UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

L2 speech learning of European Portuguese /l/ and /ɾ/ by L1-Mandarin learners: Experimental evidence and theoretical modelling

Chao Zhou

Orientadoras: Profª. Doutora Maria João dos Reis de Freitas
Profª. Doutora Anabela Alves dos Santos Rato
Profª. Doutora Paula Fikkert

Tese especialmente elaborada para obtenção do grau de Doutor em Linguística

2021

UNIVERSIDADE DE LISBOA

FACULDADE DE LETRAS

L2 speech learning of European Portuguese /l/ and /ɾ/ by L1-Mandarin learners: Experimental evidence and theoretical modelling

Chao Zhou

Orientadoras: Profª. Doutora Maria João dos Reis de Freitas
Profª. Doutora Anabela Alves dos Santos Rato
Profª. Doutora Paula Fikkert

Tese especialmente elaborada para obtenção do grau de Doutor em Linguística

Júri:

Presidente: Doutora Ana Maria Martins, Professora Catedrática e Diretora da área de Ciências da Linguagem, da Faculdade de Letras da Universidade de Lisboa

Vogais:

- Doutora Silke Hamann, Professora Auxiliar da Faculdade de Humanidades da Universidade de Amsterdão
- Doutora Laura M. Colantoni, Professora Catedrática do Departamento de Espanhol e Português da Universidade de Toronto
- Doutor João Manuel Pires da Silva e Almeida Veloso, Professor Associado com Agregação da Faculdade de Letras da Universidade do Porto
- Doutora Maria João dos Reis de Freitas, Professora Associada com Agregação da Faculdade de Letras da Universidade de Lisboa, orientadora
- Doutor João Luís Marques Veríssimo, Professor Auxiliar da Faculdade de Letras da Universidade de Lisboa

Universidade de Lisboa - Bolsa de Doutoramento em Ciências da Linguagem 2017

2021

Dedication

To my dearest grandma 王传英 and grandpa 周长春 for their unconditional love, for teaching me to be diligent, tenacious and grateful.

Acknowledgements

There are many people who helped me during this extraordinary four-year journey and it is literally impossible to thank them all within a few pages. But I would like to express my most sincere gratitude to those who contributed greatly to the elaboration of this thesis.

First and foremost, my deepest thanks go to my supervisors, Prof. Maria João Freitas, Prof. Anabela Rato and Prof. Paula Fikkert.

Prof. Maria João brought me into the world of phonology and language acquisition six years ago. To her I owe most of what I know about phonological acquisition. With endless patience and constant encouragement, she has taught me to develop rigorous work and to cultivate free thinking. Throughout these years, she has never stopped being supportive either in research or in life. I am grateful for all her invaluable guidance, which has helped me to grow as a researcher.

Prof. Anabela has always been present since the very beginning of this project. She has been an example of critical thinking and commitment to science. I cannot thank her enough for our countless meetings, where we discussed literally every tiny detail of the experimental design. Without her dedication and trust for the last four years, I could not have finished this thesis.

Prof. Paula deeply shaped my view of phonological acquisition through her work when I was still a Master student. I am very grateful for having her in my advisory team. Her insightful feedback and suggestions pushed me to sharpen my thinking and brought this thesis to a higher level.

I am also grateful to Prof. Silke Hamann and Prof. Paul Boersma, who opened the door of their Lab at Amsterdam for me.

Prof. Silke is not only a leading scientist, a great teacher, but also a wonderful friend, who is always ready to listen and help. I am deeply indebted to her for all the patience, encouragement and mentoring. Her rigor and creativity in research and generosity in life have been a great inspiration for me.

Prof. Paul taught me everything I know about statistics, which changed fundamentally the way I design experiments and interpret data. The fact that his work shapes greatly my understanding of phonetics and phonology is clearly evidenced by this thesis.

Os meus agradecimentos também vão para os meus colegas e amigos da Universidade de Lisboa. Muito obrigado:

Aos meus colegas do Gabinete 7 e da Cave, Aida Cardoso, Alice Jesus, Ana Espírito Santos, Nathália Rodrigues, Nádia Canceiro, Nuno Matos, Patrícia Costa, Rita Santos, Rita Valadas e Silvana Abalada, pela vossa companhia, paciência, ajuda e por me fazerem sentir bem integrado desde o primeiro dia.

À Diana Oliveira, pela amizade, pela cumplicidade e pelo apoio no tratamento dos dados.

À Marisa Cruz, pela disponibilidade, pela partilha e pela incansável ajuda ao longo destes anos.

À Paula Luegi, pela generosidade, pela paciência e pela discussão sobre o desenho experimental.

Ao Rodrigo Pereira pela ajuda na gravação dos estímulos e na transcrição dos dados e também pelas discussões estimulantes sobre a Fonologia, especialmente sobre as consonantes róticas.

Aos meus amigos chineses que viviam em Lisboa, Shanyi Lao, Xinyi Li, Xinyi Zhang, Tianran Zheng, Weizhe Wang, Wenjun Gu, Yi Zheng, cuja companhia me ajudou a superar a saudade de casa.

I would like to express my gratitude to the researchers at the UvA speech lab, Camilla Horslund, Klaas Seinhorst, Mirjam de Jonge and Dirk-Jan Vet, for their help and kindness during my stay in Amsterdam.

I am also grateful to the members of my Project Defence Committee: Prof. Laura Colantoni, Prof. Isabel Pereira, Prof. Armanda Costa, whose suggestions helped improve the experimental design of this thesis greatly.

I would like to thank Prof. Gueorgui Hristovsky and Prof. Celeste Rodrigues for their guidance on phonological theory and for their kind help during my PhD study.

Special thanks go to Prof. Thaïs Cristófaro Silva, for showing her interest in my research, for many insightful suggestions and for sending me her wonderful book on Acoustic Phonetics.

I am indebted to the Centre of Linguistics of University of Lisbon, for financing my participation in conferences/workshops, which has been a crucial part of my academic training.

This thesis was developed with the great support of the University of Lisbon. I am deeply grateful for the PhD grant (Bolsa de Doutoramento em Ciências da Linguagem, 2017), which allowed me to focus on this PhD project.

Finally, I’d like to thank my family for their love and support, for providing me the best education they could and for creating opportunities for me to chase my dreams.

Table of Contents

List of tables

List of figures

Abstract

It has been long recognized that the poor distinction between /l/ and /ɾ/ is one of the most perceptible characteristics in Chinese-accented Portuguese. Recent empirical research revealed that this notorious L2 speech learning difficulty goes beyond the confusion between two L2 categories, as L1-Mandarin learners' acquisition of Portuguese /l/ and /ɾ/ seems to be subject to the interaction among different prosodic positions, speech modalities and representational levels. This thesis aims to deepen our current understanding of this L2 speech learning process, by exploring what constrains the development of L2 phonological categories across syllable positions and how different modalities interact during this process. To achieve this goal, both experimental tasks and theoretical modelling were employed.

The first study of this thesis explores the role of cross-linguistic influence and orthography on L2 category formation. In order to elicit cross-linguistic influence directly, a delayed-imitation task was performed with L1-Mandarin naïve listeners. This task examined how the Mandarin phonology parses the Portuguese input ([l], [ɾ]) in intervocalic onset and in word-internal coda position. Moreover, whether orthography plays a role during the construction of L2 phonological representation was tested by manipulating the input types that were given in the experiment (auditory input alone vs. auditory + written input). Our study shows that naïve Mandarin listeners' responses corroborated with that of L1-Mandarin learners, suggesting that cross-linguistic influence is responsible for the observed L2 prosodic effects. Moreover, the Mandarin [ɻ] (a repair strategy for /ɾ/) occurred almost exclusively when the written form was given, providing evidence for the cross-linguistic interaction between phonological categorization and orthography during the construction of L2 categories.

In the second study, we first investigate the interaction between speech perception and production in L2 speech learning, by examining whether the L2 deviant productions stem from misperception and whether the order of acquisition in L2 speech perception mirrors that in production. Secondly, we test whether L2 phonological categories remain malleable at a mid-late stage of L2 speech learning. Two perceptual experiments were performed to test L1-

Mandarin learners on their discrimination ability between the target Portuguese form and the deviant form employed in L2 production. Expanding on prior research, in this study, the perceptual motivation for L2 speech difficulties was assessed in different syllable constituents (onset and coda) and at both segmental and suprasegmental levels (structural modification). The results demonstrate that some deviant forms observed in L2 production indeed have a perceptual motivation ([w] for the velarised lateral; [l] and [ɾə] for the tap), while some others cannot be attributed to misperception (deletion of syllable-final tap). Furthermore, learners confused the intervocalic /l/ and /ɾ/ bidirectionally in perception, while in production they never misproduced the lateral (/ɾ/ → [l], */l/ → [ɾ]), revealing a mismatch between two speech modalities. By contrast, the order of acquisition (/ɾ/$_{coda}$ > /ɾ/$_{onset}$) was shown to be consistent in L2 perception and production. The correspondence and discrepancy between the two speech modalities signal a complex relationship between L2 speech perception and production. To assess the plasticity of L2 categories /l/ and /ɾ/, two groups of L1-Mandarin learners who differ substantially in terms of L2 experience were recruited in the perceptual tasks. Our study shows that both groups behaved similarly in terms of the discrimination performance. No evidence for a role of L2 experience was found. The implication of this null result on L2 phonological development is discussed.

The third study of the thesis aims to contribute to bridging the gap between the L2 experimental evidence and formal theories. Adopting the Bidirectional Phonology and Phonetics Model, we formalise some of the experimental findings that cannot be elucidated by current L2 speech theories, namely, the between and within-subject variation in L2 phonological categorization; the interaction between phonological categorization and orthography during L2 category construction; and the asymmetry between L2 perception and production.

Overall, this thesis sheds light on the complex nature of L2 phonological acquisition and provides a formal account of how different modalities interact in shaping L2 speech learning. Moreover, it puts forward testable predictions for future research and suggestions for improving foreign language teaching/training methodologies.

Keywords: L2 acquisition, phonology, Portuguese, Mandarin, liquids

Resumo

É bem conhecido o facto de as trocas associadas a /l/ e /ɾ/ constituírem uma das caraterísticas mais percetíveis no português articulado pelos aprendentes chineses. Recentemente, estudos empíricos revelam que a dificuldade por parte dos aprendentes chineses não se restringe à discriminação moderada entre as duas categorias da L2, dado que a aquisição de /l/ e /ɾ/ do português por aprendentes chineses parece estar sujeita à interação entre contextos prosódicos, entre modalidades de fala e entre níveis representacionais diferentes. Esta tese visa aprofundar a nossa compreensão deste processo da aquisição fonológica L2, explorando o que condiciona o desenvolvimento das categorias fonológicas L2 em diferentes constituintes silábicos e de que modo as modalidades interagem durante este processo, recorrendo para tal a tarefas experimentais bem como a formalização teórica.

O primeiro estudo averigua o papel da influência interlinguística e o da ortografia na construção das categorias de L2. Para elicitar a influência interlinguística diretamente, uma tarefa de imitação retardada foi aplicada aos falantes nativos do mandarim sem conhecimento de português, investigando assim como a fonologia do mandarim categoriza o *input* do português ([l], [ɾ]) em ataque simples intervocálico e em coda medial. Para além disso, a influência ortográfica na construção de representações fonológicas em L2 foi examinada através da manipulação do tipo do *input* apresentado na experiência (*input* auditivo vs. *input* auditivo + ortográfico). Os resultados da situação experimental em que os participantes receberam *input* de ambos os tipos replicaram o efeito prosódico observado na literatura, evidenciando a interação entre categorização fonológica e ortografia na construção das categorias de L2.

No segundo estudo, investigamos a interação entre a perceção e a produção de fala na aquisição das líquidas do PE por aprendentes chineses e a plasticidade destas categorias fonológicas, respondendo às questões seguintes: 1) as produções desviantes de L2 resultam da perceção incorreta? 2) a ordem da aquisição em L2 é consistente na perceção e na produção? 3) as categorias da L2 permanecem maleáveis numa fase intermédia da aquisição? Duas tarefas percetivas foram conduzidas para testar a capacidade percetiva dos aprendentes nativos do mandarim em relação à discriminação entre a forma alvo do português e as formas desviantes utilizadas na produção. No presente estudo, a motivação percetiva das dificuldades em L2 foi testada nos constituintes silábicos diferentes (ataque simples e coda) e nos níveis segmental e suprassegmental (modificação estrutural). Os resultados demonstram que algumas formas desviantes que os aprendentes chineses produzem têm uma motivação percetiva (i.e. [w] para a lateral velarizada; [l] e [ɾə] para a vibrante alveolar), enquanto outras não podem ser analisadas como casos de perceção incorreta (como é o caso do o apagamento da vibrante em coda). Para além disso, na posição intervocálica, os aprendentes manifestam dificuldade na discriminação entre /l/ e /ɾ/ de forma bidirecional, mas, na produção, a lateral nunca é produzida incorretamente (/ɾ/ → [l], */l/ → [ɾ]). Tal revela uma divergência entre as duas modalidades de fala. Por contraste, mostrou-se que a ordem da aquisição (/ɾ/$_{\text{coda}}$ > /ɾ/$_{\text{ataque}}$) é consistente na perceção e na produção da L2. A correspondência e a discrepância entre as duas modalidades de fala, sinalizam uma relação complexa entre a perceção e a produção na aquisição fonológica de L2. Em relação à questão da plasticidade das categorias de L2, recrutaram-se para as tarefas percetivas dois grupos de aprendentes nativos do mandarim que se diferenciavam substancialmente em termos da experiência em L2. Não se encontrou um efeito significativo da experiência da L2. A implicação deste resultado nulo no desenvolvimento fonológico de L2 foi discutida.

O terceiro estudo desta tese tem como objetivo contribuir para a colmatação das lacunas entre estudos empíricos de L2 e as teorias formais. Adotando o Modelo Bidirecional de Fonologia e Fonética, formalizamos os resultados experimentais que as teorias atuais da aquisição fonológica de L2 não conseguem explicar, nomeadamente, a variação inter e intra-sujeitos na categorização fonológica em L2; a interação entre categorização fonológica e ortografia na construção das categorias na L2; a assimetria entre a perceção e a produção na L2.

Em suma, esta tese contribui com dados empíricos para a discussão da relação complexa entre a perceção, produção e ortografia na aquisição fonológica de L2 e formaliza a interação entre essas modalidades através de um modelo linguístico generativo. Além disso, apresentam-se predições testáveis para investigação futura e sugestões para o aperfeiçoamento das metodologias de ensino/treino da língua não materna.

Palavras-chave: aquisição de L2, fonologia, português, mandarim, líquidas

Chapter 1: Introduction

It is widely acknowledged that learning a new language (L2) in adulthood entails a considerable amount of effort. Among different grammatical components, phonology is arguably the most challenging one for L2 learners (e.g. Ortega, 2009). Even after receiving several years of formal instruction and immersing in the community where the target language is spoken, an L2 learner may master an extensive vocabulary and produce well-formed sentences, but still speak with a noticeable foreign accent (e.g. Montrul, 2014; Dollmann et al., 2020).

One of the most perceptible characteristics in Chinese-accented Portuguese is the poor distinction between /l/ and /ɾ/, which has been long observed in the pedagogical literature:

"*…a tendência é para o (**r**) pronunciar como [l] em palavras como… ´Maria´ [malía]… Acontece ainda, por vezes, que o próprio **l** é substituído por [ɾ], como em [maɾoco] por ´maluco´…O **r** final de sílaba ou de palavra não oferece dificuldades, pois que se suprime: [fesá pota] ´fechar a porta´; [ké nan ké] ´quer ou não quer´*" [The tendency is to pronounce it ([ɾ]) as [l] in words like …*Maria* "a female name" as [malía]… It even happens sometimes that the [l] is replaced with [ɾ], such as *maluco* "crazy" produced as [maɾoco]… the syllable-final or word-final [ɾ] is not problematic, because it is omitted: *fechar a porta* "close the door" [fesá pota]; *quer ou não quer* "you want it or not" [ké nan ké]] (Batalha, 1995, p.15).

Despite many similar observations (e.g. Martins, 2008; Espadinha & Silva, 2009), systematic research on this notorious L2 difficulty is a relatively recent enterprise (Zhou, 2017; Liu, 2018; Cao, 2019; Vale, 2020). These experimental studies have revealed that the complexity of this L2 speech learning process goes beyond what was documented in previous descriptive studies, namely, the confusability between /l/ and /ɾ/ (Batalha, 1995; Martins, 2008), since L1-Mandarin learners' acquisition of this novel phonological contrast seems to be shaped by the interaction between different prosodic positions, speech modalities and representational levels.

A good understanding of this L2 speech difficulty is of great importance both theoretically, to understand the underlying phonological system of novel contrasts, and practically, to improve foreign language teaching/training methodologies. This thesis thus aims to broaden the current knowledge on this L2 speech learning scenario, by examining the acquisition of European Portuguese (EP) /l/ and /ɾ/ by Mandarin-speaking learners across prosodic positions (onset vs. coda), speech modalities (perception vs. production) and learning stages (initial, intermediate and advanced). Both laboratory experiments as well as theoretical modelling were employed, with the purpose of not only adding novel empirical evidence to the literature, but also providing a detailed formal account for the attested L2 speech phenomena.

The current chapter is organized as follows. We first discuss the phonetic and phonological properties of the liquid consonants in EP and in Mandarin, respectively, followed by a presentation of some seminal work in L2 speech learning literature, highlighting several potential sources for L2 speech deviation. The chapter concludes with a review of prior research on the acquisition of /l/ and /ɾ/ by L1-Mandarin learners and an outline of the research questions that will be addressed in the following chapters of this thesis.

1.1 Liquid consonants in European Portuguese and in Mandarin

Liquids constitute a class of consonants composed of laterals and rhotics. In the case of laterals, they are widely distributed cross-linguistically — 83.2% of the documented languages comprise at least one lateral (Maddieson, 2013). Lateral consonants are characterised by their manner of articulation, i.e. the airstream flows through both sides (or only one side) of the occlusion in the vocal tract (Ladefoged & Maddienson, 1996). Their place of articulation (the place where the occlusion is formed) may vary from dental/alveolar (e.g. Mandarin *lǎo* [**l**aw] tone: 3 “old”), to palatal (e.g. Portuguese *muralha* [muˈɾaʎɐ] “wall”) and velar (e.g. Korean *dalguji* [tɐʟgud͡ʑi] “cart”).

Rhotic consonants are frequently attested as well, being present in nearly 76% of the languages. However, in contrast to laterals, the criterion for grouping rhotics into a class is still controversial. Attempts on the basis of a single articulatory parameter have proven elusive (Lindau, 1985; Ladefoged & Maddieson, 1996; Wiese, 2011), since cross-linguistic rhotics manifest high variability both in manner (e.g. English approximant [ɹ]; Spanish trill [r], French fricative [ʁ]) as well as in place of articulation[1] (e.g. alveolar in Spanish, post-alveolar in English, uvular in French and glottal in Brazilian Portuguese). Although evidence on perceptual correlates for the rhotic class has been put forward in several studies, i.e. first (F1) and second formant frequencies (F2) and trajectories (e.g. Engstrand et al., 2007; Heselwood, 2009; Howson, 2018 a; 2018b; Howson & Monahan, 2019), only a reduced number of rhotic sounds were investigated. Future research on a larger set of cross-linguistic rhotics is warranted.

[1] The most common manner of articulation for rhotics is trill (47.5%) and 83.2% of the rhotics are apical (Maddienson, 1984).

The lack of phonetic invariance leads many linguists to postulate that the motivation for characterizing rhotics into a class is mainly phonological (Sebregts, 2014; Chabot, 2019, Natvig, 2020). For example, building on the family resemblance models[2] proposed by Lindau (1985), Sebregts (2014) integrated a diachronic dimension to the model, arguing that rhotics may be united by appeals to their shared history. Chabot (2019) further dissociated the phonetic dimension from his account and advocated that rhotics form a natural class on the basis of their phonological patterning, i.e. their status as sonorants, and their procedural and diachronic stability. Comparably, according to the representational account proposed by Natvig (2020), cross-linguistic rhotics belong to the same class because they are all unspecified sonorants at the underlying level; a particular surface realisation stems from their relationship to other liquid segments of the inventory and their phonological properties in a specific language.

In the following part of this section, we introduce the phonetic and phonological characteristics of liquids in the two languages involved in this project, namely the target language EP[3] and the learners' native language Mandarin.

1.1.1 European Portuguese

EP comprises four segments of the liquid class, two laterals /l/, /ʎ/ and two rhotics /ɾ/, /ʀ/ (Mateus et al., 2005).

The EP /l/ is traditionally described as exhibiting two allophonic variants, an alveolar lateral [l] in onset and a velarised [ɫ] in coda position (Mateus & Andrade 2000; Mateus et al., 2005), see examples in (1).

[2] Lindau (1985) argued that each member in the rhotic class shares some (articulatory) property with other members but no single property shared by all exists.

[3] In this thesis, EP refers to the standard variety spoken in the Lisbon area.

(1) *non-branching onset*

lima	[ˈlimɐ]	‘lime’	bolacha	[buˈlaʃɐ]	‘biscuit’
branching *onset*					
plano	[ˈplɐnu]	‘plan’	ciclismo	[siˈkliʒmu]	‘cycling’
coda					
saldo	[ˈsaɫdu]	‘balance’	anel	[ɐˈnɛɫ]	‘ring’

Acoustically speaking, the difference between [l] and [ɫ] lies in the distance between F2 and F1 frequencies: [l] has relatively high F2 and low F1 values (larger distance between F2 and F1), whereas F1 and F2 values are closer to each other in [ɫ] (Lehiste, 1964). With regard to articulation, the realisation of [ɫ] differs from that of [l] by including increased retraction of the tongue body and/or the tongue root, instantiated by lower F2 values, and lowering of the tongue predorsum (Browman & Goldstein, 1995).

The allophonic alteration of the EP /l/ has been challenged by acoustic and articulatory studies which demonstrated that the /l/ of EP always contains a certain degree of velarisation, regardless of syllable position and adjacent context (Andrade, 1999; Marques, 2010; Martins et al., 2010; Oliveira et al., 2011). Recently, Rodrigues and colleagues (2019) contributed to the debate with new acoustic evidence, demonstrating that /l/ has indeed consistently low F2 values in EP, due to velarisation across positions; however, its third formant values (F3), another acoustic correlate of degree of velarisation, are substantially higher syllable-final position, justifying the existence of two distinct allophones of /l/.

The other lateral in the Portuguese phonological inventory /ʎ/ is palatal and can only occur intervocalically, see (2). An articulatory study making use of Magnetic Resonance Imaging (MRI) revealed that the realisation of [ʎ] requires a complete contact of the tongue blade and/or pre-dorsum with the alveolo-palatal region (Martins et al., 2010; Teixeira et al. 2012), which is more front than the place described in impressionist studies, i.e. dorso-palatal zone (Sá Nogueira, 1938; *apud* Mateus & Andrade, 2000). Regarding the phonological status of EP /ʎ/, there is no consensus in the literature as it has been analysed as 1) a singleton phoneme, just as other EP consonants (Mateus & Andrade, 2000); 2) a geminate, due to its distribution and restriction on stress assignment[4] (Wetzels, 2000); or 3) a complex segment since diachronically it corresponds to the sequence Consonant-Palatal Glide in Latin (Veloso, 2019).

(2) *non-branching onset*

gralha [ˈɡɾaʎɐ]	‘mistake’	muralha [muˈɾaʎɐ]	‘wall’

In impressionist studies, the EP /ɾ/ is considered to be a tap, whose articulation requires a very rapid tongue tip movement against the alveolar ridge (e.g. Mateus & Andrade, 2000; Mateus et al., 2005). Acoustic evidence, by contrast, has suggested that the realisation of /ɾ/ may vary hinging on adjacent consonant and prosodic positions (Jesus & Shalde, 2005; Silva, 2014). In particular, in coda as part of a cluster /ɾ.C/, the occurrence of tongue tip closure together with a supporting vowel is favoured before a stop, but a fricative realisation is more common when the following consonant also carries friction (Silva, 2014); In word-final position, /ɾ/ is many times realized as a voiceless fricative (Jesus & Shadle, 2005) and can even be omitted, especially when the following word is initiated by a consonant (Mateus & Rodrigues, 2003; Rodrigues, 2003; Rodrigues & da Hora, 2016).

It is worth noting that the phonetic realisation of /ɾ/ may also be subject to dialectal variation. For example, an acoustic study on the southern variant

[4] The EP palatal lateral cannot occur in the last syllable of proparoxytones.

spoken in Algarve (Rodrigues, 2015) showed that /ɾ/ was most frequently produced as an alveolar fricative (35%) or an approximant (25%), whereas the canonical tap realisation was hardly attested (2%). Moreover, in the northern dialect spoken by educated youngsters in Oporto city, the retroflex flap [ɽ] or the English-like approximant [ɹ] are emerging as possible variants of syllable-final rhotic (Veloso, 2015).

Regarding the distribution of /ɾ/, it can occupy all prosodic contexts, apart from word-initial position (Mateus et al., 2005), see (3).

(3) *non-branching onset*

cara [ˈkaɾɐ]	'face'	muralha [muˈɾaʎɐ]	'wall'

branching cluster

gralha [ˈɡɾaʎɐ]	'error'	estrada [ʃˈtɾadɐ]	'highway'

coda

barco [ˈbaɾku]	'boat'	mar [ˈmaɾ]	'sea'

The EP /ʀ/ is most often produced as a voiced/voiceless fricative with the place of articulation ranging from velar to uvular (Rennicke & Martins 2013; Pereira, 2020). It only occupies the non-branching onset, as illustrated in (4).

(4) simple onset

rato [ˈ ʁatu]	'mouse'	borracha [buˈ ʁaʃɐ]	'rubber'

Some researchers propose only one underlying rhotic /ɾ/ in EP, contending that the word-initial surface [ʁ] stems from a strengthening process (/ɾ/ → [ʁ]) triggered by the left edge of a prosodic word (Mateus & Andrade, 2000; Vigário, 2003; 2019). The application of such strengthening rule can be likewise extended to contexts where /ɾ/ is preceded by another consonant (e.g. ho**n**[ʁ]a).

In order to account for the intervocalic [ɾ] – [ʁ] contrast, Mateus and Andrade (2000) argued that the intervocalic [ʁ] corresponds to an underlying geminate, i.e. mu[ʁ]o as mu/ɾ.ɾ/o "punch". The existence of an underlying coda /ɾ/ in a word like mu/ɾ.ɾ/o is supported by the fact that the lexical stress cannot be assigned to the antepenultimate syllable where an intervocalic [ʁ] is present, as the syllable containing intervocalic [ʁ] (underlying /ɾ.ɾ/) is considered to be heavy by occupying two skeletal positions. Accordingly, the occurrence of the intervocalic [ʁ] can be analysed as a result of the application of the same onset phonological rule that turns the underlying /ɾ/ preceded by another consonant into [ʁ][5].

By contrast, Bonet and Mascaró (1997) proposed that EP contains two rhotic segments at the underlying level and /ʀ/[6] is the default underlying rhotic in onset position. In their proposal, the intervocalic onset tap stems from a lexically marked rhotic and the coronal realisation in branching onset and coda are driven by the sonority scale. The analysis of two underlying rhotics in EP corroborates empirical evidence from L1 phonological acquisition: Portuguese children process [ɾ] and [ʁ] differently before adult-like production: laterals ([+sonorant]) are very often produced for the target /ɾ/, whereas stops ([-sonorant]) are employed for /ʀ/ (Costa, 2010; Amorim, 2014; Amorim & Veloso, 2018; Pereira et al., 2020).

1.1.2 Mandarin Chinese

Both lateral and rhotic are present in the Mandarin inventory, but they differ from the EP counterparts in terms of phonetic realisation and distribution.

[5] In particular, during the underlying to surface mapping, the first /ɾ/, which occupies the coda of the first syllable, is erased and the second /ɾ/, being preceded by a consonant (/ɾ/ in this case), undergoes the strengthening rule and surfaces as [ʁ] (Mateus & Andrade, 2000).

[6] Bonet and Mascaró (1997)'s analysis was originally proposed for major Iberian romance languages, to which the EP pertains. They assumed that the Portuguese word-initial rhotic is a trill, which corresponds to /ʀ/ in the standard variety.

The Mandarin lateral is phonetically an alveolar [l], occupying exclusively the domain of a non-branching onset, see (3).

(3) *simple onset*

là [la] (tone: 4) 'spicy' lǎo [law] (tone: 3) 'old'

The Mandarin rhotic /ɻ/[7] is legitimate both in onset and in coda, as illustrated in (4). While in onset position it can vary between approximant and fricative realisations (Zhu, 2007; Xing, 2019), in syllable-final position it is always an approximant and resembles the rhotic in English, in as far as it varies between bunched and retroflex articulations (Jiang et al., 2019).

(4) *simple onset*

ròu	[ɻow]	(tone: 4)	'meat'	rén	[ɻen]	(tone: 2)	'person'
coda							
ér	[əɻ]	(tone: 2)	'son'	èr	[əɻ]	(tone: 4)	'two'

1.1.3 Phonetic and phonological distinction between the alveolar lateral and the rhotic in EP and in Mandarin

To to best of our knowledge, there is no prior research on the acoustic or articulatory comparison between /l/ and /ɾ/ in the standard EP[8]. An acoustic study on the southern EP variant (Rodrigues, 2015), nevertheless, has suggested that, in intervocalic position, [l] and [ɾ] differ both in spectral (F1, F2, F3 formant values and F2 formant transition) and durational dimensions ([l] is substantially longer than [ɾ]). See for instance the acoustic values in Table 1.1.

[7] Notwithstanding the fact that it is conceivable to analyse the Mandarin rhotic either as an underlying fricative /ʐ/ or an approximant /ɻ/, the latter is widely accepted in the literature (e.g. Lin, 2001; Duanmu, 2005; Lin, 2007), since there is no phonological motivation for an underlying fricative /ʐ/, which would be the only voiced obstruent in the Mandarin phonological inventory and would introduce a novel phonological contrast [± voice] to Mandarin (Duanmu, 2007).

