nonprevocalic /r/ more than function words. The findings for T2 confirm this hypothesis in that content words significantly favour rhoticity while function words do not. This can be accounted for by the lexical frequency and syllabic stress that are associated with word type. Words are always highly correlated with respect to the three dimensions. For instance, function words are lexically more frequent and unstressed relative to content words. Likewise, stressed words and infrequent words are more likely to be content words relative to function words. Both SYLLABIC STRESS and LEXICAL FREQUENCY have been repeatedly investigated and shown to predict patterns of rhoticity in world Englishes, and it has been demonstrated that nonprevocalic /r/ disfavours rhoticity in unstressed syllables and in more frequent words (e.g., Piercy 2012, Dickson & Hall-Lew 2017). As such, nonprevocalic /r/ in function words are more likely to be deleted by English speakers and learners as also found in the present study.

Stress

For the phonological predictor SYLLABIC STRESS, no statistically significant effects were found in the regression model. It was assumed that stressed words favour rhoticity more strongly than unstressed words according to patterns of rhoticity in native English speech. This hypothesis has to be rejected as SYLLABIC STRESS was eliminated by the best-fit regression model at both T1 and T2. This finding contrasts with earlier work such as Piercy (2012), Chand (2010) and Boccini (2009), who reported main effects of stress in predicting patterns of the realisation of

nonprevocalic /r/ in different English varieties. A possible explanation could be the impact of the preceding LETTER vowel for Chinese learners. As discussed earlier, the unstressed central vowel LETTER is a favouring context of rhoticity in the English speech of Chinese learners, which diverges from other English varieties primarily due to CLI from their L1. As concerns possible effects from the L1, Chinese learners were predicted to be insensitive to word stress, which does not form part of the Mandarin phonological grammar. It should also be noted that for the coding of SYLLABIC STRESS, all of the LETTER vowels belong to the unstressed category. The distinct behaviour of LETTER stands in contrast to other unstressed tokens such as function words, which usually disfavour rhoticity. As a result, the variable SYLLABIC STRESS failed to predict patterns of rhoticity due to the internal conflict between the promoting effects of LETTER tokens and the disfavouring behaviour of other unstressed ones.

8.4.3 Grammar-external variables

Gender

As discussed in chapter 5, the social meaning borne by the retroflex suffix in Mandarin was predicted to exert impact on the realisation of English rhoticity by Chinese learners. The results show that the females in the present study produced a significantly higher amount of nonprevocalic /r/ than the male learners at both T1 and T2. One possible explanation of this finding involves the meaning and function of the diminutive suffix in Mandarin, which has been described as a "smooth operator" in the Beijing vernacular speech by Zhang (2008). It conveys a sense of "endearing, casualness and intimacy", which is considered more of a feminine characteristic. Thus, it was not surprising that the female learners in this study produced rhoticity more frequently in English than the male learners regardless of their places of origin, target norms or covert norm preferences.

As it has been well established in the field of sociolinguistics that females are more likely to produce the prestigious /r/ variants than male speakers (e.g., Chand 2010, Hartmann & Zerbian 2009), it was also predicted that female learners would be inclined to follow their target norm than male learners when producing rhoticity. As shown in Figure 8.13, the female speakers targeting AmE produced more test tokens with the presence of rhoticity than those targeting BrE at T1. The difference between the female learners targeting AmE and BrE became greater after six months,

indicating the tendency of convergence towards their target norms. The male Chinese learners, on the other hand, did not follow their target norms in general. At both T1 and T2, the males targeting BrE produced a slightly higher amount of rhoticity than those targeting AmE, with a smaller difference at T2 than at T1.

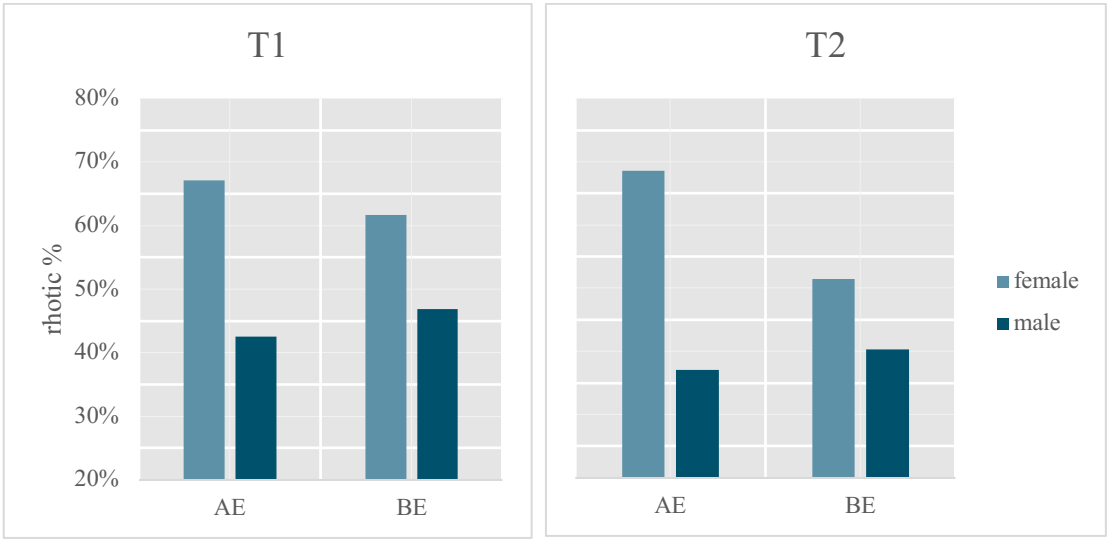


Figure 8.13 Proportions of rhotic realisations by Chinese learners on the basis of GENDER and TARGET NORM at T1 and T2

This finding accords with previous studies on rhoticity in other ESL and EFL varieties including Black South African English (Hartmann & Zerbian 2009), Singapore English (Tan & Gupta 1992), and Korean English (Kang 2013), where female learners were always more status-conscious than their male counterparts and were more likely to produce the prestigious or standard variants of nonprevocalic /r/. Male speakers, on the other hand, tended to retain the more traditional realisations, in these cases, nonrhoticity. For the same reason, Chand (2010) reported that female speakers of urban Indian English produced more nonrhoticity than their male counterparts, as the current prestigious norm in India is BrE.

L1 dialectal background

At T1, the dialectal backgrounds of the learners were found to be a significant predictor for the production of nonprevocalic /r/, the effects of which however failed to reach statistical significance at T2. Speakers from the North of China have more exposure to rhotic Mandarin dialects and produce the retroflex suffix in their L1 to

various extents on a daily basis, they were therefore predicted to be more rhotic when speaking English than those coming from the South of China, where the diminutive suffix does not exist in the dialects there. This supports Melchers and Shaw's (2003) proposal that postcolonial English speakers with substrate languages that allow rhoticity tend to favour the realisation of nonprevocalic /r/ in English. The finding is also in parallel with earlier work on the effects of L1 dialectal variation on L2 speech production (O'Brien et al. 2010).

However, the learners in the present study were shown to be less influenced by their linguistic background after the training and the six months of acquisition since no effects of their places of origin were predicted by the best-fit regression model. No significant difference with respect to the realisation of rhoticity between the Chinese learners from the South and the North of China was reached at T2. The observed pattern at T2 clusters with the effects of L1 region on rhoticity in Chinese English found by Li and Kabak (2017), who reported that Chinese teachers of English showed more convergence to their target norms rather than being influenced by L1 structures because they were highly proficient English users. Similarly, the present study also showed that NORM PREFERENCE outweighs L1 PLACE OF ORIGIN and exerted greater influence on the learners' production of nonprevocalic /r/ at later stages of L2 phonological acquisition, which is in line with the predictions of the OPM (Major 2001).

Speech formality

Two possible patterns regarding the effects of speech styles were proposed in section 5.2. On the one hand, the formality levels of tasks may interact with TARGET NORM in conditioning the production of rhoticity by Chinese learners of English. To be more specific, it was predicted that learners targeting rhotic English accents would produce more nonprevocalic /r/ in more monitered speech style and less in informal speech. Conversely, learners targeting nonrhotic accents were predicted to realise more rhoticity in casual speech other than in the formal situation. However, this hypothesis has to be rejected since no significant interaction between SPEECH STYLE and TARGET NORM was found. As shown in Figure 8.14, the Chinese learners realized a higher number of nonprevocalic /r/s in the informal picture-naming task than in the formal

passage-reading task at both T1 and T2 and independent of their norm choices. In terms of the picture naming task, all learners were highly rhotic at T1. The learners targeting BrE became much less rhotic in picture naming at T2, while those targeting AmE still retained a high level of rhotic production. As for passage reading, more rhoticity was realised by the learners targeting AmE than those targeting BrE at both T1 and T2.

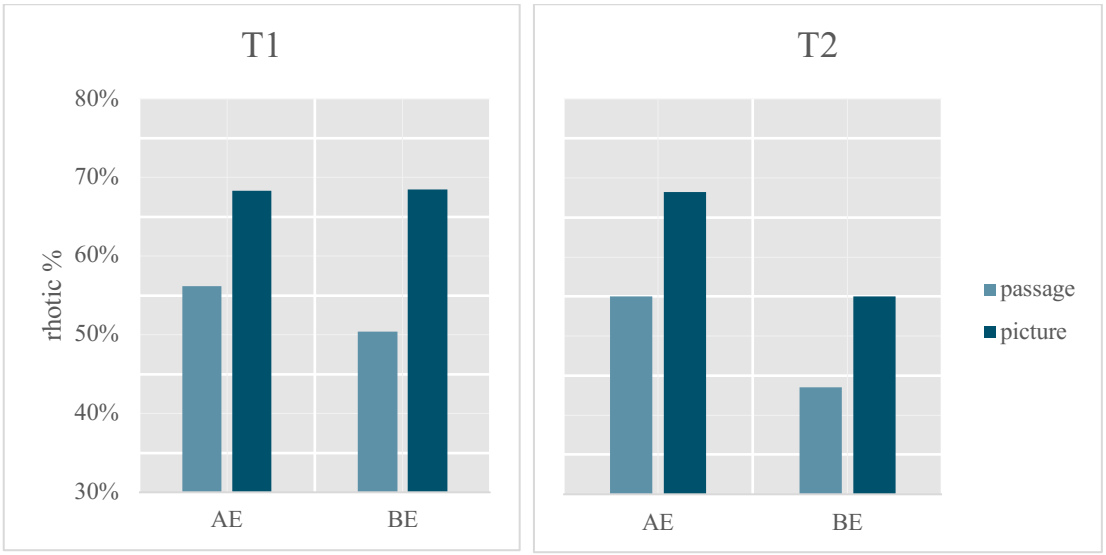


Figure 8.14 Proportions of rhotic realisations by Chinese learners on the basis of TASK *and* TARGET NORM *at T1 and T2*

On the other hand, Chinese learners of English were assumed to be more rhotic in casual speech than in formal situations due to possible stylistic transfer from L1 Mandarin regardless of their target norms. This is confirmed as the learners in the present study were found to be consistently more rhotic in picture naming than in passage reading at both data points. The difference between the proportions of rhotic realisations was more considerable at T1 than T2, which was in parallel with the effects of SPEECH STYLE revealed in the multilevel analyses. It was shown that speech formality exerted a statistically significant influence on the production of rhoticity at T1, but not after six months of further acquisition. This again supports the OPM (Major 2001) in that the effects of stylistic CLI from Maindarin was more profound at the onset of rhoticity acquisition yet became increasingly subtle over time.

It would also be interesting to investigate the possible orthographic impact on the

realisation of rhoticity by L2 learners since nonprevocalic /r/ in English is associated with orthography. In the present study, the picture naming task did not involve any orthographic input of rhoticity. The passage reading task, on the other hand, incorporated words in written form and therefore provided participants with orthographic indication of the presence of nonprevocalic /r/. The results show that the learners were significantly more rhotic in the picture naming task than in the passage reading task, suggesting no orthographic influence on the production of rhoticity by the Chinese learners. This can be accounted for by CLI from Mandarin as well. As illustrated in section 5.1, the realisation of the diminutive suffix in Mandarin does not necessarily require the presence of its written form. Rhotic Mandarin speakers simply produce the retroflex suffix habitually in casual speech regardless of orthography. However, this tentative finding should be treated with caution. Although the two tasks in the present study differ in terms of the presentation of orthography, they did not rule out other potential impacts such as lexical frequency and level of speech formality. In the future, more comparable tasks with identical test tokens should be designed for the investigation of orthography in order to control for possible influence from other constraints.

Norm preference

It was predicted that L2 learners with more positive attitude towards rhotic English accents would realise a higher amount of nonprevocalic /r/ in their English speech. This is confirmed by the results of the best-fit regression model at T2: the learners' attitude towards British and American English accents produced by highly proficient Chinese news reporters was shown to be significant in predicting their production patterns. More specifically, the learners who rated the American-oriented Chinese English accent higher realised more test tokens with nonprevocalic /r/, while those who showed more positive attitudes towards the British-oriented Chinese English accent produced a significantly less amount of rhoticity. A correlation between the self-reported target norms of the learners and their attitude towards rhotic and nonrhotic accents produced by both Chinese English news reporters was observed. As shown in the descriptive statistics, the learners aiming at AmE rated the rhotic Chinese English accent higher than those targeting BrE. Conversely, the learners that reported BrE to be their target norm generally showed more positive attitudes towards

the British-oriented Chinese English accent than those targeting AmE.

No effects, however, were observed for the learners' preference of the native American and British accents. An overall low score was received by the female native American English news reporter compared to the others. One possible reason for this is the voice quality of this speaker, which was characterized by a particularly fast pace and low pitch. This arguably diverges from the "sweet and gentle voice" that people in Asia would usually expect to hear from a pleasant female character. Starr (2015), for instance, suggests that "sweet voice" is associated with the traditional notion of Japanese femininity and voice quality operates as an authentic marker in the formation of linguistic style. Accordingly, I argue that it is the voice quality of the native American speaker that accounts for the overall dispreference to the native American accent by the Chinese learners in the present study. The native British English accent also did not interact significantly with the realisation of nonprevocalic /r/, since all of the learners rated it relatively high independent of their individual target norms.

At T1, one of the accent preference ratings showed main effects in predicting learners' production of rhoticity. In the early stages of phonological acquisition, the learners seem to be more influenced by CLI from L1 Mandarin. In other words, L1-related constraints such as stylistic patterns and L1 regional background outweighed the L2-related factors and exerted a significant impact on the learners' production of the target feature in earlier stages. After the pronunciation training and six further months of acquisition, L2-related predictors such as NORM PREFERENCE gradually overrode CLI in predicting the patterns of rhoticity produced by the Chinese learners of English. This interplay of constraints on the development of rhoticity by the learners over the course of time confirms the interrelation and relative influence of L1 and TL on IL as predicted by the Chronological Corollary of the OPM (Major 2010).

8.5 Chapter summary

In sum, the production of English rhoticity by the Chinese learners in the present study is conditioned by both grammar-internal and grammar-external predictors. Their rhoticity rates are characterized by significant inter- and intra-learner variabilities between the two data points, due to a complex interaction of various internal and

external variables over time.

The most pronounced grammar-internal variable was found to be PRECEDING VOWEL. The patterns of rhoticity regarding this aspect partially parallel with native English varieties in terms of the favouring effect of NURSE and the disfavouring effect of FORCE, indicating a possible universal influence of articulatory difficulty associated with the preceding contexts that rhoticity occurs in. The phonology of the learners' L1 Mandarin also plays a role concerning the effects of LETTER, which was shown to be a promoting context of rhoticity for the learners due to its phonological resemblance to Chinese rhoticity structure and the fact that word stress does not form part of the Mandarin phonological grammar.