[8] The phonetic distinction discussed here is restricted to intervocalic position, because a previous study (Zhou, 2017) demonstrated that L1-Mandarin learners do not confuse the EP lateral and the tap in coda position.

In the case of Mandarin, it was shown that F2 and F3 formants as well as intensity are reliable cues for discriminating [l] from [ɹ] in onset position (see Table 1.2), whereas duration plays no role (Smith, 2010).

Table 1.1: Values of acoustic parameters of European Portuguese [l] and [ɾ] in intervocalic onset position (Rodrigues, 2015)

	Duration	F1	F2	F3
[l]	92 ms	375 Hz	1048 Hz	2540 Hz
[ɾ]	33 ms	410 Hz	1541 Hz	3474 Hz

Table 1.2: Values of acoustic parameters of Mandarin [l] and [ɾ] in intervocalic onset position (Smith, 2010)

	Duration	F1	F2	F3
[l]	104 ms	376 Hz	1137 Hz	2643 Hz
[ɹ]	96 ms	372 Hz	1459 Hz	2118 Hz

Following the hierarchical organization of features proposed by Clements and Hume (1995), Mateus and Andrade (2000) assumed that the phonological feature that distinguishes Portuguese /l/ from /ɾ/ is [lateral]. Alternatively, with the purpose of minimizing the lexical storage, Andrade (1977) specified Portuguese laterals as ([-continuant]), due to the presence of occlusion during articulation, and rhotics ([+continuant]), dismissing [lateral] from the features needed in Portuguese.

As for Mandarin, based on place of articulation, Duanmu (2007) proposed [± anterior] to distinguish between Mandarin /l/ and /ɹ/, while the value of feature [lateral] is always predictable, thus, redundant at the underlying level.

1.2 The origin of L2 speech deviation

One of the central goals of L2 speech learning research is to elucidate the divergence between the target form and the learners' output. Decades of studies on L2 speech have provided converging lines of evidence that many L2 speech learning difficulties can be attributed to cross-linguistic influence (CLI), i.e. an interaction between an individual's previous linguistic knowledge and the target language, see Major (2008) and Colantoni et al. (2015) for reviews. In particular, some researchers propose that CLI shapes L2 speech perception and production by acting like a perceptual sieve, modulating how the L2 input is parsed during the construction of L2 phonological representations (e.g. Polivanov, 1931; Trubetzkoy, 1977; Flege, 1995; Escudero, 2005; Best & Tyler, 2007); while some others advocate that CLI is detectable at the articulatory level, irrespective of learners' perceptual ability and whether L2 phonological representations are target-like or not (Honikman, 1964; Zimmer & Alves, 2012). Furthermore, accumulating evidence on L2 deviations that can neither be explained by the learners' L1 nor by the target language (e.g. Altenberg & Vago, 1983; Eckman, 1984) have led some researchers to deduce that L2 phonological acquisition may be constrained by certain phonetic (e.g. Colantoni & Steele, 2008) and/or phonological universals (e.g. Eckman, 1977; 2004).

Apart from the aforementioned speech-internal factors, which have been long studied in the field, the role of orthography in L2 speech learning has attracted increasing attention during the past few years. It is no surprise that orthography has been shown to shape both L2 speech perception and production (see Basseti et al., 2015 for a review), since the overwhelming majority of adult L2 learners are exposed to auditory and orthographic input simultaneously, from the onset of L2 phonological acquisition.

The following part of this section reviews some seminal studies that have made significant contribution to our understanding of sources for L2 speech difficulties, featuring different representational levels and modalities involved in learning a foreign language. We begin the discussion with a concise description of phonetic/phonological representations and their mappings encompassed in human speech perception and production processes. This helps contextualize the processes and representational levels investigated in L2 speech research.

1.2.1 Representational levels and their mappings in speech perception and production

Research on human speech can be split into two domains: speech perception and production. Historically, these two lines of inquiry have experienced limited mutual influence, because research methodologies and data analysis are quite distinct when aimed at direct observation of overt behavior, as in speech production, or investigation of abstract cognitive computation and representations, as in speech perception. However, models for speech perception and production, which build upon empirical evidence provided by a substantial amount of studies, have suggested that the two speech modalities may share some representations and mappings, while at the same time maintaining their domain-specific properties. In this section, we will discuss the representational levels and the mappings between them first in speech perception, and then in speech production.

Speech signals are extremely complex as they entail considerable variation induced by coarticulation, physiological differences, speaking rate, emotional state and many other speaker-specific or contextual factors, hinging on which the same sound can be perceived as different segments or different sound can be identified as a single category (e.g. Eisner & McQueen, 2018 for a review). How such complex speech signals are perceived and subsequently recognized as meaningful words by listeners is of great interest to any researcher studying human speech.

Many linguists, whose research covers diverse linguistic subfields, e.g. formal phonology (Boersma, 1998, 2007, 2011; Boersma & Hamann, 2009 a), psycholinguistics (McQueen & Cutler, 1997; Ramus et al., 2010), L1 phonological acquisition (Fikkert, 2007) and L2 speech learning (Escudero & Boersma, 2004; Darcy et al., 2013; Flege & Bohn, 2021), advocate that speech perception/comprehension is not a single module but consists of at least two level mappings, as in the model depicted in Figure 1.1.

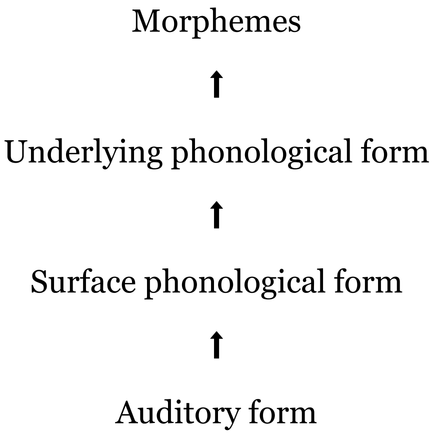


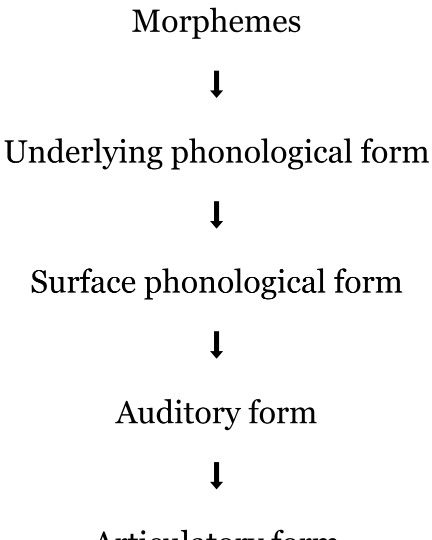
Figure 1.1: The speech perception process, based on McQueen and Cutler (1997) and Boersma and Hamann (2009a)

According to the model in Figure 1.1, in order to understand a word uttered by an interlocutor, the listener first abstracts away from noises, mapping speech-relevant acoustic information isolated from continuous speech stream, i.e. pitches, spectra, silences, transitions and durations (Auditory form) onto discrete phonological units (Surface phonological form)[9]. This mapping is designated in the literature as prelexical perception (McQueen & Cutler, 1997)

[9] The output of prelexical perception remains a matter of debate. It has been proposed to be phonemes (McClelland & Elman, 1986), features (Lahiri & Reetz, 2002), allophones (Mitterer et al., 2013), syllables (Church, 1987) or articulatory gestures (Fowler, 1986). Following Boersma (2011), we adopt the view that surface phonological form refers to a hierarchically organized tree-like structure of abstract phonological elements such as features segment, syllable, and other prosodic constituents. In principle, all these units with different sizes can be chosen as target in speech perception, which presumably hinges on the specific experimental task, see Samuel (2020) for a discussion.

or phonological categorization (Escudero & Boersma, 2004); these two terms will be used interchangeably henceforward. Subsequently, the listener accesses the intended lexical meaning by matching the perceived form with the phonological representation stored in the long-term memory (Underlying phonological form), which is linked to the semantic representation. This mapping between the two phonological forms is referred to as word/lexical recognition. The aforementioned two stages involved in speech comprehension may occur not only sequentially, but also in a parallel manner, as evident by the lexical effect on prelexical perception (Ganong, 1980; Norris et al., 2003) and by the influence of within-category differences on lexical activation (Andruski et al., 1994; McMurray et al., 2002).

Analogous modularity has also been proposed for speech production, as shown in Figure 1.2. This production model integrates the one proposed by Levelt (1989), as well as the production chain of the bidirectional perception-production model by Boersma and Hamann (2009a).



Morphemes
↓
Underlying phonological form
↓
Surface phonological form
↓
Auditory form
↓
Articulatory form

Figure 1.2: The speech production process, based on Levelt (1989) and Boersma and Hamann (2009a)

The above model depicts the modular processes involved in uttering a meaningful word. First, a speaker retrieves the underlying phonological form of an intended lexical entry from the long-term memory and translates it into a prosodically-detailed surface form [10] . This fully specified phonological representation is then converted into an auditory target form that the speaker aims to achieve, and later transformed by sensorimotor knowledge to an articulatory-motor representation, composed of a sequence of continuous gestural activities and coordination executed by the relevant muscles of articulators, including tongue, larynx, lips and lungs. Please note that in Levelt's proposal (Levelt, 1989), the surface phonological form is connected directly to the articulatory plan and no auditory form mediates between the phonetic and phonological level; however, this proposal seems to be problematic, as demonstrated by bite-block experiments[11], whereby the speakers adjusted to articulatory obstructions very fast, even if they are inhibited articulatorily (Lindblom et al., 1979; Gay et al., 1981). This provides evidence for the primacy of auditory goals in speech production (Shiller et al., 2010; see Boersma, 2011; Hamann, 2011 for more theoretical implications) and challenges the view that the articulatory form is the phonetic target as put forward by some linguists (e.g. Liberman et al., 1967; Fowler, 1986; Hale & Kissock, 2007). It has been argued that all mappings included in speech production proceed in a parallel fashion (see Melinger et al. 2014 for a review on empirical evidence), which is a crucial property that allows phonetic factors from "later" levels to influence phonological decisions from "earlier" levels. We will come back to this point in section 4.5.

One of the implications that can be drawn from the aforementioned human speech perception/comprehension mechanisms for L2 speech learning is that

[10] The surface phonological form equals the syllabified form in Levelt's terminology (Levelt, 1989).

[11] In a bite block experiment, participants are asked to produce speech sounds when the position of the jaw is fixed by a bite block.

building target-like underlying representations is a fundamental task. This is due to the fact that underlying phonological forms bridge two speech modalities by serving as target in the comprehension process and as input in production. As a result, deficit underlying forms may preclude learners from efficient speech comprehension, i.e. causing more lexical activation and reduced competition (e.g. Broersma & Cutler, 2011; Broersma, 2012; Cutler, 2015; Cook et al., 2016), and give rise to inaccurate speech production (Flege, 1995; Flege & Bohn, 2021). Accordingly, it comes as no surprise that understanding what deviates the formation of L2 underlying forms has turned into one of the main research interests of many L2 theorists and experimentalists. We review in the following section some L2 speech acquisition models, which will be reviewed for the purpose of explaining how a deviant L2 underlying form is constructed.

1.2.2 Perception-based explanation

Most current L2 speech theories acknowledge that the specification of an L2 underlying representation is contingent on phonological categorization (e.g. Flege, 1995; Best & Tyler, 2007; Escudero, 2007), due to the fact that the initial state of an underlying form should resemble how learners have perceived the auditory input[12] and the primary source for updating the underlying form also comes from prelexical perception. Optimal L2 phonological categorization is, however, not warranted, because the mapping from auditory forms to surface phonological forms is a language-specific process, which has been recurrently shown in cross-linguistic perceptual studies (e.g. Best, 1995; Flege, 1995; Escudero & Boersma, 2004). To take one example, the same auditory form of the English dental fricative [θ] is categorized as /t/ by L1-Russian listeners but

[12] In particular, when learning a novel word, one is normally presented with a pair of lexical meaning and auditory form. Then, she/he has to store what has been perceived from the auditory form in the lexicon as a temporary underlying form (Being temporary means the construction of an underlying form is gradual and subject to later changes), in order to "compare" it with future perceived forms for word recognition or to retrieve it for production.

as /s/ by L1-Japanese, despite the fact that both /s/ and /t/ are present in the Russian as well as in the Japanese phoneme inventories.

Language-specific phonological categorization implies that, during L2 speech learning, it is necessary to adjust the existing L1 category or even create a new sound category[13], as long as the L1 and the L2 differ in how to map an auditory form onto surface phonological representation. However, achieving target-like L2 categories very often imposes intractable problems, because L2 learners cannot resist resorting to their L1 prelexical mapping subconsciously and automatically when parsing the L2 input (Elvin & Escudero, 2019). Many L2 speech models have come up with a fairly adequate account for how the learners' L1 affects L2 category formation (Flege, 1995; Kuhl & Iverson, 1995; Escudero & Boersma, 2004; Best & Tyler, 2007; Strange, 2011), of which the Speech Learning Model (Flege, 1995; Flege & Bohn, 2021; henceforth: SLM) and the Perceptual Assimilation Model-L2 (Best & Tyler, 2007; henceforth: PAM-L2) are the two most widely tested ones.

The SLM proposes that the degree of distortion in L2 phonological categorization hinges on the perceived dissimilarity between an L2 sound and its closest L1 category. In particular, on the basis of a high number of cross-linguistic studies (see Flege, 1995 for a review), Flege and colleagues hypothesize that the relationship between L1 and L2 sounds exists on a continuum ranging from "identical" over "similar" to "new". For illustration, let us consider three prototypical types of L2 sounds for the moment: 1) **Identical** L2 sounds are deemed easy to master as they are exactly the same as L1 sounds and straight transfer from L1 to L2 will result in target-like performance immediately; 2) **New** sounds refer to those L2 sounds that do not resemble any (pre-existing) L1 category and, compared to identical sounds, new sounds

[13] The "sound category" we used here corresponds to the surface phonological form in Figure 1.1. Since this is the recurring terminology in L2 speech literature, we will use it when reviewing the L2 speech models.

require extra-learning of some novel aspects, but notable L1 interference is not expected due to a high degree of L1-L2 disparity; 3) **Similar** sounds, by contrast, are the most difficult since they are different but close enough to be regarded as "instantiations" of L1 categories. As a result, the formation of novel categories is impeded, at times even blocked, and learners have to rely on a composite L1-L2 category (diaphone) in L2 speech, which inevitably gives rise to imprecise L2 perception and production.

In sum, according to the SLM, the greater the perceived phonetic dissimilarity of a novel sound from its closest L1 counterpart is, the more likely this L2 sound will be acquired. It is important to note that the aforementioned predictions formulated in the SLM concern the learnability of L2 surface phonological forms (which are called "language-specific phonetic categories" in the SLM), as it is explicitly informed in Flege and Bohn (2020, p. 9-10) that it differs from the materials at the lexico-phonological level (underlying forms).

The SLM's predictions have been extensively investigated in studies on different L1-L2 pairs, among which the most well-known and representative one is the acquisition of English /l/ and /ɹ/ by L1-Japanese (e.g. Goto, 1971; Best & Strange, 1992; Iverson et al., 2003). Both English /l/ and /ɹ/ are novel categories for Japanese natives; the SLM, however, anticipates that /l/ would be more demanding to learn than /ɹ/, because the English lateral is perceived as a good exemplar of the Japanese category /ɾ/ (small perceived dissimilarity), while [ɹ] is not (large perceived dissimilarity) (Best & Strange, 1992). This was borne out in a longitudinal study whereby L1-Japanese learners' perceptual and production development of /l/ and /ɹ/ were examined (Aoyama et al, 2004).

Another well-recognized perception-based theory, PAM-L2, expands upon the Perceptual Assimilation Model (PAM) (Best, 1995)[14], which was developed

[14] In line with the direct realist view (Fowler, 1986) and Articulatory Phonology (Browman & Goldstein, 1989, 1992), PAM (-L2) assumes that non-native auditory forms are perceived as articulatory gestures, which are the building blocks of L2 lexical-phonological representation.

to account for how adult listeners perceive unfamiliar non-native sounds. Unlike the SLM, whose theoretical postulations are formulated on the basis of the relationship between individual L1 and L2 sounds, the PAM-L2 predicts the likelihood of forming a novel sound category, relying on a typology of diverse ways in which L2 contrasts can be assimilated to L1 sounds. In particular, if two L2 sounds are assimilated to two different L1 categories (**Two Category assimilation**), the creation of novel L2 categories is not likely to happen, because L1 categories will function fairly well for discriminating an L2 contrast, and thus no additional learning will be likely to be triggered. On the other hand, when both L2 phonemes are assimilated to the same L1 category but with a perceived difference in phonetic goodness-of-fit (**category-goodness assimilation**: one L2 sound is perceived as a good exemplar of that L1 category, whereas the other L2 sound as a poor exemplar), the learning of a novel category becomes possible; PAM-L2 reasons that, as long as learners are able to discern the difference between two L2 sounds (e.g. category-goodness assimilation), they will eventually recognise that this perceived difference signals a difference in meaning and a new phonological category would be created under such lexical pressure for the L2 sound regarded as the poor exemplar. By contrast, when both L2 phonemes are assimilated to the same L1 category and learners fail to detect the phonetic difference between them (**single-category assimilation**), a new L2 phonological category is unlikely to develop. The PAM-L2 also admits two other major types of categorization patterns, namely the “uncategorized assimilation” (an L2 sound is not categorized as any L1 category) and the “non-assimilable” (an L2 sound is heard as non-speech), and puts forward specific predictions on their learnability with respect to different assimilation types (see Faris et al., 2016 for empirical test on “uncategorized assimilation”).

To summarize, the SLM and the PAM-L2 converge on the idea that the construction of an L2 underlying form hinges on how the L2 sound is categorized, under CLI, and an imprecise underlying form will lead to inaccurate L2 production. This tight perception-production link has been advocated by the empirical evidence that L2 learners with better phoneme discrimination ability also produce the same contrasts with more accuracy (e.g. Perkell et al., 2004; Bion et al., 2006; Rauber et al. 2010; Brunner et al., 2011).

However, dissenting from the SLM, where the interrelatedness between the specification of L2 underlying forms and L2 phonological categorization is taken for granted, the PAM-L2 acknowledges the possibility that the L2 surface and underlying phonological forms may not resemble, leading to a divergence between L2 perception and production. For instance, a concrete example was put forward in Best and Tyler (2007): L1-Engish learners of French are able to perceptually detect the difference between English [ɹ] and French [ʁ] (distinction at the surface phonological level), but they still replace the French [ʁ] with native [ɹ] in the production of lexical items, presumably due to the fact that L1-English learners "equate the lexical-functional category /r/ across two languages (p.26)." Despite the fact that the conceivable distinction between the two phonological levels has been brought up in the paper where the PAM-L2 is officially introduced (Best & Tyler, 2007), it has received little attention in studies designed to test the PAM-L2, as they focus overwhelmingly on the differences in terms of discrimination accuracy across different assimilation types.

This intriguing and important research question on L2 category learnability has been carried on, though, in another line of research by psycholinguists (e.g. Weber & Cutler, 2004; Cutler et al., 2006; Darcy et al., 2012; Darcy et al., 2013; Kojima & Darcy, 2014; Amengual, 2016). On the one hand, some researchers have encountered cases where L2 learners maintain two separated L2 categories at the lexical level despite inaccurate prelexical categorization (Weber & Cutler, 2004; Cutler et al., 2006). Weber and Cutler (2004) performed an eye-tracking study, whereby L1-Dutch learners of English were asked to hear some English words containing /ɛ/ or /æ/ and then to match them to the pictured meanings. These two sounds are perceptually confusable for L1-Dutch listeners, who often categorized both English [ɛ] and [æ] as Dutch /ɛ/. It was displayed that, upon hearing a word with a novel L2 category /æ/ (e.g. panda /p**æ**nda/), L1-Dutch listeners partially looked at the picture of the competitor, a word comprising /ɛ/ (e.g. pencil /p**ɛ**nsɪl/), because the first syllable of the two words are perceptually ambiguous (both perceived as /p**ɛ**n…/) and two lexical entries were co-activated momentarily. While, upon hearing the familiar L1 category /ɛ/ (e.g. pencil /p**ɛ**nsɪl/), no temporary activation of /æ/ (e.g. panda /p**æ**nda/) was attested. Similar L2 asymmetric lexical access was documented by Cutler et al. (2006), who examined how the English contrast /l/-/r/ is coded lexically by L1-Japanese learners employing the same eye-tracking paradigm. Participants saw two pictures instantiating two English liquids respectively ("rocket" and "locker"), and in the meantime they heard either [l]ocker or [ɹ]ocket. In line with Weber and Cutler's results, when hearing [ɹ]rocket, participants looked towards the picture of "locker" for a short period, whereas their eyes did not fixate on the picture of "rocket" when the auditory stimulus was [l]ocker. Such asymmetric lexical activation, which was not manifested by native controls in either of the studies, suggests that the two contrasts in question (/æ/-/ɛ/ and /l/-/r/) are not fully merged in the L2 learners' lexicon and the contrast is somehow preserved. Otherwise, fixations would have been symmetrical. On the basis of the results from the above two experimental studies, it has been reasoned that the lexical distinction, which apparently does not come from categorization ability, might be achieved by speech-external/metalinguistic information, i.e. instruction that the two L2 sounds in question are supposed to be distinct (see Cutler, 2015 for a detailed discussion).

On the other hand, some researchers contributed to the debate by demonstrating that L2 lexical-phonological representation may still not be target-like even though learners already manifest accurate phonological categorization (e.g. Darcy et al., 2012; Darcy et al., 2013; Daidone & Darcy, 2014; Kojima & Darcy, 2014; Cook et al., 2016, Amengual, 2016; Simonchyk, 2017). Darcy and colleagues (2013), for instance, studied the processing of Japanese geminate/singleton contrasts and German front/back rounded vowel contrasts by L1-English learners. In their study, the L2 phonological categorization was tested with an ABX discrimination task[15], whereby the L1-English participants did not behave differently from the Japanese and German controls, indicating that the L2 phonological categorization resembles the target; however, the English natives were found to diverge from the Japanese and German controls in a lexical decision task by displaying asymmetrical lexical access, which is taken as evidence that the lexical representations of the L2 sound pairs in question though not fully merged are not target-like yet. Even more compelling evidence in support of the possibility that the two phonological forms involved in speech comprehension are not interdependent was put forward by Cook et al. (2016), who examined how the Russian contrast /t/ -/k/ is lexically coded in the lexicon of L1-English learners. Although /t/ -/k/ are not perceptually confusable for English natives, L1-English participants performed significantly worse from the Russian controls in a lexical priming task, suggesting that a high accuracy in phonological categorization will not guarantee accurate lexical distinction of a confusable L2 contrast. One possible explanation is that the development of L2 underlying representation might not occur with phonological categorization in tandem.

[15] During the ABX discrimination task, participants will hear 3 stimuli in sequence on each trial and will be asked to indicate whether the third stimulus (X) is more similar to the first (A) or the second (B).

The aforementioned theoretical and experimental studies have provided some insight into how the construction of L2 underlying forms may or may not be deviated by CLI on L2 phonological categorization. This conceivable dissociation between the two phonological levels opens up a possibility that L2 perception and production may not be always paired, highlighting that L2 underlying representation and phonological categorization need to be examined separately (Curtin et al., 1998; Escudero, 2005; Gor, 2015; Melnik, 2019).

Another implication of speech perception and production mechanisms for L2 speech learning is that the two speech modalities are inherently different by encompassing distinct phonetic components, in particular, only auditory forms in perception and both auditory and articulatory forms in production (see the comparison between Figure 1.1 and Figure 1.2). The existence of a production-specific representational level implies that some deviations observed in (L2) phonological acquisition may solely stem from articulatory difficulties, irrespective of phonological representations (Buchwald & Miozzo, 2011; 2012). In the next section, we provide overviews of two possible sources for the articulatory imprecision in L2 speech, namely CLI (Honikman, 1964) and phonetic (articulatory) universals (Colantoni & Steele, 2008).

1.2.3 Production-based explanation

The articulatory form is a continuous representation, composed of a sequence of continuous gestural activities and coordination executed by the relevant muscles of articulators, including tongue, larynx, lips and lungs. This imposes vexing problems for linguistic analysis, since the position of an active articulator may vary drastically depending on the segmental context. Nevertheless, it is plausible to assess the articulatory characteristics of a given language by tracking down a default rest position (Perkell, 1969), where "the set of postural configurations that the vocal tract articulators tend to be deployed from and return to in the process of producing fluent and natural speech" (Ramanarayanan et al., 2013, p. 510). This default articulatory position has been shown to vary across languages (Gick et al., 2004; Wilson & Gick, 2014), shedding insight into the fact that the articulatory settings are language-specific, i.e. there might be gestural differences with respect to the same phonological segment shared by two languages. Apart from this, languages may employ some articulatory movements that are not used by others (e.g. ejectives and clicks). The implication of the above-mentioned two facts has been underlined in the Articulatory Settings Theory, proposed by Honikan (1964). In particular, CLI is expected to occur at the articulatory level, when the learners' L1 and the target language differ in terms of articulatory settings.

Empirical evidence in support of CLI on the articulatory level can be found in various studies targeting different L1-L2 language pairs. Zimmer and Alves (2012) examined the production of English final stops by L1-Portuguese (Brazilian) learners and they reported that the closure duration in syllable-final stops was found to be significantly longer than in onset stops. Given that Brazilian Portuguese does not allow any stops in coda, Zimmer and Alves reasoned that Brazilian learners articulated longer closure syllable-final stops as articulatory compensation for the vowel that usually follows a stop consonant in their L1, suggesting that learners still articulate an L2 sound within their L1 articulatory settings. More straightforward evidence for CLI on L2 articulation was provided by Święciński (2013). With the aid of electromagnetic articulography, Święciński observed that, for advanced L1-Polish learners, a significant difference was found for tongue positions between their Polish and English productions, while no such differences existed for those learners, whose English was identified as elementary, indicating that L2 learners initially employ their L1 articulatory settings to produce L2 sounds and gradually acquire more target-like gestural configuration.

Apart from CLI, an L2 articulatory form may also be subject to universal articulatory constraints, which militates against articulatory effort (Kirchner, 1998) or favours salient phonetic parameters and positions (Ohala & Kawasaki, 1984; Strange, 1992). In phonetic research, it is well known that certain articulatory combinations, speech sounds or sequences can be realized with more ease than others. For instance, the well-recognized "Aerodynamic Voicing Constraint" (Ohala, 1983; 2011) states that stops are more likely to be voiceless, due to the fact that the oral occlusion makes it difficult to maintain the oral-subglottal air pressure difference, required for voicing. This is particularly true when the stop is followed by another stop or a pause (Colatoni et al., 2015). Therefore, the aerodynamic constraint on voicing provides a fairly adequate account for final obstruent devoicing observed in the L2 English production by L1-Hungarian learners (Altenberg & Vago, 1983), which cannot be attributed to CLI since neither Hungarian nor English allows a final devoicing rule.

Another articulatory constraint on L2 speech learning has been put forward in Colantoni and Steele (2008), whereby the L2 acquisition of the Spanish /ɾ/ and the French /ʁ/ by L1-English speakers was examined across different prosodic contexts. It was shown that English learners mastered both novel rhotics faster in intervocalic onset than in word-internal coda. Colantoni and Steele attribute this L2 onset-coda asymmetry to the fact that, when learning a novel segment, L2 learners usually target the most salient environment, i.e. [V_V]. The intervocalic onset position is considered to be more salient than the internal coda, where less transitional cues are available and consonant-consonant coarticulation may be required.

It is worth noting that, apart from a phonetic approach, the two aforementioned L2 phenomena can also be predicted by a phonological universal, proposed by Eckman (1977; 2004), formulated as the Markedness Differential Hypothesis (MDH). The MDH, which relies on the notion of *typological markedness*[16] (Greenberg, 1978), puts forward the following predictions (Eckman, 1977, p.321):

a. Those areas of the target language that differ from the native language and are more marked than the native language will be difficult*;*

b. The relative degree of difficulty of the areas of difference of target language that are more marked than the native language will correspond to the relative degree of markedness;

c. Those areas of the target language that are different from the native language but are not more marked than the native language will not be difficult.

The MDH thus provides a tool to predict and account for the L2 final-devoicing and the prosodic position effect, which cannot be simply explained by CLI. In particular, the observed preference for voiceless stops and onset position can be attributed to the fact that voiceless stops are unmarked with respect to their voiced counterparts and onset position is unmarked in comparison with coda position.

However, studies testing the MDH do not always provide support for it (e.g. Cichocki et al., 1999; Colantoni & Steele, 2008). For instance, the MDH predicts that, for L1-English learners, the Spanish alveolar rhotic should be easier than the French uvular rhotic, because the coronal place is argued to be universally unmarked (e.g. Lahiri & Reetz, 2010); however, the exactly opposite acquisition pattern (French uvular rhotic > Spanish alveolar rhotic) was reported in the cross-linguistic study performed by Colantoni and Steele (2008). Additionally, the MDH fails to account for the non-simultaneous mastery of all phonetic parameters involved in the acquisition of a novel segment. In particular, when learning a novel segment, a voiced uvular fricative, it has been shown that L2 learners target its manner, before place and voicing (Colantoni & Steele, 2008 for L1 English - L2 French; Zhou, 2017 for L1 Mandarin- L2 Portuguese). By

[16] A structure X is typologically marked relative to another structure, Y, (and Y is typologically unmarked relative to X) if every language that has X also has Y, but every language that has Y does not necessarily have X.

contrast, the phonetic approach postulates that manner is more salient than other properties (Steriade, 1999), offering an adequate explanation for the observed data.

Up till this moment, we have seen that L2 speech learning can be subject to CLI as well as to some universals. However, it is important to be aware of the fact that L2 speech is multimodal. In order to fully understand L2 speech learning, some sources external to speech should be considered, i.e. orthography. In the following section, we provide an overview of the orthographic influence on L2 phonological acquisition, which has attracted increasing attention in the literature over the past few decades.

1.2.4 Orthographic influence on L2 speech learning

L2 speech learning by literate adults is different from child L1 acquisition in may ways, one of which lies in the sequence in the development of speech and literacy.

Speech precedes literacy in L1 acquisition, implying that meaning is initially linked to phonological representation. Afterwards, as literacy is attained, children must learn to match distinctive visual symbols to units of sound. The mastery of this sound-symbol mapping allows them to access the words already present in their spoken lexicon and later to establish mappings between meaning and symbols, known as reading (for a review, see Ziegler & Goswami, 2005). Since the development of reading is mediated by phonology (learning the sound-symbol mapping precedes the symbol-meaning mapping), it comes as no surprise that phonology is actively involved in learning to read by children (see for instance, Van Orden & Goldinger, 1994; Frost, 1998; Ziegler et al., 2001). Moreover, the onset of reading has been shown to impose constraints on the children's phonological development (e.g. Goswami et al., 2005) and the orthographic effect can be observed in adults' perception and production in a variety of experimental paradigms (e.g. Ziegler & Ferrand, 1998;

Perre & Ziegler, 2008), as well as in the phonological information stored in the lexicon (e.g. Ziegler et al., 2003; Taft, 2006; Nevins & Vaux, 2007; Ranbom & Connine, 2007). These findings all together can be taken as evidence in favour of the interactive relationship between phonology and orthography.

For literate adults, it seems to be efficient and common practice to learn a new language having both auditory and written material at their disposal, in particular for those in instructed settings. This indicates that adult L2 speech learning is, in fact, multimodal by nature[17]. Although the orthographic influence on L2 speech is to be expected, it is rather surprising that the widespread interest in orthographic effects on L2 phonology solely surfaced two decades ago (see Bassetti et al., 2015 for a review). Recent research has demonstrated that orthographic input can affect L2 phonological development in various ways: the exposure to orthographic input may facilitate speech perception, production and word learning (e.g. Escudero et al., 2008; Showalter & Hayes-Harb, 2013), hinder target-like acquisition (e.g. Bassetti, 2007; Hayes-Harb et al., 2010) or manifest mixed or null effects (e.g. Escudero & Wanrooij, 2010; Simon et al., 2010).

Recall the case presented in Section 1.2.2, where L2 learners maintain a distinction between two categories at the lexical level, in spite of inaccurate phonological categorization (Weber & Cutler, 2004; Cutler et al., 2006). These data suggest that that the lexical distinction might be achieved through a speech-external source, e.g. orthography, since the perceptually confusable L2 sounds (/ɛ/ - /æ/ in Weber & Cutler and /l/ - /ɹ/ in Cutler et al.) are represented by distinct graphemes. Whether the lexical distinction can be cued by orthography was explicitly tested in Escudero et al. (2008), where L1-Dutch speakers were asked to learn 20 English pseudo-words, constituting 10 minimal pairs with the confusable contrast /ɛ/ - /æ/. During the learning phase, half of

[17] See also Navarra & Soto-Faraco (2007) for the impact of another visual information (lip movements) on L2 phonological categorization, which reminds of the famous McGurk effect (McGurk & MacDonald, 1976).

the participants were assigned to the auditory condition, where they received auditory input of the test items, whereas the other half participated in the auditory-orthographic condition, where both auditory and orthographic forms were given; In the testing phase, upon hearing words either containing /ɛ/ or /æ/, the participants from the auditory condition looked towards pictures representing words with /ɛ/ and words with /æ/ randomly, suggesting that /ɛ/ and /æ/ were stored as a single segmental category in the L2 lexicon. In contrast, learners from the auditory-orthographic condition manifested the same asymmetric lexical activation as reported in previous studies (e.g. Weber & Cutler, 2004; Cutler et al., 2006; Darcy et al., 2013), evidencing that a lexical distinction has been implemented due to the presence of orthography.

Although the access to orthographic information may help L2 learners to resolve, at least partially, perceptual confusability during the construction of lexical-phonological representation (e.g. Escudero et al, 2008; Cerni et al., 2019), in some cases it can display a hindering effect (Escudero & Wanrooij, 2010; Hayes-Harb et al., 2010). Hayes-Harb et al. (2010) conducted a novel word learning experiment with L1-English speakers, similar to the one in Escudero et al. (2008), manipulating the L1-L2 orthographic congruency. In particular, one group of participants received orthographic forms of words that were congruent with L1 English spelling conventions (e.g. <fa**z**a> - [fa**z**ə]), while participants in the other group were exposed to orthographic forms that were incongruent with English grapheme-phoneme convention (e.g. <fa**z**a> - [fa**ʃ**ə]). Participants were later asked to perform a lexical decision task (picture-auditory matching paradigm) and results showed that the *incongruent* group performed significantly worse than the *congruent* group. This results together with the similar findings in Escudero et al. (2014) suggest that the orthographic effect on L2 speech learning is contingent on the congruence of phoneme-grapheme mappings across L1 and L2. In particular, presenting orthographic information during L2 phonological acquisition has a positive effect when the non-native grapheme-phoneme correspondences are congruent with native one, while having a negative effect when they are incongruent.

The L1-L2 orthographic incongruence can not only slow down the construction of an L2 contrast at the lexical level (Bassetti, 2007; Escudero et al., 2014; Rafat, 2016), but it may even determine the specification of an L2 underlying representation. For instance, cue integration (comparable to the McGurk-like effect; McGurk & MacDonald, 1976) was reported in Rafat and Steveson (2018), where L1-English learners produced the Spanish target po[lj]o ("chicken") by integrating the auditory (po[j]o) and the orthographic information (<pollo>) from the input. It has also been demonstrated in Bakker et al. (2014) that a phonological representation created solely on the basis of orthographic input (e.g. through grapheme-phoneme convention) can compete with familiar words during spoken word recognition, indicating lexical consolidation across modalities[18].

In contrast to many studies examining the orthographic effect on L2 phonological acquisition, Escudero and Wanrooij (2010) reported mixed results regarding the cross-linguistic orthographic influence. They examined the effects of L1 orthographic input on the L2 perception of five Dutch vocalic contrasts by L1-Spanish learners, /a-ɑ/, /i-ɪ/, /i-y/, /ɪ -ʏ/ and /y-ʏ/. After being first examined on their perception of these Dutch contrasts without written input, the Spanish participants were given Dutch spellings of the auditory stimuli during an XAB discrimination task. In particular, the participants heard a vowel (X) and were instructed to indicate whether it was similar to the second (A) or third auditory (B) token accompanied by their orthographic representation in Dutch. The results revealed that the orthography manifested different effects as significant improvement in perception was only found for

[18] It has been also shown in Bakker et al. (2014) that auditorily acquired novel words entered into the competition in the written modality as well.

/a-ɑ/, but not for other contrasts. Escudero and Wanrooij attributed this result to the alignment of both the acoustic and orthographic information, because the durational difference between /a/ and /ɑ/ is cued orthographically (<a> for /a/ and <aa> for /ɑ/), in contrast to other difficult sound pairs under study such as the contrast /i-ɪ/, where the acoustic difference cannot be reinforced by the presence of orthography, since both sounds can be represented by <i>.

As reviewed above, to date, the overwhelming majority of the studies on L2 orthographic effect have focused on the acquisition of phonemic contrasts in the target language, while the role of orthography in learning L2 phonological process, which is as important as the phoneme inventory for phonological acquisition, has not been addressed. The only exception is Barrios and Hayes-Harb (2020), where the orthographic influence on the L2 learning of German final devoicing (e.g. e.g. [vɛɐ̯bən] verb.plural; [vɛɐ̯p], verb.singular; <verb>) by L1-English learners was investigated. In order to simulate the real learning scenario, both items that exhibit morphophonological voicing alternation (e.g. [kʁap]-singular, [kʁabən]-plural, <krab>-singular and <kraben>-plural) and those that do not (e.g. [tʁap]-singular, [tʁapən]-plural, <trap>singular and <trapen>-plural) were included in the experimental design. It was found that, irrespective of input types (auditory vs. auditory + orthographic), learners could infer the morphologically conditioned voicing alternation in the acoustic input to acquire the investigated L2 phonological process to some extent; however, this learning effect is only statistically significant in learners who were only exposed to the auditory forms, not in those who had access to orthography. This evidences that the conflicting information between modalities ([kʁa**p**] and < kra**b**>) hinders the L2 acquisition of phonological process. In particular, learners who were exposed to written forms during word learning tend to simply follow the orthographic cue (e.g. < kra**b**> <kra**b**en>), producing the underlying voiced forms, both singular and plural, as surface voiced forms, i.e.

*[kʁa**b**] and [kʁa**b**ən], ignoring the voicing alternation present in the acoustic input to some extent.

Despite the important role that orthography plays in L2 phonological acquisition, to our best knowledge, no current L2 theories incorporate the orthographic effect into formalisation; however, in order to fully understand L2 speech learning, which is multimodal by nature, it is necessary to take orthography into consideration.

1.3 Difficulties in the acquisition of European Portuguese /l/ and /ɾ/ by L1-Mandarin learners

The documentation of Chinese learners' struggle with EP /l/ and /ɾ/ can be traced back to Batalha (1995). It comes as a surprise, though, that it was not until recently that this notorious difficulty was investigated systematically in experimental studies (Zhou, 2017, Liu 2018; Cao, 2019; Vale, 2020).

Cao (2018) deliberated on how L1-Mandarin learners perceive /l/ and /ɾ/ in intervocalic position and on whether L2 perception accuracy is mediated by L2 experience. Twenty Mandarin-speaking participants who studied EP in a formal setting for 2.9 years were recruited in China (Group 1) and another twenty were tested in Portugal, where they were in a Portuguese language immersion program for 9 months after having received two years of formal instruction in China (Group 2). Results from an AX discrimination task indicated that the Mandarin speaking participants failed to reliably discern the difference between /l/ and /ɾ/ in an AX discrimination task (M= 0.49) and L2 perceptual discrimination seemed to be mediated by the immersion experience: learners that studied abroad outperformed those without immersion experience (Group 1: M=0.41 and Group 2: M=0.58). On the other hand, a forced-choice identification task further revealed that both EP /l/ and /ɾ/ impose perceptual difficulty (M=0.72 for /l/ and M=0.62 for/ɾ/) and L2 phonological development might be shaped inversely by L2 experience, which speeds up the target-like formation of /ɾ/ (Group 1: M=0.53 and Group 2: M=0.7), while hinders the development of /l/ (Group 1: M=0.83 and Group 2: M=0.63). Note that no test on statistical significance was performed in Cao (2018), precluding confirmation on any observed effect pertaining to the L2 experience.

The perceptual distortion with intervocalic /l/ and /ɾ/ was likewise reported in Vale (2020), where the L2 perceptual categorization of eleven L1-Mandarin subjects who studied EP for 2 years was examined (M=0.47 for /l/ and M=0.45 for/ɾ/). In addition, it was found that the categorization accuracy was higher in certain adjacent vocalic contexts (e.g. /a_a/, M=0.51) than others (e.g. /i_i/, M=0.44). Again, the lack of statistical analysis does not allow any conclusion to be drawn regarding the effect of vocalic context on L2 phonological categorization.

In contrast to perceptual studies, production studies by Zhou (2017), where all EP liquids were assessed, and by Liu (2018), who solely focused on the rhotics, took into account the prosodic positions where liquids may occur.

Fourteen L1-Mandarin learners with homogenous L2 experience (2-year formal instruction + 3 months immersion) participated in a picture-naming task in Zhou (2017) and the results indicated that /l/ was never mispronounced in onset position (Mean Accuracy: 1), while very often vocalised syllable-finally (in 61% of the cases); As for /ɾ/, learners produced more target-like instantiations in coda than in onset (M = 0.69 in coda and M= 0.39 in onset, $\chi^2(1) = 9.87$, *p*=0.002[19]) and, when failing to produce [ɾ], learners used [l] exclusively in onset, whereas in coda they deleted the segment, inserted a schwa (and thus create a new onset) or replaced it with [l] or [ɻ].

The above-mentioned onset-coda asymmetry with respect to L2 production of /ɾ/ was replicated in Liu (2018), where both a picture-naming and a text reading task were performed. It is worth noting that, in addition to [l] or [ɻ], Liu also attested that her participants, who only had received formal instruction of EP for less than a year, replaced the syllable-final /ɾ/ with coronal stops [d, t, t^h] and [n].

[19] The statistical analysis was performed in Zhou et et al. (2020), where the data from Zhou (2017) were reanalysed.

In sum, in onset position, L1-Mandarin learners confuse the EP /l/ and /ɾ/ bidirectionally in speech perception, both [l] → /ɾ/ and [ɾ] → /l/ (Cao, 2018; Vale, 2020) and the category development seems to be subject to L2 experience (Cao, 2018) and adjacent segmental context (Vale, 2020). Nevertheless, in L2 production, the difficulty was restricted to /ɾ/; /l/ → [ɾ] was not attested (Zhou, 2017); in coda, the confusability between /l/-/ɾ/does not hold, because the EP lateral tends to be vocalized as [w] (Zhou, 2017) and /ɾ/ may undergo both segmental or structural modifications (Zhou, 2017; Liu, 2018).

In this thesis, we concentrate on the L2 phonological acquisition of EP /l/ and /ɾ/ in intervocalic onset and word-internal coda positions. The branching onset was not considered, due to the fact that L1-Mandarin learners did not manifest difficulty with this novel syllable structure, i.e. no structural repairs such as epenthesis or deletion were reported (Zhou, 2017).

Word-final coda was not included either, due to the questionable syllable status of the EP liquids in this position. Morales-Front and Holt (1997), for instance, have argued that word-final /l/ occupies syllable nucleus. This is evidenced by the fact that, when followed by a tautosyllablic consonant (e.g. fricative /s/, forming a plural form), /l/ surfaces as [j] (e.g. anima/l/ → anima[i]s). As for /ɾ/, based on the evidence that the word-final tap, but not the word-internal one, blocks the unstressed vowel reduction, Carvalho (2006) postulated that the EP /ɾ/ occupies the 'true' coda position word-internally, whereas an onset preceding an empty nucleus word-finally. This may explain why the insertion of schwa is only possible in word-final position. Moreover, the word final tap (of a verb) in EP is the infinitive marker. The impact of interface between phonology and morphology on L2 phonological acquisition is beyond the scope of the current project.

1.4 Outline of the following chapters

As reviewed in 1.3, although previous studies on the L2 acquisition of EP /l/ and /ɾ/ by L1-Mandarin learners are still scarce, the complexity of this L2 speech learning process was reported and several pertinent questions were raised, whose answers may contribute substantially to our understanding of the underlying mechanisms behind L2 phonological acquisition. This thesis aims to further investigate these questions, exploring possible explanations, by making use of both laboratory experiments and theoretical modelling. The thesis is organized as three independent but closely connected studies:

The first study looks into what underlies the prosodic effect observed in the acquisition of Portuguese /l/ and /ɾ/ by L1-Mandarin learners. As reviewed in section 1.2, there is consensus that most divergence between the learners' output and the target form can be attributed to CLI. Therefore, in chapter 2 we first examined whether CLI contributes to the L2 prosodic effects, by testing how naïve Mandarin speakers, without any knowledge of Portuguese, parse the Portuguese speech input. We also manipulated the types of input (auditory vs. auditory + orthographic) in our experimental task to assess the conceivable orthographic effect. This study on naïve phonological categorization has also the applied benefit of outlining the initial state of L2 speech learning and allowing to generate predictions on subsequent L2 phonological development.

The second study examines how speech perception and production interact in the L2 speech learning of Portuguese /l/ and /ɾ/ across different prosodic contexts and whether L2 experience shapes this process at a non-initial learning stage (learning length: 2-8 years). In chapter 3, we performed two perceptual experiments to test L1-Mandarin learners on their discrimination ability between the target Portuguese forms and the deviant forms that they often employ in production. The first goal of this study is to set apart perception-induced difficulties and production-based errors, as we reasoned that a deviant form would have a perceptual basis if learners failed to discriminate it from the target form reliably. Expanding on prior perceptual studies (Cao, 2018; Vale, 2020), we investigated the potential perceptual confusability across syllable constituency and took both segmental replacement as well as structural modifications into account. Moreover, we investigated whether L2 perception and production coincide with respect to the developmental sequence of the EP tap. A correlation would be a strong indicator towards the fact that L2 segmental acquisition is not only subject to the relationship between L1 and L2 categories, but also constrained by more abstract phonological restrictions. The second objective is to explore the plasticity of L2 phonological representations of EP /l/ and /ɾ/ at a mid-late learning stage. In order to achieve this, we recruited two groups of L1-Mandarin learners of EP differing substantially from each other in terms of L2 experience. If these categories remain malleable, we would expect to attest a difference between two groups of participants.

Apart from the experimental studies, in chapter 4, we provide a formal account for several L2 speech phenomena emerged during the acquisition of EP /l/ and /ɾ/ by L1-Mandairn learners, namely the between-subject (4.3.1) and within-subject variations (4.3.2) in L2 phonological categorization, prosodic effect in L2 phonological categorization (4.3.3), the interaction between phonological categorization and orthography during the construction of L2 underlying representations (4.4) and the L2 perception-production asymmetry (4.5). Our theoretical modelling within the Bidirectional Phonology and Phonetics Model (Boersma & Hamann, 2009a; Boersma, 2011; Hamann & Colombo, 2017) not only offers a fairly adequate account for all aforementioned L2 phenomena, which other L2 models cannot explain, but also put forward testable predictions for future studies.

Finally, in chapter 5 we revisit the findings of this thesis and discuss remaining question and directions for further research.

Chapter 2: Naïve categorization of European Portuguese /l/ and /ɾ/

2.1 Introduction

Recent research (Zhou, 2017; Liu, 2018) has shown that the acquisition of the European Portuguese (EP) lateral /l/ and rhotic /ɾ/ by L1-Mandarin learners is constrained by prosodic context: learners almost never mispronounce EP /l/ in onset, but in coda they frequently vocalize it as [w]. EP /ɾ/ is replaced with [l] in onset, while in coda learners delete the segment, insert a schwa (and thus create an onset), or substitute it either with [l], a coronal stop [t, t^h] or the Mandarin rhotic [ɻ].

Decades of studies on L2 speech have led researchers to converge on the idea that most divergence between the learners' output and the target form can be attributed to CLI (cross-linguistic influence), i.e. an interaction between the learner's mother tongue and the target language, which has constituted the core assumption in most, if not all, L2 speech theories (Flege, 1995; Kuhl & Iverson, 1995; Escudero & Boersma, 2004; Best & Tyler, 2007; Strange, 2011).

In the present chapter, we first explored whether CLI can also explain the observed prosodically-conditioned repair strategies applied by Mandarin learners in the production of the EP liquids. To achieve this, we tested with a delayed imitation task how L1-Mandarin speakers without any knowledge of Portuguese (henceforth: naïve listeners) parse EP /l/ and /ɾ/ in different prosodic positions. If CLI is indeed responsible for the position-dependent treatment of EP liquids, then we expect to observe similar prosodically-conditioned repairs by naïve Mandarin speakers. The imitation results of the imitation task will also outline the initial state of L2 speech learning of the EP /l/ and /ɾ/, serving as the point of departure for understanding and modelling L2 phonological development (see chapter 4).

For the previously-reported replacement of EP /ɾ/ by the Mandarin rhotic [ɻ] in coda position, which is similar to the realisation of [ɻ] for the Spanish tap by L1-Mandarin learners (Patience, 2018), we furthermore tested whether this stems from CLI via phonological categorization or via orthography: From the point of categorization, the EP /ɾ/ and the Mandarin /ɻ/ could be argued to share acoustic-perceptual cues, such as first (F1) and second formant frequencies (F2) and trajectories (Howson, 2018), or phonological features (Hall, 1997), which would result in EP /ɾ/ being perceived as Mandarin /ɻ/. Evidence for this comes from a non-native perception experiment where various types of rhotics, i.e. [r, ɻ, ʀ], were treated as one class (Howson & Monahan, 2019). Alternatively, the replacement with [ɻ] could be driven by orthography, since both Mandarin /ɻ/ (in Pinyin) and Portuguese /ɾ/ are represented by the grapheme <r>, and L2 adult learners receive written input from the onset of L2 speech learning. If this modality of CLI were the reason for the replacement, we would expect a Mandarin /ɻ/ only to occur if naïve listeners were presented with written forms from which they could deduct the phoneme equivalence. In order to test whether the replacement with the Mandarin rhotic stems from perception or orthography, we manipulated the input types (auditory only vs. auditory together with orthography) that were provided in our experiment.

2.2 Method

As reviewed in Bohn (2017), attempts to assess CLI in L2 speech learning include comparing the phonetic or phonemic symbols used for transcribing the target sounds of the two languages (e.g. Briere, 1968; Collins & Mees, 1984), analysing acoustic properties (e.g. Bohn & Flege, 1990; Escudero & Boersma, 2004) and observing directly how L2 sounds are mapped to L1 categories (e.g. Flege, 1995; Best & Tyler, 2007). In the current study, we adopt the last approach to explore whether CLI is responsible for the prosodic effect on L2 production of EP /l/ and /ɾ/, by testing how the EP liquids are categorized by naïve Mandarin listeners across prosodic contexts (intervocalic onset and word-internal coda) in a delayed imitation task.

This task was deemed especially suited for the present study for the following reasons. First, delayed imitation responses were shown to be mediated by phonological processing (Schouten, 1977), and L2 imitation was shown to be strongly related to L2 phonological categorization (Llompart & Reinisch, 2019). Accordingly, we reason that naïve imitators will only produce what they perceive and thus naïve imitation reflects how unfamiliar sounds (e.g. EP liquids) are mapped to L1 phonological categories. Second, in comparison with other direct measures of CLI, an imitation task avoids using orthography as response labels, and also saves participants from a large number of trials, needed, for example in a graded rating perceptual similarity task (Flege et al., 1994)[20].

To test whether the use of the Mandarin rhotic is perceptually or orthographically driven, two experimental conditions were created. In the auditory condition, naïve listeners merely received auditory input containing

[20] In this experimental approach, listeners are asked to compare instantiations of L1 and L2 sounds and rate them on a scale ranging from “very similar” to “very dissimilar”; however, this task has not been widely used, presumably due to the fact that participants may have to provide ratings of perceptual similarity on a very large number of stimulus pairs. For instance, Flege and colleagues (1994) asked their participants to compare three Spanish vowels with seven English vowels in one phonetic context and thus the participants had to rate 405 vowel pairs in total.

the target segments in different syllable positions (onset and coda). In the orthographic condition, both auditory and written forms of test items were presented simultaneously. We refrained from including a condition where participants would only receive orthographic input, because we deemed this uninformative, as naïve listeners would simply apply their native grapheme-phoneme conversion, namely <r> → /ɹ/. For the two conditions used, we reasoned that if the replacement of EP /ɾ/ with [ɹ] occurred in the auditory condition, it would provide evidence for perception-based CLI, while orthography-based CLI would be supported if the replacement with [ɹ] only appeared in the orthographic condition.

2.2.1 Stimuli

Test materials comprised the Portuguese liquids /l/ and /ɾ/ in intervocalic onset position and word-internal coda position. Sixteen pseudo-words were created, where target /l/ and /ɾ/ were always in a stressed syllable. Intervocalically, the liquid appeared between vowel /a/ (e.g. *pa*/l/*afa*, *pa*/ɾ/*afa*); in syllable-final position, the target liquid followed the vowel /a/ and preceded either the voiceless bilabial stop (/p/) or the voiceless labiodental fricative (/f/) (e.g. *ta*/ɾ/*pa*, *ta*/l/*fa*).

Three female native Portuguese speakers from the Lisbon area were recorded reading the test items, resulting in 48 tokens: two liquids (/l/ and /ɾ/) × two positions (intervocalic onset and word-internal coda) × four stimuli per position × three speakers, see Table 2.1. Recordings were made in a sound proof booth to a Tascam DR-100mkIII recorder. The recordings were digitized at an audio sampling rate of 44.1 kHz. All recorded sound files were adjusted to the average intensity of 70 dB in Praat 6.1.05 (Boersma & Weenink, 2019).

Table 2.1: Stimuli for the delayed imitation task[21]

/l/$_{onset}$	palafa	falapa	talafa	calapa
/l/$_{coda}$	talpa	falpa	palfa	calfa
/ɾ/$_{onset}$	parafa	farapa	tarafa	carafa
/ɾ/$_{coda}$	tarfa	farpa	parfa	tarpa

2.2.2 Participant

Twenty-three L1-Mandarin listeners were recruited for the experiment. Four of them were excluded as they reported having studied Portuguese for a short period of time (on average 12 weeks, 4 hours per week). Nineteen participants, who were on average 24.73 years old (SD=3.28), were considered for data analysis: 10 were students at the Jiangsu Normal University and were tested in China, and 9 were recruited in Lisbon and had lived there for less than a month. Their background questionnaires indicated no Portuguese learning experience, no fluency in or regular use of another language than English, and no history of hearing, speech or language impairments.

2.2.3 Procedure

A delayed imitation task was set up using Microsoft PowerPoint. In the first part of the experiment, the 48 test items were presented only auditorily in random order and subjects were asked to imitate as closely as possible the word that they had heard after being cued by the written instruction 请重复 'Please repeat' on a computer screen. The temporal interval between offset of the sound stimulus and onset of the written instruction was set to 1200 ms, with the purpose of encouraging phonological categorization rather than merely acoustic mimicry (Escudero et al., 2009). In the second part, the written form of each test item was presented on the screen simultaneously with its auditory

[21] *Farpa* is actually a meaningful Portuguese word, but this is not relevant here since all participants had no knowledge of Portuguese.

form. The order of the two parts of the experiment was not counterbalanced among participants as we deemed the presence of orthography to considerably effect the lexical storage of the words, and therefore also influence later, auditory-only presentations of the test items.

All auditory stimuli were presented to subjects via headphones at a comfortable listening level. Participants' imitations were recorded individually in a quiet room. The task took each participant about 20 minutes to complete.

2.2.4 Data preparation and analysis

Recordings were examined in Praat. All target segments were identified through a visual analysis of waveform and spectrogram together with an auditory evaluation. The presence of [l] or [ɾ] was determined through changes in intensity and formants. [l] was differentiated by having a longer duration than [ɾ] (Rodrigues, 2015), and [ɻ] by having a low F3 (Smith, 2010). A stop was marked when a closure phase and burst noise was present. An epenthetic vowel was determined on the basis of the presence of a voice bar and non-lowered formants in the spectrogram. The lack of an abrupt post-vocalic F3 transition was used as indication that the lateral was vocalized in coda (Colantoni et al., 2015).

All coding was performed by a Mandarin native speaker with near-native proficiency in EP, and checked individually by two trained Portuguese phoneticians. The annotations were then extracted and used to calculate the frequency of occurrence of each segmental realization. Note that for three participants, the imitation data of EP /l/ in the orthographic condition were lost due to a technical problem.

As we assumed that naïve imitators can only produce a segment they have perceived, we interpreted the imitation responses in the following section as the output of categorization. We return to the possible mismatch between imitation and categorization in Section 2.4.

2.3 Results

2.3.1 Lateral

The EP alveolar lateral in intervocalic position was consistently categorized as /l/ by all participants, see Figure 2.1. There was no difference between the auditory condition (M=0.99), see left of Figure 1, and the orthographic condition (M=0.99), see right of Figure 1.

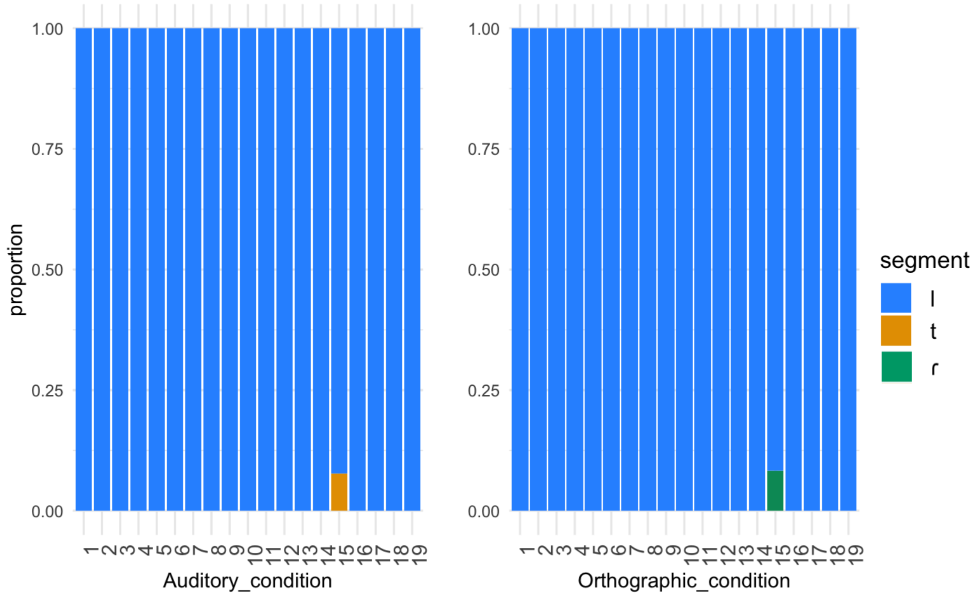


Figure 2.1: Categorization of EP /l/ in onset position split by participants

Figure 2.2 shows that in coda, the EP /l/ was most often identified as a vocalized segment (coded as *u*), with M=0.77 for the auditory condition (left) and M=0.85 for the orthographic condition (right). Again, there was no considerable difference between the two conditions. We did find, however, variation between (and within) subjects, with replacement of the target segment by a non-velarised lateral (/l/ or /lə/), replacement by other segments (/t, s, x/, coded *other*), or deletion (coded as *0*).