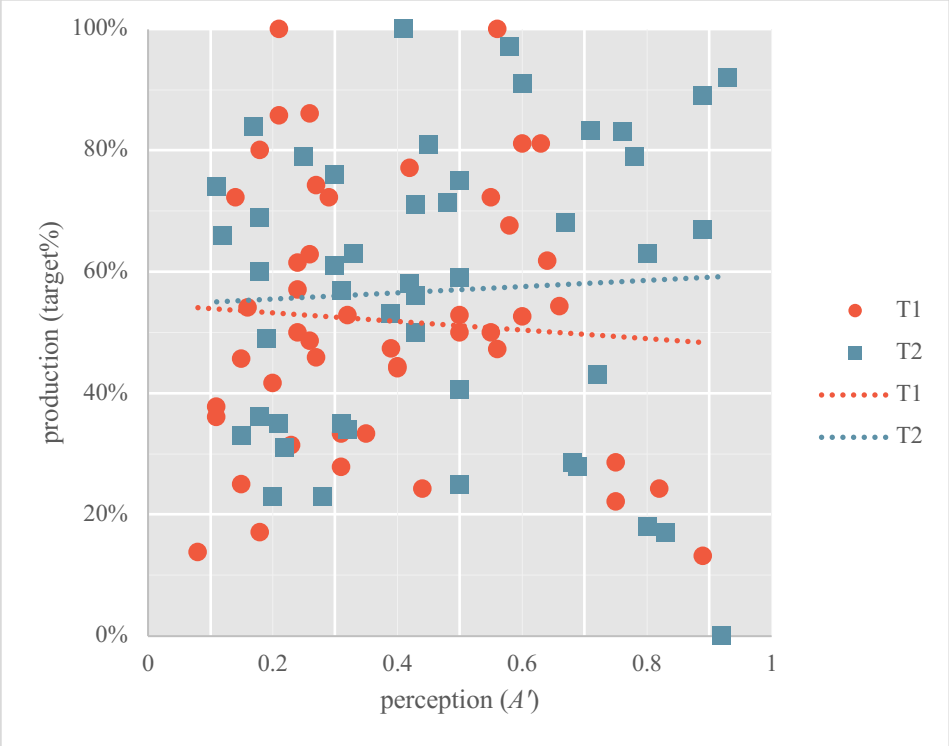
In addition, grammar-external factors also exerted significant impact on the production of rhoticity by Chinese learners. Interestingly, the weights and magnitude of L1- and TL-related factors underwent change over the course of phonological acquisition. To be more specific, the patterns of rhoticity were significantly controlled by L1 dialectal variation and L1 stylistic variation at initial stages of rhoticity acquisition, which were further overridden by their target norm preferences at T2. The interplay of L1- and TL-related external predictors at T1 and T2 further underlines the acquisition process predicted by the Chronological Corollary of the OPM (Major 2001).

9. The perception-production link

This chapter is concerned with the perception-production link in the acquisition of rhoticity by Chinese learners of English and how it changes over time. Section 9.1 illustrates the overall relationship between the learners' perception and production performance from a bird's eye perspective. Section 9.2 turns to individual learners divided into four learner groups on the basis of their L1 backgrounds and target norms. Following a group comparison in terms of correlation patterns and relative difficulty of rhoticity realisation in section 9.3, section 9.4 zooms into the perception-production link in the six phonetic contexts. Results are further discussed in Section 9.5 and Section 9.6 closes with a chapter summary.

9.1 Overall relationship between perception and production

In order to identify the underlying nature of the relationship between the perception and production of rhoticity in L2 soicophonological acquisition, the performance of the 47 Chinese learners of English in the present study before and after the six months of pronunciation training is illustrated in Figure 9.1. Generally speaking, no strong correlation can be observed between the two domains at either data point. With respect to T1, there seems to be a tendency of a very minor negative correlation between the two. At T2, on the other hand, a rather slightly positive link between perception and production was found, which indicates a marginal change of the relationship over time. A Pearson's correlation coefficient was computed to assess the perception-production link. Results show that there was no significant correlation between the two variables at either T1 ($r = 0.068$, $p = 0.64$) or T2 ($r = 0.052$. $p = 0.72$). Overall, the realisation of rhoticity in the speech of the L2 learners does not seem to be closely associated with their discrimination ability of the presence and absence of English nonprevocalic /r/.



target% = percentages of absent nonprevocalic /r/ for learners targeting BrE and percentages of present nonprevocalic /r/ for those targeting AmE

Figure 9.1 Relationship between the perception (A') and production (target%) of rhoticity by the 47 Chinese learners at T1 and T2

Furthermore, the learners' development in their perception and production performance from T1 to T2 is shown in Figure 9.2. There seems to be a slightly positive tendency indicating that the learners who improved more in perception accuracy also improved more in production. However, results of a Pearson's correlation showed that this correlation did not reach statistical significance ($r = 0.164$, $p = 0.27$).

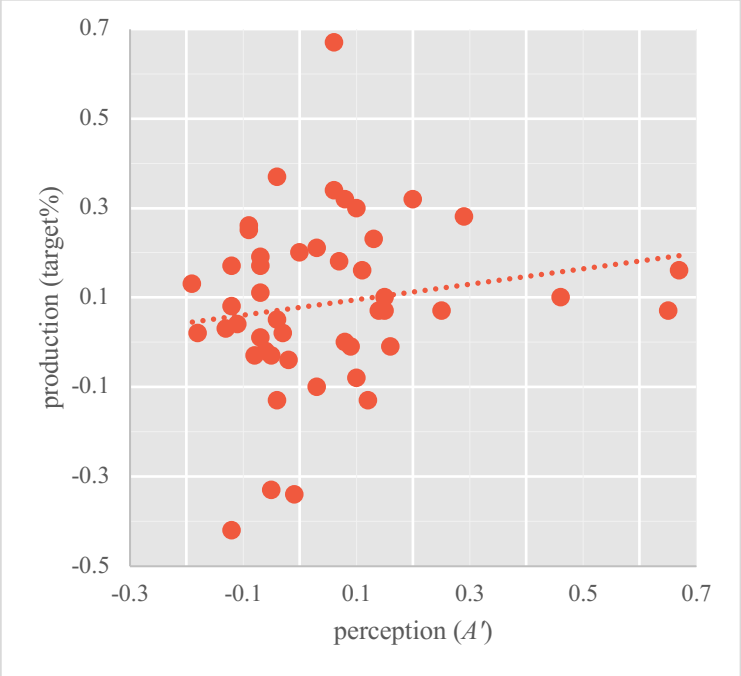


Figure 9.2 Relationship between the development in perception and production of rhoticity by the 47 Chinese learners from T1 to T2

Despite the lack of correlation, it is still unclear how much these findings could inform the perception-production link concerning the potential variability conditioned by different sets of structural and social factors. As indicated in the previous chapters, the perception and production of English rhoticity by the Chinese learners in the present study vary across individuals over the course of time. Moreover, both grammar-internal and -external factors were shown to modulate the two domains. Therefore, the perception-production mapping of rhoticity is further examined from the individual, group, and contextual perspectives in the followings.

9.2 Zooming into individual learners in four groups

Although no clear link was observed with respect to the overall relationship between the perception and production of rhoticity for all learners, the individual variability and potential group patterns of correlation are worth investigating. Therefore, the perception and production performance of each learner at both data collection times are presented at the individual level and at the group level in the following subsections. The 47 learners were classified into four groups according to their regional backgrounds in China and self-stated target norms (see Table 9.1): learners from the rhotic regions in China targeting AmE (the R-AmE group), learners from the

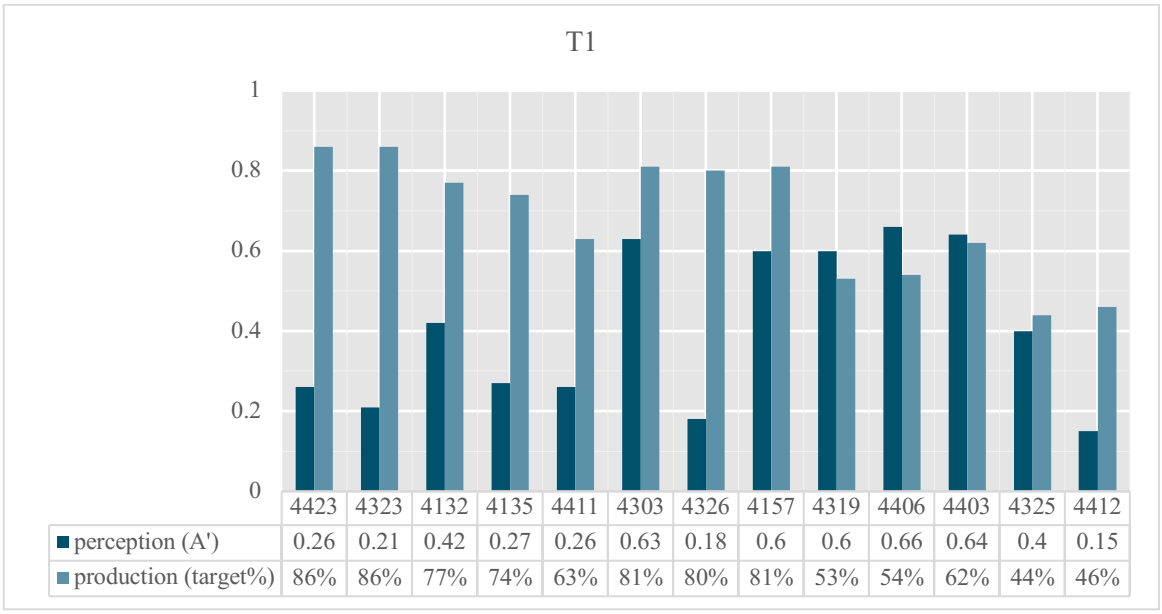
nonrhotic regions targeting BrE (the NR-BrE group), learners from the rhotic regions targeting BrE (the R-BrE group), and learners from nonrhotic regions targeting AmE (the NR-AmE group). As described in chapter 5, the retroflex diminutive suffix in Mandarin is predominantly used in the North of China and does not exist in most Southern Chinese dialects except for Chengdu and Chongqing, where the suffix is used when referring to specific proper names. Thus, the four learners from Chengdu and Chongqing in the present study were categorized into the R-AmE and the R-BrE groups according to their respective target norms, although they originate from the largely nonrhotic South of China.

Table 9.1 Four learner groups classified on the basis of their L1 regions and L2 targets

Learner group	L1 Mandarin	TL English	No. of learners
R-AmE	rhotic	American English	13
NR-BrE	nonrhotic	British English	9
R-BrE	rhotic	British English	14
NR-AmE	nonrhotic	American English	11

9.2.1 The R-AmE group

Figure 9.3 presents the discrimination ability and the realisation of rhoticity in English achieved by the 13 learners of the R-AmE group. At T1, five learners were shown to be perceptually sensitive to English rhoticity (participants 4303, 4157, 4319, 4406, and 4403), who also produced more than half of the test tokens with the presence of nonprevocalic /r/ following their target norm. It seems that the learners who are perceptually sensitive to English rhoticity are also more likely to follow their target norm in production. Learners 4325 and 4412, on the other hand, only produced 44% and 46% of the tokens rhotic and were also found to be insensitive to English rhoticity in perception. The remaining six learners successfully followed their norm in production, however, failed to discriminate English rhoticity. It can be concluded that accurate discrimination can facilitate accurate production of the target /r/ variant. Nevertheless, accurate perception is not the only variable conditioning the production of English rhoticity, since half of the learners in the R-AmE group were target-like in production independent of their discrimination ability.



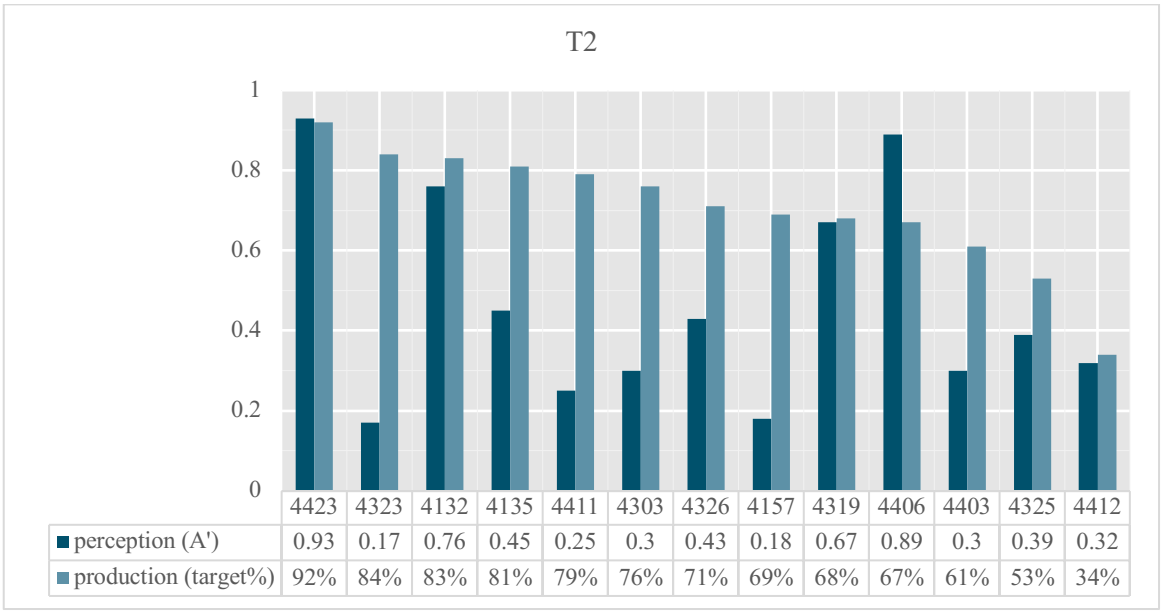


Figure 9.3 Discrimination ability and proportions of target production of English rhoticity across the 13 learners of the R-AmE group at T1 and T2

In terms of T2, twelve out of the 13 participants successfully produced more than half of the tokens rhotic, with only four of them being perceptually sensitive. The only participant that failed to achieve target-like production was learner 4412, who was also perceptually insensitive to English rhoticity at both T1 and T2. The patterns for the R-AmE group revealed that it is not difficult for Chinese learners from a rhotic dialectal background to articulate English nonprevocalic /r/ irrespective of their perceptual sensitivity. It is also worth noting that although most of the L2 learners in

this group showed a relatively high degree of rhoticity in their speech, no one was found to be categorically rhotic. The highest rate of rhoticity was achieved by learner 4423 from Chengdu (92%), who also had the best discrimination ability among all participants (0.93).

9.2.2 The NR-BrE group

Figure 9.4 illustrates the perception and production of (non)rhoticity by the Chinese learners from a nonrhotic L1 background targeting BrE. With regard to T1, participant 4131 was shown to be the only learner in this group that was perceptually sensitive to English rhoticity, and who was categorically nonrhotic in speaking English as well, illustrating a potential perception-production correspondence. The other learner found to not produce nonprevocalic /r/ categorically was participant 4413, who, on the other hand, displayed a nearly complete insensitivity in discriminating English rhotic-nonrhotic contrasts. The distinct behaviours of these two learners revealed that accurate perception is not a must for accurate production, especially when L1 and TL structures do not differ.

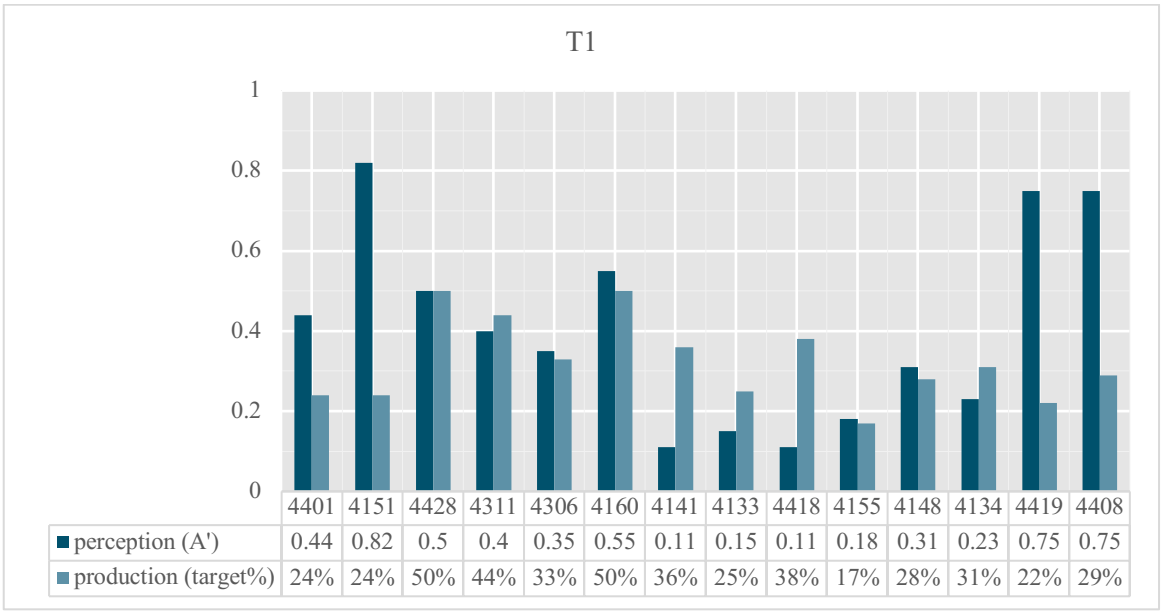
Figure 9.4 Discrimination ability and proportions of target production of English rhoticity across the 9 learners of the NR-BrE group at T1 and T2

Concerning T2, seven out of the nine participants successfully followed their BrE norm in producing more than half of the English tokens with the absence of nonprevocalic /r/. However, only two of them were shown to be sensitive in perceiving English rhoticity, which further suggests that the production of (non)rhoticity is not necessarily constrained by the perceptual discriminability for the learners speaking a nonrhotic L1 Mandarin dialect. Participants 4301 and 4150, who were not target-like in production, were found to be perceptually insensitive to rhoticity in English as well. It is noteworthy that, in contrast to the R-AmE group, it is possible for the speakers from the South of China to achieve categorical nonrhoticity

such as participants 4413 at both T1 and T2. The results revealed that it was less difficult for the learners from the nonrhotic L1 backgrounds in China to acquire nonrhoticity in English in comparison with the learners of the R-AmE group.