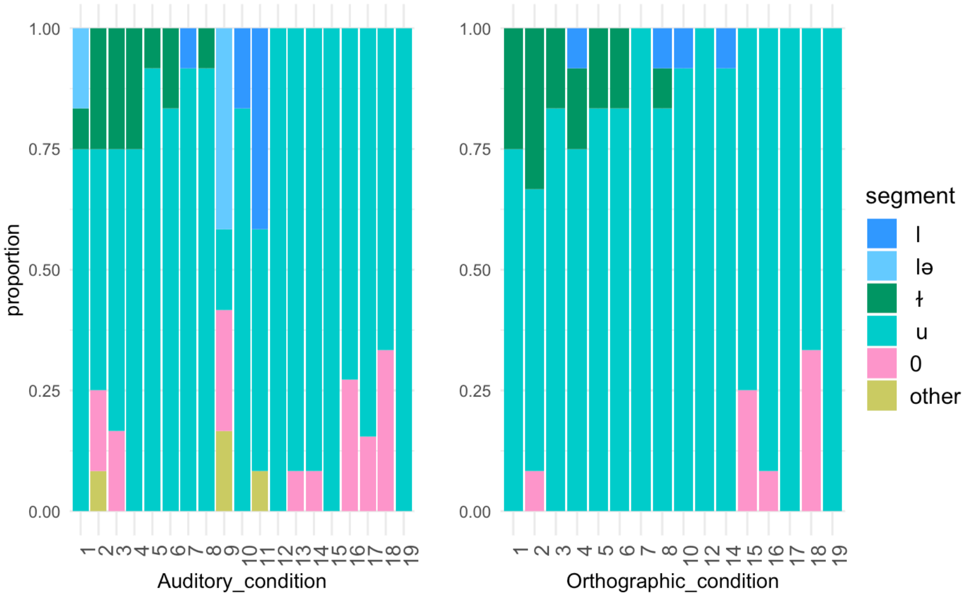


Figure 2.2: Categorization of the EP /l/ in coda position split by participants.

2.3.2 Tap

Turning to the EP tap /ɾ/, and looking first again at intervocalic onset position, this sound was predominately processed as /l/ (M=0.76) and sometimes as coronal stop /t/ or /t^{h}/ (M=0.19) in the auditory condition, cf. Figure 2.3 left. In the orthographic condition, the use of /l/ for target /ɾ/ was still prevailing (M=0.79), but it was followed by /ɹ/ (M=0.12) and a coronal stop (/t/ or /t^{h}/, M=0.05), cf. Figure 2.3 right.

Figure 2.3 also reveals that there was notable between-subject variation. In the auditory condition, some listeners constantly identified [ɾ] as a lateral (listeners 1, 2, 4, 5, 6, 7, 8, 9, 10, 11 and 14), while others perceived the tap either as /l/ or stop (listeners 3, 12, 13, 15, 16, 17, 18 and 19). In the orthographic condition, Figure 3 on the right, listeners also used these two types, but some listeners additionally categorized [ɾ] as /ɹ/. Listeners 2 and 5, for instance, categorized [ɾ] solely as /ɹ/. As this only happened in the orthographic condition, we can conclude that it was only the orthographic cue that triggered their categorization as /ɹ/, and that participants 2 and 5 disregarded the auditory information in this condition completely.

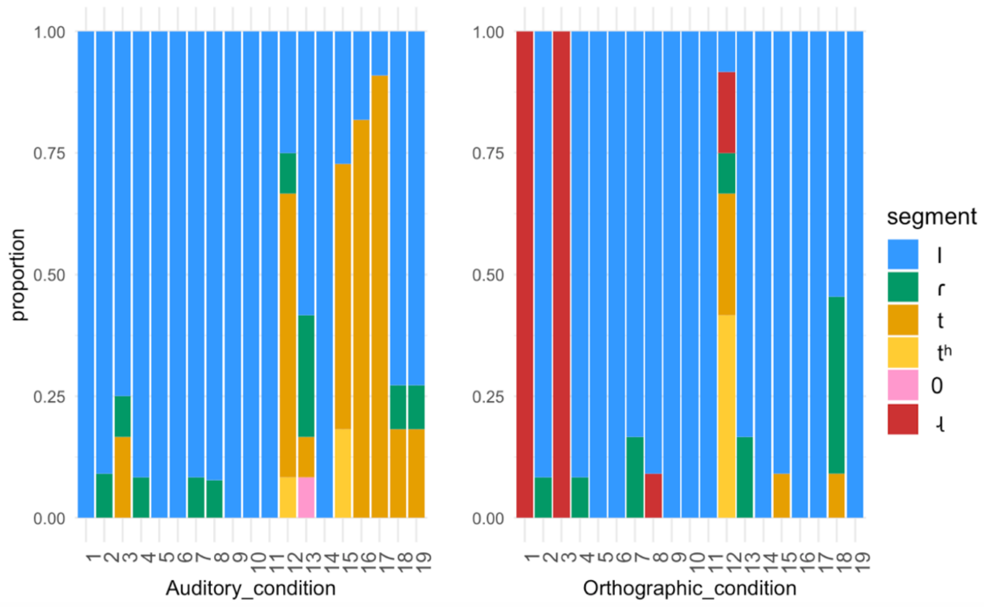


Figure 2.3: Categorization of the EP /ɾ/ in onset.

Syllable-final [ɾ] in the auditory condition, cf. Figure 2.4 on the left, despite being deleted in some cases (coded as *0*, M=0.12), was most often identified as lateral (/l/ or /lə/, M=0.5), and less often as a coronal stop (/t/, /tə/ or /t^h/, M=0.25) or some other segment (e.g. /s, ʂ/, coded as *other*, M < 0.04 each). In the orthographic condition, presented on the left side of Figure 2.4, post-vocalic [ɾ] was also deleted in some cases (M =0.2), but most often assigned to a lateral (/l/ or /lə/, M=0.48), followed by the Mandarin rhotic (M=0.18) and a coronal stop (M=0.06).

In Figure 2.4 we can see that the categorization of syllable-final EP [ɾ] by naïve Mandarin listeners manifests large between- and within-subject variation. In the auditory condition (left), participants can be grouped into three types: Type I employing predominantly a lateral (listeners 4, 5, 8, 9, 10, 11, 13, 14, 18, 19), Type II mainly using a stop (listeners 1, 3, 16, 17), and Type III alternating between lateral and stop (listeners 2, 12, 15); All three types manifested variation with respect to the insertion of an epenthetic vowel; Mandarin /ɻ/ only occurred in two instances. In the orthographic condition (right), six listeners categorized the EP [ɾ] as Mandarin rhotic in several instances (listeners 1, 3, 9, 11, 13, 15).

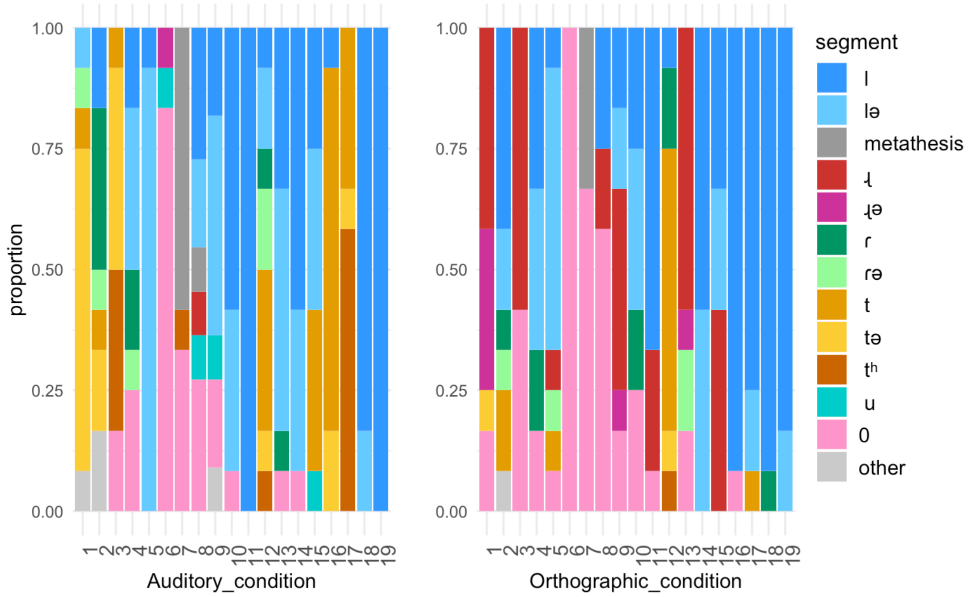


Figure 2.4: Categorization of the EP /ɾ/ in coda.

2.3.3 Statistical analysis

In order to examine whether orthography indeed accounts for the emergence of the L1 rhotic /ɻ/, we built a generalized linear mixed-effects model using lme4 package (Bates et al., 2015) in R (R Core Team) on the imitation results by the listeners who produced [ɻ] in either auditory or orthographic conditions (listeners 1, 3, 5, 6, 8, 9, 11, 12, 13, 15). The outcome of the model is the presence of [ɻ] (with 1 for present and 0 for absent). The model has Condition (with contrast-coded two levels auditory and orthographic) as predictor, and random intercepts and slopes for Participants and Stimuli. The model comparison using likelihood ratio test revealed a significant effect of Condition ($\chi^2(1) = 8.688$, *p* = 0.0032), indicating that the use of the Mandarin rhotic is due to orthographic influence.

In addition, a visual inspection of Figure 2.3 and Figure 2.4 shows that the number of coronal stop answers decreased from auditory to orthographic condition. We thus performed an exploratory analysis of the orthographic influence on the use of the coronal stop. Another generalized linear mixed-effects model was run on the imitation results by listeners who categorized EP [ɾ] as coronal stops (listeners 1, 2, 3, 5, 7, 12, 13, 15, 16, 17, 18, 19). The outcome of the model is the presence of coronal stops (with 1 for present and 0 for absent). The model has Condition (with contrast-coded two levels auditory and orthographic) as predictor, and random intercepts and slopes for Participants and Stimuli. A main effect of Condition ($\chi^2(1) = 7.362$, $p = 0.0067$) was found, which may be explained by the fact that the naïve Mandarin listeners categorized EP [ɾ] as a coronal stop to a lesser extent in the orthographic condition than in the auditory condition.

2.4 Discussion

To return to our first research question, we hypothesized that the previously-reported L2 prosodic effect on the acquisition of EP /l/ and /ɾ/ by L1-Mandarin learners (Zhou, 2017; Liu, 2018) could be accounted for by CLI. The results of the delayed imitation task with naïve Mandarin listeners largely replicated the modifications employed by L1-Mandarin learners, as shown in Table 2.2.

Table 2.2: Imitation results by Mandarin naïve speakers and repair strategies by L1-Mandarin learners of EP (C[ə] = schwa epenthesis, ∅ = deletion)

	/l/$_{\mathbf{vcv}}$	**/l/$_{\mathbf{vc}}$**	**/ɾ/$_{\mathbf{vcv}}$**	**/ɾ/$_{\mathbf{vc}}$**
Naïve	[l]	[w]	[l]	[l], [t,t^{h}], [ɻ], C[ə], ∅
Learners	[l]	[w]	[l]	[l], [t,d,t^{h}][22], [ɻ], C[ə], ∅

These results thus support our hypothesis. For the EP lateral, the different categorization outputs across prosodic contexts can be attributed to the allophonic variation in EP. Intervocalically, this sound was assimilated to the Mandarin alveolar lateral, because no detectable differences in acoustic realization seem to exist between the two. In post-vocalic position, [ɫ] was identified as /w/, presumably due to their similar spectral configuration (low F2). A similar syllable-final [ɫ]-vocalization was also reported for L1-Mandarin learners of English (He, 2015), evidencing the notorious difficulty in mastering the dark /l/ by Mandarin speakers. These results corroborate one of the most important assumptions put forward in the SLM, namely that the mapping between L1 and L2 categories occurs at the allophonic level (see also Mitterer et al., 2018 for allophones as the basic units in L1 prelexical perception).

[22] The use of alveolar stops was only reported in Liu (2018), where L2 beginners were tested, but not in Zhou (2017), where intermediate learners participated.

EP /ɾ/ in onset position was assimilated to native /l/ as this seems to be perceptually the most similar native category. In the categorization of coda [ɾ] we found large variation, which we attribute to the native phonotactic restriction that only [ɻ] or nasals are allowed in coda position (Duanmu, 2005; Lin, 2007). An assimilation of the tap to /l/ is therefore less preferred in this position. Instead, the listeners often replace it by native [ɻ] or plosives, or employ structural modifications such as [ə] epenthesis or deletion to accommodate the unfamiliar /ɾ.C/ sequence. We will return to the structural repairs of syllable-final tap in chapter 3.

In response to the second research question (whether the use of the Mandarin rhotic is perceptually or orthographically driven), Mandarin [ɻ] occurred almost exclusively when the written form was provided in the input, indicating that the use of the L1 rhotic is due to orthographic influence. This finding calls for a revision of the interpretation that cross-linguistic equivalence of phonetically-distinct rhotics is driven by phonological identity (Paradis & LaCharité, 2005). Adherents of traditional phonological accounts usually dismiss orthographic explanations by criticising that only some of the observed changes can be accounted for by orthography, whereas other equally likely candidates clearly do not yield to orthographic influence (Paradis & Prunet, 2000). For instance, the syllable-final [ɫ] was not categorized as /l/ in accordance with the written input <l>. Instead, it was predominately identified as vocalised.

Our data suggest two feasible responses to this. First, even though the written form is available to all listeners, the reliance on orthographic cues is individual-specific (see the right side of Figures 2.3 and 2.4). Learners may thus manifest a different weighting of auditory vs. orthographic cues (Hazan et al., 2010). Second, the notable within-subject variation (see left side of Figure 4) exhibited by certain listeners suggests that these listeners failed to consistently map syllable-final [ɾ] to any existing L1 category, reminiscent of the "uncategorized" L2-to-L1 mapping scenario established in PAM-L2 (Best & Tyler, 2007; Faris et al., 2016), presumably because syllable-final EP /ɾ/ displays larger allophonic variability (Silva, 2014) and less acoustic information, due to a lack of CV transition, in comparison with onset /ɾ/. It is therefore likely that during multimodal L2 speech learning, in the cases where auditory and orthographic input compete with each other, learners shift their attention to orthography when the auditory information is less consistent or insufficient.

Our exploratory analysis showed that stop responses for /ɾ/ decreased significantly with the presence of orthography, which might be explained by the fact that listeners' cue weighing strategies were altered by the written input, as demonstrated by McGuire (2014). In particular, listeners who categorized /ɾ/ as a stop seem to give more weight to its brief closure cue than to its formant structure cue, otherwise a sonorant consonant, characterized by steady formants, would have been perceived. The simultaneous presentation of the orthographic form <r>, corresponding to a sonorant sound in Mandarin, seems to avert listeners' attention away from the closure cue. This finding, together with McGuire's (2014), suggests that the auditory-orthographic cue competition and integration occur at a sub-phonemic level, in support of the view that acoustic information is mapped to phonological features in speech categorization (Lahiri & Reetz, 2010; Chládková et al., 2015; Monahan, 2018). This orthographic influence also accounts for the developmental path of L2 rhotic as the exposure to the written form clearly aids Mandarin natives to dismiss plosives as a possible variant for the target rhotic, which elucidates why the stop deviant is only observable in L2 production by beginners (Liu, 2018; learning length: less than 8 months), but not by intermediate learners (Zhou, 2017; learning length: 26 months).

Since all our participants spoke English as L2, and this also holds for the Mandarin learners of EP in prior research, one may wonder whether knowledge of English plays a role in the acquisition of the EP liquids. English has a similar grapheme-phoneme conversion as Mandarin with respect to the rhotic, <r> → /ɹ/. Therefore, it is not possible to keep the influence of the two languages apart. One can only speculate that the existence of a similar grapheme-phoneme mapping in L1 and an earlier acquired L2 would encourage its application to a new language. With respect to the acquisition of the EP tap /ɾ/, which does occur intervocalically in American English, Patience (2018) showed that the mastery of [ɾ] in L2 English does not necessarily aid its acquisition in an L3.

A methodological limitation of using an imitation task to assess L2-to-L1 category assimilation is that it measures production rather than the perception output which one would like to directly tap into[23]. In our interpretation we consistently ignored the possible role that articulatory restrictions might have in accounting for the imitation output. Nevertheless, we could account for all the observed L2 prosodic effects in the acquisition of EP liquids. Future studies will need to test whether an account including L2 articulatory restrictions is superior to the one we provided here.

[23] For instance, the potential articulatory influence can be controlled if one uses a perceptual assimilation task, in which naïve listeners are asked to match the L2 sounds they hear with their L1 categories. However, we deem that the delayed imitation task is more efficient as naïve listeners might have to face too many category labels (e.g. all allophones from his previous inventory) more than once, if L2-to-L1 assimilation were investigated in different positions.

Chapter 3: L2 perceptual development of European Portuguese /l/ and /ɾ/

3.1 Introduction

As reviewed in 1.3, previous studies on L2 perception (Cao, 2018; Vale, 2020) and production (Zhou, 2017; Liu, 2018) of EP /l/ and /ɾ/ by L1-Mandarin learners differ substantially in the structures that were investigated. This divergence in the literature obscures the origin of some deviations observed in this L2 speech learning process. Finding out in which modality these L2 learning difficulties originate is of great importance not only to understand the relationship between perception and production during L2 phonological acquisition, but also to guide the development of teaching/training methodologies, which would be more efficient if the modality where the difficulty occurs could be targeted. The first goal of this chapter therefore was to mind the gap between prior perceptual and production studies, by testing whether L1-Mandarin learners' deviant productions of Portuguese /l/ and /ɾ/ has a perceptual basis across prosodic contexts.

3.1.1 L2 Perception of /l/ and /ɾ/ across prosodic contexts

Prior research has shown that L1-Mandarin learners' production of EP /l/ and /ɾ/ is constrained by prosodic positions (Zhou, 2017; Liu, 2018). In particular, they do not have difficulty in producing /l/ in onset, but very often vocalise it in coda; regarding the non-target-like production of /ɾ/, learners use [l] in onset, while employing both segmental (e.g. [l][24]) and structural repairs (e.g. epenthesis and deletion) in coda. These L2 prosodic effects have been replicated

[24] The use of approximant [ɻ] (Zhou, 2017; Liu, 2018) and of alveolar stops [t/d/t^h] (Liu, 2018; chapter 2) has also been reported; however, these segmental repairs were not examined in the current chapter because it has been shown that [ɻ] is triggered by orthography (chapter 2); stops are only used by some beginners (Liu, 2018; chapter 2) and can be dismissed rapidly when the written input is given (chapter 2).

in naïve Mandarin categorization of EP /l/ and /ɾ/ (chapter 2), suggesting that CLI poses constraints on L2 category learning.

Major theories in the field of L2 speech (Flege, 1995; Escudero & Boersma, 2004; Best & Tyler, 2007) advocate that non-native segmental learning difficulties stem from misperception. This was partially confirmed for the deviations produced by L1-Mandarin learners of EP, as Cao (2018) and Vale (2020) both reported that these learners encounter difficulty in perceptually detecting the difference between /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$. To our best knowledge, no existing perceptual studies concerned the syllable-final position, thus whether the use of [w] for /l/$_{\text{coda}}$ and [l] for /ɾ/$_{\text{coda}}$ are perceptually driven remains an open question.

In contrast to the perception-based hypothesis put forward in aforementioned L2 speech models, some analysts warned that L2 deviant productions do not necessarily mirror deficits in L2 phonological categorization (e.g. Honikman, 1964; Colantoni & Steele, 2008). To give an example, even if L1-Mandarin learners could categorize the EP syllable-final [ɫ] accurately and established a target-like phonological representation for it, they might still vocalize it in production due to motor control issues, as the realization of [ɫ] demands a coordination between a coronal and a dorsal gesture, which is entirely novel to Mandarin natives. This speculation is plausible since it has been shown that the articulatory factor alone is sufficient to trigger /l/-vocalization (Recanses & Espinosa, 2010).

Moreover, current L2 speech models are engaged in exploring only issues pertaining to learning barriers at the segmental level (consonant and vowels), while how L2 perception and production interact beyond segmental level is still absent from their theoretical formulation. On the contrary, empirical research has revealed that the puzzling cross-modality relationship in L2 speech extends to suprasegmental level, which will be discussed in the following two paragraphs.

On the one hand, L2 structural repairs may arise out of perception. A seminal study demonstrating structural restriction on speech perception was conducted by Dupoux and colleagues (1999). They showed that L1-Japanese listeners perceived an "illusory vowel", upon hearing an illegitimate sequence, e.g. [ebzo] perceived as /ebuzo/. The insertion of an epenthetic vowel to accommodate a structure that does not respect L1 phonotactic constraints seems to constitute a prevailing feature of L2 speech perception, as it so far has been attested in learners who are native speakers of Korean (Kabak & Idsardi, 2007), Brazilian Portuguese (Cardoso, 2011; Dupoux et al., 2011; Cabrelli et al., 2019), Spanish (Cuetos et al., 2011), English, and Mandarin (Durvasula et al., 2018), to mention a few. Therefore, the Mandarin phonotactic grammar, which only allows nasals or [ɻ] in coda (Duanmu 2006, Lin, 2007), is likely to give rise to a perceptual restoration of EP syllable-final /ɾ/, e.g. ca[ɾ]ta "*letter*" reconstructed as ca/ɾə/ta[25]. If perceptual epenthesis is encoded in the learners' mental lexicon, an inserted vowel will be expected in L2 production as previously attested by Matthews and Brown (2004), Davidson et al. (2007) and Darcy and Thomas (2019). Apart from epenthesis, the L1-L2 structural differences can likewise lead to perceptual deletion. Steele (2009), for instance, reported that, in an identification task, L1-Mandarin learners perceptually simplified the French stop-liquid clusters (e.g. auditory stimulus [bekʁe], learner's response *bécaie*;), which can be explained by the fact that the Mandarin phonotactic grammar does not allow any kind of consonantal clusters. Comparably, Davidson and Shaw (2012) found that English-speaking natives manifested reduced sensitivity to the difference between [tmafa] – [mafa], implying that listeners many times failed to detect the existence of the initial stop in an illegal cluster. One may argue that the above-mentioned instances can be attributed to the low acoustic salience of the deleted segment

[25] In many cases, the epenthesis employed by L1-Mandarin learners to repair EP syllable-final tap is accompanied by segmental change as well, [ɾ] → [lə] (Zhou, 2017), generating a well-formed structure conforming to Mandarin phonotactics. However, the segmental change is due to the perceived similarity between [l] and [ɾ] (chapter 2). In the current study, segmental and structural repairs are examined separately.

rather than the L1 phonotactic restrictions. However, a study on the L2 perception of English [h], which is less audible than [ʁ] and [t][26], refutes the acoustic account. Mah et al. (2016) used the mismatch negativity in event-related potentials (ERPs) to investigate how French speaking informants perceive the English word-initial /h/ both at acoustic and phonological levels. A mismatch negativity response was only elicited at the acoustic level, indicating that the perceptual deletion of English word-initial /h/ was due to the French phonotactics, which "silences" the word-initial /h/, instead of the lack of acoustic saliency. Melink and Peperkamp (2019) further revealed that the perceptual deletion of English word-initial /h/ is mirrored in the L1-French learners' lexicon (e.g. [ˈʌzbənd] was accepted as a real word "husband"), which seems to justify why the word-initial /h/ is often omitted in the English words produced by French natives.

On the other hand, the emergence of structural repairs can be restricted to production. A good illustration also comes from L1-Japanese. It has been reported that Japanese native speakers produce different epenthetic vowels to break an illegal cluster, [o] after a coronal stop and [u] elsewhere (e.g. Polivanov, 1931), due to the fact that the sequence coronal stop + [u] is not allowed by the Japanese phonotactics. Monahan et al. (2009) performed a perceptual experiment to test whether Japanese listeners perceive different illusory vowels across contexts and observed that they only illusorily epenthesized [u], but not [o]. This asymmetric perceptual epenthesis leads Monahan and colleagues to reason that the phonotactic rule that triggers the insertion of [o] (e.g. */tu/) is only active in production, not in perception (Kabak & Idsardi, 2003; Ramus et al., 2010)[27]. Additionally, structural repairs may solely stem from imprecise articulation, thus not induced by phonological restriction. As reviewed in 1.2.2,

[26] In comparison with [ʁ] from Steele (2009) and [t] from Davidson & Shaw (2012), [h] lacks frication and burst respectively.

[27] We remain skeptical about the claim that phonological rule/constraint can be modality-specific, since the divergent patterns between perception and production can stem from the bidirectional use of the constraints and constraint ranking (Smolensky, 1996; Boersma & Hamann, 2009). We will return to this issue in chapter 4.

each language has a particular articulatory setting of articulators, which characterizes the articulatory gestures and gestural coordination involved in the realization of an individual segment (Honikman, 1964). When the L1 and L2 articulatory settings do not resemble each other, subtle articulatory adjustments or novel articulatory motor controls are necessary for target-like production. For instance, the realization of EP /ɾ/ requires a ballistic movement of the tongue toward the dental/alveolar region (Mateus et al., 2005). The non-mastery or unsuccessful implementation of this rapid articulatory gesture, which is unfamiliar to Mandarin speakers, might effect the omission of /ɾ/, regardless of the quality of the phonological representation of /ɾ/. Apart from deletion, gestural mistiming was shown to give rise to epenthesis as well (Davidson, 2005, 2006; Funatsu & Fujimoto, 2012). By acoustically measuring the transitional vowels inserted in illegal consonant clusters by English-speaking natives, Davidson (2006) demonstrated that the epenthetic vowels were substantially distinct from the lexical schwa, both in terms of duration and formant values. This acoustic disparity suggests that the vowel insertion occurs after phonological computation, in support of the idea that epenthesis is driven by gestural mistiming. Davidson's postulation was further borne out in an electromagnetic articulograph study performed by Funatsu and Fujimoto (2012), who provided direct articulatory evidence for articulation-based epenthesis: the insertion of a vocalic element to break the illegal consonant clusters by both Japanese and German speakers is reportedly driven by the unseemly timing between the articulatory movement and vocal fold vibration. In particular, when the first consonant is voiceless, epenthesis occurs if the vocal fold vibration initiates before the onset of the second consonant. Provided that the first consonant is voiced, the interruption of vocal fold vibration between two consonants will trigger vowel insertion.

To summarise, the divergence between the existing perceptual and production studies on L1-Mandarin learners' difficulties with the EP /l/ and /ɾ/

not only lies in the prosodic positions under investigation, but also relates to the level of analysis (segmental and suprasegmental). This chapter thus sets out to fill this gap by testing L1-Mandarin learners on their discrimination ability between the target Portuguese form and the deviant form that they often employ in L2 production. Expanding on prior research which predominately focuses on the confusability at segmental level in one particular position (e.g. intervocalic onset), we looked into the potential L2 perception-based difficulties across syllable constituency (onset and coda) and took both segmental as well as structural modifications into account.

3.1.2 L2 onset-coda asymmetry - /ɾ/$_{\text{coda}}$ > /ɾ/$_{\text{onset}}$

When acquiring a novel sound, L2 learners normally target syllable onset before coda position (Flege, 1989, Rogers & Dalby, 2005; Waltmunson, 2005; Bent et al., 2007; Colantoni & Steele, 2008; Cheng & Zhang, 2015). This onset-coda asymmetry with respect to the acquisition order has been attested both in L2 perception and production. For instance, by testing L1-Mandarin learners' perception of English stop contrasts, Flege (1989) observed that learners were more accurate syllable-initially than syllable-finally. Comparably, in a cross-linguistic production study, Colantoni and Steele (2008) reported that L1-English learners stabilize both Spanish alveolar tap and French uvular fricative in intervocalic onset position before coda. More compelling evidence on the privileged status of syllable onset in L2 phonological acquisition was put forward by Cheng and Zhang (2015), who assessed L1-Mandarin learners' performance with 20 English consonants across prosodic contexts. Their results indicated that Mandarin speakers had higher accuracy rates syllable-initially than syllable-finally for all segments both in perception and in production. This cross-modality asymmetry between onset and coda, echoing the same developmental sequences widely attested in L1 phonological acquisition (e.g. Fikkert, 1994; Freitas, 1997), can be attributed to the universal salience of syllable onset in terms of accessibility and learnability (Ohala, 1996; Carlisle, 2001).

Despite the converging evidence reviewed above, an opposite developmental pattern was encountered in the acquisition of EP tap by L1-Mandarin learners. It has been shown that they produced more target-like instantiations of /ɾ/ in coda than in onset (Zhou, 2017; Liu, 2018). This rarely reported acquisition order clearly requires explanations other than the syllable onset saliency. One study on L1 phonological development seems to offer us a clue. Cohen (2015) performed a longitudinal study with two Hebrew toddlers from the onset of speech until the completion of rhotic acquisition and he attested that the Hebrew rhotic is fully stabilized in coda before in onset position, in stark contrast to other Hebrew consonants (Ben-David, 2001). Cohen (2015) attributed this unusual developmental pattern to the different degrees of phonetic consistency of the Hebrew rhotic across prosodic contexts. In particular, the more allophonic variation a segment manifests, the less phonetic consistency it has. Since the Hebrew rhotic manifests more allophonic variation in onset (Cohen, 2013), the greater phonetic consistency may expedite the development of a phonological representation in coda. However, the consistency hypothesis would predict an acquisition order (onset > coda) that is opposite to what has been observed, since EP /ɾ/ displays less variation in onset (71% tap, 28% approximant, 0.6% fricatives) than in coda (31.35% tap, 30% tap + supporting vowel, 25 % approximant, 4% approximant, 3% stops, 3% approximant + supporting vowel, 2% deletion) (Silva, 2014).