9.2.3 The R-BrE group

Figure 9.5 shows the performance in perceiving and producing English rhoticity by the Chinese learners from the rhotic L1 regions targeting BrE. At T1, it was particularly difficult for the rhotic Mandarin speakers to suppress rhoticity in speaking English although they targeted a nonrhotic L2 norm - no learners produced more than half of the test tokens without nonprevocalic /r/. Four learners were perceptually sensitive to discriminating English nonprevocalic /r/ (participants 4151, 4160, 4419, and 4408), indicating a tendency of perception preceding production for the R-BrE learners at the onset of rhoticity acquisition. In general, no learners were shown to be accurate in both perception and production of English (non)rhoticity at T1.



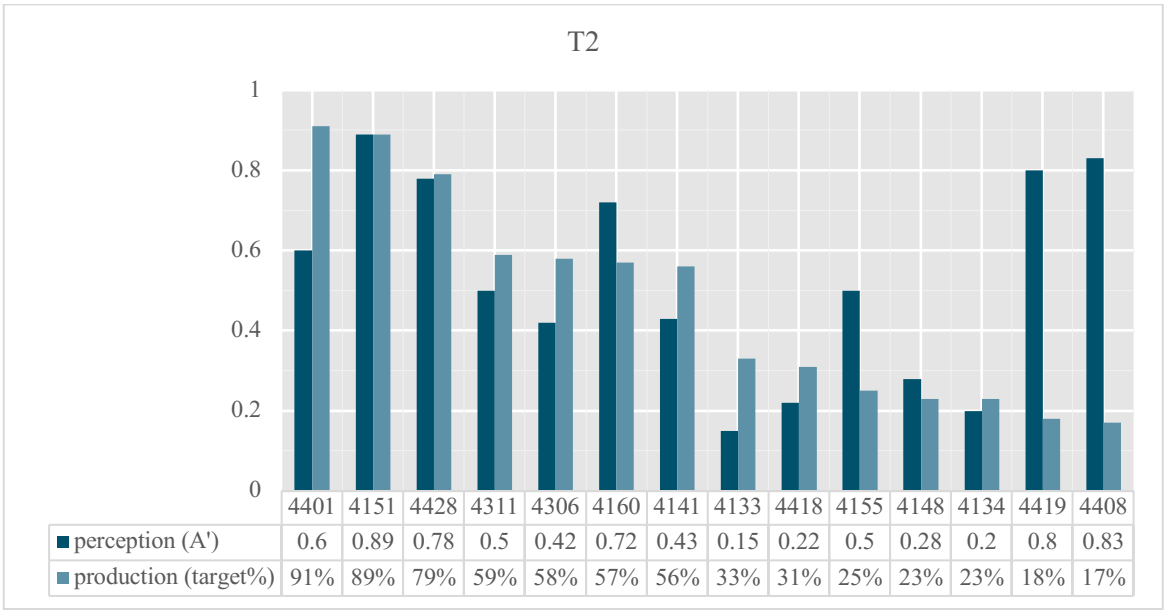


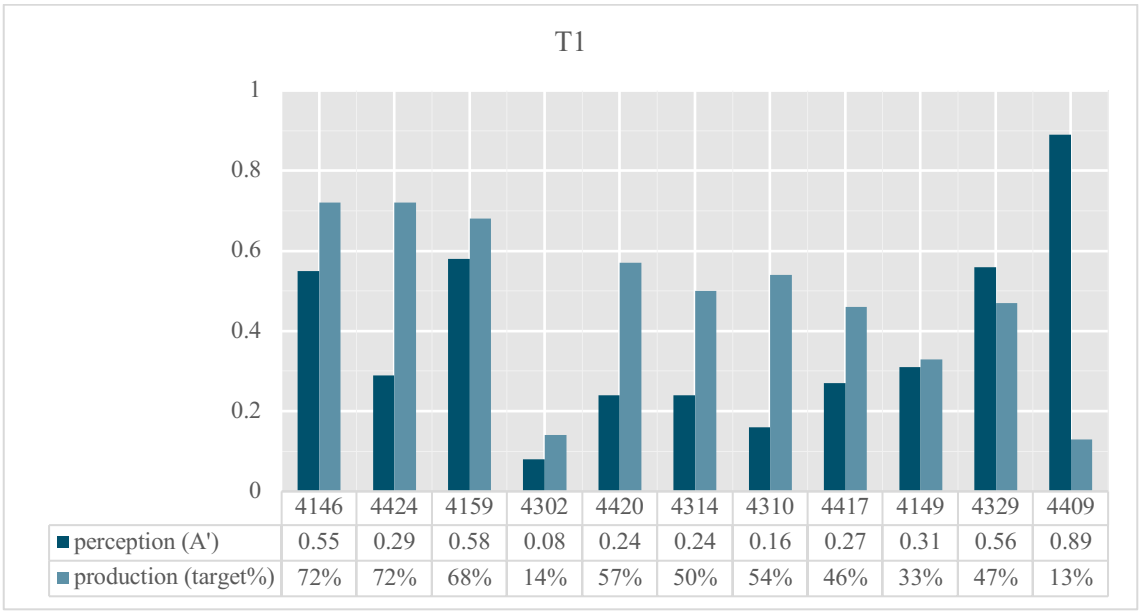
Figure 9.5 Discrimination ability and proportions of target production of English rhoticity across the 14 learners of the R-BrE group at T1 and T2

After the six months of phonological training, half of the (seven out of 14) learners in the R-BrE group were found to be target-like in producing more than half of the English tokens nonrhotic, most of whom (five out of seven) were perceptually sensitive to English rhoticity as well. Thus, there seems to be a positive link between the perception and production of (non)rhoticity by these five successful learners in this group. However, participants 4419 and 4408, who were very sensitive in distinguishing English rhoticity, failed to suppress the production of nonprevocalic /r/

when producing English speech, indicating a relative difficulty for the learners in the R-BrE group to acquire the production of English nonrhoticity. The remaining five learners were neither perceptually sensitive to English rhoticity nor capable of producing consistent nonrhoticity at T2.

9.2.4 The NR-AmE group

The perception and production of rhoticity by the learners of the NR-AmE group are presented in Figure 9.6. At T1, four learners were found to be perceptually sensitive, two of whom (participants 4146 and 4159) also realised more than half of the tokens rhotic. Learners 4329 and 4490, by contrast, were good at discriminating English rhoticity, but did not accurately produce the feature following their target norm. Participant 4302 showed a complete insensitivity to English rhoticity with an *A'* score of 0.08 and produced only 14% of English /r/ tokens with the presence of rhoticity. Similarly, learners 4314, 4310, and 4417 were neither sensitive in perception nor target-like in production.



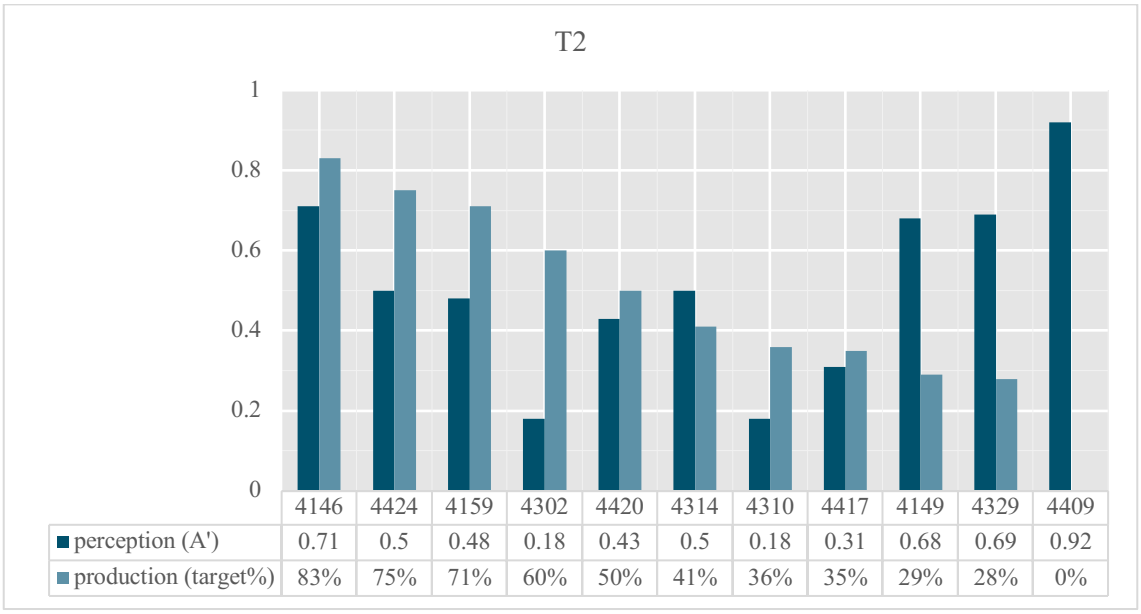


Figure 9.6 Discrimination ability and proportions of target production of English rhoticity across the 11 learners of the NR-AmE group at T1 and T2

Concerning T2, four learners (4146, 4424, 4159, 4302) successfully followed their target norm in production. Only one of them (learner 4146) was perceptually sensitive to rhoticity. Learners 4149, 4329, and 4409 were able to accurately distinguish the presence or absence of nonprevocalic /r/ in English. However, coming from nonrhotic regions in China, it was still particularly challenging for them to produce the rhotic /r/ variant, especially for participant 4409, who had very high *A'* scores (0.89 at T1 and 0.92 at T2) but produced none of the test tokens target-like at T2. Comparing with the

other groups, it seems to be most difficult for the learners speaking a nonrhotic L1 to produce rhoticity in their L2, regardless of their discrimination ability.

9.3 Group comparison and relative difficulty of production

In this section, the four groups of learners were compared in terms of their perception and production performance at both data collection times. Whether a learner is classified as perceptually sensitive to English rhoticity and/or target-like in production is determined based on the following criteria in Table 9.2.

Table 9.2 Criteria for the classification of discrimination and production performance

	Performance	**Criteria**
Perception	sensitive	$A' > 0.5$
	insensitive	$A' =< 0.5$
Production	target-like	target% > 50%
	non-target-like	target% =< 50%

The performance of the Chinese learners in the present study were further classified into four performance categories based on a combination of their *A'* scores and percentages of target production following the criteria described in Table 9.2: *neither, both, only production*, and *only perception*.

- *neither*: insensitive in perception and non-target-like in production
- *both*: sensitive in perception and target-like in production
- *only production*: insensitive in perception but target-like in production
- *only perception*: sensitive in perception but non-target-like in production

The number of participants of each performance category in the four learner groups at T1 and T2 are illustrated in Figure 9.7. Among the 47 participants, 27 at T1 and 23 at T2 performed consistently in discriminating and producing English rhoticity. That is to say, they were either both sensitive in perception and target-like in production, or inaccurate in both. These learners could be assumed to show a positive link between the perception and production of rhoticity. To be more specific, some of the learners that were perceptually sensitive to English rhoticity were also successful in producing their target /r/ variants. Conversely, those that were inaccurate in realising rhoticity tended to show insensitivity in perception as well. At T1, 19 participants were found in the *neither* group, which decreased to twelve at T2. Only eight learners were

accurate in both perception and production at T1, which increased to eleven after the six months of learning, showing a slight improvement in their overall performance over time.

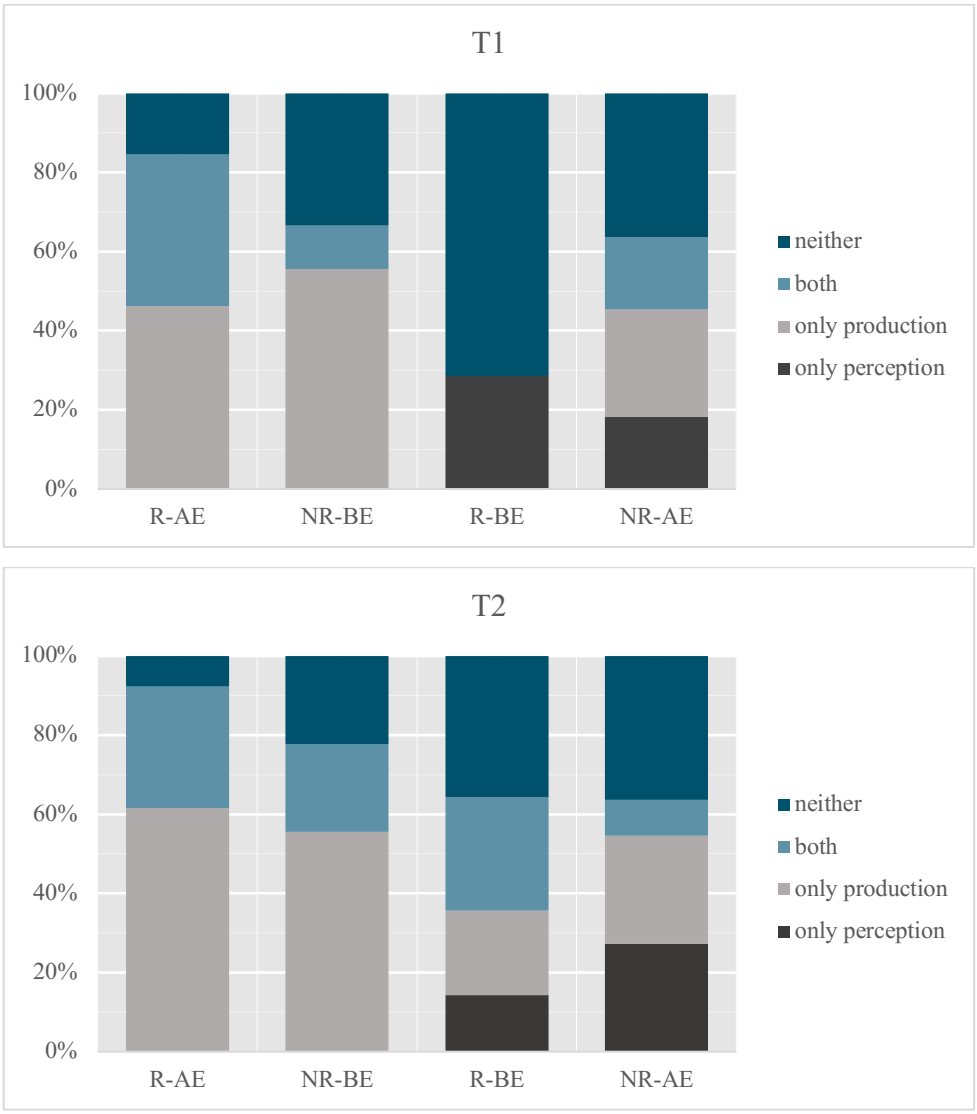


Figure 9.7 Perception and production performance of the learners in the four groups at T1 and T2

In contrast to the learners exhibiting a link between the perception and production of rhoticity, a rather negative correlation was observed for the rest of the learners, who were either found to be sensitive in perception or target-like in production, but not in both. More specifically, 20 Chinese participants showed inconsistent performance in their perception and production at T1, which increased to 24 after six months of phonological training. Six participants at T1 and five at T2 were only capable of

discriminating English coda /r/ but failed to produce their target /r/ variants, showing a pattern of perception preceding production. It is worth noting that all of them belonged to the R-BrE and the NR-AmE learner groups. By contrast, no one in the R-AmE and the NR-BrE learner groups was found to be part of the *only perception* category at both T1 and T2.

In comparison with the relatively small number of learners exhibiting perception preceding production, more learners showed a tendency of production without perception. More specifically, 14 learners at T1 and 19 learners at T2 were accurate in production instead of perception. Most of them belong to the R-AmE and NR-BrE learner groups, who targeted the /r/ variants in the same category occurring in their respective L1 Mandarin dialects. In particular, no one in the R-BrE group showed to be target-like in production at T1. The most remarkable changes over time could also be identified with this group. At T1, only two performance categories were observed, namely *neither* and *only perception*. More diverse perception-production patterns were identified for this group after the six months of pronunciation training.