Another plausible explanation was proposed by Zhou (2017), who argued that the acquisition of /ɾ/ is boosted by the reduced cross-linguistic interference in coda position. Specifically, the interferer Mandarin /l/, whose existence hinders the acquisition of /ɾ/ due to the high degree of perceptual similarity, is banned syllable-finally in the Mandarin phonotactic grammar (Duanmu, 2005;

Lin, 2007). The native phonotactics thus should make it easier for learners to overcome the L1 interference in coda position.

In the current study, we assessed the degrees of L1 interference on the acquisition of EP /ɾ/ by comparing the degrees of perceptual confusability between [l]-[ɾ] across syllable positions. We hypothesize that L1-Mandarin learners should discriminate the target EP /ɾ/ from its most similar L1 category [l] better in coda than in onset position, due to the Mandarin phonotactic restriction on syllable-final [l]. If perceptual and production evidence converge, it would provide strong evidence for the fact that, during the acquisition of EP tap, L1-Mandarin learners experience less L1 interference in coda than in onset, which gives rise to the observed onset-coda asymmetry.

3.1.3 Plasticity of L2 phonological representations of /l/ and /ɾ/

Decades of studies on L2 speech learning have shown that adult learners often struggle to master certain novel sounds. And the rare optimal attainment of these L2 categories have lead some to advocate the existence of a "critical period" for L2 speech learning (Lenneberg, 1967), after which the capacity of forming a target-like L2 phonological representation is lost due to a lack of plasticity. Such claim is explicitly refuted by most L2 speech models (e.g. Flege, 1995; Escudero, 2005; Best & Tyler, 2007), which assume uniformly that target-like L2 category formation is possible across lifespan.

Evidence for evaluating these competing postulations can be found either in research employing laboratory trainings, or in studies of naturalistic learning, in which an L2 is acquired through the use in daily life. Results from training studies have generally supported the view that L2 categories remain malleable at all ages (see Sakai & Moorman, 2018 for a meta-analysis), because L2 category refinement can be observed immediately after a few training sections (Wong, 2013; Rato, 2014; Oliveira, 2020).

On the contrary, mixed findings can be found in studies of naturalistic learning. Despite the general assumption that L2 sound categories become more refined as a function of more experience with the target language (see Bohn, 2017 for a review on supporting evidence), it has been demonstrated that the plasticity of L2 phonological representation may have some limits. For instance, Dupoux and colleagues (2008) tested how 39 adult L1-French learners of Spanish with varying L2 experience[28] process a lexical stress contrast, e.g. /múmi-mumí/, which is missing in the French phonology. Their results indicated that French learners encountered much difficulty in encoding the target-like suprasegmental contrast in their phonological representations, which does not seem to be malleable by more L2 experience in the naturalistic learning[29]. The limited role that L2 experience plays in non-native phonological development, especially at the mid-late stage, might be elucidated by the fact that

> "... the time window for this L2 learning may be brief and occur early in L2 acquisition; it may possibly be curtailed by increases in learning higher-order aspects of the L2, such as an expanding lexicon and the acquisition of morphological and syntactic structure. In other words, the focus of attention and learning may shift away from the phonetic level as the learners focus increasingly on higher levels of linguistic structure (Best & Tyler, 2007; p. 26-27)."

Among all previous studies on the acquisition of EP /l/ and /ɾ/ by L1-Mandarin learners, Cao (2018) was the only one that has investigated the role of L2 experience (quality, i.e. two groups of participants with comparable learning experience but differing with respect to the immersion experience) and

[28] 14 participants were identified as beginners, 14 intermediate and 11 advanced learners, based on their length of residence in a Spanish-speaking country and self-reported use of Spanish.

[29] In Dupoux and colleagues (2008), the quantitative difference between groups was determined on the basis of their Spanish language background (age/place/manner of acquisition, length of residence in a Spanish-speaking country) and their current usage of Spanish (visits to Spanish-speaking countries, private and professional usage of Spanish).

no effect of L2 experience on perceptual accuracy was reported. In the current study, we further looked into the conceivable effect of L2 experience both in terms of quality and quantity at a mid-late stage of L2 speech learning.

3.2 Method

In the current chapter, we first explored whether the deviant productions of EP /l/ and /ɾ/ articulated by L1-Mandarin learners across prosodic contexts are rooted in misperception (RQ1). L1-Mandarin learners' perceptual ability was investigated first in an AXB discrimination task. We reasoned that if they fail to discriminate reliably between a target form and the respective deviant form, this will indicate perceptual motivation for that imprecise production. Moreover, we furthermore examined whether /ɾ/$_{\text{onset}}$ and /l/$_{\text{onset}}$ were interchanged (merged or overlapped) in a forced-choice identification task.

Moreover, a comparison with respect to the discrimination accuracy of the contrast [l]-[ɾ] between onset and coda position would allow us to assess the degree of L1 interference during the acquisition of EP (RQ2).

In order to explore the plasticity of L2 phonological representations (RQ3), two groups of L1-Mandarin learners differing substantially in L2 experience were recruited to participate in the perceptual experiments. If L2 experience did play a role, learners receiving more formal instruction and spending more time in a Portuguese-speaking country would score better than those with reduced L2 experience. The three research questions are summarized as follow:

RQ1: whether the deviant productions articulated by L1-Mandarin learners across prosodic positions stem from the inaccurate perception?

RQ2: whether L1-Mandarin learners experience less L1 interference in coda than in onset during the acquisition of EP /ɾ/?

RQ3: whether L1-Mandarin learners' phonological representations of /l/ and /ɾ/ are malleable in a mid-late stage of L2 speech learning?

3.2.1 Participants

Sixty-one L1-Mandarin learners of EP and 10 native speakers of EP completed in the perceptual experiment. All listeners were recruited from Lisbon.

The inclusion criteria for Chinese participants were as follows: (1) they had to be native speakers of Mandarin who, regardless of the Chinese city where they were raised, considered Mandarin as their dominant language[30]; (2) they had to have no fluency in or regular use of another language than English. All participants completed a language background questionnaire which ensured that they met the inclusion criteria (Appendix I).

L1-Mandarin participants were divided into two groups, based on their experience with EP. The intermediate-level group consisted of 31 learners (mean age = 20.3 years, SD=0.59), all of whom were third-year college students majoring in Portuguese at a Chinese university, studied EP for two years in a classroom setting in China, and were immersed in a Portuguese language course in Lisbon for 2 months at the moment of testing. Thirty learners (mean age=24.6 years, SD=1.5) in the advanced-level group were either enrolled in a Master degree course in Portugal or worked in Lisbon after obtaining a Master degree from a Portuguese university. All advanced learners completed a 4-year bachelor degree in Portuguese from a Chinese university and they reported having spoken Portuguese for 5.54 years on average (SD = 1.2). In order to attain reasonable effect size regarding our research question on L2 experience, we deliberately recruited and selected these learners who form two groups with a notable difference both in quality and quantity of L2 experience. Many studies in the literature took the length of residence as an indicator of overall amount

[30] Many studies in the literature avoided the potential problem caused by Chinese dialects by recruiting participants who were raised in Beijing area and/or only acquire Mandarin as L1. However, we included all Chinese participants as long as they consider Mandarin as their dominant language. The only-Mandarin-speaking learners do not have the representative language profile for the Chinese learners of EP, who are generally spread all over China. Moreover, the dialects spoken by the participants of this study, namely Zhongyuan Mandarin, Cantonese, Sichuanese, Wu, Gan, Xiang do not report to have a tap or any other rhotic consonant that might resemble with the Portuguese tap. Although /l/-/n/ distinction does not exist in Sichuanese and Xiang, the early exposure to Mandarin may mediate this difficulty (Johnson & Song, 2016).

of L2 input, but the length of residence has been argued to be a poor estimate (Flege, 2021). Therefore, in the current study the L2 experience is determined both with respect to the length of residence in an immersion setting (intermediate group: 2 months; advanced group: 1.54 years), during which learners participated in a course taught in Portuguese regularly[31], and years of formal instruction (intermediate group: 2 years; advanced group: 4 years). Therefore, if L2 experience indeed plays a crucial role in shaping L2 phonological development, we expect to detect it[32].

Regarding experience with English, most L1-Mandarin participants had begun learning English around the age of seven in China, but none of them was ever immersed in an English-speaking environment.

Ten Portuguese controls who were all born and educated in Portugal also participated. They were living in Lisbon at the time of testing and were either master or PhD students at the University of Lisbon. They were on average 29 years old (SD=1.5).

No participants reported hearing, speech or any other language impairment. They all gave informed consent at the beginning of the study.

3.2.2 Materials and recording

Stimuli for the AXB discrimination task were pairs of trisyllabic pseudo-words. The target segment was always in a stressed syllable and vowels /a/ and /i/ were used in adjacent vocalic contexts and counterbalanced across stimuli.

In test word pairs, the target consonants (/l/ and /ɾ/) alternated with deviant forms attested in L1-Mandarin learners' production of EP (Zhou, 2017; Liu, 2018; chapter 2):

[31] Flege & Liu (2001) demonstrated that the length of residence may be a useful estimate of quantity of L2 input only for those who have the opportunity and the need to use the target language.

[32] Other methods of assessing and quantifying L2 experience are also available, such as Cumulative Use Index (Flege, 2021) or self-reported use of L2 (Miatto et al., 2019). We reasoned that the two groups of participants in the present study differ substantially both in length of residence and amount of formal instruction, which should be enough to effect L2 category development.

RQ1. Perceptual basis for L2 deviant production:

[ɫ]$_{\text{coda}}$ - [w], [ɾ]$_{\text{onset}}$ - [l], [ɾ]$_{\text{coda}}$ - [l], [ɾ]$_{\text{coda}}$ - [ɾə] and [ɾ]$_{\text{coda}}$ - [∅]

RQ2. Degrees of L1 interference across syllable contexts:

[ɾ]$_{\text{onset}}$ - [l] and [ɾ]$_{\text{coda}}$ - [l];

Fillers were word pairs containing easily discriminable contrasts (/l-k/, /t-s/, /t/-/k/) for Mandarin listeners.

There were 120 trials in total, consisting of 80 test trials: four trials per contrast × four counterbalancing orders (AAB, ABB, BBA, BAA) × five contrasts. In addition, there were 40 fillers: 4 contrasts × 10 times. See Table 3.1.

Table 3.1: Stimuli for the AXB discrimination task

contrast	Test pairs			
[ɾ] – [ɾə] (coda)	ta[ɾ]pa – ta[ɾə]pa	pa[ɾ]fa – pa[ɾə]fa	fi[ɾ]pa – fi[ɾə]pa	si[ɾ]pa – si[ɾə]pa
[ɾ] – [∅] (coda)	ta[ɾ]pa – ta[∅]pa	pa[ɾ]fa – pa[∅]fa	fi[ɾ]pa – fi[∅]pa	si[ɾ]pa – si[∅]pa
[l] – [ɾ] (onset)	pa[l]afa– pa[ɾ]afa	fa[l]apa – fa[ɾ]apa	pi[l]ifa – pi[ɾ]ifa	si[l]ifa – si[ɾ]ifa
[ɾ] – [l] (coda)	ta[ɾ]pa – ta[l]pa	pa[ɾ]fa – pa[l]fa	fi[ɾ]pa – fi[l]pa	si[ɾ]pa – si[l]pa
[ɫ] – [w] (coda)	ta[ɫ]pa – ta[w]pa	pa[ɫ]fa – pa[w]fa	fi[ɫ]pa – fi[w]pa	si[ɫ]pa –si[w]pa
Fillers	pa[t]afa-pa[s]afa	pa[t]afa-pa[k]afa	pa[l]afa-pa[k]afa	fa[t]apa- fa[k]afa

The 12 test items created for the identification task were trisyllabic pseudo-words words, comprising the target segments /l/ and /ɾ/ in stressed intervocalic position. The adjacent vowels were either /a/ or /i/ and counterbalanced across stimuli. Fillers contained voiceless stops, which are present in both the Mandarin and the Portuguese inventories. There were in total 24 test tokens: six words per segment × two segments (/l/ and /ɾ/) × two repetitions and 12 fillers, see Table 3.2.

Table 3.2: Stimuli for the identification task

Segment	Test words					
/l	pa[l]afa	fa[l]apa	ta[l]afa	ti[l]ifa	si[l]ipa	pi[l]ipa
/ɾ/	pa[ɾ]afa	fa[ɾ]apa	ta[ɾ]afa	ta[ɾ]afa	si[ɾ]apa	pi[ɾ]apa
Fillers	pa[t]afa	pa[k]afa	ta[t]afa	ti[k]afa	si[t]ipa	pa[t]ipa

A male native Portuguese phonetician was recorded reading all stimuli. Recordings were made in a sound-proof booth with a Zoom H4n pro recorder, and a Shure SM58 microphone. They were digitized at an audio sampling rate of 44.1 kHz. All recorded sound files were adjusted to the average intensity of 70 dB in Praat 6.1.05 (Boersma & Weenink, 2019). For the AXB discrimination task, two renditions were obtained for each pseudo-word, so that the audio stimulus for A, for example in a triplet AAB, were actually instantiated by two acoustically different tokens.

It is worth mentioning that the duration and spectral properties of the inserted vowel (ta[ɾə]pa) in the experimental stimuli were measured to verity that it contains acoustic values comparable to previous studies on perceptual epenthesis. The measurement was performed using visual cues from the spectrogram and waveform visualised in Praat. Table 3.3 summarizes the duration of epenthetic vowels, as well as their formant values (F1 and F2). The beginning of the inserted vowel following a coda rhotic was determined at the point of a sharp increase in intensity coinciding with the onset of a periodic waveform with regular formant structure. The end of the vowels was marked when the formant structure disappears. The formant values were extracted in at the midpoint of the vowel steady-state.

Table 3.3: Descriptive statistics of acoustic measurements for the inserted vowel in the experimental stimuli

	Mean Duration (ms)	Mean F1 (Hz)	Mean F2 (Hz)
	80.5 (*SD*=10.5)	324 (*SD*=15.8)	1775 (*SD*=62.9)

Note: *SD* = standard deviation

The duration range of the inserted vowels is comparable to those (mean duration: 96 ms; range: 49 – 159 ms) utilized in Darcy and Thomas (2019). And the vowel duration in our stimuli is also substantially longer than the inserted vocoids (mean duration: 38 ms, *SD* = 12.9) produced by Mandarin speakers who do not perceive an illusory vowel between two consonants (Guan, 2019). Taken together, these data suggest that our stimuli contained clearly perceptible, unambiguous vowels following the Portuguese rhotic.

3.2.3 Procedure

Participants were tested in a quiet room. The experiment was set up and run in OpenSesame 3.2.8 (Mathôt et al., 2012), with auditory stimuli presented through Sony noise cancelling headphones WH1000XM3. Participants first completed the AXB discrimination task and then the identification task. The two perceptual tasks together took about 20 minutes to complete.

During the AXB task, participants were presented with three auditory stimuli in sequence, and were required to indicate whether the second (X) was more similar to the first (A) or to the third (B) by pressing the corresponding buttons on a keyboard. Stimulus presentation counterbalanced across trials (AAB, ABB, BBA, BAA). Within each trial, the inter-stimulus interval (ISI) was set to 1200ms in order to encourage judgment at phonological level, rather than acoustic comparison (Escudero et al., 2009). After a short practice (4 trials), the task ran with 4 blocks, each of which contained 20 test trials and 10 filler trials. The test trails were balanced across blocks. Participants were given self-paced interval between blocks to avoid fatigue.

As for the identification task, listeners were presented with a single auditory stimulus each time and were required to assign a label to the stimulus by choosing one of the four orthographically represented alternatives, which were composed of target segments (/l/ and /ɾ/) and two distractors containing either /k/ or /t/. For instance, after hearing the auditory form [fɐˈlapɐ], learners were asked to choose the correct response from <falapa>, <farapa>, <facapa> or <fatapa>.

3.3 Results

3.3.1 AXB discrimination task

The results of discrimination accuracy by 10 native Portuguese and by 61 L1-Mandarin learners are presented in Figure 3.1. Visual inspection suggests that, apart from the contrast involving deletion ([ɾ]$_{coda}$ – [∅]; M= 0.98), L1-Mandarin learners were less accurate, i.e., [ɫ]$_{coda}$ - [w] (M =0.56), [ɾ]$_{onset}$ - [l] (M = 0.72), [ɾ]$_{coda}$ - [l] (M = 0.85) and [ɾ]$_{coda}$ - [ɾə] (M =0.84), than native controls, who reached ceilings in all test contrasts.

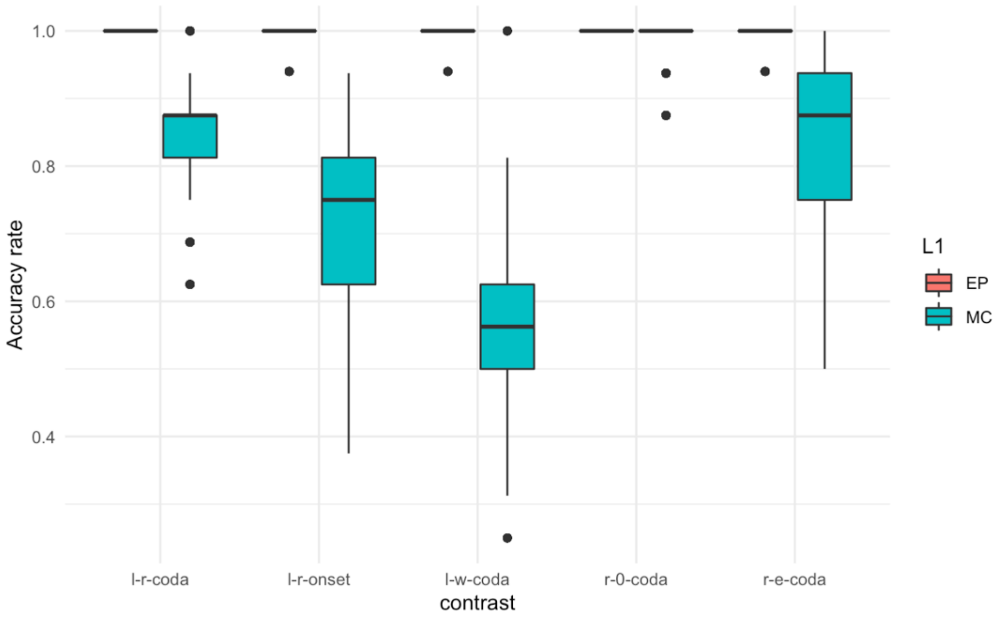


Figure 3.1: Accuracy results in the AXB discrimination task (the results of native controls are always presented at the left side of each condition, instantiated by solid lines on the top; [ɾ]$_{coda}$ - [l] coded as *l-r-coda*, [ɾ]$_{onset}$ - [l] coded as *l-r-onset*, [ɫ]$_{coda}$ - [w] coded as *l-w-coda,* [ɾ]$_{coda}$ - [∅] coded as *r-0-coda*, [ɾ]$_{coda}$ - [ɾə] coded as *r-e-coda*)

The accuracy data was then analysed in several generalized linear mixed-effects models, using the lme4 package (Bates et al., 2019) in R. All *p*-values were obtained via likelihood ratio tests.

In order to answer RQ1 on whether the deviant productions by L1-Mandarin learners stem from inaccurate perception, several mixed-effects models were built on the results for the following contrasts: [ɫ]$_{\text{coda}}$ - [w], [ɾ]$_{\text{onset}}$ - [l], [ɾ]$_{\text{coda}}$ - [l], [ɾ]$_{\text{coda}}$ - [ɾə] and [ɾ]$_{\text{coda}}$ - [∅]. Each model had Native Language (with contrast-coded at two levels, Portuguese and Mandarin) and Preceding Vowel (with contrast-coded at two levels, A and I) as predictors. Random intercepts for Participant and Trial, together with random slopes for Preceding Vowel by Participant and for Native Language by Trial were also included (see 3.1).

(3.1) Accuracy ~ Native Language+ Preceding Vowel + (Preceding Vowel | Participant) + (Native Language| Trial)

Table 3.4: Results of the models built for RQ1

condition	Effect	df	Chisq	*p*.value
[ɫ]$_{\text{coda}}$ - [w]	Native Language	1	22.198	<0.0001 ***
	Preceding Vowel	1	9.4424	0.0021 **
[ɾ]$_{\text{onset}}$ - [l]	Native Language	1	5.4197	0.02 *
	Preceding Vowel	1	7.4962	0.0062 **
[ɾ]$_{\text{coda}}$ - [l]	Native Language	1	6.3663	0.012 *
	Preceding Vowel	1	6.0233	0.014*
[ɾ]$_{\text{coda}}$ – [ɾə]	Native Language	1	7.0393	0.008**
	Preceding Vowel	1	0.7427	0.39
[ɾ]$_{\text{coda}}$ – [∅]	Native Language	1	2.4156	0.12
	Preceding Vowel	1	1.3596	0.2436

The models' results were listed in Table 3.4. A main effect of Native Language was found for all contrast, except for the one [ɾ]$_{\text{coda}}$ – [∅]. This indicates that the use of [w] for /l/$_{\text{coda}}$, [l] for /ɾ/ across syllable contexts and epenthetic form for /ɾ/$_{\text{coda}}$ have a perceptual basis, while the deletion of syllable-final tap is restricted to production.

Regarding RQ2 on whether L1-Mandarin learners experience less L1 interference syllable-finally than syllable-initially, we built another model on the accuracy results for contrasts [ɾ]$_{\text{onset}}$ - [l] and [ɾ]$_{\text{coda}}$ - [l], which had Position (with contrast-coded two levels, onset and coda), Proficiency (with contrast-coded two levels, intermediate and advanced) and Preceding Vowel (with contrast-coded two levels, A and I) as predictors. The model also included random intercepts for Participant and Trial, random slopes for Position/Preceding Vowels by Participant and for L2 Experience by Trial, see (3.2).

(3.2) Accuracy ~ Position+ Preceding Vowel + L2 Experience + (Position + Preceding Vowel | Participant) + (L2 Experience | Trial)

Results of this model can be found in Table 3.5. A main effect of position was found, which confirmed our hypothesis that, during the acquisition of EP tap, L1-Mandarin learners experience less cross-linguistic interference in coda than in onset, which may explain why /ɾ/ is mastered in coda before onset.

Table 3.5: Results of the model built for RQ2

Effect	df	Chisq	*p*.value
Position	1	4.9614	0.026*
L2 Experience	1	1.0062	0.32
Preceding Vowel	1	11.703	<0.001***

In order to test whether the L2 phonological representations of /l/ and /ɾ/ become more accurate with more exposure to the target language (RQ3), four other models were developed on accuracy results solely by L1-Mandarin learners. These models had L2 Experience (with contrast-coded two levels, intermediate and advanced) and Preceding Vowel (with contrast-coded two levels, A and I) as predictors. Random intercepts for Participant and Trial, together with random slopes for Preceding Vowel by Participant and for L2 Experience by Trial were also included, see (3.3).

(3.3) Accuracy ~ L2 Experience + Preceding Vowel + (Preceding Vowel | Participant) + (Native Language| Trial)

As shown in Table 3.6, where models' results are summarized, no significant effect of L2 Experience was found in any of the models. This seems to suggest that more exposure to the target language does not contribute to the refinement of L1-Mandarin learners' phonological representations of EP /l/ and /ɾ/ at a mid-late stage of L2 speech learning.

Table 3.6: Results of the models built for RQ3

condition	Effect	df	Chisq	*p*.value
[ɫ]$_{coda}$ - [w]	L2 Experience	1	1.4982	0.22
	Preceding Vowel	1	10.204	0.0014 **
[ɾ]$_{onset}$ - [l]	L2 Experience	1	0.0098	0.92
	Preceding Vowel	1	7.5177	0.0061 **
[ɾ]$_{coda}$ - [l]	L2 Experience	1	1.8612	0.17
	Preceding Vowel	1	6.0333	0.014*
[ɾ]$_{coda}$ –[ɾə]	L2 Experience	1	0.0233	0.8787
	Preceding Vowel	1	1.6896	0.1936

In addition to the results relevant for our research questions, it is worth noting that a main effect of adjacent vocalic contexts was found in the model for [ɫ] - [w], as well as for [l] – [ɾ] both in onset and in coda (with a Bonferroni-corrected alpha level of 0.025). This indicates that L1-Mandarin learners discriminate [ɫ] - [w] better in /i_i/ context than in /a_a/ context, while an inverse context effect exists for [l] – [ɾ]. Particularly, learners showed higher discrimination accuracy between [l] – [ɾ] in /a_a/ context than in /i_i/ context.

3.3.2 Forced-choice identification task

As illustrated in Figure 3.2, in stark contrast to the EP natives, who always labelled all target segments correctly, L1-Mandarin learners had difficulty in categorizing both /l/ and /ɾ/ in intervocalic position, suggesting that, although /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$ are not encoded as homophones (accuracy rate higher than chance level), neither of their phonological representations are target-like, diverging from what was observed in L2 production (Zhou, 2017).

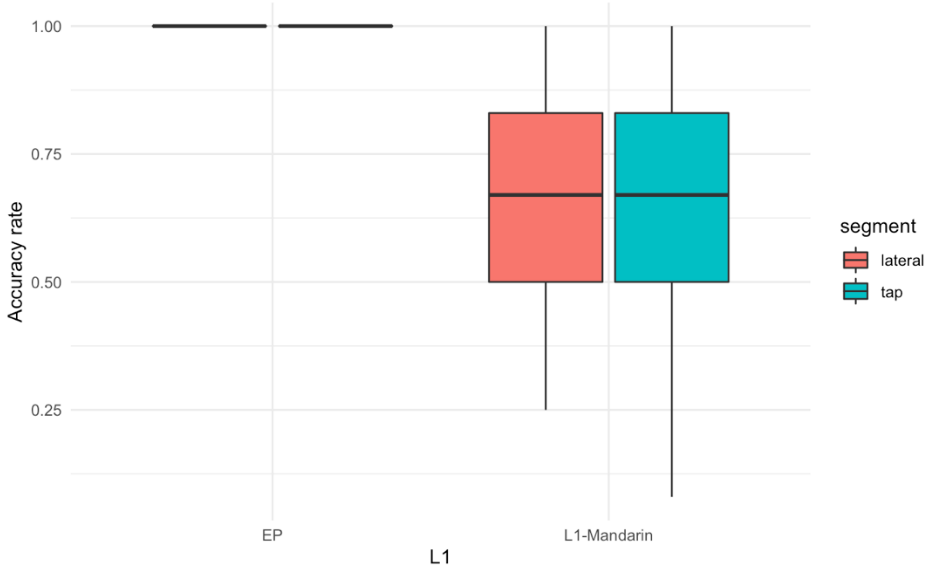


Figure 3.2: Accuracy results on the categorization of /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$ by EP native speakers and L1-Mandarin learners

Pertaining to RQ2 on whether L2 phonological representations develop as a function of more L2 experience with the target language, a visual inspection in figure 3.3 suggests that /l/$_{\text{onset}}$ improves with more exposure to the L2 input, whereas /ɾ/$_{\text{onset}}$ does not.

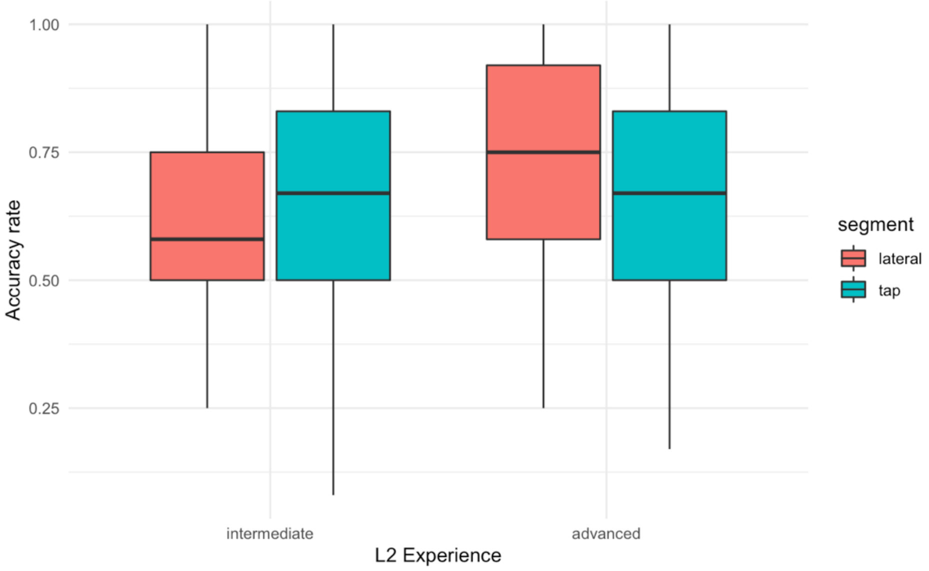


Figure 3.3: Accuracy results on the categorization of /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$ by learners with different L2 experience

We built a generalized linear mixed-effects model on the learners' identification results of /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$. The model had Segment (with contrast-coded two levels, L and R), L2 Experience (with contrast-coded two levels, Intermediate and Advanced) and Adjacent Vowel (with contrast-coded two levels, A and I) as predictors. Random intercepts for Participant and Trial, together with random slopes for Segment by Participant, Adjacent Vowel by Participant and L2 Experience by Trial were also included, see (3.5).

(3.5) Accuracy ~ Segment * L2 Experience * Adjacent Vowel+ (Segment | Participant) + (Adjacent Vowel | Participant) + (L2 Experience| Trial)

Table 3.7: Results of the models built for identification accuracy

Effect	df	Chisq	*p*.value
L2 Experience	1	3.31	0.07
Segment	1	0.33	0.56
Adjacent Vowel	1	10.07	0.002**
L2 Experience * Segment	1	3.49	0.06
L2 Experience * Adjacent Vowel	1	0.06	0.80
Segment * Adjacent Vowel	1	8.46	0.004**
L2 Experience * Segment * Adjacent Vowel	1	4.76	0.03 *

The model's results were summarized in Table 3.7. Again, no main effect of L2 Experience was found, suggesting that more L2 experience does not necessarily boost the development of L1-Mandarin learners' phonological representations of /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$.

Nevertheless, a main effect of Adjacent Vowel was found, indicating that, in general, the participants had better categorization accuracy in /a_a/ than in /i_i/ context. This corroborates what has been observed in the discrimination results. In addition, a significant effect of the interaction between Segment type and Adjacent Vowel further revealed that the /a_a/ context favours the identification of /ɾ/ (/aɾa/: M=0.79; /ala/: M=0.68), whereas /i_i/ context facilitates the categorization of /l/ (/iɾi/: M=0.53; /ili/: M=0.63).

3.4 Discussion

In the current chapter, we aimed to explore the perception-production interaction across prosodic contexts during the acquisition of EP /l/ and /ɾ/ by L1-Mandarin learners, as well as the plasticity of their L2 phonological representations, by investigating the following research questions:

RQ1: whether the deviant productions articulated by L1-Mandarin learners across prosodic positions stem from the inaccurate perception?

RQ2: whether L1-Mandarin learners experience less L1 interference in coda than in onset during the acquisition of EP /ɾ/?

RQ3: whether L1-Mandarin learners' phonological representations of /l/ and /ɾ/ are malleable in a mid-late stage of L2 speech learning?

Regarding the RQ1, results of an AXB discrimination task revealed that L1-Mandarin participants failed to reliably discern the differences between the target form and the segmental repair they often employ in production: [ɫ]$_{\text{coda}}$ - [w], [ɾ]$_{\text{onset}}$ - [l] and [ɾ]$_{\text{coda}}$ – [l]. This indicates that these L2 deviant productions have a perceptual motivation, in accordance with the prediction of current L2 speech theories: inaccurate L2 (segmental) perception leads to imprecise L2 (segmental) production (Flege, 1995; Escudero, 2005; Best & Tyler, 2007).

In the case of the use of L2 structural modifications for syllable-final tap, a divergence between L2 perception and production was attested. In particular, the epenthesis is perceptually driven whereas the segmental deletion is restricted to production. The insertion of an illusory vowel cannot be simply understood as a modification conforming to the L1 phonotactics (no /ɾ/ in coda) or to the universal syllabic constraint favouring format CV, since deleting the tap would likewise lead to structural well-formedness, i.e. /CV.CV/. Additionally, if the L1 structural requirement is the only underlying force, the participants would be expected to show a preference for segmental deletion as repair strategy, thus creating a disyllabic word[33], conforming to the Mandarin minimal word constraint that favours words of two syllables (Broselow et al., 1998). Given that both epenthesis and deletion are possible options to accommodate an illicit L2 structure, the deletion has, however, an apparent drawback as it would lead to a complete loss of segmental information. We thus speculate that the employment of an epenthetic schwa in L2 perception is a compromise between achieving the structural well-formedness and maximally maintaining the input information.

The next question is how the perceived illusory schwa appears in the L2 learners' production. The first possibility concerns the L2 perception-production loop formulated in the SLM. Namely, the perceived form with an epenthetic schwa (e.g. the target auditory form ca[ɾ]ta perceived as ca/ɾə/ta, "letter") is mapped to the L2 lexicon and later retrieved in production. This hypothesis was supported by a recent study conducted by Darcy and Thomas (2019), who demonstrated that L1-Korean learners of English encoded the perceptual epenthesis in the L2 lexicon[34], i.e. the English word "blue" represented lexically as |bʊlu:|. Another possibility is in line with the theoretical reasoning of generative phonology: the epenthetic schwa is not listed in the L2 lexicon, but inserted by the production grammar. For instance, according to *Richness of the Base* in the Optimality Theory (Prince & Smokensky, 1993), there is no restriction on the lexicon (the input in production) and all statements on the surface structure (the output) are achieved by grammar. Future studies tapping into the lexical level are needed to evaluate these two competing hypotheses.

[33] All test words containing a syllable-final tap are disyllablic. Therefore, deletion will lead to a disyllabic word while epenthesis gives rise to a trisyllabic word.

[34] Darcy & Thomas (2019) put forward an alternative explanation for why L1-Korean learners accepted an epenthetic form [bʊ ˈlu:] as a real word "blue": the phonolexical representation might not be fully specified with respect to CV Skeleton, timing slots or syllabic structure (p.15). We stay sceptical with this possibility as whether the structural information is part of the lexicon is still a matter of debate.

The high discrimination accuracy between the target form and the form where the syllable-final tap is deleted (/Vɾ.C/ - /V∅.C/) indicated that the omission of [ɾ] in L2 production cannot be attributed to misperception. What first comes to mind is that L2 learners developed distinct grammars for perception and production (Ramus et al., 2010). Although assuming separated grammars seems to be a rather simple solution, it might run the risk of not being theoretically detailed enough, thus generating inaccurate predictions (see Boersma, 2012 for a discussion). Instead, we present two alternative explanations. First, the omission of /ɾ/ might be a result of articulatory imprecision[35]. The Portuguese tap imposes great articulatory complexity since it stipulates a ballistic movement of the tongue tip and a constriction towards the pharynx (Berti, 2010; Barberena et al., 2014; Barberena et al., 2019)[36], and L1-Mandarin learners might need extra time and effort to master this novel gestural coordination. It is therefore very likely that they sometimes delete the tap in word-internal coda position, where consonant-to-consonant co-articulation increases articulatory difficulty. The second explanation pertains to the fact that two paralinguistic processes targeted by perception and production experiments involve different mappings: in the perception experiment, only the mapping from auditory to phonological surface form is triggered, while the production task also involves mapping of the lexical form onto the phonological surface form (Boersma & Hamann, 2009 b). We will come back to how a single grammar accounts for the asymmetry between L2 speech perception and production in section 4.5. The first hypothesis suggests that the segmental deletion occurs at the articulatory level, while the second argues for omission at the phonological level. These two hypotheses are testable

[35] We assume that speech perception and production differ in terms of the representational levels involved (e.g. the articulatory level is only activated in production, not in perception) but use the same grammar, which can be instantiated by the employment of the same set of constraints and same constraint ranking.

[36] All studies cited pertaining to the articulatory characteristics of the Portuguese tap are based on Brazilian Portuguese, since, currently, to our best knowledge, no comparable studies exist for EP. However, the potential articulatory differences between the EP tap and the Brazilian Portuguese one does not invalidate our argument that the articulation of this segment is challenging for Mandarin native speakers.

as the former one predicts the existence of /ɾ/ at the phonological level and, accordingly, co-articulation traces should be left in the adjacent segments. By contrast, the phonological deletion would not affect adjacent segments (Buchwald & Miozzo, 2011; 2012).

Another mismatch between L2 perception and production was found pertaining to the confusability between /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$. In particular, the results of the identification task demonstrated that, in contrast to L2 production (Zhou, 2017), learners manifested bidirectional perceptual confusability between /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$, corroborating earlier findings (Cao, 2018; Vale, 2020). Recall that the EP /l/$_{\text{onset}}$ was consistently assimilated to /l/ by naïve Mandarin listeners (chapter 2), implying that the reuse of the L1 lateral category should suffice for target-like production of the EP /l/$_{\text{onset}}$ from the onset of L2 learning (the *identical scenario* in the SLM). The acquisition of a novel sound category is, nevertheless, constrained by the presence of other segments from the same repertoire: apart from /l/$_{\text{onset}}$, the EP /ɾ/$_{\text{onset}}$ was likewise assimilated to the Mandarin lateral category (the *similar scenario* in the SLM). According to the SLM, the perceptual equivalence between /l/ and /ɾ/ will inevitably lead to the formation of a composite category (diaphone) in a common phonological space where L1 and L2 sound categories co-exist (Flege, 1995; Flege & Bohn, 2021) and this perceptual linkage will reinforce the two categories to eventually resemble one another, resulting in the perceptual boundary shift of /l/.

Alternatively, the "deterioration" of /l/$_{\text{onset}}$ as a function of L2 experience could be argued to be due to L1-like novel category creation (Escudero & Boersma, 2004). As reviewed in Section 1.1.3., acoustically speaking, the EP alveolar lateral and tap differ with respect to both formant values and duration (Rodrigues, 2015), while segmental duration does not cue any phonological contrast in Mandarin (Duanmu, 2007; Lin,2007; Smith, 2010). Therefore, the perceptual learning of EP /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$ by native Mandarin speakers is comparable to the learning scenario investigated in Escudero and Boersma (2004): native Spanish speakers, who do not use durational information in their L1, acquired the Southern British English /i/-/ɪ/, a vowel contrast differing in two acoustic dimensions (formants and duration). Since both /i/ and /ɪ/ are assimilated to a single Spanish category /i/, Escudero and Boersma reasoned that, in order to perceive /i/ and /ɪ/ as two categories, L1-Spanish learners could in principle split the /i/ category into two novel vowels, i.e. boundary shift on spectral dimension, or form a new length contrast that does not exist in the learners' L1, i.e. category formation on durational dimension /short/-/long/. Their computational simulation and experimental results pointed in the same direction that L1-Spanish learners chose the new length distinction over splitting an L1 category. The preference for temporal cues over spectral cues has been reported in many L2 acquisition studies, even if the learners' L1 does not rely on a durational cue for phoneme distinction (e.g. Bohn, 1995; Flege et al., 1997). Two possible explanations have been put forward in the literature to account for this counterintuitive finding. On the one hand, Bohn (1995) speculated that duration serves as a universal source for phonological distinction that learners can resort to if the L1 has insufficient spectral distinctions to separate two L2 categories (*Desensitization Hypothesis*). On the other hand, Escudero and Boersma (2004) argued that learners target an acoustic dimension that is not phonological informative in their L1, because category creation (novel length contrast) is more natural, being an L1-like acquisition strategy (Boersma et al, 2003), whereas category split has not been reported as a mechanism used by children.

Escudero and Boersma's proposal diverges from Bohn's *Desensitization Hypothesis*[37] by assuming that everything starts from scratch: no duration category nor duration-to-category mapping exist for L2 phonological grammar (E&B:p. 575). If Escudero and Boersma were correct about how novel category

[37] Bohn (1995) postulated that learners start with a pre-existing duration category, the one corresponding to their L1 category duration.

learning proceeds, L1-Mandarin learners of EP, who would also be expected to choose category creation (using durational cue) over splitting the old category into two (adjusting spectral cue boundary). In particular, they would first access an L1 acquisition mechanism, distributional learning[38] (Maye et al., 2002), to detect the two peaks in the binomial distribution of duration in the input, which allow them to build two abstract categories from the duration continuum[39]; then they would start learning the association between acoustic cues and the two novel categories[40]. Consequently, before developing target-like duration-to-category mapping, learners will not always be able to categorize /l/$_{\text{onset}}$ and /ɾ/$_{\text{onset}}$ accurately.

Although both Flege's and Escudero and Boersma's explanations for the perceptual "deterioration" of /l/$_{\text{onset}}$ as a function of L2 experience are plausible, they do not predict L2 perception and production to diverge. A straightforward solution to this mismatch is again to postulate that learners have developed distinct phonological grammars for L2 perception and production (Ramus et al. 2010). However, in section 4.5, we will argue against this distinct-grammar view by showing that the mismatch can emerge from an L2 phonological grammar, which is identical in the two speech modalities, and that L2 perception-production asymmetry emerges due to the fact that the two paralinguistic processes targeted by perception and production experiments involve different mappings: in the perception experiment only the mapping from auditory to phonological surface form is triggered, while the production task also involves the mapping of the lexical form onto the phonological surface form.

[38] Distributional learning refers to the ability of tracking statistical distribution of auditory tokens in the input.

[39] See Gulian et al. (2007) and Nixon (2020) for experimental and computational evidence that L2 learners are able to acquire a new phonological contrast by applying distributional learning to acoustic cues

[40] This can be achieved through the Gradual Learning Algorithm (Boersma & Hayes, 2001). See Boersma et al. (2003) for L1 and Escudero & Boersma (2004) for L2 cue-category (auditory to phonological surface form) mapping learning respectively.

RQ2 concerns the phonotactics-based explanation for the onset-coda asymmetry with respect to the developmental sequence of /ɾ/ across syllable contexts. A significant effect of syllable position confirmed that in L2 speech L1-Mandarin learners do experience less L1 interference of /l/ in syllable-final position than in syllable-initial position, presumably due to a phonotactic constraint from the learners' L1 that bans [l] in coda position[41] (Duamnu, 2007; Lin, 2007). This phonotactically-conditioned L2 perception can thus contribute to the accelerated acquisition of /ɾ/ in coda. This result, together with findings from chapter 2, corroborates earlier findings on cross-linguistic phonotactic restriction in L2 phonological categorization (e.g. de Jong et al., 2009; Li & Zhang, 2017; Park & de Jong, 2017; Rasmussen & Bohn, 2019), suggesting that L2 segmental acquisition is not only subject to the relationship between L1 and L2 categories, but also constrained by more abstract phonological restrictions, namely the learners' L1 phonotactics. Our findings can in no way be regarded as evidence against the universal salience of onset position, especially in L1 and L2 phonological acquisition. What perception and production data suggest instead is that CLI may sometimes override universals in L2 speech learning. Future studies should examine the dynamic interaction between CLI and universals during L2 phonological development.

As for RQ3, both discrimination and identification results suggest that more exposure to the target language does not seem to contribute to the refinement of L2 phonological representations in a mid-late stage of learning. In contrast to studies where only self-reported length of residence was used to quantify L2 input, the two groups of participants in the current study differed substantially not only in the time they have spent in an immersion language course in Portugal (0.2 years vs. 1.59 years), but also in the years of formal instruction they have received in China (2 years vs. 4 years), minimizing thus the possibility that the difference in terms of the amount of L2 input is not

[41] The alveolar lateral [l] is not allowed in coda by English, the other L2 either.

sufficient for category development. Note that the experimental finding does not imply that more L2 experience in naturalistic learning is not beneficial for all learners, but rather suggests that it might not be a determining source for the L2 phonological optimal attainment at the group level. It is totally conceivable that some learners have developed target-like L2 categories, whereas others haven't yet. This is supported by the notable within-group variance in terms of perception accuracy (see Figure 3.1). This inter-speaker variability could be attributed to different learning strategies or to a domain-general auditory processing ability (Saito et al., 2020). What underlies this variability goes beyond the scope of this study and calls for further research. In the following two paragraphs, we will speculate what might make L2 experience in general not as helpful as one may think at a mid-late stage of L2 speech learning.

The null effect of L2 experience at a mid-late learning stage is in accordance with the PAM-L2's prediction that the refinement of L2 phonological representation might only occur in a very short period of time at the onset of learning, because the learning task at later stages involves other grammatical structures (e.g. morphosyntactic, semantic and pragmatic), which would avert learners' attention away from the phonetic-phonological level. The crucial role that learners' attention plays in L2 category development has been supported by laboratory training studies, where learners' perceptual performance improves immediately after a few training sessions in a brief period of time, irrespective of their proficiency and ages (Wong, 2013; Rato, 2014; Oliveira 2020; see Sakai & Moorman, 2018 for a meta-analysis). These studies normally employ a High Variability Perceptual Training technique (HVPT), which aims to direct learners' attention to the critical acoustic differences between the two confusable non-native categories (Lively et al., 1993; 1994; Wong, 2012; 2014). Antoniou and Wong (2016) tested the role of learners' attention on L2 category learning (Voice Onset Time of prevoiced and of unaspirated stops) by manipulating the variation of an irrelevant acoustic feature (lexical tone) in the auditory stimuli (with vs. without variation; more variation of the irrelevant feature implies less attention that learners can pay to the critical feature) during their HVPT with English learners of Hindi. Their results showed that participants who were trained with stimuli that varied in the irrelevant feature were outperformed by those that were trained with stimuli that held the irrelevant feature constant, suggesting that learners' attention to the critical acoustic feature is essential to the L2 perceptual development.

The L2 speech learning outside laboratory is more attentionally demanding as it entails further variation of more acoustic information[42] that is irrelevant to the categories under acquisition. Therefore, the null effect of L2 experience attested in the current study might be due to the learners' reduced attention to the phonetic information in a mid-late L2 speech learning phase. Future perceptual training studies on /l/ and /ɾ/ across syllable contexts may be promising as numerous studies have demonstrated that the perceptual training gain can be maintained over time (e.g. Nobre-Oliveira, 2007; Wang, 2008; Rato, 2014), generalized to untrained items or talkers (e.g. Nobre-Oliveira, 2007; Aliaga-Garcia, 2010, Rato, 2014) and extended to L2 production (e.g. Rato, 2014).

In addition to the lack of attention to the phonetic differences, what further constraints L2 category refinement could be the underspecified lexical representation. It has been well acknowledged that category learning benefits from both "bottom-up" and "top-down" processes (e.g. McCandliss et al, 2002; Boersma et al., 2003; Boersma, 2012; Nixon, 2020): the former refers to the ability of tracking statistical distribution of auditory tokens in the input, known as *Distributional Learning* (Maye et al., 2002), while the latter stands for feedback on auditory categorization (e.g. Ganong, 1980). The top-down

[42] For instance, variation induced by noise, gender, dialects, social status, to make some example.

information has been argued to be a crucial trigger in L2 phonological acquisition (McCandliss et al, 2002; Escudero & Boersma, 2004; Boersma & Escudero, 2008) and the lack of such feedback could be even detrimental (Fuhrmeister & Myers, 2017). One fundamental source for top-down feedback is the lexicon, which is conceptualized as a supervisor for achieving a more accurate phonological representation (Boersma et al., 2003). Particularly, if a learner detects an error in their speech, perhaps due to the semantic violation denoted by sentential context (e.g. a perceived *pu/l/o*, which means jump, is not the intended adjective in the sentence: *O ar aqui é mais pu|ɾ|o* [The air here is cleaner]), they will adjust the category boundary in order to accommodate the auditory input more accurately in the future (know as error-driven or lexicon-driven learning).

A considerable amount of studies has demonstrated, however, that an L2 lexical representation is normally far from being fine-grained and it is fuzzy not only for a difficult L2 category (Amengual, 2016; Darcy et al., 2012; Darcy et al., 2013; Kojima & Darcy, 2014), but also for a perceptually non-confusable category (Cook et al., 2016). A fuzzy lexical representation can be understood as phonologically underspecified[43] and compatible with both the target form and its confusable counterpart. Consequently, a mismatch between the perceived form (either target or not) and the lexical form will not occur, providing thus no informative top-down information and not triggering lexicon-driven category learning. Our identification results did point toward this possibility as the /l/ category, which is assumed to be target-like at the underlying level (Zhou, 2017), benefits more from increasing L2 experience than the /ɾ/ category, which can be realised either as an alveolar lateral or a tap, though this difference is statistically marginal.

[43] An alternative interpretation for the fuzzy L2 lexical representation is the co-existence of multiple underlying forms connected to the single lexical entry (John & Cardoso, 2017). Testing the underspecified account and the multi-representational account goes beyond the scope of this study, but they do not conflict with respect to the fact that a fuzzy lexical representation does not yield a mismatch between the underlying form and the perceived surface form, thus not triggering category learning.

In addition to the three research questions explored in this study, significant effects of preceding vowel were found on the perceptual discrimination between [ɫ] - [w], as well as between [l] – [ɾ]. These results indicate that L1-Mandarin learners distinguish between [ɫ] - [w] better when the target segment is preceded by /i/ than by /a/, while the discrimination between [l] and [ɾ] is facilitated when the target liquid follows /a/. This facilitating contextual effect on speech perception has been long observed in the literature (e.g. Liberman et al., 1952; Mann & Repp, 1980; 1981; Mann, 1986) and can be understood as the enhancement of critical acoustic cue for categorizing the target segment (see Stilp, 2020 for a review).

Regarding the contrast [ɫ]-[w], although no data of EP is available, acoustic evidence of American English indicates that the F1 and F2 frequencies for [w] and [ɫ] coincide to a large extent[44], while F3 for /w/ is somewhat lower than F3 for [ɫ] (Lehiste, 1964; *apud* Recasens, 1996). Based on this acoustic characteristic, we speculate that the observed facilitating effect of adjacent vowel [i] can be attributed to the fact that the preceding [i] enhances the acoustic difference between [w] and [ɫ] in terms of F3. In particular, in comparison with the EP [a], which normally has F3 values of 2333 Hz (Escudero, 2009; male speaker), the EP [i] with relatively high F3 (2774 Hz) leads to a steeper F3 transition to [w], whose F3 values are presumably close to that of [u] (2315 Hz). Please see the comparison in terms of F3 transition slope between the right side of Figure 3.4 and the right side of Figure 3.5. This more notable downward[45] F3 transition from [i] to [w] (in comparison with the F3 transition from [a] to [w]) might help L1-Mandarin learners to better distinguish [w] from [ɫ] in a perceptual task, because it contrasts more with the

[44] The high degree of overlap in terms of F1 and F2 values elucidates why L1-Mandarin learners struggle with their distinction.

[45] One may argue that the F3 transition from [a] to [w] could in principle be also downward. It is necessary to emphasize that what facilitates the perception of [w] is not only the direction of F3 transition, which is in contrast to the F3 transition for [ɫ], but also the slope of this transition.

upward F3 transition (see the left side of Figure 3.4 and 3.5) from [i] to [ɫ] (F3 of [ɫ]: 3054; Rodrigues et al., 2019).

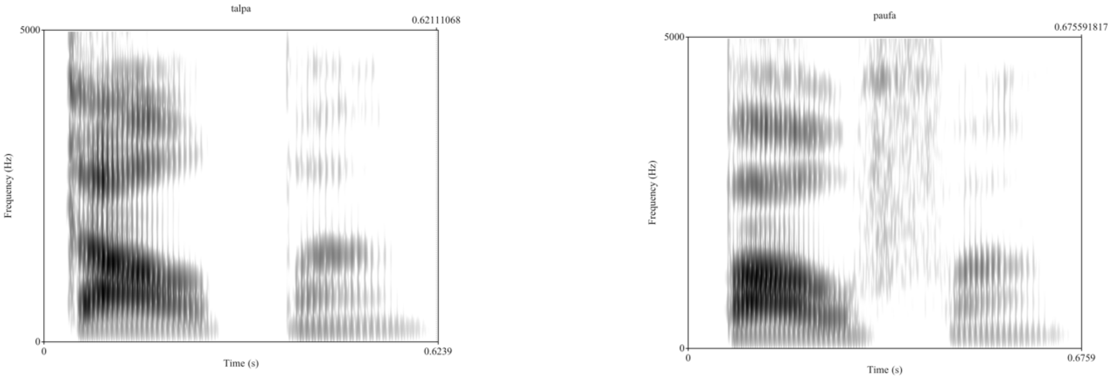


Figure 3.4: Spectrograms of [aɫ] (left) and of [aw] (right) produced by a male Portuguese speaker

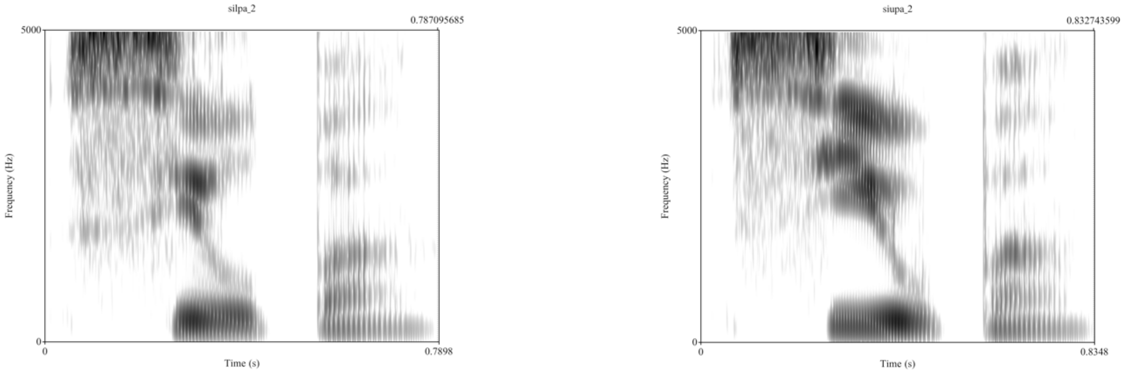


Figure 3.5: Spectrograms of [iɫ] (left) and of [iw] (right) produced by a male Portuguese speaker

Accordingly, the facilitating effect of preceding [a] on the discrimination between [l] and [ɾ] might also be due to the enhancement of acoustic difference between the two liquids. As reviewed in 1.1.3, no prior research exists for the acoustic comparison between /l/ and /ɾ/ in the standard EP. Nevertheless, the acoustic studies on the southern EP variant (Rodrigues, 2015) and on the Rioplatense Spanish (Guirao & García Jurado, 1991) suggest that [l] and [ɾ] might differ both in spectral (F1, F2, F3 formant values and F2 formant transition) and durational dimensions. Whether the vocalic quality affects the duration of the following consonant is unclear, but it seems more conceivable that vowels modulate the formant information of the following liquids.

In comparison with a preceding [i], the overall facilitating effect of [a] on the perceptual discrimination between [l]-[ɾ] can be attributed to the fact that [a] leads to more salient formant transition slopes than [i] does, thus empathizing the formant differences between [l]-[ɾ]. For instance, the F1 and F3 for the front vowel [i] are more close to the formant values of coronal liquids than [a] does.

The facilitating effect of [i] on the identification of /l/ might be due to the salient F2 transition, as illustrated in the left side of figure 3.7. However, what makes the context [a_a] favours the categorization of /ɾ/ remains unclear to us.

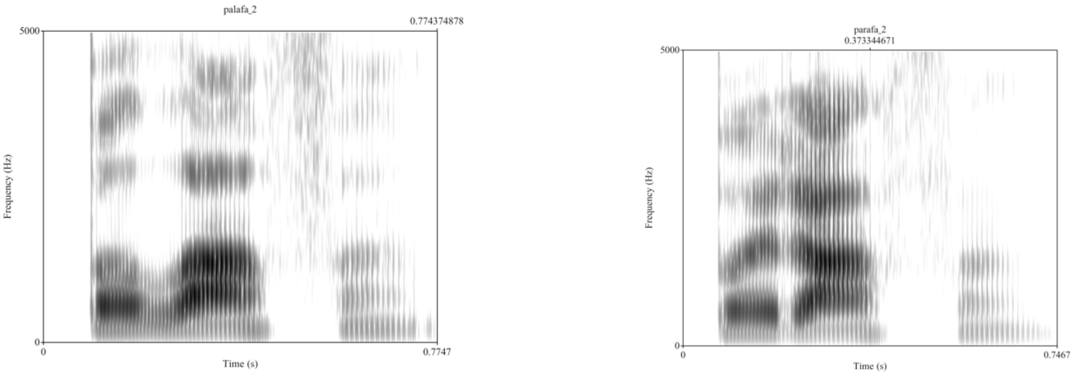


Figure 3.6: /ala/ (left) and /aɾa/ (right) produced by a male Portuguese speaker

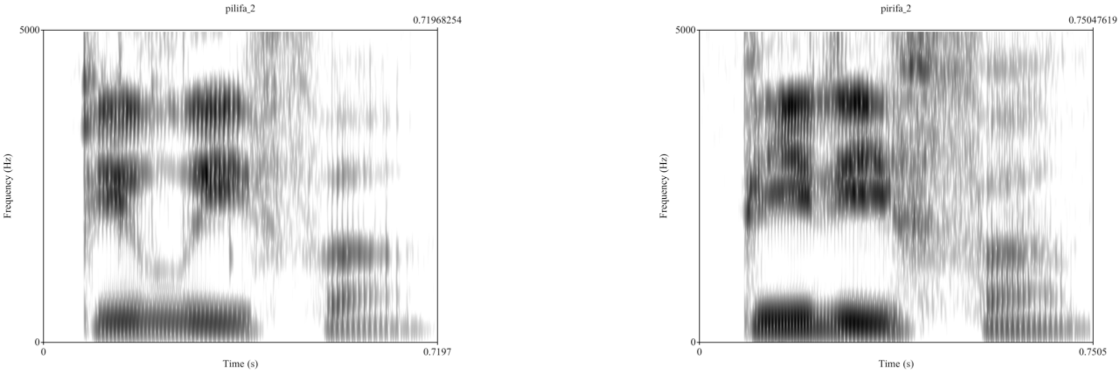


Figure 3.7: /ili/ (left) and /iɾi/ (right) produced by a male Portuguese speaker

To sum up, in this chapter, we first investigated the interaction between speech perception and production in L2 speech learning, by examining whether the L2 deviant productions stem from misperception and whether the order of acquisition in L2 speech perception mirrors that in production. Secondly, we tested whether L2 phonological categories remain malleable at a mid-late stage of L2 speech learning. Results demonstrated that, although L1-Mandarin learners perceptually confuse the target Portuguese segments ([ɫ] and [ɾ]) with some deviant forms they tend to produce (e.g. [w] for the velarised lateral; [l] and [ɾə] for the tap), some imprecise production cannot be attributed to misperception (deletion of syllable-final tap). On the other hand, the order of acquisition (/ɾ/$_{\text{coda}}$ > /ɾ/$_{\text{onset}}$) was shown to be consistent in L2 perception and production. The correspondence as well as discrepancy between the two speech modalities signal a complex relationship between L2 speech perception and production. Regarding the question on the plasticity of L2 phonological categories, no main effect of L2 experience was found, which suggests that L2 experience seems to play no role in the refinement of these non-native phonological categories at a non-initial stage of L2 speech learning.