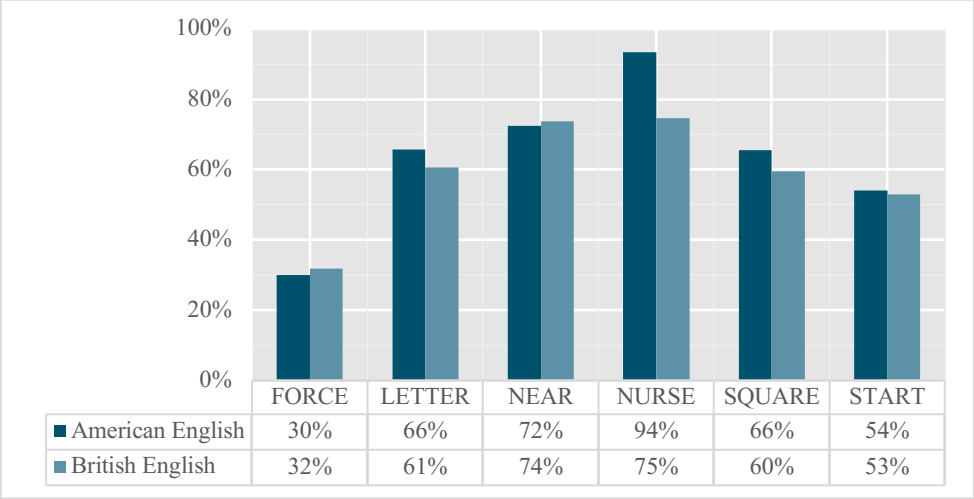
The relative difficulty of the production of English rhoticity in the four learner groups can be observed in Figure 9.7 as well. Concerning the R-AmE group, where the learners spoke rhotic Mandarin and targeted American English, the production of rhoticity in English was very easy to be acquired. Only two participants at T1 and one participant at T2 failed to produce more than half of the English tokens with the presence of nonprevocalic /r/. Similarly, it was also relatively easy for the nonrhotic Mandarin learners of the NR-BrE group to retain nonrhoticity in their English speech, with only three participants at T1 and two at T2 not achieving their British English target norm in producing more tokens with nonprevocalic /r/. Furthermore, two participants exhibited categorical nonrhoticity in this group.

In the R-BrE group and the NR-AmE group, on the other hand, much higher levels of difficulty in following their respective target norms were observed. In particular, no learners in the R-BrE group successfully suppressed the realisation of nonprevocalic /r/ to follow their nonrhotic British target at T1. After six months of phonological training, half of them produced more than half of the English tokens nonrhotic,

suggesting that it was not particularly difficult for the rhotic L1 speakers to acquire nonrhoticity in the L2 English, although no learner managed to do so before the training course. With regard to the NR-AmE group, five out of eleven learners were accurate in production at T1, which, however, decreased to four at T2. Therefore, no improvements can be observed regarding the production of rhoticity from T1 to T2 for this group of learners, indicating the difficulty for nonrhotic Mandarin learners to acquire English rhoticity.

9.4 Zooming into various phonetic contexts

As shown above, the relationship between the perception and production of English rhoticity varies distinctly at both individual and group level over time. Many factors are responsible for explaining the variability in the results. As discussed in chapters 7 and 8, context-dependent variation was observed in both domains. The preceding vocalic context in which nonprevocalic /r/ occurs is a significant phonological factor affecting both the perception and production patterns of English rhoticity by the Chinese learners. Therefore, the vocalic influence is considered an integral part of explaining the perception-production link. Figure 9.9 zooms into each preceding vowel environment and demonstrates the acquisition of rhoticity by the learners across the six phonetic contexts at both data points. The proportions of rhotic realisations and the discrimination scores of each context are illustrated. Here, the rhotic production is presented instead of the target realisation because the learners' targeting different norms did not have a statistically significant effect on their production performance especially for T1 (see Figure 9.8).



	FORCE	LETTER	NEAR	NURSE	SQUARE	START
American English	30%	66%	72%	94%	66%	54%
British English	32%	61%	74%	75%	60%	53%

Figure 9.8 Proportions of the production of rhoticity in various phonetic contexts by the learners with different target norms at T1

As shown in Figure 9.9, apparent inconsistency between perception and production was found in the context of NURSE. At T1, 84.6% of the NURSE tokens were realised with the presence of nonprevocalic /r/, suggesting that a stressed central vowel is a particularly favourable context for rhotic production. However, the learners rather showed a complete insensitivity to distinguishing rhoticity in this context, with an average *A'* score amounting to merely 0.21. This finding remained similar for T2, with a high proportion of rhotic production (73%) and a low discrimination score (0.26).

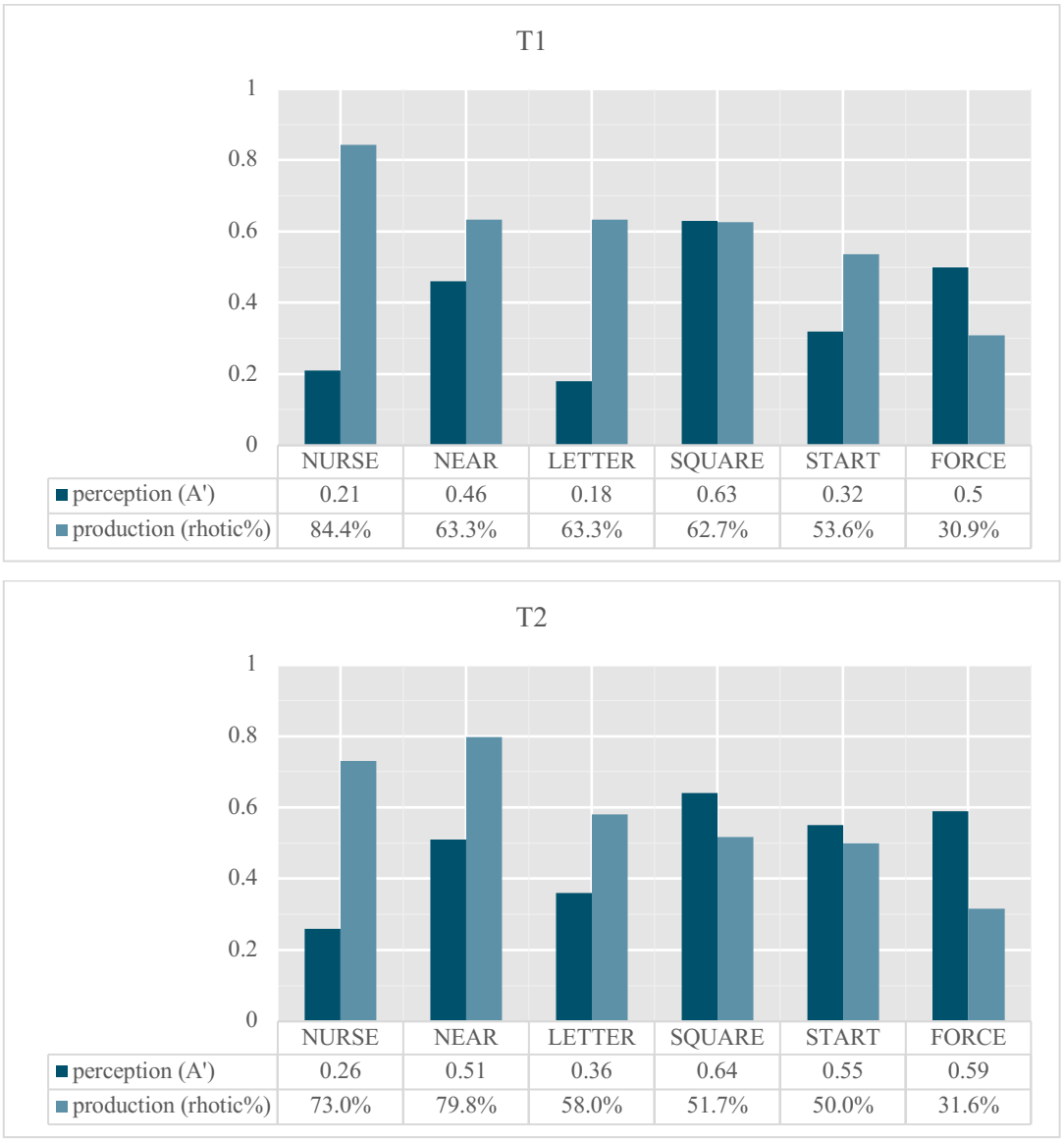


Figure 9.9 Discrimination ability and rhoticity production in the six phonetic contexts at T1 and T2

The mean performance concerning the perception and production of rhoticity in the context of NURSE poses a challenge to modelling the perception-production link. The individual perception and production performance for NURSE by the 47 learners at both data points is further illustrated in Figure 9.10. Most of the learners clustered together along the accurate production and inaccurate perception area, supporting the view of production without perception for this context.

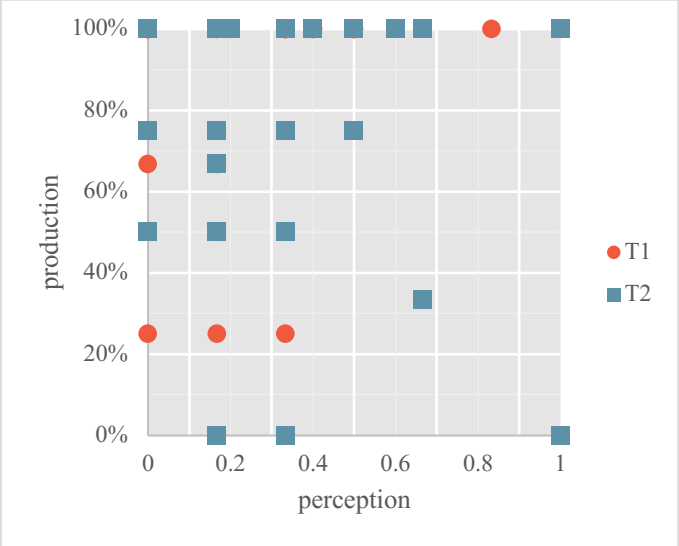


Figure 9.10 Perception and production of NURSE *at T1 and T2*

Similarly, rhoticity in the context of an unstressed central vowel also showed inconsistency between the perception and production by the Chinese learners. LETTER was found to promote the realisation of rhoticity, with 63.3% and 58% of tokens being produced at T1 and T2, respectively. However, the discrimination of rhoticity in this context was shown to be particularly challenging for the learners, with the *A'* scores amounting to only 0.18 at T1 and 0.36 at T2. Like the context of NURSE, the learners in the present study had enormous difficulties in discriminating the rhotic-nonrhotic contrasts in the context of LETTER despite relatively high rhotic production.

SQUARE was shown to be the only context supporting a perception-production link at both T1 and T2. The learners displayed similarly accurate discrimination ability (0.63 at T1 and 0.64 at T2) and rhotic productions (62.7% at T1 and 51.7% at T2) of nonprevocalic /r/ in this context.

In terms of NEAR, it can be observed that the learners were perceptually sensitive to rhoticity in this context at T2, who was also found to promote the realisation of nonprevocalic /r/ with about 80% of the NEAR tokens being realised. However, the evidence in favour of a perception-production link was not found for NEAR at T1, since the learners exhibited insensitivity to English rhoticity in terms of perception.

Concerning the context of START, the Chinese learners were better at producing

nonprevocalic /r/ than discriminating it at the onset of rhoticity acquisition. After the six months of pronunciation training, they became perceptually more sensitive to rhoticity in this context (0.55) and produced half of the START tokens rhotic.

The context of FORCE, in contrast, offered evidence of perception without production regarding English rhoticity for the Chinese learners in the present study. Regarding perception, it was relatively easy for the learners to perceive the difference between the rhotic and nonrhotic realisations of FORCE with the *A'* scores amounting to 0.5 at T1 and 0.59 at T2. However, only around 30% of the FORCE tokens were realised with the presence of nonprevocalic /r/ at both times of data collection.

9.5 Discussion of results

The objective of this chapter was to explore the relationship between the perception and production of English rhoticity by Chinese learners and how it changes over time. From a bird's eye perspective, no direct and consistent correlation between the two domains can be observed at either data collection time. To be more specific, production can be accurate with no discrimination ability, and accurate perception does not necessarily correlate with target-like production. It is possible for an L2 learner of English to produce rhoticity in a highly target-like manner, regardless of whether the learner is capable of distinguishing a rhotic-nonrhotic English contrast or not. Therefore, the possible perception-production link was further examined on the individual, group, and contextual levels.

In terms of individual learners, distinct patterns that both confirm and challenge the assumption of a perception-production link were found. A few L2 learners showed both accurate perception and production of English nonprevocalic /r/, while the others were incompetent in both, providing evidence supporting a correlation between the two. Conversely, some learners' performance seem to challenge the concept of a perception-production link: it was found that the discrimination sensitivity of some learners increases as their production ability decreases from T1 to T2, and vice versa, which provides further evidence that there is no uniform patterns between the perception and production of rhoticity. However, it is yet not legitimate to jump to a conclusion that there is no relationship between the two domains. It was found that accurate discrimination tends to facilitate the accurate production of the target /r/

variants in some cases. Among the learners that were perceptually sensitive to rhoticity, only six of them at T1 and five at T2 did not realise their target production. Target-like production of rhoticity, in contrast, does not predict perceptual sensitivity, as 14 learners at T1 and 19 learners at T2 were able to produce more than half of the English tokens following their target norms without perceiving the contrast accurately. All in all, the correlation patterns between the production and perception of rhoticity vary across individual learners, which is not surprising as the high degree of individual variability is considered one of the most salient features of L2 speech acquisition. Apart from factors such as language attitude and amount of English usage, the Chinese learners in the present study also differ in terms of places of origin in China and target norms. Therefore, the perception-production link within four learner groups was investigated.

The present findings show that the perception-production patterns for each learner group are even more complex. In general, the learners acquiring variant pattern of rhoticity that has no counterpart in their L1 Mandarin dialect are more likely to display a tendency of perception without production, while those targeting a structure that exists in their L1 tend to show production without perception. Possible explanations for these findings involve the relative acquisition difficulty in terms of CLI and markedness. As predicted by the CAH (Lado 1957) and the MDH (Eckman 1977), acquiring a different and implicationally more marked L2 structure would be difficult, while the acquisition of the same and a less marked L2 category would be relatively easy. Accordingly, it was predicted that it would be the most challenging for nonrhotic L1 speakers to acquire the production of rhoticity. This hypothesis is confirmed by the present study. It was particularly difficult for the learners from the South of China to acquire English rhoticity. With respect to the R-AmE and the NR-BrE group, the production of their target /r/ variants are relatively easy, which constitutes further support for the CAH and the MDH.

The group findings also partially confirm Major and Kim's (1996) SDRH, which postulates that dissimilar target language structures are acquired at faster rates than similar ones, with markedness operating as a mediating factor that slows down acquisition rate. In the present study, it was found that while at initial stages major

difficulties were presented for both rhotic Mandarin speakers acquiring nonrhotic English and nonrhotic Mandarin speakers targeting rhotic English, the first group acquired their target production of rhoticity faster than the learners of the latter group, indicating that rhoticity might be a more marked structures than nonrhoticity, which tends to slow the rate of acquisition relative to less marked structures. Moreover, little improvement over time was found for the learners targeting a more marked rhotic variant.

In terms of context-dependent variation, furthermore, the relationship between the perception and production of English rhoticity was found to vary distinctly across the six preceding vowel environments. More specifically, it was found that the contexts of LETTER and NURSE contradicted the assumption of a perception-production link in providing evidence of production before perception. The environment of FORCE, conversely, yields evidence of perception without production. The contexts of SQUARE and NEAR, by contrast, were shown to be in favour of a perception-production correlation in presenting similar discrimination and production performances. The context of START revealed a rather inconsistent performance in terms of perception and production of rhoticity of the learners over the course of time, who were found to be better at producing nonprevocalic /r/ than perceiving it at T1, but more accurate in perception than production at T2.

The diverging context-dependent patterns observed can be accounted for by the extent of articulatory difficulty, crosslinguistic dissimilarity, and acoustic salience of each preceding vocalic environment that nonprevocalic /r/ occurs in. To begin with, the contexts of NURSE and LETTER provide evidence of production without perception. It is relatively easy for speakers to produce rhoticity in these contexts. As discussed in section 8.4.2, the promoting effect of a preceding NURSE vowel on the production of rhoticity has been confirmed in the present study as well as by a number of sociolinguistic work on rhoticity in ENL varieties (e.g., Irwin & Nagy 2010, Piercy 2012). The context of LETTER also favours the production of rhoticity by Chinese learners because of its structural resemblance with the retroflex suffix in Mandarin Chinese (Li & Kabak 2017). The perception of rhoticity in NURSE and LETTER vowels, on the other hand, is rather difficult due to its crosslinguistic similarity to the

Mandarin retroflex suffix, as predicted by the SLM (Flege 1995). Moreover, considering the acoustic cues that are indexed to different contextual information, NURSE and LETTER are acoustically less salient relative to the other phonetic contexts due to their shorter duration and less extent of formant modulation. Thus, the relationship between the perception and production of rhoticity in NURSE and LETTER can be explained by the interplay of the acoustic salience, crosslinguistic dissimilarity, and articulatory complexity that are inherent in terms of phonetics and phonology.

The context of FORCE, by contrast, offers evidence of perception before production, which can also be explained by the operation of the articulatory difficulty and acoustic salience of this context, and its crosslinguistic dissimilarity to Mandarin. In terms of production, a preceding back vowel has consistently been found to disfavour rhoticity relative to front vowel environments in world Englishes such as Scottish English (Dickson and Hall-Lew 2017) and New York City English (Becker 2014), which could be explained by the articulatory difficulty involved. This is borne out as well in the present study and for proficient Chinese English speakers (Li & Kabak 2017). Concerning the acoustic salience of this sequence, it is characterised by dramatic movements of both F2 and F3 as well as a relatively long duration. This context was therefore shown to be perceptually easier for the Chinese learners to perceive, but rather difficult for production. It seems to be the persistent difficulty in producing the /r/ variants in the context of FORCE that constrains a possible promoting impact of its perception. In other words, the articulatory difficulty of this sequence exerted a greater impact on its production than its accurate perception by the learners. In terms of possible crosslinguistic impact from L1, the FORCE vowel does not exist in Mandarin. Also, the process of rhoticisation in the rhotic Mandarin usually yields an *r*-coloured vowel instead of a sequence of vowel plus a following /r/ as in American English. Thus, FORCE appears to be a dissimilar structure for Chinese learners, which, according to the CAH (Lado 1957), would be more difficult for the learners to acquire.

The contexts of SQUARE and NEAR did not show a large discrepancy between perception and production. Being front vowel environments, they are more promoting of the realisation of the following nonprevocalic /r/ than FORCE but less than NURSE,

which is in line with earlier sociolinguistic work on rhoticity in Boston English (Nagy and Irwin 2010) and Dorset English (Piercy 2012). Concerning perception, SQUARE and NEAR are acoustically more salient than central vowels due to their longer duration and more dramatic formant modulation. It is therefore not surprising that the Chinese learners in the present study were accurate in both perception and production of rhoticity in these contexts at T2.

9.6 Chapter summary

To summarise, the present findings underline that the relationship between the perception and production of English rhoticity in L2 speech learning is a particularly complex phenomenon that can only be explained from a multi-faceted perspective. Evidence both supporting and contradicting the perception-production relationship predicted by Flege's SLM (1995) has been observed. The correlation patterns differ distinctly across individual speakers, learner groups, and phonetic contexts. Indeed, no existing L2 acquisition models tested in the present study in isolation can satisfactorily account for such complex variations. But the patterns observed are not entirely random either. The context-dependent patterns can be explained by a variety of phonological and phonetic factors such as crosslinguistic dissimilarity, acoustic salience, and articulatory complexity, whereas the group patterns are affected by sociolinguistic factors including L1 dialectal regions and target norm variation. The perception-production mismatch is not surprising concerning the fact that perception and production constitute rather separate mechanisms. As revealed by the multivariate analyses of the learners' perception and production performance in chapters 7 and 8, the perception of rhoticity is only conditioned by the preceding vowel environments in which nonprevocalic /r/ occurs, indicating that it is a process that mostly requires listeners to adjust their perceptual expectation in light of the vocalic influence. However, their production is affected by both phonetic/phonological environments and external factors including L1 region, speech style, target norm preference, and gender. Moreover, the weights and magnitude of the factors change over time, with L1-related factors outperforming L2 in initial stages and L2-related factors gaining significance in later stages. Perception is only one of the many variables that can possibly affect production. The perception-production mapping as predicted in the SLM is therefore not always consistently observed in the present study.

10. General discussion and outlook

In this chapter, the results presented and discussed in chapters 7, 8, and 9 are first reviewed in the light of the three research questions posed in section 5.2. To that end, section 10.1 discusses the effects of CLI from Mandarin on the acquisition of English rhoticity by Chinese learners. Section 10.2 deals with the relationship between the perception and production of rhoticity by the learners and how it changes over time under the control of both grammar-internal and -external variables. Section 10.3 is concerned with the predictors and patterns shared by native and non-native English rhoticity. In section 10.4, the implication of the present findings for models of L2 phonological acquisition is further discussed and a combined framework that aims to predict and explain the developmental patterns of the perception and production of an L2 sociophonological feature is proposed. Section 10.5 closes this chapter with concluding remarks and gives an outlook on future research.

10.1 Types of CLI from L1 Mandarin

The first research question of the present study asked how the effects of the L1 operate on the perception and production of nonprevocalic /r/ in English over the course of L2 speech learning. In the review of the theoretical models describing the impact of CLI on SLA in chapter 2, several gaps and unresolved questions had been identified. Specifically, it remained unclear whether the dissimilarities or the similarities between learners' L1 and TL systems posed major challenges to the acquisition of an L2 phonology. Furthermore, current theories failed to predict possible transfer opportunities of the sociolinguistic aspects that are inherent in the L1 and TL of learners. In addition, the effects of context-dependent CLI beyond the segmental level are largely absent in the theories reviewed in chapter 2. The present study attempted to address these gaps by testing the effects of L1 Mandarin on L2 English rhoticity. As presented in chapter 4, nonprevocalic /r/ in English is a sociophonological variable undergoing diachronic and synchronic changes and is subject to a multitude of grammar-internal and –external factors. Interestingly, in Mandarin, the diminutive retroflex suffix is also a salient variable that is indexical of dialectal, stylistic, and social variation. By testing Chinese learners acquiring English rhoticity over time, the questions uncovered can be thoroughly investigated in the study at hand.

As demonstrated empirically, the Chinese learners displayed the following influence from Mandarin on their L2 English speech: To begin with, L1 context-dependent transfer with regard to the preceding phonological/phonetic contexts that nonprevocalic /r/ occurs in was observed for both perception and production, albeit with distinctive patterns. To illustrate, the English NURSE and LETTER vowels that bear more phonetic resemblance to the Mandarin retroflex suffix [ɚ] were promoting of the presence of rhoticity in the English speech produced by the Chinese learners. Dissimilar TL rhoticity sequences such as /r/ following NEAR, FORCE, and SQUARE, by contrast, did not favour the production of rhoticity at both T1 and T2. This finding also confirms that, in terms of the preceding vowel contexts of nonprevocalic /r/, similar TL categories are easier for learners to produce than dissimilar TL aspects. It reveals that it is the dissimilarities between the L1 and the TL systems that lead to major acquisition difficulties for L2 learners acquiring the production of English rhoticity, as predicted by the CAH (Lado 1957) and the MDH (Eckman 1977). Predictions of the SLM (Flege 1995) and the SDRH (Major & Kim 1996) that recognise similar TL categories as being more challenging for learners cannot be confirmed for the acquisition of rhoticity production. The perception of English rhoticity, on the other hand, exhibited a completely different pattern at both data points. Specifically, the Chinese learners were more sensitive and accurate in perceiving rhoticity in the dissimilar contexts than in the similar ones, which supports the predictions of the SLM (Flege 1995).

The other L1 phonological aspect that has shown impact on the production of rhoticity by the Chinese learners is word stress, which does not form part of Mandarin phonological grammar. In native English varieties, stressed syllables were found to favour rhoticity more than unstressed syllables (e.g., Piercy 2012). The Chinese learners in the present study, however, did not differ in their rhoticity rates between stressed and unstressed syllables at both data collection times. For example, both NURSE and LETTER vowels were promoting of rhoticity in the speech of the Chinese learners, despite their differences in terms of word stress, i.e. the NURSE vowel always occurring in stressed and the LETTER vowel always occurring in unstressed syllables in English. As Mandarin does not permit lexical stress, Chinese learners can be assumed to be insensitive to word stress when producing English rhoticity due to CLI

from Mandarin phonology. This finding further corroborates Li and Kabak's (2017), who reported that the effects of stress did not reach statistical significance in determining the production of rhoticity by highly proficient Chinese English speakers.

Moreover, gender was found to condition the patterns of rhoticity by the Chinese learners of English at both data collection times. Female learners produced more rhoticity than males. A possible explanation for this finding is the social meaning of the suffix [ɚ] in Mandarin. As mentioned in chapter 5, the Mandarin diminutive suffix carries an implication of endearment and fondness and functions as a "smooth operator" (Zhang 2008: 201) in the casual speech, which is assumed to be more closely related to feminine attributes. Thus, this might prompt Chinese female learners to favour English rhoticity more than their male counterparts. However, it should be noted that viewing the Mandarin [ɚ] as a gender-biased feature comes from my observation and intuition as a linguistically-trained native Mandarin speaker. The social aspects of Mandarin rhoticity in different dialect regions of China are yet severely understudied. Therefore, future research focussing on the sociolinguistic variation of the diminutive suffix in Mandarin and speakers' language attitude towards this stylistic feature should be conducted to offer further empirical evidence for the intuitive connection between Mandarin rhoticity and gender.

While the effects of phonological CLI and gender-specific social CLI from Mandarin are significant at both data collection times, the relevance of L1 stylistic impact differs between T1 and T2. In the present study, the Chinese learners produced more rhotic tokens in the relatively informal picture-naming task than in the more formal passage reading task, regardless of the learners' target norms, L1 dialect regions, or possible effects of orthography. This finding can be accounted for by the production contexts of the retroflex suffix in L1 Mandarin, where rhoticity constitutes a stylistic feature of colloquial language and is favoured in informal situations. Moreover, the present finding is in line with what Li and Kabak (2017) reported for highly proficient English speakers in China. In their study, the formality levels of different tasks exerted a significant albeit weak effect on the rate of rhoticity, with informal contexts favouring the presence of nonprevocalic /r/. The stylistic influence from L1 found in the present study however declined at T2, suggesting that its effect is not as persistent as L1 phonological influence when the learners became more proficient over time.

L1 dialectal variation is the other factor that predicted the patterns of rhoticity by the Chinese learners only at the first time of data collection. Only then did the learners from rhotic dialectal regions in China favoured rhoticity in their speech more than those speaking nonrhotic Mandarin dialects. As concerns perception, no significant impact was observed from L1 dialectal background. The present findings support earlier work demonstrating the effects of L1 dialectal variation on speech production (O'Brien et al. 2010, Kautzsch 2017), while contradicting those showing that native dialect influences L2 speech perception (Chladkova 2011, Escudero & Williams 2012, Escudero et al. 2012). In this study, the L1 dialectal influence operates on the English speech of the learners only at T1, but was eliminated by the best-fit regression model at T2, like L1 stylistic influence. This change over time supports the predictions of the OPM (Major 2001), which claims that the IL of L2 learners tends to be L1-dominant only at early states of acquisition and that the effects of L1 gradually decrease over time.

10.2 The relationship between perception and production

The second research question on the acquisition of English rhoticity by L2 learners concerns the relationship between speech perception and production and the underlying forces conditioning the two domains. In chapter 9, the relationship between perception and production of L2 rhoticity was investigated at individual, learner group, and contextual levels.

At first glance, the hypothesis of a perception-production correlation as predicted by the SLM (Flege 1995) has to be rejected as no general link between the perception and production of rhoticity was found for the Chinese learners. The relationship between perception and production differed distinctly across individual speakers at both data collection points, which might be accounted for by the different self-stated target norms and L1 dialectal backgrounds of the participants. As such, the learners were further divided into four learner groups on the basis of their TL norms and L1 regional backgrounds. Different perception-production relationships were found in each learner group. More specifically, learners acquiring a dissimilar TL structure (e.g., learners from a nonrhotic L1 region aiming at AmE) are more likely to show a tendency of perception preceding production, while those targeting a similar TL

feature (e.g., learners from a nonrhotic L1 region aiming at BrE) tend to display accurate production without accurate perception. Thus, the learners' perception-production relationships were shown to be conditioned by the level of dissimilarity between their L1 and TL sounds. Crucially, the relationship seems to depend largely on the learners' production performances, the patterns of which could be predicted by the CAH (Lado 1957) and the MDH (Eckman 1977). In particular, same TL categories are more easily produced than different TL categories. When acquiring a similar TL category, the learners' production tends to be accurate and therefore more likely to exceed their perception. By contrast, different TL categories tend to cause learning difficulty in production, which leads to perception preceding production. Therefore, L1-TL differences are a key factor determining the perception-production relationship of English rhoticity by different learner groups.

Furthermore, the relationship between the perception and production of English rhoticity was found to vary considerably across different preceding contexts, which could be explained by their different degrees of acoustic salience, crosslinguistic dissimilarity, and articulatory difficulty. To be more specific, contexts that are acoustically less salient, perceptually similar, and require less articulatory effort for L2 learners, such as NURSE and LETTER vowels, are more likely to display a pattern of accurate production without accurate perception. NURSE and LETTER vowels are promoting of the production of rhoticity also due to their phonetic resemblance to the Mandarin retroflex suffix [ɚ]. On the other hand, the context of preceding FORCE, which was found to be acoustically salient and perceptually dissimilar for the Chinese learners in this study but articulatory challenging for both native English speakers (e.g., Becker 2014) and Chinese English speakers (Li & Kabak 2017), is promoting of the relationship of perception preceding production. Therefore, the perception and production relationship of rhoticity by the Chinese learners in the present study is dependent on the interplay of acoustic salience, perceived crosslinguistic dissimilarity, and articulatory difficulty of various phonetic contexts that nonprevocalic /r/ occurs in. An intuitive link between speech perception and production by L2 learners as predicted by the SLM (Flege 1995) cannot be upheld for the Chinese learners in this study on a context-dependent level either.

Instead of a straightforward link between speech perception and production, the

patterns of rhoticity in the study at hand are particularly complex. In terms of the patterns of different learner groups, the relationship is conditioned by the L1 dialects and the TL norms of the learners. Concerning various phonetic contexts that /r/ occurs in, the relationship is controlled by the interplay of degrees of articulatory complexity, crosslinguistic dissimilarity, and acoustic salience that are inherent in each preceding environment. The relationship between perception and production for individual learners, however, seems to be rather unsystematic over time. By investigating factors conditioning the perception and production of rhoticity using a multivariate approach, it was found that the perception and production of rhoticity are conditioned by different grammar-internal and -external factors. To begin with, the perception of rhoticity was found to be only conditioned by the preceding vowel context. The production of rhoticity, on the other hand, is not only constrained by preceding vowel environment, but also other phonological and grammar-external factors, the relevance of which further differ distinctly between the two data collection points. To be more specific, at T1, the production patterns of rhoticity were constrained by the L1 dialectal background of the learners and the stylistic variation of Mandarin rhoticity, but the influence of these two factors did not surface anymore after six months of acquisition. The individual norm preference of the learners, in turn, constituted a main effect determining the patterns of rhoticity at T2. In addition, gender was shown to have a persistent social effect on production with females being significantly more rhotic than male speakers at both time points. Thus, it is not surprising that no perception-production correlation of rhoticity could be found at an individual level, since there is significant variability in both domains across speakers controlled by different predictors over time.

While the SLM (Flege 1995) proposes a correlation between speech perception and production that is largely determined by the perceived crosslinguistic dissimilarity between the L1 and TL sound systems, the present study shows that the perception and production of English rhoticity are shaped by a multitude of factors. Both grammar-internal and -external indicators affect and alter speech perception and production. As such, speech perception and production are rather two mechanisms controlled by different phonetic, phonological, and sociolinguistic factors. Various sources contribute to the varying patterns and gradience of the production of rhoticity, whereas its perception is mainly affected by the preceding context. Considering the

significant and persistent influence of phonological context on the acquisition of rhoticity regarding both perception and production, the acquisition of rhoticity should always be investigated alongside its preceding vowel environments. In other words, /V(r)/ sequences should be examined instead of a single nonprevocalic /r/.

Moreover, no systematic patterns could be found for the changes of the perception-production link over time. The two domains rather underwent different pathways over the course of L2 phonological acquisition. As regards perception, the learners improved significantly from T1 to T2 in terms of their discrimination ability. The results showed, however, that the Chinese learners did not master the perception of rhoticity in all phonetic contexts at once. Instead, they acquired sensitivity towards English rhoticity context by context, from acoustically more salient and perceptually dissimilar ones to less salient and more similar ones. This finding is consistent with earlier work that has revealed that sounds are not equally perceived in all contexts (Wright 2004). Therefore, by analysing different parameters acoustically and crosslinguistically, we can discover specific challenges that L2 learners might encounter when acquiring the perception of a particular sound in an L2.