Chapter 4: Formalising the interaction between speech perception, production and orthography in L2 phonological acquisition of European Portuguese /l/ and /ɾ/

4.1 Introduction

In the previous two chapters, we have experimentally examined how L2 phonological categories /l/ and /ɾ/ are constructed across different prosodic positions (onset vs. coda) and represented in two learning stages (initial vs. mid-late). Several intriguing findings were reported, namely, variations in L2 phonological categorization (chapter 2), interaction between phonology and orthography during L2 category creation (chapter 2), and asymmetry between L2 perception and production (chapter 3).

As reviewed in 1.2, current L2 speech theories (e.g. Flege, 1995; Best & Tyler, 2007; Honikman, 1964; Colantoni & Steele, 2008) only include some aspects of non-native phonological acquisition, thus not providing a comprehensive account for neither of the aforementioned phenomena. This chapter sets out to bridge this gap by formalising our experimental results, namely the interaction between speech perception, production and orthography within one generative linguistic model, the Bidirectional Phonology and Phonetics Model (henceforth: BiPhon; Boersma, 2007; Boersma & Hamann, 2009a; Boersma, 2011).

The present chapter is structured as follows. Section 4.2 introduces the basic architecture and crucial assumptions of the BiPhon model. Section 4.3 deals with the formalisation of the variations in L2 phonological categorization, i.e. 4.3.1 between-subject variation, 4.3.2 within-subject variation, 4.3.3 variation as a function of prosodic context. Section 4.4 shows how phonological categorization and orthography interact in shaping the construction of L2 categories. In section 4.5, we provide a formal account for the mismatch between L2 perception and production without assuming distinct grammars between modalities.

4.2 Bidirectional Phonology and Phonetics Model

BiPhon is a model of phonetics and phonology which assumes multi-level representations and aims to account for both speech comprehension and production. The connections between representational levels can be formulated with ranked constraints, much like in classical Optimality Theory (Prince & Smolensky, 1993; henceforth: OT), with weighted constraints of Harmonic Grammar (Legendre et al., 1990) or with weighted connections of Neural Network (Boersma et al. 2020).

In the present study, we adopt the OT version of the model (henceforth: BiPhon-OT[46]). OT is a well developed generative framework of phonology, and BiPhon-OT further expands on the classical OT in several aspects, which makes it a suitable tool to model the experimental findings of this thesis.

1. BiPhon-OT is stochastic.

In the classical OT (also known as Strict OT), the phonological grammar is represented by a set of violable constraints, which are discretely ranked with respect to each other. This discreteness implies that the decision-making in Strict OT is categorical: the violation of higher ranked constraints is always more fatal than the disobedience to lower ranked ones. For instance, as illustrated in tableau (1), given two possible candidates, A and B, the candidate B that violates Constraint 2 (indicated with a violation mark "*") turns out to be more harmonic, since the other candidate, A, is penalized by a higher ranked Constraint 1. The violation of Constraint 1 is deemed as fatal (indicated with an exclamation mark "!") due to the strict ranking *Constraint 1 > Constraint 2*. The grey boxes indicate that candidate A is no longer in the running, even if it does

[46] BiPhon-OT has successfully been implemented to account for or various phonological processes (Boersma, 2006a for Prototype effect; Boersma, 2006b for McGurk effect; Boersma, 2008 for Licensing by cue; Boersma & Hamman, 2008 for auditory dispersion; Boersma & Hamann, 2009b for loanword adaptation; Hamann, 2009b for sound change; Azevedo-Quintanilha, 2016 for vowel epenthesis in native speech), L1 phonological acquisition (Boersma et al., 2003 for category creation; Hamann et al., 2012 for phonotactics formation in the lexicon; Vogele, 2020 for phonological disorders) and L2 speech learning (Escudero & Boersma, 2004 for L2 speech perception, Hamann 2009a for L2 speech variation).

not disobey the lower ranked Constraint 2 that B violates. As a result, candidate B is chosen as the output (indicated with a pointing finger "☞") in example (1).

(1)

Input	Constraint 1	Constraint 2
A	*!	
☞ B		*

However, the categorical behaviour implies that strict OT lacks certain flexibility, which makes it unable to account for some real-world data, i.e. variations that have been frequently reported by studies on L2 phonological acquisition. To give an example, before fully mastering the English coda stops, Brazilian learners of English may sometimes insert an epenthetic vowel [i] in their English production, e.g. /dɒg/ → [dɒgi]. The L2 acquisition of the word-final stop does not take place across the board, but proceeds gradually, which means that the co-occurrence of the correct form [dɒg] and the epenthetic form [dɒgi] can last for some time in learners' production (Cardoso, 2011; John & Cardoso, 2017). The alternation between two possible forms can be easily formalised in the BiPhon model, which adopts the stochastic version of OT (Boersma, 1998; Boersma & Hayes, 2001).

In contrast to the strict OT, the stochastic OT proposes a continuous scale of constraint strictness whereby each constraint is assigned with a value that can be temporarily altered by a random positive or negative value (noise) at each evaluation time. Accordingly, each constraint does not have a single fixed value, but rather is associated with a range of values. Two possible scenarios can thus be imagined. First, if the value range of constraint C1 and that of C2 do not overlap, the ranking scale merely recapitulates the typical categorical ranking as in the strict OT, as shown in Figure 4.1. In this case, no variation is expected.

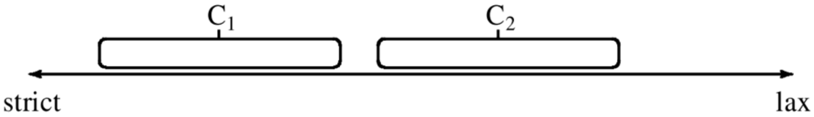


Figure 4.1: The instance of categorical ranking in Stochastic OT (strict direction signals high-ranked; Boersma & Hayes, 2001: 47)

The other scenario refers to a situation in which the ranges of two constraints partially overlap, as in Figure 4.2. Since it is possible to choose a value (named as *selection point* in stochastic OT) from anywhere within the range of a constraint, two types of rankings can be yielded. If a value of C2 were chosen from the left part of its range and a value of C3 from the right part, C2 would be ranked above C3. However, if a ranking value of C2 were chosen from the rightmost part of the range and the value of C3 from the leftmost, then C3 would outrank C2.

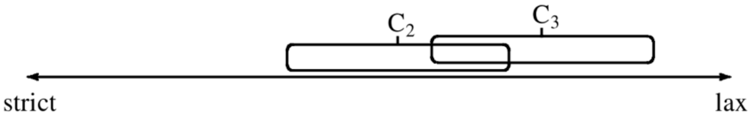


Figure 4.2: The instance of categorical ranking in Stochastic OT (strict direction signals high-ranked; Boersma & Hayes, 2001: 48)

As shown above, the Stochastic OT makes it possible to simulate the relative frequency of each ranking type (C2 > C3 or C3 > C2) and, consequently, model the variation in the output.

2. BiPhon-OT employs the Gradual Learning Algorithm (GLA).

Another advantage of BiPhon-OT over the original proposal by Prince and Smolensky (1993) lies in its associated learning algorithm. The strict OT speculates that, at least for children, markedness constraints which favour unmarked structures are initially ranked higher than faithfulness constraints that demand the input and the output to be isomorphic (e.g. Tesar & Smolensky, 1998; but see Hale & Reiss, 1998 for the opposite initial ranking)[47]. On the assumption of this initial state, the child has to learn that certain marked structures are permitted in the native language (e.g. coda). In the traditional OT, this grammatical learning refers to constraint re-ranking, which can be achieved by lowering the corresponding markedness constraints (e.g. NoCoda) via the Error-Driven Constraint Demotion learning algorithm (Tesar & Smolensky, 1998)[48]. In particular, as shown in tableau (2), when a child or a learner discovers that their output D mismatches the input (the error detection[49] is represented by a check mark on the correct candidate C), they will adjust the current constraint ranking by demoting (the direction of constraint movement is signalled by an arrow "→") constraints that penalize the correct output.

(2)

Input	Constraint 3	Constraint 4	Constraint 5	Constraint 6
√ C (the correct output)	*→			*→
☞ D (learners' output)		*	*	

One main criticism to the application of this learning algorithm in strict OT is that it may lead to an abrupt change in the grammatical development (Fikkert & De Hoop, 2009), which does not conform to the realistic gradual learning curves (Boersma & Hayes, 2001).

[47] This assumption is not necessary for learning language-specific cue knowledge, e.g. the ranking of cue constraints (Boersma et al., 2003), as it only "affect(s) the amount of input data and computation needed, but do(es) not materially affect the final outcome" (Boersma & Hayes, 2001: 51); however, having initial rankings may be crucial to learn phonotactic restrictions represented by structural constraints in BiPhon (Boersma & Hayes, 2001; Hamann et al., 2012).

[48] An alternative for constraint re-ranking is through Promotion (Bernhardt & Stemberger, 1998), which will be not further discussed here as it faces the same problems as Constraint Demotion.

[49] "Although this may sound like a case of explicit learning, one way to interpret the process is that the cognitive system is alerted when it hears something unexpected, i.e., an input that violates the current grammar. However, this does not have to reach the learner's consciousness" (Fikkert & De Hoop, 2009:319).

By contrast, the Gradual Learning Algorithm (GLA), associated to the Stochastic OT (BiPhon-OT), introduces small changes in the ranking values of constraints with every learning step. In particular, in the GLA, the amount of adjustment to the constraint ranking value at each learning step is contingent on a numerical plasticity, which is set to be reasonably small (See Boersma & Hayes, 2001 for the discussion on implications of choosing different plasticity values). In this way, an L2 learner's phonological grammar that starts with being not target-like will gradually approach the target grammar and the alternation between the target form and the deviant form is anticipated.

The GLA is also error-driven, but different from the Error-Driven Constraint Demotion learning algorithm, the GLA not only demotes but also promotes constraints. To give an example, the GLA is triggered each time a learner detects an error in their speech, perhaps due to the semantic violation denoted by sentential context (e.g. a perceived *pu/l/o*, which means jump, is not the intended adjective in the sentence: *O ar aqui é mais pu|ɾ|o* [The air here is cleaner]). Consequently, the GLA will raise the constraints penalizing the currently selected output (the unintended one) and, at the same time, will move those constraints favouring the wrong winner downwards.

3. BiPhon-OT assumes multi-level representations and bidirectionality.

In a considerable amount of studies employing the classical OT, the model is considered to contain two levels of representation, an underlying level and a surface level. The underlying representation refers to the highly abstract phonological information (e.g. only contrastive information) in the speakers' lexicon, while the surface representation contains both hidden structures (e.g. syllable boundaries) and overt phonetic information (e.g. acoustic cues and articulatory gestures). In this approach, it is impossible to separate phonetics and phonology, as (at least) some phonological information and some phonetic details are intertwined at the surface level. Although some researchers may claim that the surface representation does not entail phonetic details and the phonetic representation is yield through a universal automatic phonetic implementation (e.g. Hale & Kissock 2007). However, this view has been seriously challenged by cross-linguistic perception studies (see 1.2.2), which demonstrate that the mapping between phonology and phonetics is language-specific. Moreover, the classical OT model has been basically employed to deal with production-only process, i.e. how an underlying representation (input) is converted through a set of constraints into a surface representation (output). Nevertheless, it has recurrently been shown that perception involves phonology (see Boersma & Hamann 2009 for a review), thus a comprehensive phonological model needs to integrate both modalities into its formulation.

In stark contrast to the classical version of OT, BiPhon-OT takes up the aforementioned challenges. First, BiPhon-OT employs a modular approach, assuming multi-level representations, which are compatible with major psycholinguistic models (see 1.2.1), and establishing a clear distinction between phonetics and phonology. Figure 4.3 illustrates representational levels proposed in BiPhon and their connections[50]. The BiPhon model assumes two discrete phonological representations, an Underlying Form (UF) and a Surface Form (SF); and two continuous phonetic representations, an Auditory Form (AudF) and an Articulatory Form (ArtF). UF is a stored phonological form linked to the lexicon (e.g. a morpheme); SF is a prosodically detailed representation which also contains prosodic constituents; i.e., features, segments, syllables and feet. AudF is a continuous representation of sound which consists of acoustic information such as noise, pitch, spectrum and duration. ArtF is a continuous representation of the articulatory gestures, e.g. tongue and lip movements, jaw depression.

Second, all constraints in BiPhon-OT are bidirectional, which means that the same set of constraints and constraint rankings are used in perception and production.

[50] In this study, we simply ignore the morpheme level and the lexical constraints, which are irrelevant for our formalisation (see, Apoussidou, 2007 and Boersma, 2011 for their possible influence on phonological processes).

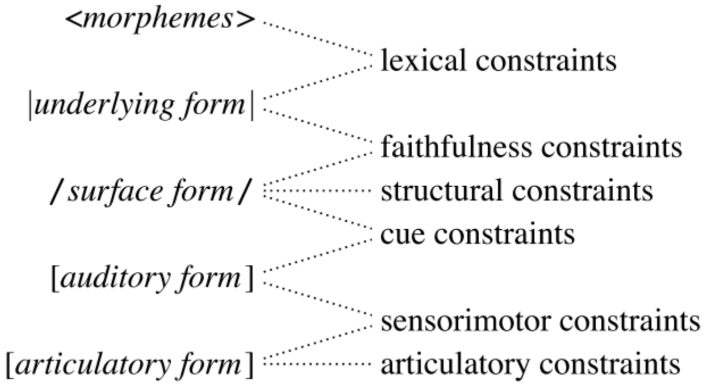


Figure 4.3: Multi-level representations connected by OT-constraints in BiPhon (Boersma, 2009)

In the comprehension direction[51], the prelexical phonological categorization in the BiPhon model refers to the mapping from continuous auditory information (AudF) to a discrete phonological surface representation (SF)[52]. The relation between AudF and SF is expressed by cue knowledge, formalised as cue constraints (Escudero & Boersma, 2003; 2004; Boersma, 2009). The SF, i.e., the output of phonological categorization, undergoes the evaluation of structural constraints, which reflects language-specific phonotactic well-formedness. During word recognition, the mapping between SF and UF is evaluated by faithfulness constraints, which penalize any mismatch between the two phonological forms.

[51] The perception and production architecture assumed in BiPhon is mostly compatible with the current generative and psycholinguistic models. The divergence between BiPhon and other models has been discussed in 1.2.1 and 1.2.2.

[52] Note that there is another line of research, namely the Direct Realist Theory of Speech Perception (Fowler, 1986) and the Motor Theory of Speech Perception (Liberman et al., 1967), which advocate that the speech signal is interpreted in terms of articulatory gestures. The primacy of the articulatory representation in speech perception process has been strongly defended with the evidence that listening to speech evokes neural responses in the motor cortex (e.g. Fadiga et al., 2002, Watkins et al., 2003; Wilson et al., 2004). Nevertheless, a recent study by Cheung et al. (2016) revealed that the neural patterns evoked during listening differs substantially from those during speech articulation and the structure of neural responses during listening was organized along acoustic features similar to auditory cortex, rather than along articulatory features as during speaking, suggesting that the perceived auditory input is not represented as articulatory gestures in motor cortex. Furthermore, a model whereby speech perception relies on articulatory gestures is not compatible with the fact that infants' perception precedes production (e.g. Jusczyk, 1997), which would not be possible if perception hinges on the availability of articulatory representations.

In the production direction, the aforementioned two processes execute inversely (UF → SF → AudF). UF is converted into SF by means of faithfulness and structural constraints. The translation from AudF to ArtF is evaluated by sensorimotor constraints, which express the speakers' knowledge of the relation between sound and articulation, and the output of the whole production chain, ArtF, is further subject to articulatory constraints, which militates against articulatory effort.

Whether the mappings between representational levels proceed in a sequential (serial) or parallel (interactive) fashion has been a matter of vigorous debate (e.g. Norris et al., 2000; Boersma, 2009). The BiPhon model proposes cross-level parallelism in both directions, which entitles a straightforward account for the Ganong effect[53] (Ganong, 1980) in perception and gradient phonetic influence on discrete phonological decisions in production (Kirchner, 1998) without adopting process-specific mechanisms (Boersma, 2005; 2006; 2011; Boersma & van Leussen, 2017). In particular, in BiPhon-OT, given an AudF as input, the perception grammar (constraint ranking) does not decide on a group of singular candidates (e.g. candidate A, candidate B, candidate C), but evaluates a paired of SFs and UFs (e.g. SF_a-UF_a, SF_a-UF_b, SF_b-UF_a, SF_b-UF_b). In this way, the mapping from AudF to SF and the one from SF to UF are interactive, allowing high-level representation (UF) and constraint (faithfulness constraints) to influence low-level processing (AudF to SF). We come back to the parallelism assumption in 4.5, whereby it becomes crucial to model the mismatch between L2 perception and production.

(4) BiPhon-OT takes orthographic influence into consideration

Although the mutual influence between phonology and orthography has been the research topic in many experimental studies[54], orthography has been long ignored in the formulation of formal phonological theories. One exception

[53] The "Ganong effect" is the tendency to perceive an ambiguous speech sound as a phoneme that would complete a real word.

[54] see 1.2.4 for a review of orthographic influence on L2 speech.

is BiPhon-OT. Apart from its machinery that allows modelling of phonetic and phonological phenomena, BiPhon-OT also integrates a reading/writing grammar to account for the orthographic influence (Hamann & Colombo, 2017; Hamann, 2020).

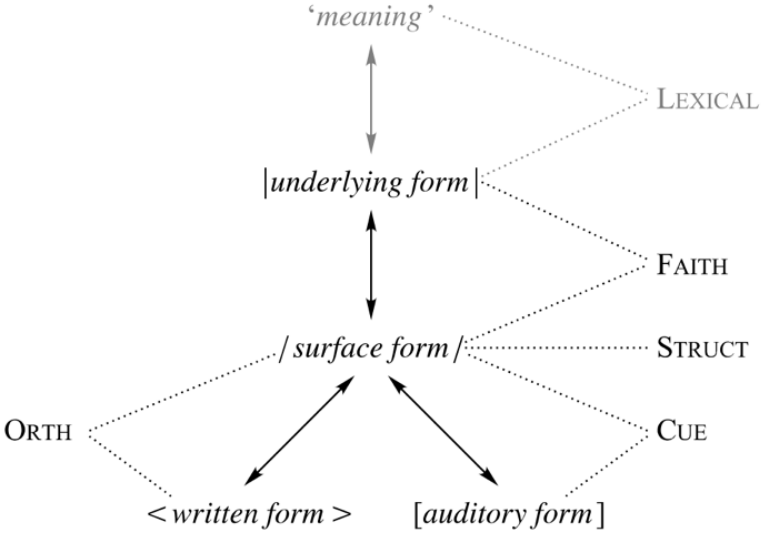


Figure 4.4: The integration of orthographic influence via orthographic constrains (ORTH) in BiPhon (Hamann & Colombo, 2017, p.701)

In BiPhon-OT, the orthographic effect is modelled through orthographic constraints, which represent the conversion knowledge between grapheme and phoneme. It is assumed that these orthographic constraints compete with structural constraints in selecting SF in perception, while interacting with faithfulness constraints in regulating SF in production, as shown in Figure 4.4.

In sum, BiPhon-OT manifests several important characteristics, which lead us to deem that it is suitable to model the interaction between speech perception, production and orthography manifested in the experimental results of this thesis. In the following sections of this chapter, we formalise some experimental findings obtained in chapter 2 and 3 within BiPhon-OT. We begin with the variation in L2 phonological acquisition of EP /ɾ/.

4.3 Formalising variations in L2 phonological acquisition of EP /ɾ/

In chapter 2, considerable variability was attested in naïve Mandarin listeners' categorization of the EP tap. In the present section, we set out to formalise three types of variation observed at this initial stage of L2 phonological acquisition: 1) between and 2) within-subject variations in L2 phonological categorization and 3) variation as a function of prosodic context. We will then argue how these patterns in naïve phonological categorization lead to the prosodic effect in late L2 learners' production of /ɾ/ (Zhou, 2017; Liu, 2018).

4.3.1 Between-subject variations in L2 phonological categorization

An individual-based analysis of the imitation responses in chapter 2 showed that the naïve Mandarin listeners can be grouped into two types, on the basis of their categorization patterns pertaining to the EP onset tap: Type I systematically identified the tap as /l/ and Type II mapped [ɾ] onto either /l/ or an alveolar stop [t/t^{h}], see Figure 4.5.

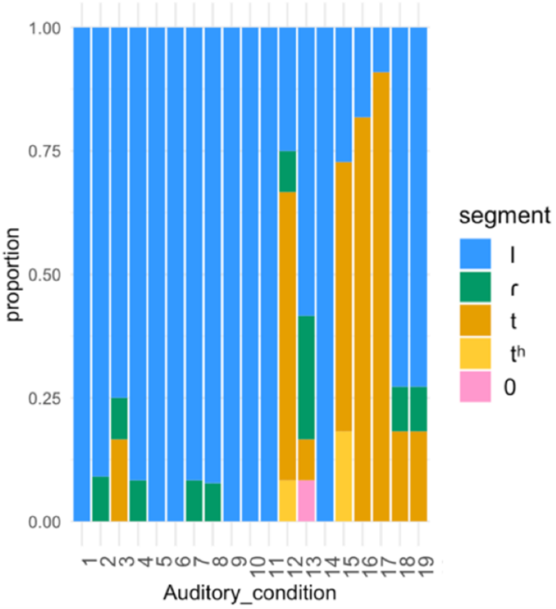


Figure 4.5: Naïve Phonological categorization of the EP onset tap

To be able to explain the differences between the EP native speakers and the L1-Mandarin learners of EP, we first need a clear depiction of how the EP native listeners reliably categorize the two auditory forms $[\text{l}]_{\text{Aud}}$ and $[\text{ɾ}]_{\text{Aud}}$ as two separate categories. In BiPhon-OT, phonological categorization[55] refers to the process of mapping an AudF onto an abstract SF, which can be formalised by means of negative *cue constraints* (Boersma, 1998; 2009; Escudero & Boersma, 2003; 2004).

(3) *Cue constraints*

"A value *x* on the auditory continuum *y* should not be perceived as the phonological category *z*" or short "$[x]_{\text{AudF}}$ is not perceived as $/z/_{\text{SF}}$".

As stated in (3), cue constraints constitute the phonetics-phonology interface in the BiPhon model as they allow any arbitrary connection between a discrete phonological category (e.g. feature, segment, syllable) and a continuous auditory dimension. In a modelling whereby only one auditory dimension or two different phonological categories are considered, the positively formulated cue constraint would perform equally well. However, as long as two (or more) auditory continua and more than two categories are involved, it is crucial to have negatively formulated cue constraints; otherwise the highest ranked positive cue constraint would always determine the output, thus no cue integration (e.g. two acoustic cues are relevant for categorizing certain SF, such as relying on F1 and F2 to categorize three different vowels /e/, /o/ and /a/) is allowed[56]. Another reason for having cue constraints negatively-formulated is due to OT's exclusion mechanism. In particular, the worst candidate is excluded first, and the search for the best candidate continues. This allows for two surface

[55] The mapping from AudF to SF is also subject to structural constraints, but we will leave them out for the moment as they are only needed from section 4.3.3.

[56] Boersma and Escudero (2008) performed a computational simulation on how F1 and F2 integrate in the L1 phonological acquisition of Dutch vowel inventories. They showed that a grammar with negatively formulated cue constraints can account for 78.2% of the real data, whereas a grammar with positive cue constraints was much worse and scored only 44.9%.

forms being able to win (variation), where neither of them is ideal. With a positive formulation, only one winner is possible.

Recall that the EP alveolar lateral and the tap differ from each other in terms of both spectral (formant values and formant transition) and durational dimensions (Rodrigues, 2015). For simplicity, the auditory event that will be used as input in our formalisations to represent AudF is restricted to the F3 formant values, which are typically around 2542 Hz for the EP tap. Accordingly, in our formalisation, we use the cue constraint "2542 Hz (is not perceived as) not /l/"[57] to formalise the fact that a segment with F3 values of 2542 Hz is not an alveolar lateral in EP.

Each cue constraint that maps an auditory event onto a phonological category has antagonistic cue constraints that map the same value onto other categories of the respective language. That is to say, for the auditory event [F3: 2542Hz], we also have the cue constraint "2542 Hz not /ɾ/". Comparably, for the perfectly acceptable F3 values for the EP /l/, 2692 Hz, there will be two cue constraints, "2692 Hz not /l/" and "2692 Hz not /ɾ/".

Since 2542 Hz is a prototypical F3 value for /ɾ/ in EP, the constraint "2542 Hz not /ɾ/" should be much lower ranked than its antagonist "2542 Hz not /l/" in Portuguese native perception grammar. In line with this reasoning, the cue constraint "2692 Hz not /ɾ/", therefore, outranks "2692 Hz not /l/". The perception grammar represented by the four above-mentioned cue constraints accurately maps the auditory events [2542 Hz] ($[ɾ]_{Aud}$) and [2692 Hz] ($[l]_{Aud}$) onto their corresponding categories in EP accurately, as shown in tableaux (4) and (5), respectively.

[57] The cue constraints used here map continuous auditory dimension directly onto segmental categories, in stead of features. As a matter of fact, the use of distinctive features, i.e. [+lateral] for laterals and [-lateral] for liquids (Mateus & Andrade, 2000), would work equally well here and it will become crucial when we formalising the interaction between phonological categorization and orthography in section 4.4.2. The use of segmental labels can be viewed as a simplification.

(4) *EP [ɾ]$_{Aud}$ categorized by the EP perception grammar as /ɾ/*

[2542Hz] ([ɾ]$_{\text{Aud}}$)	[2542Hz] not/l /	[2692Hz] not /ɾ/	[2542Hz] not /ɾ/	[2692Hz] not /l/
/l/	*!			
☞ /ɾ/			*	

(5) *EP [l]$_{Aud}$ categorized by the EP perception grammar as /l/*

[2692 Hz] ([l]$_{\text{Aud}}$)	[2542Hz] not/l/	[2692Hz] not /ɾ/	[2542Hz] not /ɾ/	[2692Hz] not /l/
☞ /l/				*
/ɾ/		*!		

Following the Full Transfer Hypothesis (Schwartz & Sprouse, 1996), which is adapted to L2 speech perception by Escudero and Boersma (2004), we assume that at the initial stage of L2 speech learning, learners simply use a copy of their L1 perception grammar, i.e. the phonological categories, and cue constraints from Mandarin, to parse the EP sounds. Therefore, a Mandarin perception grammar is constructed and applied to the EP input.

For the Mandarin naïve listeners, they have not acquired the tap category yet, which means no cue constraints that refer to /ɾ/ exist at this initial state. When receiving an auditory input [F3: xxxx Hz], the cue constraints relevant for decision making are presumably those that refer to the Mandarin liquid categories, "xxxx Hz not /l/" and "xxxx Hz not /ɻ/". The relative constraint ranking between "xxxx Hz not /l/" and "xxxx Hz not /ɻ/" is contingent on the distance between this auditory event and the F3 value range of the Mandarin /l/ and /ɻ/. In particular, since the F3 value of 2542 Hz (prototypical for the EP tap) falls within the acoustic value range for the Mandarin /l/, which is typically realised with F3 of 2643 Hz, and stays far away from the range for the Mandarin /ɹ/, whose F3 values are around 2118 Hz (Smith, 2010). The respective cue constraints for this auditory event [F3: 2542 Hz] are ranked as follows in the Mandarin grammar: "2542 Hz not /ɹ/" > "2542 Hz not /l/". The Mandarin perception grammar, instantiated in tableau (6)[58], successfully simulates the Type I naïve listeners, who consistently categorized the EP [ɾ]$_{\text{Aud}}$ as /l/$_{\text{onset}}$, as shown in Figure 4.5.

(6) *EP [ɾ]$_{\text{Aud}}$ categorized by the Mandarin perception grammar as /l/*

[2542 Hz]	[2542Hz]	[2118Hz]	[2542Hz]	[2118Hz]
([ɾ]$_{\text{Aud}}$)	not/ɹ/	not /l/	not /l/	not /ɹ/
☞ /l/			*	
/ɹ/	*!			

Type II naïve Mandarin listeners mapped the tap sometimes onto /l/ and sometimes onto a stop category (/t/ or /t^{h}/), displaying between-subject (different from Type I listeners) and within-subject variations (alternation between a lateral and a stop by Type II listeners). We first formalise the between-subject variation, explaining what leads some listeners to categorize the EP [ɾ]$_{\text{Aud}}$ as /t/ or /t^{h}/.

We attribute this variation to the presence of multiple cues in the input and to learners' individual cue-weighting strategies. Multiple acoustic cues often contribute to a single phonological distinction in speech and listeners can combine different sources available in input to help resolve ambiguity, i.e. voicing contrast in EP is cued by Voice Onset Time, vowel duration (Lousada, 2006; Pape & Jesus, 2014) and stop duration (Veloso, 1997). In the case of the

[58] Tableau (6) represents an L1-Mandarin perception grammar, which will categorize the [ɹ]$_{\text{Aud}}$ (F3: 2118 Hz) as the Mandarin rhotic category, due to the cue constraint ranking "2118 Hz not /l/" > "2118 Hz not /ɹ/".

EP tap, apart from formant structures and segmental duration, its acoustic form may also involve a brief silence caused by tongue tip closure (Silva, 2014). When presented with multiple cues whose configuration is entirely novel, non-native listeners may manifest individual differences in cue use and weighting (e.g. Schertz et al., 2015). We thus assume that the use of alveolar stops in phonological categorization stems from the fact that some listeners weigh closure cue over spectral cue. Tableau (7) shows how this works in BiPhon-OT.