Production, on the other hand, did not display significant changes with respect to the overall proportion of target production by the Chinese learners between T1 and T2. The change is rather associated with the varying significance of the predictors determining rhoticity. More specifically, the productions of the learners at T1 were significantly controlled by L1 related aspects such as L1 stylistic and dialectal variation. At T2, norm preference emerged as a main factor, with the L1-related effects being eliminated. As such, I argue that longitudinal research should be conducted within a multivariate approach in order to examine the variation of underlying forces controlling the surface patterns over time.

In terms of the developmental sequence of rhoticity production in different phonetic contexts, the Chinese learners did not show the same pattern as found for their perception. Articulatory constraints seem to be rather persistent in controlling the rhoticity rates by the learners at both T1 and T2. It is therefore speculated that L2 learners need to manipulate articulatory parameters to produce nativelike sounds successfully, which is not only conditioned by the perception of the contexts. This

finding however did not yield support for Colantoni and Steele's (2008) study, who found that French learners tend to first acquire the production of the rhotics with the most salient properties. This discrepancy can be attributed to the different scopes of the two studies. More specifically, Colantoni and Steele focussed on the acquisition of different phonetic types of rhotics, while the present study investigated context-dependent variation of rhoticity, namely the different preceding phonological contexts that nonprevocalic /r/ exists in.

10.3 Predictors and patterns shared by native and non-native English speech

The third research question that the present study set out to answer was to what extent native and non-native English speech share universal patterns of rhoticity and its predictors. Rhoticity is known as a salient feature distinguishing spoken English varieties, the gradience of which is subject to a multitude of phonological, stylistic, and sociolinguistic factors. To explore the similarity between rhoticity in native and non-native speech, factors that were shown to condition rhoticity in world Englishes were also investigated in the present study for L2 learners. As reviewed in chapter 4, rhoticity in native English is mainly constrained by the grammar-internal factors preceding vowel, stress, syllable structure, word type, as well as the grammar-external factors speech style, substrate language, target norm, and gender. All these factors were explored in the present study for Chinese learners of English as well.

The results show that native and non-native rhoticity share the following predictors, both grammar-internal and -external: First, the production of nonprevocalic /r/ is controlled by the preceding vowel. Similar patterns were found for rhoticity in ENL varieties and in the L2 speech by the Chinese learners. Specifically, a preceding stressed central NURSE vowel was shown to be the most promoting context of rhoticity, while a back FORCE vowel strongly disfavours rhoticity in both native and non-native English speech. It appears that the articulatory difficulty of sounds exerts a universal impact on native and non-native speakers. However, the patterns of L2 rhoticity concerning preceding vowels in the present study are not entirely consistent with native rhoticity. A preceding unstressed central vowel LETTER, where rhoticity was disfavoured by native speakers of English due to word stress, was found to favour rhoticity by the Chinese learners of English. As discussed earlier, since word stress

does not form part of Mandarin phonological grammar, the learners in the present study are assumed to be insensitive to this prosodic feature. Thus, it is not surprising that stress does not operate on the production of rhoticity by Chinese learners the way it does in native English speech.

Another phonological predictor shared by native and non-native rhoticity is word type. Content words were found to favour the presence of nonprevocalic /r/ more than function words in both native English varieties and for the Chinese learners at T2. The fact that word type did not reach statistical significance at the first time of data collection confirms the tenets of the OPM (Major 2001). As claimed by Major, the speech of L2 learners is L1-dominant at the initial state of acquisition, and gradually reaches the idealized state where the TL is completely mastered by the learners. The Chinese learners in the present study did not display a preference for rhoticity in content words at T1, but gradually approached this TL pattern over time.

Since the substrate language and the language attitude of ESL speakers have been shown to condition the production of rhoticity in ESL varieties (Melchers & Shaw 2003), L1 dialectal variation and the norm preference of the individual Chinese learners were investigated in the present study. The latter was tested using a verbal-guise test as a supplement to the self-stated target norms of the learners. Results show that both factors exerted significant influence on the production of rhoticity by the learners, indicating that the two predictors are shared by native and non-native speakers. However, the relevance of the two factors varied between the two data points, with L1 dialectal variation showing main effects at T1, which was then replaced by the norm preference of the learners at T2. It reveals that the Chinese learners are more controlled by their L1 dialectal background at initial stages of acquisition, the effects of which decrease over the course of L2 phonological acquisition. The learners exhibit patterns that are more controlled by their TL preference at a later point of learning, which again yields support for the OPM (Major 2001). This finding further corroborates the study by Li and Kabak (2017), who found that the patterns of rhoticity produced by English teachers in China are subject to their self-stated target norms rather than L1 dialect regions as they are highly proficient English speakers. In their study, it was speculated that L1 influence might override target norm and exert a stronger impact on rhoticity in EFL learner varieties than for

highly proficient Chinese English speakers, which can in turn be confirmed by the findings of the present study.

Speech style is another predictor for the realisation of rhoticity in world Englishes. In the present study, it also had an impact on rhoticity in the Chinese learners' speech. The pattern, however, is not entirely consistent with findings on rhoticity in ENL and ESL. The Chinese learners produced rhoticity more frequently in informal situations at T1, which reflects the stylistic pattern of rhoticity in their L1 Mandarin, where rhoticity is a feature of colloquial language. However, after six months of acquisition, this stylistic influence from L1 fails to play a statistically significant role in conditioning the production of rhoticity by the Chinese learners. This shows that native and non-native speech do share the same stylistic predictor, the patterns of which however do not necessarily align. Moreover, the patterns of non-native speech are changing over the course of phonological acquisition with time and learners' L2 proficiency, moving from L1-controlled pattern to TL-oriented pattern and gradually approximate to native-like performance.

As demonstrated in the present study, rhoticity in native and non-native speech does share similar phonetic, phonological, stylistic, and social predictors. This finding resonates with the dynamic equational approach proposed by Kabak (2019), who argues that the variable and gradient phonological patterns in language change and SLA are controlled by the same set of internal and external constraints, although the significance of the variables may differ from one another. While similar predictors are shared by native and non-native rhoticity, the present study shows that the directionality or patterns modulated by those predictors does not always align. To be more specific, it is possible for native and non-native speech to have the same pattern, if the realisation of the feature is conditioned by a universal constraint across languages such as articulatory complexity. However, a divergent pattern in the L2 is likely to emerge when learners' native language exerts an influence. The effects of CLI from L1 are most prominent at earlier stages of phonological acquisition and gradually decline as L2 learners' proficiency increases over time. It should be mentioned that the changing weights of L1- and TL-related predictors can be best captured within a variationist approach. However, "relatively few SLA researchers availed themselves of the potential offered by the methodological and analytical tools

developed in variationist sociolinguistics" (Bayley 2005: 1). Thus, such an approach that takes into account the varying effects of both grammar-internal and -external factors on speech perception and production should be encouraged and further promoted in the field of SLA.

10.4 Implications for models of L2 speech acquisition

Given that the phenomenon investigated in the present study is phonologically and sociolinguistically variable in both the learners' L1 Mandarin and L2 English, no existing models can offer a fully comprehensive account for the complex L2 patterns observed in the present study. To begin with, different types of CLI from Mandarin were found to operate on L2 English rhoticity, including L1 transfer of structural, stylistic, dialectal, and social variations. As concerns the relationship between perception and production, no straightforward and consistent link between the two domains could be found at individual, group, or contextual levels, which shows that a perception-production correlation as claimed by the SLM (Flege 1995) cannot be upheld here. Perception and production are rather two independent mechanisms controlled by the interplay of both grammar-internal and -external constraints over the course of L2 phonological acquisition. They have their own forces and are not governed by the same rules. With regard to the predictors and patterns shared by native and non-native rhoticity, the results demonstrate that similar predictors are shared by the two, the patterns of which nonetheless do not always parallel each other.

Based on this summary, none of the models discussed in the theoretical chapters can make entirely accurate predictions for the acquisition of rhoticity by Chinese learners of English. The CAH (Lado 1957) and the MDH (Eckman 1977) propose that L1-TL differences are the main source of learning difficulties for L2 learners, which is generally confirmed by the present findings. However, they do not make any predictions regarding the acquisition of L2 speech structures beyond the segmental level or transfer opportunities of L1 dialectal, stylistic, and social variabilities. The SLM (Flege 1995) and the MSA (Colantoni & Steele 2008) posit a close relationship between speech perception and production, which is controlled by individual learners' perceived crosslinguistic dissimilarity of sounds. The results of the present study yield contradicting evidence for the role of perception on the production as predicted by the SLM and the MSA, indicating that perception and production are conditioned by

different structural and sociolinguistic factors. Moreover, the OPM (Major 2001) is only successful in capturing the changing constraints on the production of rhoticity as learner's proficiency increases over time, but fails to predict the rather consistent effects of the preceding vowels on rhoticity at both data points, its resemblance to native English rhoticity patterns, as well as the developmental sequence of perception.

Based on the outcome of this study, I propose that in predicting the developmental patterns of an L2 phonological feature, three essential questions should be asked. The first one is concerned with the phonetic properties of the TL sound. More specifically, do the TL feature and its surrounding phonetic and phonological contexts involve different degrees of acoustic salience and articulatory complexity? If not, the perception and production of the specific TL aspect could possibly be accounted for by perceived crosslinguistic dissimilarity as predicted by the SLM (Flege 1995), according to which similar TL aspects would be relatively difficult for learners to acquire, while dissimilar ones would be easier. As such, a perception-production correlation might be observed, and the production would be shaped by the learners' perception of the TL feature. According to the predictions of SDRH (Major & Kim 1996) and the Similarity Corollary of the OPM (Major 2001), dissimilar TL structures are expected to be acquired first, followed by similar TL categories at later stages of acquisition due to their greater difficulty for learners.

If the TL phonological variable or the contexts it occurs in differ with respect to their acoustic salience and articulatory difficulty, then a context-dependent perception-production link is likely to emerge, as learners' performances in the two domains are additionally controlled by the two phonetic factors respectively. In terms of the developmental sequence of perception, learners seem to acquire the perception of the TL variable context by context, from acoustically more salient ones to less salient ones. In terms of the developmental pattern of production, learners are expected to first acquire the articulatorily simpler ones and later the articulatorily difficult ones.

The second question that should be addressed is whether the TL feature is sociophonologically variable in the TL varieties. In other words, does the TL feature exhibit any diachronic and synchronic variation and change across different varieties?

More crucially, is it sensitive to any structural and sociolinguistic variables? If so, both grammar-internal and grammar-external predictors that have previously been found or that are expected to exert impact on the realisation of the feature should be investigated in non-native production as well. The results of the present study demonstrate that native and non-native speech share similar predictors, but not necessarily similar patterns, due to the influence of learners' native languages, their proficiency level, and as suggested by the MDH (Eckman 1977) and the OPM (Major 2001), language universals. In particular, the patterns of the L2 sociophonological feature conditioned by different parameters are expected to approach TL patterns as learners' proficiency increases over time. According to the OPM, the patterns are assumed to be influenced by language universals during intermediate acquisition stages, the impact of which first increases and then decreases. However, it is also possible for learners to bypass the influence of language universals as claimed by the OPM.

The last question concerns possible sources and types of CLI from the L1 of learners. Does the TL feature exist in the L1? If so, does it show any kind of variation in the L1? Various aspects about the nature of the sound in the L1 could possibly lead to CLI, including but not restricted to L1 structural variation, L1 dialectal variation, L1 stylistic variation, and L1 social variation. A contrastive analysis of differences of the feature between the L1 and the TL should be investigated in detail to shed light on possible transfer opportunities from L1 to learners' L2. All of them could possibly lead to different acquisition patterns of L2 learners. As predicted by the OPM (Major 2001), the influence of CLI from the L1 would decrease as the proficiency of learners increases, and the IL of learners is expected to be increasingly determined by learners' TL over time, like the patterns exhibited by the Chinese learners in the present study.

It is noteworthy that a perception-production correlation as predicted by the SLM can only surface when the feature to be acquired does not show any context-dependent acoustic, articulatory, and sociophonological variability in both L1 and TL. Otherwise, a context-dependent perception-production relationship would tend to emerge because the perception of the TL feature was additionally controlled by the acoustic salience of its contexts, while its production might be determined by its articulatory complexity. If the feature further involves sociophonological variation in both the TL

and the L1, a perception-production mismatch will be observed. The production of the TL feature can be conditioned by both grammar-internal and -external predictors, the significance of which varies over the course of acquisition. The perception of the feature is however not affected by the same factors influencing production, leading to a severe perception-production asymmetry when acquiring an L2 sociophonological feature.

As none of the existing models can offer a comprehensive account of the findings in the present study, a preliminary framework, the *L2 Sociophonological Development Model (SPDM)* is proposed to predict and explain the patterns of L2 sociophonological acquisition. Figure 10.1 provides a schematic representation of the SPDM, which depicts both perception and production over the course of L2 speech learning.

The basic assumption of the SPDM is that speech perception and production will not align with each other in acquiring a sociophonological feature. Instead, perception and production are separate domains with different underlying forces behind their variation. As concerns perception, apart from the influence of the degree of perceived crosslingusitic dissimilarity between the L1 and the TL structures of learners as suggested by the SLM (Flege 1995), the perception of an L2 sociophonological feature can be additionally shaped by the acoustic salience of the phonetic/phonological contexts the TL feature exists in. As shown in figure 10.1, the SPDM predicts that learners are more accurate in perceiving the TL feature in acoustically salient and perceptually dissimilar contexts than in acoustically less salient and perceptually more similar contexts. Regarding the development of perception over time, learners are expected to acquire the TL feature context by context from dissimilar and acoustically salient contexts to similar and less salient ones.

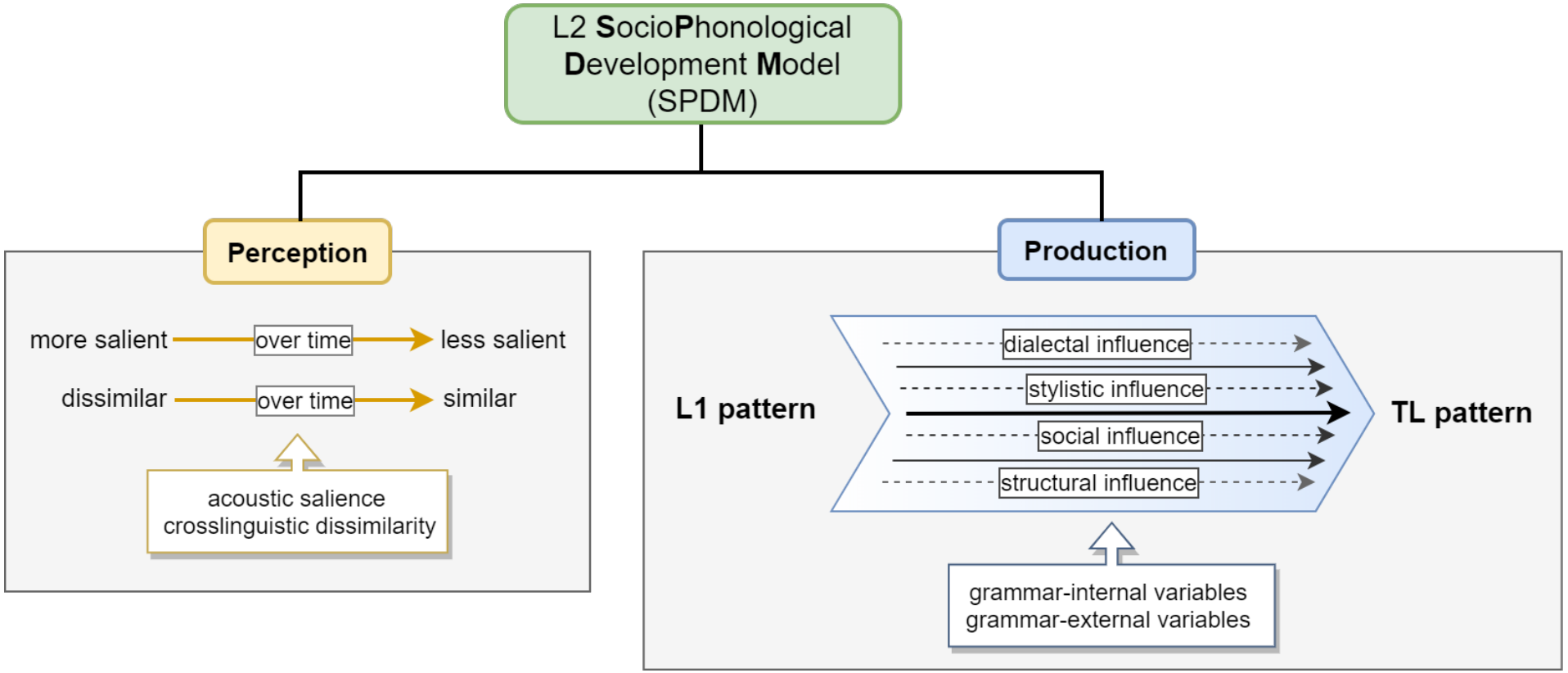


Figure 10.1 A schematic diagram of L2 Sociophonological Development Model (SPDM)

In terms of production, the SPDM suggests a multi-layer developmental continuum describing learners' L2 speech patterns. In general, their production performance would move on a continuum from L1 to the other end of the continuum, TL, as their proficiency increases over time. Crucially, the continuum does not only represent the phonetic or phonological patterns of the TL feature itself, but also its sociolinguistic aspects. To be more specific, the continuum is constrained by a multitude of grammar-internal and -external variables depending on the nature of the sociophonological feature to be acquired, such as its dialectal, stylistic, structural, and social variations in both the L1 and the TL of learners. A dialectal continuum, i.e. a developmental continuum under dialectal influence, for example, displays the competing role of L1 dialectal background and TL norm variation at the two ends of the scale. L2 learners' speech is expected to be more subject to L1 dialectal effects at earlier stages of acquisition, and approximate to their respective TL norm preference over time. As concerns a stylistic continuum, L2 learners are predicted to first adopt the stylistic patterns used in their L1 when producing the TL feature. The speech patterns of the learners will move on the continuum and gradually approach a stage when TL stylistic influence overrides L1 patterns and plays a more decisive role in shaping the learners' L2. This would be the same for other sociolinguistic forces, where L2 learners are expected to move on the respective developmental continua from L1-related patterns to TL-related patterns. Moreover, it is possible for L2 learners to drift along the continuum over the course of L2 learning due to the influence of language universals or other learner-specific variables, such as the change of language attitude, motivation, increased language input, and other related factors. According to the SPDM, an ideal L2 learner will master TL patterns on every layer of the L1-TL multi-layer continuum, approximating thus ultimately native speakers in their spoken competence.

10.5 Concluding remarks and outlook

The present study investigated the perception and production of English rhoticity by Chinese learners before and after a six-month training course and had two main objectives. The first objective was to shed light on our understanding of L2 sociophonological acquisition. This aim was achieved by this study from three perspectives: First, by focusing on the acquisition of an L2 sociophonological feature, the influence of different types of CLI across time was explored. Second, the

relationship between the perception and production of the sociophonological feature and the underlying factors conditioning the two domains were investigated. Third, the universal and divergent predictors and patterns between the speech produced by native and non-native English speakers were examined. The second objective of this study was to make theoretical contribution to existing hypotheses and models of L2 phonological acquisition and to develop a new model that offers an explanatory and predictive account for the acquisition of an L2 sociophonological feature.

The following main findings were reported and discussed: To begin with, multiple types of CLI were identified. Transfer occurred from L1 structural, stylistic, dialectal, and social variations beyond the phonetic and the phonological level. Second, a perception-production link as predicted by previous theoretical frameworks such as the SLM (Flege 1995) cannot be observed for the data of the present study, given that speech production is not solely affected by speech perception when an L2 sociophonological feature is acquired. Instead, there were different forces behind the variation and development of the two domains. In general, learners' perception and production of English rhoticity changed greatly over the course of acquisition, albeit with both inter- and intra-learner variability. Their perception improved over time and exhibited a context-sensitive acquisitional pattern depending on the relative acoustic salience and the crosslinguistic dissimilarity of the phonetic and phonological contexts, while the development of their production performance can be best described as being controlled by a complex interplay of both grammar-internal and -external factors. Third, native and non-native rhoticity were found to share universal predictors. More precisely, similar factors such as preceding vowel, gender, and speech style have been suggested to shape rhoticity in the learners' L2 and the synchronic variability of rhoticity that leads to language change in native English varieties. The patterns of rhoticity predicted by these universal predictors in the L2 speech of learners were, however, not always in parallel with those observed in native English, but rather developed on an L1-TL continuum over time due to CLI and different proficiency levels of L2 learners.

These findings are based on a meticulous study design. First, the data used in the present study were collected by testing 47 learners from 24 provinces of mainland

China, representing a wide spectrum of the dialects spoken in the country. Furthermore, a relatively balanced number of learners were selected on the basis of their L1 dialect regions and TL norms to ensure the comparability of the data across learner groups. A verbal-guise language attitude test was additionally conducted as a supplement to the learners' self-stated target norms. In particular, recruiting the same learners for both speech perception and production tests and for both times of data collection ruled out the potential interference of learner differences. By testing learners that were mostly homogeneous in terms of different learner-related variables (such as age, proficiency level, age of acquisition, motivation, L2 input, amount of L1 and L2 usage, and linguistic experience), the study eliminated variation between learners to the largest extent and allowed sociolinguistic variation, the primary focus of this study, to surface. In addition, the analysis was carried out using a variationist approach, which includes the examination of both grammar-internal and -external factors predicting the learners' perceptual and production performance. With respect to the object of inquiry and the language combination chosen for the present study, investigating rhoticity in L1 Mandarin and L2 English offers particularly valuable insights into L2 sociophonological acquisition especially in terms of the effects of CLI on the sociolinguistic level. What makes the interpretation of the present findings more challenging, however, is the scarcity of empirical work on the sociolinguistics of rhoticity in Mandarin. As already mentioned in chapter 5, the social status of the diminutive suffix [ɚ] in China remains largely unknown in regions outside Beijing. Although various authors have suggested that the suffix is subject to phonological, stylistic, and social variables, it is still unclear how they may interact with each other and modulate the variable realisation of rhoticity in Mandarin. Thus, future empirical studies focussing on the grammar-internal and -external factors conditioning Mandarin rhoticity produced by Chinese speakers from all over China are immediately needed, as our investigation of L2 rhoticity would certainly benefit from fine-grained descriptions of rhoticity in both the L1 and the TL of learners.

In the review of the theoretical L2 models describing the development of the effects of CLI and the relationship between speech perception and production, it has been revealed that none of the existing theories can provide a fully comprehensive account for the findings of the present study. Therefore, a preliminary model, the SPDM, was

proposed in an attempt to explain and predict the developmental patterns of the perception and production of an L2 sociophonological feature by language learners. As discussed in section 10.4, this model contradicts the traditionally intuitive relationship of L2 speech perception leading production. Instead, a clear perception-production mismatch emerged to the acquisition of an L2 sociophonology, which is constrained by a complex network of both grammar-internal and -external factors. As such, a multi-layer acquisition continuum conditioned by a multitude of variables has been advanced to describe the development of L2 speech, which takes a general pathway from L1 patterns to TL patterns with the possibilities of drifting along the continuum over the course of L2 phonological acquisition.

By testing the development of the perception and production of English rhoticity by Chinese learners over time, the study at hand provides a valuable piece to our understanding of L2 sociophonological acquisition, a research area that is severely understudied in comparison with other aspects of SLA. As the SPDM was developed based on the empirical findings of the present study, future research is necessary to test the predictive and explanatory power of the model and to provide further empirical evidence that will support or falsify the current version of the SPDM. This study opens routes for future research to investigate the acquisition of rhoticity by L2 learners with other language combinations. For instance, native English speakers acquiring L2 Mandarin rhoticity would form an instructive and comparative test case. Furthermore, exploring the acquisition of other L2 sociophonological structures that are preferably subject to fine-grained sociolinguistic variation like rhoticity is worth recommending. While the present study focussed exclusively on adult experienced learners of English over the period of six months, future longitudinal studies should include young/adolescent L2 learners too, and extended periods of examination. This could also involve dense data collection (cf. Ortega & Han 2017) to gain a more accurate insight into the process of L2 sociophonological acquisition and the influencing factors thereof. From a broader perspective, future studies should extend the inquiry into other areas such as bilingualism and multilingualism, which together with SLA are expected to fully capture the nature and dynamism of sociophonology in language development.

11. References

Allport, A., MacKay, D. G., Prinz, W., & Scheerer, E. (1987). *Language Perception and Production: Relationships Between Listening, Speaking, Reading, and Writing.* London: Academic Press.

Archibald, J. (2009). Phonological feature re-assembly and the importance of phonetic cues. *Second Language Research*, *25*(2), 231-233.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL–International Review of Applied Linguistics in Language Teaching*, *44*(3), 231-250.

Bao, Z. (1996). The syllable in Chinese. *Journal of Chinese Linguistics*, *24*(2), 312-354.

Bayley, R. (1996). Competing constraints on variation in the speech of adult Chinese learners of English. *Second language acquisition and linguistic variation*, *97120*.

Bayley, R. (2005). Second language acquisition and sociolinguistic variation. *Intercultural communication studies*, *14*(2), 1.

Bayley, R., & Preston, D. R. (Eds.). (1996). *Second language acquisition and linguistic variation* (Vol. 10). John Benjamins Publishing.

Becker, K. (2014). (r) we there yet? The change to rhoticity in New York City English. *Language Variation and Change*, *26*(2), 141-168.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International 5:9/10*, 341-345.

Bohn, O., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*, 131–158.

Bonnici, L. M. (2010). *Variation in Maltese English: The interplay of the local and the global in an emerging postcolonial variety*. University of California, Davis

Bradlow, A. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287–308): Amsterdam: Benjamins.

Broselow, E., & Kang, Y. (2013). Second language phonology and speech. *The Cambridge handbook of second language acquisition*, 529-554.

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, *14*(2), 136-193.

Buchstaller, I. (2006). Social stereotypes, personality traits and regional perception displaced: attitudes towards the "new" quotatives in the U.K. *Journal of Sociolinguistics, 10*, 362–81.

Cardoso, W. (2007). The variable development of English word-final stops by Brazilian Portuguese speakers: A stochastic optimality theoretic account. *Language Variation and Change*, *19*(3), 219-248.

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

Chan, A. Y. (2014). The perception and production of English speech sounds by Cantonese ESL learners in Hong Kong. *Linguistics*, *52*(1), 35-72.

Chand, V. (2010). Postvocalic (r) in urban Indian English. *English World-Wide*, *31*(1), 1-39.

Chao, Y. R. (1965). *A grammar of spoken Chinese*. University of California Press.

Chen, P. (1999). *Modern Chinese: history and sociolinguistics*. Cambridge University Press.

Chiba, Reiko, Matsuura, Hiroko and Yamamoto, Asako. 1995. Japanese Attitudes toward English Accents. *World Englishes*, 14(1): 77–86.

Chládková, K., & Podlipský, V. (2011). Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners. *The Journal of the Acoustical Society of America*, *130*(4), EL186-EL192.

Colantoni, L., & Steele, J. (2007). Acquiring /ʁ/ in context. *Studies in Second Language Acquisition*, *29*(3), 381-406.

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*(3), 489-534.

Colantoni, L., Steele, J., Escudero, P., & Neyra, P. R. E. (2015). *Second language speech*. Cambridge University Press.

Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass*, *5*(3), 126-139.

Delattre, P., & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, *6*(44), 29-68.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research* (Vol. 42). John Benjamins Publishing Company.

Dickson, V., & Hall-Lew, L. (2017). Class, gender, and rhoticity: The social stratification of non-prevocalic /r/ in Edinburgh speech. *Journal of English Linguistics*, *45*(3), 229-259.

Dinnsen, D. A., & Eckman, F. (1975). A functional explanation of some phonological typologies. *Functionalism.* 126-134.

Dowd, J., Zuengler, J., & Berkowitz, D. (1990). L2 Social Marking: Research Issues1. *Applied Linguistics*, *11*(1), 16-29.

Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages* (Doctoral dissertation, Massachusetts Institute of Technology).

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University. Press.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*, 315–330.

Eckman, F. R. (1996). A functional-typological approach to second language acquisition theory. *Handbook of second language acquisition*, 195-211.

Eckman, F. R. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition*, *26*(4), 513-549.

EI-Dash, L., & Tucker, R. (1975). Subjective reactions to various speech styles in Egypt. *International Journal of the sociology of language*, *6*, 33-54.

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.

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.

Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *The Journal of the Acoustical*

Society of America, *92*(2), 736-757.

Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America*, *108*(1), 343-356.

Fairbanks, G. (1960). *Voice and articulation drillbook,* 2nd ed. New York: Harper.

Feng, G. (2001). Existential and universal anchoring: A case study of Chinese diminutive affixation. *Manuscript. Los Angeles: University of Southern California*.

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

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

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

Flege, J. E. (2016). *The role of phonetic category formation in second language speech acquisition*. Paper presented at the 8th International Conference on Second Language Speech, Aarhus University, Denmark.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1730-1741.

Gendrot, C. (2014). Perception et réalisation du /r/ standard français en finale de mot. In *JEP 2014*, 193-201.

Gick, B. (2002). An X-ray investigation of pharyngeal constriction in American English schwa. *Phonetica*, *59*(1), 38-48.

Gick, B., & Wilson, I. (2006). Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. *Laboratory phonology*, *8*, 635-659.

Gleason, H. A. Jr. (1961). *An introduction to descriptive linguistics,* 2nd ed. New York: Holt, Rinehart and Winston.

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.

Gut, U. (2009). *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German* (Vol. 9). Peter Lang.

Hansen Edwards, I. G. (2008). Social factors and variation in production in L2 phonology. In Jette G. Hansen Edwards and Mary Zampini,（Eds). *Phonology and Second Language Acquisition*. Amsterdam, Philadelphia: Benjamins, 251–82.

Hartmann, D., & Zerbian, S. (2010). Rhoticity in Black South African English–a sociolinguistic study. *Southern African linguistics and applied language studies*, *27*(2), 135-148.

Hartmann, J. (2020). Mobile anglophone Africans' attitudes to different Englishes. Manuscript in preparation.

Heeringa, W. & Van de Velde, H. (2018). "Visible Vowels: A Tool for the Visualization of Vowel Variation." In *Proceedings CLARIN Annual Conference 2018, 8 - 10 October, Pisa, Italy.* CLARIN ERIC.

Heffner, R. M. S. (1950). *General phonetics*. Madison: University of Wisconsin Press.

Heselwood, B. (2009). Rhoticity without F3: Lowpass Filtering, F1-F2 relations and the perception of rhoticity in 'NORTH-FORCE', 'START' and 'NURSE' words. *Leeds working papers in linguistics and phonetics*, *14*, 49-64.

Heselwood, B., & Plug, L. (2011). The Role of F2 and F3 in the Perception of Rhoticity: Evidence from Listening Experiments. In *ICPhS*, 867-870.

Hickey, R. (2014). Vowels before /r/ in the history of English. *Contact, Variation and Change in the History of English*. Amsterdam: John Benjamins, 95-110.

Himmel, C. M. (2018). Phonetic and phonological variability in the L1 and L2 of late bilinguals: The case of /r/ and /l/ (Unpublished PhD dissertation), Julius-Maximilians-Universität Würzburg, Germany.

Hiraga, Y. (2005). British attitudes towards six varieties of English in the USA and Britain. *World Englishes, 24*, 289–308.

Honey, J. (1998) Sociophonology. In Coulmas, F. (Ed.), *The Handbook of Sociolinguistics*. Blackwell Publishing.

Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. Fry, P. MacCarthy, N. C. Scott, & J. Trim (Eds.), *In honour of Daniel Jones* (pp. 73–84). London: Longman.

Huang, T. (2010). Er-suffixation in Chinese monophthongs: phonological analysis and phonetic data. In *Proceedings of the 18th International Conference on Chinese*

Linguistics (IACL-18) (Vol. 1). 331-344.

Hultzen, Lee S. (1950). Symbol for the nonsyllabic postvocalic *R* of General American. *Quarterly Journal of Speech, 36*, 189-201.

Hyman, L. (1977). On the nature of linguistic stress. in L. Hyman (ed.), *Studies in stress and accent*, Southern California Occasional Papers in Linguistics *4*, 37-82.

Irwin, P., & Nagy, N. (2007). Bostonians /r/ speaking: A quantitative look at (R) in Boston. *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 11.

Isbell, D. R. (2016). The perception-production link in L2 phonology. *MSU Working Papers in Second Language Studies*, *7*(1).

Kabak, B. (2019). A dynamic equational approach to sound patterns in language change and second language acquisition: The (un)stability of English dental fricatives illustrated. In: Ruth Moehlig-Falke and Beatrix Busse (eds.), *Patterns in Language and Linguistics. New Perspectives on a Ubiquitous Concept. Topics in English Linguistics* [TiEL] 104, 221-254. Berlin, Boston: Mouton de Gruyter.

Kang강현석. (2013). Internal and External Constraints on Rhoticity in Korean English. *사회언어학*, *21*, 1-27.

Kautzsch, A. (2017). *The Attainment of an English Accent: British and American Features in Advanced German Learners*. Peter Lang.

Kenyon, J. S. (1950). *American Pronunciation,* 10th ed. Ann Arbor: Wahr.

Kim, C. W., & Park, S. G. (1995). Pronunciation problems of Australian students learning Korean. *Australian Review of Applied Linguistics. Supplement Series*, *12*(1), 183-202.

Kerswill, P. (2002). A dialect with “great inner strength”? The perception of nativeness in the Bergen speech community. In D. Long & D. R. Preston (Eds), *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins, pp. 151–73.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038-1054.

Kramsch, C. (2007). Classic Book Review. Re-reading Robert Lado, 1957. *Linguistics Across Cultures. Applied Linguistics for Language Teachers. International Journal of Applied Linguistics*, *17*(2), 241-247.

Kuecker, K., Lockenvitz, S., & Müller, N. (2015). Amount of rhoticity in schwar and in vowel+/r/ in American English. *Clinical linguistics & phonetics*, *29*(8-10), 623-629.

Labov, W. (1966). *The social stratification of English in New York city*. Washington DC: Center for Applied Linguistics.

Labov, W. (1994). *Principles of linguistic change: Internal factors*. Vol. 1. Malden: Blackwell.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

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

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2013). Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics*, *41*(3-4), 198-210.

Lee, W. S. (2005). A Phonetic Study of the "er-hua" Rimes in Beijing Mandarin. In *Ninth European conference on speech communication and technology*.

Lehiste, I. (1962). *Acoustical characteristics of selected English consonants* (No. 9). Michigan Uni Ann Arbor.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1). https://doi.org/10.18637/jss.v069.i01

Lenth, R. V. (2018). emmeans: Estimated marginal means, aka least-squares means. Retrieved from https://CRAN.R-project.org/package=emmeans

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

Li, Z., & Kabak, B. (2017). Rhoticity in Chinese English: An experimental investigation on the realization of the variant (r) in an Expanding Circle variety. *Alicante Journal of English Studies 30*, 61-91.

Lin, Y. H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology* (Doctoral dissertation), University of Texas at Austin.

Lindau, M. (1978). Vowel features. *Language*, *54*(3), 541-563.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer, Dordrecht.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *The Journal of the acoustical society of America*, *99*(3), 1683-1692.

Liu, A. (2017). An acoustic study of Mandarin rhotic suffix. *ExLing 2017*, 69.

Lockenvitz, S., Kuecker, K., & Ball, M. J. (2015). Evidence for the distinction between "consonantal-/r/" and "vocalic-/r/" in American English. *Clinical linguistics & phonetics*, *29*(8-10), 613-622

Love, J., & Walker, A. (2013). Football versus football: Effect of topic on /r/ realization in American and English sports fans. *Language and speech*, *56*(4), 443-460.

Ma, L. (1997). Mandarin diminutive formation: An optimality analysis. *Working Papers of the Linguistics Circle*, *14*, 47-60.

Ma, L. (1998). Issues in Mandarin Er-Suffixation and the implications for an Optimality Analysis. *Proceedings of the Ninth North American Conference on Chinese Linguistics,* Vol. 2, p. 190-207.

Major, R. C. (1986). The ontogeny model: Evidence from L2 acquisition of Spanish r. *Language Learning*, *36*(4), 453-504.

Major, R. C., & Kim, E. (1996). The similarity differential rate hypothesis. *Language Learning*, *46*(3), 465-496.

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

McKenzie, R. M. (2008). Social Factors and Non-native Attitudes towards Varieties of Spoken English: A Japanese Case Study. *International Journal of Applied Linguistics*, 18(1): 63–88.

Melchers, G., & Shaw, P. (2003). *World Englishes*. Routledge.

Miller & Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America.* 27: 338-352.

Moulton, W. (1962). *The sounds of English and German*. Chicago: The University of Chicago Press.

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge University Press.

Nagy, N. (2008). Perceptual frequency and formant frequency in R speech. *LabPhon11*.

Nagy, N., & Irwin, P. (2010). Boston (r): Neighbo (r) s nea (r) and fa (r). *Language*

Variation and Change, *22*(2), 241-278.

Newmark, L. (1966). How not to interfere with language learning. *International journal of American linguistics*, *32*(1), 77-83.

O'brien, M. G., & Smith, L. C. (2010). Role of first language dialect in the production of second language German vowels. *IRAL-International Review of Applied Linguistics in Language Teaching*, *48*(4), 297-330.

Ohala, J. J., & Kawasaki, H. (1984). Prosodic phonology and phonetics. *Phonology*, *1*, 113-127.

Oller Jr, J. W., & Ziahosseiny, S. M. (1970). The contrastive analysis hypothesis and spelling errors. *Language learning*, *20*(2), 183-189.

Ortega, L., & Han, Z. (2017). *Complexity theory and language development: In celebration of Diane Larsen*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Patience, M. (2018). Acquisition of the Tap-Trill Contrast by L1 Mandarin–L2 English–L3 Spanish Speakers. *Languages*, *3*(4), 42.

Piercy, C. (2012). A transatlantic cross-dialectal comparison of non-prevocalic /r/. *University of Pennsylvania Working Papers in Linguistics*, *18*(2), 10.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of phonetics*, *29*(2), 191-215.

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from http://www.pstnet.com.

Rosenfelder, I. (2009). Rhoticity in educated Jamaican English: An analysis of the spoken component of ICE-Jamaica. In *World Englishes–Problems, Properties and Prospects: Selected Papers from the 13th IAWE Conference* (pp. 61-82).

Schützler, O. (2010). Variable Scottish English Consonants: The Cases of /ʍ/ and Non-Prevocalic /r/. *Research in language*, *8*, 1-17.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*(1-4), 209-232.

Shariatmadari, D. (2006). Sounds difficult? Why phonological theory needs' ease of articulation'. *School of Oriental and African Studies Working Papers in Linguistics*, *14*, 207-226.

Simon, E., Debaene, M., & Van Herreweghe, M. (2015). The effect of L1 regional variation on the perception and production of standard L1 and L2 vowels. *Folia Linguistica*, *49*(2), 521-553.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic inquiry*, *27*, 720-731.

Snodgrass, J. G., Levy-Berger, G., & Haydon, M. (1985). *Human experimental psychology* (Vol. 395). New York: Oxford University Press.

Starr, R. L. (2015). Sweet voice: The role of voice quality in a Japanese feminine style. *Language in Society*, *44*(1), 1-34.

Stevens, K. N. (1999). *Acoustic Phonetics*. MIT Press, Cambridge, MA.

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics*, *3*(4), 215-233.

Storme, B. (2018). The adaptation of French liquids in Haitian. *Journal of Pidgin and Creole Languages*, *33*(2), 386-410.

Sundkvist, P., & Gao, M. (2016). Rhoticity in Yunnan English. *World Englishes*, *35*(1), 42-59.

Tan, C. H., & Gupta, A. F. (1992). Post-Vocalic /r/ in Singapore English.

Tan, Y. (2012). To r or not to r: social correlates of /ɹ/ in Singapore English.*, 2012(218), pp. 1-24. Retrieved 29 Oct. 2019, from doi:10.1515/ijsl-2012-0057*

Thomas, C. K. (1958). *An introduction to the phonetics of American English. 2nd* ed. New York: Ronald Press.

Thomas, E. R. (2007). Phonological and phonetic characteristics of African American Vernacular English. *Language and Linguistics Compass 1*(5):450–475.

Tian, J. (2007). An optimality theory analysis of diminutive suffixation of Beijing Chinese. *Working Papers of the Linguistics Circle*, *19*(1), 217-232.

Trudgill, P., & Hannah, J. (2013). *International English: A guide to the varieties of standard English*. Routledge.

Ulbrich, C., & Ordin, M. (2014). Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics*, *45*, 26-42.

Wang, L., & He, N. (1985). Beijinghua er-huayun de tingbian shiyan he shengxue fenxi. *[Discrimination Experiments and Phonetic Analysis of r-rhymes in Beijing], in L. Tao et al.(eds.), Beijing yuyin shiyanlu*, 27-72.

Wang, J. Z. (1993). The geometry of segmental features in Beijing Mandarin (Doctoral dissertation). Newark, DE: University of Delaware.

Werker, J. F., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, *25*(1), 121-133.

Woolley, D. E. (1978). The perception of the English R vowels. *Linguistics*, *16*(210), 5-22.

Wright, R. (2004). A review of perceptual cues and cue robustness. *Phonetically based phonology*, 34-57.

Young-Scholten, M. (2004). Prosodic constraints on allophonic distribution in adult L2 acquisition. *International Journal of Bilingualism*, *8*(1), 67-77.

Yu, A. C. L. (2019). On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 2.

Yuan, J. (1989). *Hanyu Fangyan Gaiyao* [Outline of Chinese Dialects], 2nd edn. Beijing: Wenzi Gaige Chubanshe.

Zhang, J. (2000). Non-contrastive features and categorical patterning in Chinese diminutive suffixation: Max [F] or Ident [F]?. *Phonology*, *17*(3), 427-478.

Zhang, Q. (2008). Rhotacization and the 'Beijing Smooth Operator': The social meaning of a linguistic variable 1. *Journal of Sociolinguistics*, *12*(2), 201-222.

Appendices

Appendix A: Background Questionnaire

A1 Questionnaire—English version

Name_____ Student ID_____ Gender_____ Age____

1. How old were you when you started learning English? _____

2. Where is your hometown? Where did you grow up? _____

3. What Chinese dialect(s) do you speak besides Mandarin? _____

4. What foreign language(s) can you speak besides English? _____

5. Do you have any experience abroad? If yes, please indicate the country and length of residence_____

6. Do you know the differences between American English and British English? Please give some examples.

_____

7. Which kind of Standard English would you like to speak? British English or American English?_____

8. How many hours do you learn/speak/use English per week?_____

9. Which kind of media do you regularly have exposure to? How many hours per week?_____

☐ American English media (American TV series, movies, VOA, etc.)

☐ British English media (British TV series, movies, BBC, etc.)

☐ Both

☐ no English media

A2 Questionnaire—Chinese version

姓名_____ 学号_____ 性别_____ 年龄_____

1. 你从几岁开始学习英语? _____

2. 你的出生地是哪里? _____

3. 除普通话外，你还会说哪种（些）中国方言? _____

4. 除英语外，你还会说哪种（些）外语？ _____

5. 你有过出国经历吗？如有，在哪些国家？停留多久？

_____

6. 你知道英式英语和美式英语发音的区别吗？请举例说明。

_____

7. 你更喜欢（想学）英音还是美音？_____

8. 你每周大概花几小时学习／使用英语？_____

9. 你更喜欢或者更经常看／听哪个（些）国家的英语新闻，电影电视或娱乐节目？

（可多选） 每周几小时？_____

☐ 美国（美音） (美剧，美国电影，VOA 等)

☐ 英国（英音） (英剧，英国电影，BBC 等)

☐ 其它（如 TED 等）

☐ 不看英语类节目

Appendix B: Materials

B1 Perceptual stimuli

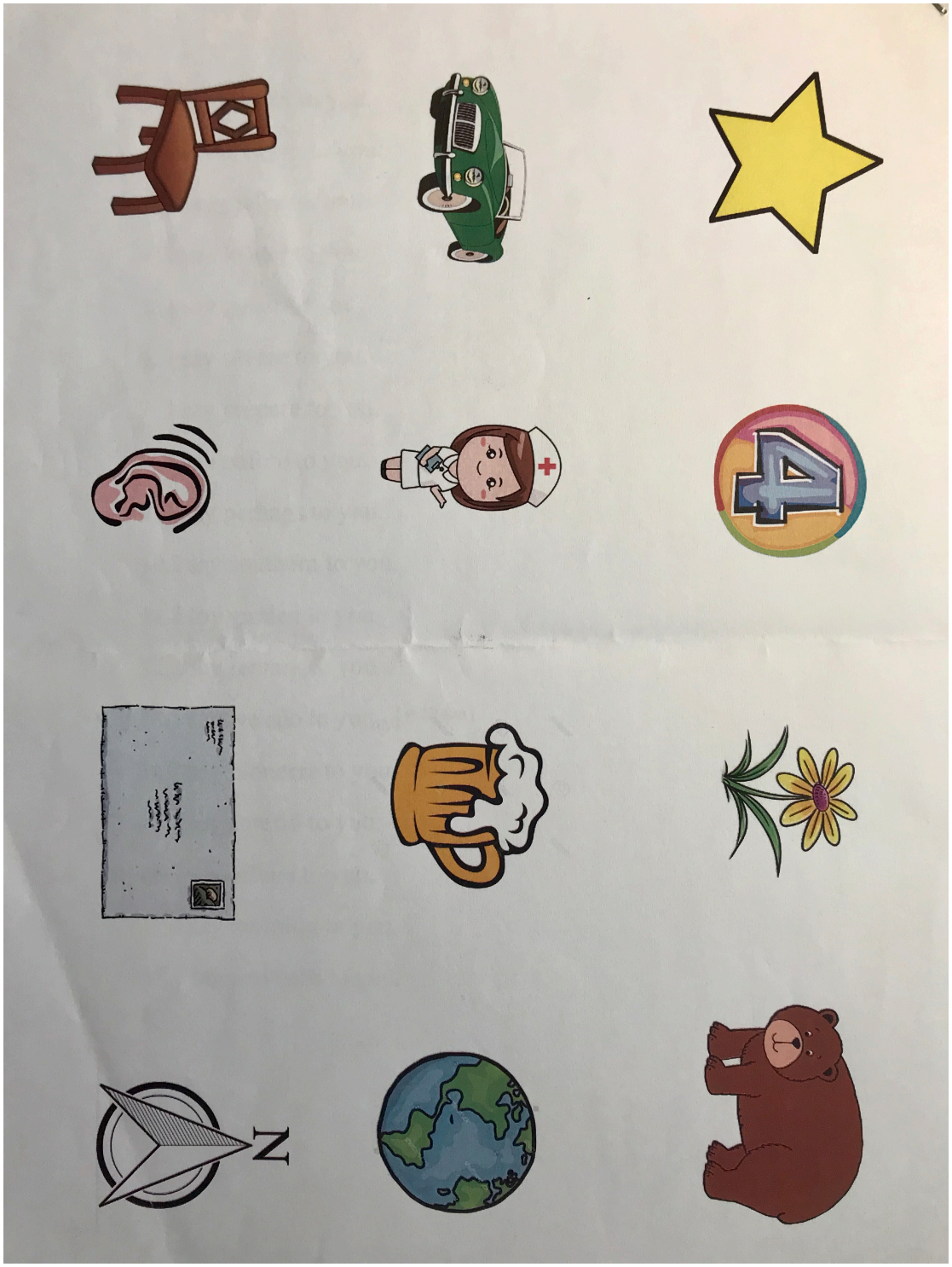
Test stimuli:

I say affairs to you.
I say before to you.
I say careful to you.
I say concern to you.
I say garden to you.
I say guitar to you.
I say louder to you.
I say morning to you.
I say newborn to you.
I say perfect to you.
I say weirdo to you.
I say perhaps to you.
I say pioneers to you.
I say prefer to you.
I say prepare to you.
I say remark to you.
I say severe to you.
I say southern to you.

Distractor stimuli:

I say alive to you.
I say arrive to you.
I say began to you.
I say begin to you.
I say begun to you.
I say breathe to you.
I say breeze to you.
I say came to you.
I say game to you.
I say crumble to you.
I say grumble to you.
I say label to you.
I say table to you.
I say sudden to you.

B2 Picture naming task



B3 English reading passage

The boy who cried wolf:

There was once a poor shepherd boy who watched his flocks in the fields next to a dark forest near the foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran down to the village shouting “Wolf, wolf.” As soon as they heard him, the villagers all rushed from their homes, full of concern for his safety, and two of them stayed with him for a while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful. However, not long after, a wolf was looking for a change in its usual diet of chicken and duck, so it actually did come out from the forest and began to threaten the sheep. Racing down to the village, the boy of course cried out even louder than before, but as all the villagers were convinced that he was trying to fool them a third time, nobody bothered to come and help him. And so the wolf had a feast.

B4 Mandarin reading passage

大好的周末天很不错，小赵决定收拾收拾自家菜园，出门摆摊赚点零花钱。家中的一儿一女尚是年幼，闹着要一起出门。小女儿最是爱美，出门前要妈妈给扎起了小辫。小小的人就像飞出了牢笼的小鸟，这摸摸，那瞧瞧，哪还能闲得住。不出一会，又转起了圈，嘴里还念叨着：一，二，三。