(7) *EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as an alveolar stop (silence cue > formant cue)*

[2542 Hz] ([ɾ]$_{\text{Aud}}$) [silence]	[2542 Hz] not /ɹ/	[silence] not /+son/ /+ con/	*[2542 Hz] not /-son/	[2542 Hz] not /l/
/ɹ/	*!	*		
/l/		*!		*
☞/t/			*	
☞/t^{h}/			*	

In the Mandarin perception grammar illustrated in tableau (7), the input are auditory events ([F3: 2542 Hz] and [silence]) of a canonical tap, which is the most prevalent phonetic realisation of the EP /ɾ/ in intervocalic position (Silva, 2014). The cue constraint "2542 Hz not /ɹ/" occupies the highest position of the constraint ranking, simulating the Mandarin cue knowledge that a F3 value of 2542 Hz is not very likely a Mandarin rhotic /ɹ/. Due to its highest ranking, the EP [ɾ]$_{\text{Aud}}$ was never perceptually identified as /ɹ/, in line with the experimental results in chapter 2 (auditory-alone condition).

The next constraint makes reference to the auditory form [silence], which is activated here, because of the presence of the corresponding acoustic information in the input; this cue constraint formalises that a brief silence caused by tongue tip closure in the auditory input signals a stop category[59] in Mandarin, thus disfavouring candidates with features like /+sonorant/ or /+continuant/[60] (/sonorant / and /continuant/ simplified as /son/ and /con/ for the sake of space).

The third constraint which militates against perceiving a steady formant structure as /-son/ does not play any role in choosing the optimal output due to its relative low position in the current constraint ranking. As a result, the EP [ɾ]$_{\text{Aud}}$ is categorized as a Mandarin alveolar stop (either /t/ or /t^{h}/[61]) by an L1-Mandarin listener who pays more attention to the stop-like cue than to the formant cue, instantiated by the crucial constraint ranking "[silence] not /+son/ or /+ con/" > "[2542 Hz] not /-son/".

If an L1-Mandarin listener weighs a spectral cue over a closure cue, the crucial cue constraint ranking will be the opposite as "[2542 Hz] not /-son/ > "[silence] not /+son/ or /+ con/", shown in tableau (8), and will lead the EP [ɾ]$_{\text{Aud}}$ to be categorized as /l/.

(8) *EP [ɾ]$_{\text{Aud}}$ categorized by the Mandarin perception grammar as /l/ (formant cue > silence cue)*

[2542 Hz] ([ɾ]$_{\text{Aud}}$) [silence]	[2542 Hz] not /ɹ/	*[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	[2542 Hz] not /l/
/ɹ/	*!		*	
/t/		*!		
/t^{h}/		*!		
☞ /l/			*	*

[59] See Liberman et al. (1981) for the silence as an important cue for perceiving a stop.

[60] In traditional phonological analysis, the features are usually written within "[ ]". However, in the current study, where a modular view of phonetics and phonology is adopted, we put all abstract phonological categories (e.g. feature, segment syllable) that occur at the surface phonological level between "/ /", in line with works conducted within BiPhon (e.g. Boersma, 2011).

[61] The use of /t^{h}/ is much less frequent than that of /t/, however, our data does not permit a statistical confirmation. We speculate that the relative limited number of /t^{h}/ might be due to the lack of aspiration noise in the input.

This Mandarin perception grammar in (8) can be seen as the same grammar in (5) and (6). In particular, in each tableau the listed constraints may vary depending on the input information as only the relevant constraints that participate in decision making are included; however, the constraint ranking remains the same across these tableaux.

A comparison between tableau (7) and (8) shows that the between-subject variation (Type I vs. Type II) with respect to the categorization of the EP tap can be formalised as different rankings of cue constraints, which express learners' individual cue weighting strategies.

4.3.2 Within-subject variations in L2 phonological categorization

In fact, as reviewed in the last section, Type II listeners are actually not consistent with their categorization of the EP tap as they sometimes identified it as /l/, and sometimes as an alveolar stop. This within-subject variation can be formalised in BiPhon-OT as constraint re-ranking. As introduced in 4.2, BiPhon adopts the stochastic version of OT, which assumes that each constraint is assigned with a value on a continuous scale of constraint strictness and this value can be temporarily altered by a random positive or negative value (noise) at each evaluation time. This assumption allows the grammar to yield different outputs via temporary constraint re-ranking, simulating variations in the data. In particular, since the relative ranking between the two cue constraints "[silence] not /+son/ or /+ con/" and "[2542 Hz] not /-son/" decides whether the Mandarin perception grammar parses the EP $[ɾ]_{Aud}$ as an alveolar stop or a lateral, as shown in tableaux (8) and (9), respectively, if these two constraints are ranked closely to each other in the Mandarin perception grammar[62] (constraint values range partially overlapping as illustrated in Figure 4.6), the evaluation noise added at each evaluation time will give rise to two types of

[62] It is totally conceivable that the constraint value ranges of these two constraints partially coincide with each other, because the ranking between these two is not informative for categorizing the Mandarin sounds, whose auditory forms do not have the acoustic cue configuration as manifested by the EP tap. In other words, these two cue constraints are not expected to compete in the decision-making process in Mandarin.

constraint ranking, thus yielding both types of output. Tableau (9) shows how this works in BiPhon-OT.

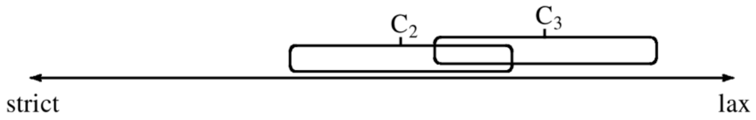


Figure 4.6: Constraint value ranges of “[silence] not /+son/ or /+ con/” and “[2542 Hz] not /-son/” are partially overlapped

In tableau (9), the dash line indicates that the relative ranking between the two constraints is not fixed and it can vary at each evaluation time. Therefore, both the lateral and the stop categories can surface as the output of phonological categorization.

(9) *EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as either /l/ or an alveolar stop*

[2542 Hz] ([ɾ]$_{Aud}$) [silence]	[2542 Hz] not /ɻ/	*[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	[2542 Hz] not /l/
/ɻ/	*!		*	
☞/t/		*		
☞/t^h/		*		
☞ /l/			*	*

In this section, we have shown that the within-subject variation (Type II) can be modelled as the probabilistic re-ranking of two constraints whose ranking value are (partially) overlapped. In stochastic OT, although it is possible to compute the exact constraint value and evaluation noise, quantitatively simulating the occurrence rate of each output, i. e. [ɾ]$_{\text{Aud}}$ is perceived 70% of the time as /l/, 30% of the time /t/, the present study sets out to show how these frequently attested variations, especially in L2 speech learning, are modelled within a linguistic model, instead of capturing the numerical tendencies in the data. We leave the quantitative simulation for further studies.

4.3.3 Variation in L2 phonological categorization as function of prosodic context

As revealed in chapter 2, there were apparent prosodic effects in naïve phonological categorization of the EP /ɾ/. Particularly, in onset position, Mandarin naïve listeners categorized the EP tap as either an alveolar lateral or a stop, while the repair strategies applied to the syllable-final tap are more diverse. Hence, the participants can be grouped into three types: Type I employing predominantly a lateral, usually accompanied with a schwa; Type II mainly using a stop (/t/, /tʰ/ or /tə/); Type III alternating between lateral, stop and structural modification (epenthesis[63]), see Figure 4.7.

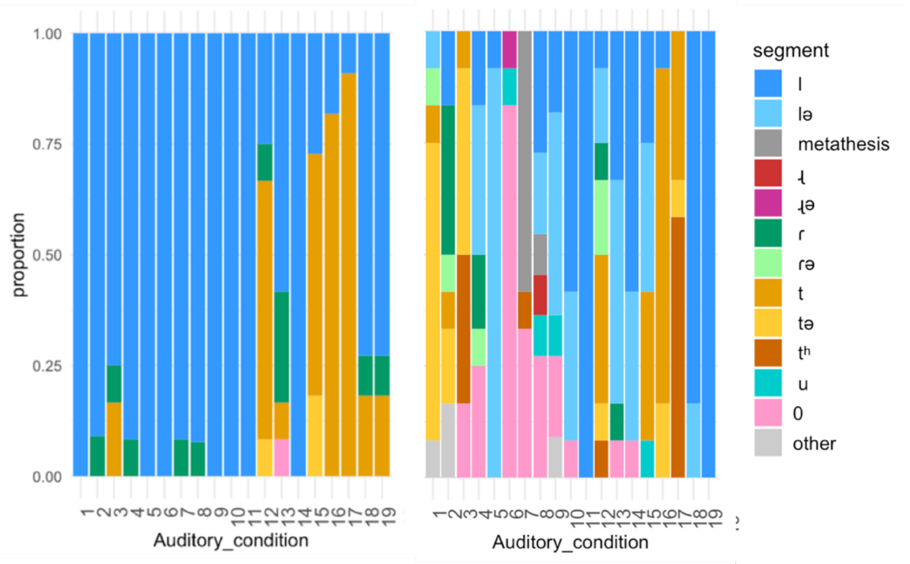


Figure 4.7: naïve categorization of the EP tap in onset (left) and in coda (right)

[63] Some naïve Mandarin listeners also employed segmental deletion; however, the experimental results of chapter 3 suggested that this was not due to perceptual deletion. Therefore, we do not attribute the segmental deletion to inaccurate phonological categorization.

It is obvious that the major difference between onset and coda in L2 phonological categorization lies in the employment of structural repairs. We argued that this can be attributed to the Mandarin phonotactic restriction, which only allows /ɹ/ and nasals, /n/ and /ŋ/, in syllable-final position (Duanmu, 2007; Lin, 2007).

Apart from the relationship between L1 and L2 categories, it has long been observed that non-native speech perception is constrained by the learners' L1 phonotactics (Polivanov, 1931); nevertheless, the current L2 speech theories have not yet provided an account for how segmental material interacts with phonotactic requirements in L2 speech learning. In this section, we formalise the interaction between segmental information and phonotactic restrictions in non-native phonological categorization by means of cue constraints and structural constraints within the framework of BiPhon-OT.

Structural constraints, which express language-specific phonotactic knowledge, have been shown to interact with cue constraints in various speech perception processes, see Boersma (2006) on the McGurk effect, Boersma (2007) on h-aspiré in French, Boersma (2009) on non-native phonological perception and Boersma and Hamann (2009b) on loanword adaptation. In the current section, a structural constraint is needed to represent the Mandarin structural restriction on the syllable-final phoneme distribution: no consonants other than /ɹ/ and nasals, /n/ and /ŋ/, are permitted in coda (Duanmu, 2007; Lin, 2007). Since the segmental material involved in our formalisation includes only the alveolar lateral and stop, we specify the structural constraint simply as "*/l./and/t./", in which the asterisk indicates a prohibition and the dot marks the syllable boundary.

How the Mandarin perception grammar parses the EP tap in onset and in coda position is formalised in tableaux (10) and (11), respectively. In (10), where the EP tap is placed intervocalically[64], the Mandarin grammar shows little tolerance for segmental omission[65], reflected by the highest ranked cue constraint penalising the perception of an auditory event (e.g. [F3: 2542 Hz]) as nothing at the surface phonological level ("/ /"). The second cue constraint rules out the candidate with a Mandarin rhotic /ɻ/, which typically bears F3 values quite different from those of the EP tap. A listener's cue weighting is determined by the ranking of the following two constraints, as we have already seen in the previous section: since "[2542 Hz] not /-son/" > "[silence] not /+son/ or /+con/", the lateral turns out to be the winner. The structural constraint, which is only ranked at the fifth place, does not play a role in choosing the optimal candidate.

(10) *EP [peɾafɐ]$_{Aud}$ categorized by the Mandarin perception grammar as /pa.la.fa/ in onset*

[2542 Hz] ([ɾ]$_{\text{Aud}}$) [silence]	[2542 Hz] not / /	[2542 Hz] not /ɻ/	[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	*/l./ /t./	[ ] not /ə/	[2542 Hz] not /l/
/pa.a.fa/	*!						
/pa.ɻa.fa/		*!		*			
/pa.ta.fa/			*!				
/pa.tʰa.fa/			*!				
☞ /pa.la.fa/				*			*

In coda, the same Mandarin perception grammar (11) cannot choose the optimal output with the first four cue constraints as in onset position, because

[64] The input in the OT tableaux only contains the auditory form of the target segment ([ɾ]$_{\text{Aud}}$), which is a simplification of the (pseudo)word form. It was the segmental sequence (VCV or VCC) that provided information on which syllable constituent (onset or coda) the target tap occupies.

[65] Segmental deletion implies a complete loss of the segment contained in the input and it might thus be considered by listeners as a bad strategy in phonological categorization. See 3.4 for a detailed discussion.

the fourth cue constraint “[silence] not /+son/ or /+ con/” penalises the surface lateral, irrespective of which syllable constituent it occupies (indicated by parenthesis on the exclamation mark). Both /l./ and /.l/ are equally harmonic until the fifth structural constraint comes into play. Due to the fact that the structural constraint against syllable-final /l/ and /t/ is ranked above the cue constraint “[ ] not /ə/”, which penalises generating a surface schwa that does not have any acoustic correlate in the input, the candidate with an epenthetic vowel is thus chosen as the most harmonic one.

(11) *EP [paɾfɐ]$_{Aud}$ categorized by the Mandarin perception grammar as /pa.lə.fa/ in coda*

[2542 Hz] ([ɾ]$_{Aud}$) [silence]	[2542 Hz] not / /	[2542 Hz] not /ɻ/	[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	*/l./ /t./	[ ] not /ə/	[2542 Hz] not /l/
/pa.fa./	*!						*
/paɻ.fa./		*!		*			
/pat.fa./			*!		*		
/pat^{h}.fa./			*!		*		
/pa.tə.fa./			*!				*
/pal.fa./				*(!)	*!		*
☞/pa.lə.fa./				*(!)		*	*

The alternation between a surface form with an epenthetic vowel and a form without an inserted vowel, as manifested by Type I listeners, can be formalised straightforwardly as the unfixed ranking (signalled by the dash line) between the structural constraint and the cue constraint against the illusory vowel, see tableau (12).

(12) *EP [paɾfɐ]$_{Aud}$ categorized by the Mandarin perception grammar as either /pal.fa/ or /pa.lə.fa/ in coda*

[2542 Hz] ([ɾ]$_{Aud}$) [silence]	[2542 Hz] not / /	[2542 Hz] not /ɻ/	[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	*/l./ /t./	[ ] not /ə/	[2542 Hz] not /l/
/pa.fa./	*!						*
/paɻ.fa./		*!		*			
/pat.fa./			*!		*		
/pat^{h}.fa./			*!		*		
/pa.tə.fa./			*!				*
☞ /pal.fa./				*(!)	*(!)		*(!)
☞/pa.lə.fa./				*(!)		*(!)	*(!)

Diverging from the Type I listeners, Type II listeners showed a preference for the alveolar stop when categorizing the EP tap across prosodic contexts. As we argued in the last section, this can be attributed to individual cue-weighting strategies, expressed by the relative ranking between cue constraint "[2542 Hz] not /-son/" and "[silence] not /+son/ or /+con/". As shown in tableau (13), as long as the constraint "[2542 Hz] not /-son/" is outranked by "[silence] not /+son/ or /+con/", L1-Mandarin listeners will perceive the EP tap as an alveolar stop in onset position.

Similar to Type I listeners, in coda position, the alternation between /t/ and /tə/ can be interpreted as the unfixed ranking between the Mandarin structural constraint against syllable-final stop and the cue constraint disfavouring perceptual epenthesis, see tableau (14).

(13) *EP [peɾafe]$_{Aud}$ categorized by the Mandarin perception grammar as /pa.ta.fa/ or /pa.t^{h}a.fa/ in onset*

[2542 Hz] ([ɾ]$_{Aud}$) [silence]	[2542 Hz] not / /	[2542 Hz] not /ɻ/	[silence] not /+son/ /+ con/	[2542 Hz] not /-son/	*/l./ /t./	[ ] not /ə/	[2542 Hz] not /l/
/pa.a./	*!						
/pa.ɻa./		*!	*				
/pa.la./			*!				*
☞/pa.ta./				*			
☞/pa.t^h a./				*			

(14) *EP [paɾfe]$_{Aud}$ categorized by the Mandarin perception grammar as either /pat.fa/, /pat^h.fa/ or /pa.tə.fa/ in coda*

[2542 Hz] ([ɾ]$_{Aud}$) [silence]	[2542 Hz] not / /	[2542 Hz] not /ɻ/	[silence] not /+son/ /+ con/	[2542 Hz] not /-son/	*/l./ /t./	[ ] not /ə/	[2542 Hz] not /l/
/pa.fa./	*!						
/paɻ.fa./		*!	*				
/pal.fa./			*!		*		*
/pa.lə.fa./			*!			*	*
☞/pat.fa./				*(!)	*(!)		
☞/pat^h.fa./				*(!)	*(!)		
☞/pa.tə.fa./				*(!)		*(!)	

Regarding the behaviour of Type III listeners, whose categorization of the EP tap manifested variation both in terms of segmental type and structural accommodation, we provide the folloing formalisation: in their perception grammar, the ranking of the two constraints that determine cue weighting and ranking of the two that decides whether an illusory vowel is inserted are unfixed, as in (15).

(15) The categorization of *EP [paɾfɐ]$_{Aud}$ by the Mandarin naïve listeners who showed variation in terms of both segmental type and structural accommodation*

[2542 Hz] ([ɾ]$_{Aud}$) [silence]	[2542 Hz] not / /	[2542 Hz] not /ɻ/	[silence] not /+son/ /+ con/	[2542 Hz] not /-son/	*/l./ /t./	[ ] not /ə/
/pa.fa./	*!					
/paɻ.fa./		*!	*			
☞/pal.fa./			*(!)		*(!)	
☞/pa.lə.fa./			*(!)			*(!)
☞/pat.fa./				*(!)	*(!)	
☞/pat^h.fa./				*(!)	*(!)	
☞/pa.tə.fa./				*(!)		*(!)

The formalisation in the current section revealed that variation as a function of prosodic context, during non-native phonological categorization, may stem from the interaction between structural and cue constraints. It should be noted that a single constraint ranking (i.e. a single grammar) was constructed to account for the prosodic effect and no position-specific ranking was needed in our formalisation.

Up to this point, we have formalised the variation observed in naïve Mandarin listeners' categorization of the EP tap, both in terms of speaker (4.3.1 and 4.3.2) and of prosodic contexts (4.3.3). However, as summarised in table 4.1, L2 production by late learners still diverge from naïve categorization (only the auditory input is given), with respect to the decreasing use of the alveolar stop (only used by beginners in reduced context and not by intermediate learners at all) and prosodically-conditioned employment of the Mandarin rhotic (e.g. only in coda). In the following section, we will argue that these differences between acquisition stages stem from the interaction between cross-linguistic phonological categorization and orthographic influence.

Table 5.1: Imitation results by Mandarin naïve speakers (auditory condition) and repair strategies by L1-Mandarin learners of EP (C[ə] = schwa epenthesis, ∅ = deletion)

	/ɾ/$_{\text{vcv}}$	/ɾ/$_{\text{vc}}$
Naïve (auditory condition)	[l], [t]	[l], [t,t^{h}], C[ə], ∅
Beginners (Liu, 2018)	[l]	[l], [t,d,t^{h}], [ɻ], C[ə], ∅
Intermediate (Zhou 2017)	[l]	[l], [ɻ], C[ə], ∅

4.4 Formalising the interaction between L2 phonological categorization and orthography

Adult L2 speech learning is inherently multimodal, since learners are generally exposed to both auditory input and written input simultaneously from the very beginning. In chapter 2, naïve Mandarin listeners performed an imitation task to test test how the Mandarin perception grammar parses the EP liquids across prosodic contexts. Listeners' responses were elicited in two experimental conditions: in the first one, only the auditory form was presented; in the second one, the written form was presented together with the auditory form.

Results demonstrated that 1) the Mandarin rhotic occurred almost exclusively when the written input was given, providing direct evidence that the use of the Mandarin /ɹ/ is orthographically driven; 2) responses with an alveolar stop decreased substantially from the auditory condition to the orthographic condition, suggesting that the presence of orthography affects the cue-weighting strategy in non-native phonological categorization.

Although it has been long attested in the literature that L2 speech perception and production are subject to influences induced by orthography, particularly when L1 and L2 grapheme-phoneme relations are incongruent (see 1.2.3), no current L2 speech learning model takes the cross-linguistic orthographic influence into its theoretical formulation. In this section, we formalise the orthographic effect on non-native phonological categorization of EP tap by Mandarin speaking natives, following the work by Hamann and Colombo (2017) and Hamann (2020), who proposed an Optimality-Theoretic reading/writing grammar that can be integrated to the BiPhon model.

According to Hamann and Colombo (2017), a reading grammar refers to the language-specific mapping of written forms onto SFs and it is formalised as *orthographic constraints*. In principle, there are two possible mappings of the written form (dual-route model; Coltheart et al., 1993): one is the pre-lexical route, where the graphemes are mapped onto SFs, while the other one is the lexical route, also called as direct access, where the written form is matched directly to UF. In the present study, we only address the orthographic constraints that express the pre-lexical mapping for the following two reasons: First, constraints that map written forms onto SFs avoid duplication of phonological knowledge in the lexicon (see 5.3 in Hamann & Colombo, 2017 for a detailed discussion); Second, the observed orthographic effect apparently emerged at a pre-lexical level, since naïve listeners did not have a Portuguese lexicon yet. The general types of orthographic constraint are listed in (16).

(16) *Orthographic constraints* (Hamann & Colombo, 2017; p. 690)
a) <γ>/P/: Assign a violation mark to every grapheme <γ> that is not mapped onto the phonological form /P/ and vice versa.
b) *<γ>/ /: Assign a violation mark to every grapheme <γ> that is mapped onto an empty segment in the SF.
c) *< >/P/: Assign a violation mark if the absence of a grapheme is mapped onto the phonological form /P/.

All orthographic constraints evaluate the relationship between the input (written form) and output (SF)[66]. Constraint (16a) express the knowledge of the grapheme-phoneme conversion of a certain language and is violated if a mapping does not conform to such conversion, e.g. the written form <s> is mapped onto /m/ in English. Constraints (16b) and (16c) together simulate the "one letter – one sound" orthographic principle proposed by Wiese (2004). Please note that the orthographic constraints employed by Hamann and Colombo (2017) map graphemes onto allophones (position-specific category), while, in the present study, we assume that the SF which orthographic constraints target is a tree-like prosodically-detailed structure (e.g. feature,

[66] The input and output are inverse in a writing grammar.

segment, syllable, feet; Nespor & Vogel, 1986). The implication of this will be elaborated later when we deal with the orthographic influence on the decreasing use of the alveolar stop.

In line with the Full Transfer Hypothesis adapted to L2 speech (Escudero & Boersma, 2004), we build a multimodal grammar, integrating the Mandarin perception grammar and Mandarin grapheme-phoneme conversion knowledge, and apply the EP auditory and written input to it. The Mandarin perception grammar consists of the same constraint ranking as we used in the last two sections and the Mandarin grapheme to phoneme mapping is formalised as an orthographic constraint "<r> /ɹ/", which is violated if the grapheme <r> is mapped to any SF other than /ɹ/.

It should be noted that some naïve Mandarin listeners resort to the Mandarin /ɹ/ only when presented with the written input. This difference between the auditory condition and orthographic condition can be illustrated when comparing tableaux (17) and (18).

(17) EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as /l/
(auditory input only)

[2542 Hz] [silence]	<r> /ɹ/	[2542 Hz] not /ɹ/	*[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	[2542 Hz] not /l/
/ɹ/		*!		*	
/t/			*!		
/t^h/			*!		
☞ /l/				*	*

(18) EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as /ɹ/
(both auditory and orthographic input presented)

[2542 Hz] [silence] <r>	<r> /ɹ/	[2542 Hz] not /ɹ/	*[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	[2542 Hz] not /l/
☞/ɹ/		*		*	
/t/	*!		*		
/t^h/	*!		*		
/l/	*!			*	*

In the Mandarin multimodal grammar, illustrated in (17), orthography does not affect the phonological categorization as no written input is given and the orthographic constraint is thus irrelevant, in spite of its high ranking. By contrast, due to the presence of a written form (<r>), along with the auditory form in the input, the orthographic influence emerges as illustrated in (18), because of the high ranked orthographic constraint, which penalises all SFs, expect /ɹ/.

A Mandarin listener having the constraint ranking in (17) and (18) weighs orthographic cues over auditory cues, expressed by the fact that cue constraints are outranked by an orthographic constraint. For those listeners whose categorization performance was not affected by the presence of orthography, their grammar should be like the constraint ranking in (19), where the orthographic constraint is not decisive, ranked below cue constraints.

Note that the relative ranking between the orthographic constraint and the cue constraints is irrelevant in the learners' L1, Mandarin, because no incongruence between the auditory and written input is expected, i.e. <l> with [l] and <r> with [ɹ]. Only if there is an incongruence, the difference with respect to the weighting between auditory cue and orthographic cue will give rise to different categorization outputs as in the case of L2 phonological categorization.

(19) EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as /l/ (both auditory and orthographic input presented)

[2542 Hz] [silence] <r>	[2542 Hz] not /ɹ/	*[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	[2542 Hz] not /l/	<r> /ɹ/
/ɹ/	*!		*		
/t/		*!			*
/t^h/		*!			*
☞/l/			*	*	*

The two tableaux (18) and (19) show how the interaction between orthography and phonological categorization accounts for the orthographic influence attested in chapter 2. In the following analysis, we further elaborate how this interaction may explain the gap between naïve phonological categorization and L2 production.

4.4.1 Prosodically-conditioned use of L1 /ɹ/ for the L2 tap

The first notable divergence between naïve categorization and L2 production lies in the fact the use of the Mandarin rhotic /ɹ/ by late learners for the target EP tap that is prosodically conditioned, namely restricted to coda position by L2 learners (Zhou, 2017; Liu, 2018).

Despite the fact that the L1 rhotic /ɹ/ was observed both in onset and in coda position in Mandarin naïve listeners' responses (chapter 2), there was a clear tendency for naïve listeners to be influenced more by orthography in syllable-final position, that is to say that they identified more instances of the syllable-final EP tap as the Mandarin /ɹ/, see Figure 4.7.

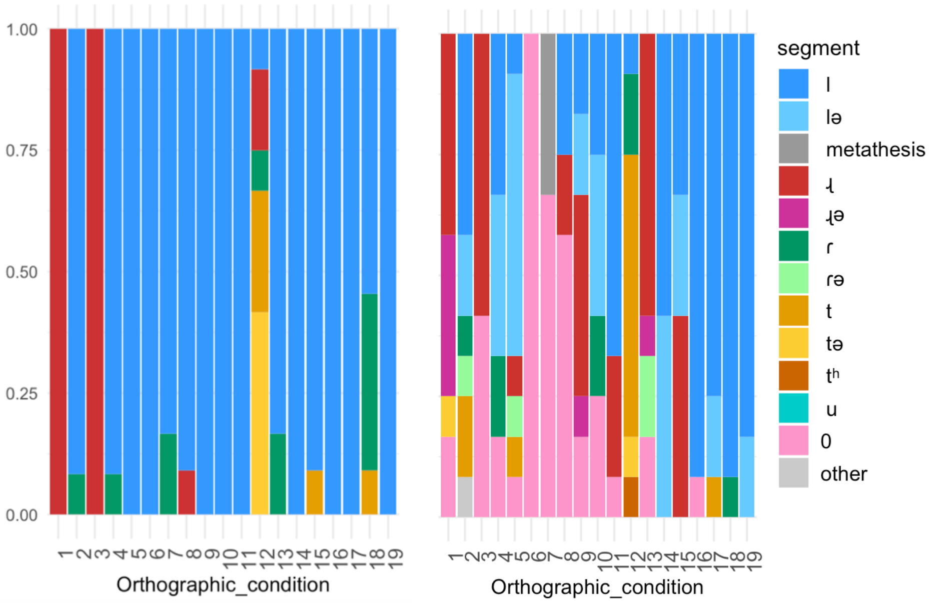


Figure 4.8: Naïve Categorization of the EP /ɾ/ in onset (left) and in coda (right).

Apart from this, the considerable within-subject variation in certain listeners (see right side of Figure 4.7) suggests that they failed to consistently map syllable-final [ɾ] to any existing L1 category, reminiscent of the "uncategorized" L2-to-L1 mapping scenario established in PAM-L2 (Best & Tyler, 2007; Faris et al., 2016), presumably because syllable-final EP /ɾ/ displays larger allophonic variability (Silva, 2014) and contains less acoustic information, due to the lack of consonant-to-vowel (CV) transition, in comparison to onset /ɾ/, where both CV as well as vowel-to-consonant (VC) transition are present.

It is therefore likely that during multimodal L2 speech learning, in the cases where auditory and orthographic information compete with each other, learners shift their attention to orthography when the auditory information is less consistent or insufficient (e.g. in coda position). This optimal utilization of cues during speech categorization has long been observed, particularly when informative/primary cues become occluded or degraded (Scharinger et al., 2014). If this were true, in intervocalic onset, where the auditory cues are clear and reliable (represented by both CV and VC formant transitions), orthography would not play a role, illustrated in tableau (20).

(20) EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as /l/
(both auditory and orthographic input presented)

[2542 Hz]$_{\text{VC}}$ [2542 Hz]$_{\text{CV}}$ <r>	[2542 Hz]$_{\text{CV}}$ not /ɹ/	<r> /ɹ/	[2542 Hz]$_{\text{VC}}$ not /ɹ/	[2542 Hz]$_{\text{VC}}$ not /l/	[2542 Hz]$_{\text{CV}}$ not /l/
/ɹ/	*!		*		
☞/l/		*		*	*

It is worth noting that the cue constraint "[2542 Hz] not /ɹ/" splits[67] into two in tableau (20): one representing the cue knowledge pertaining to the CV transition *"[2542 Hz]$_{CV}$ not /ɹ/"* and the other expressing the listeners' use of the VC transition *"[2542 Hz]$_{VC}$ not /ɹ/"*. The cue constraint referring to the CV transition is ranked higher than the one concerning VC transition, because it has been shown that, when both transitions are present, CV transition contributes more to the identification of an intervocalic consonant (e.g. for stops, see Tartter et al., 1983). Consequently, the high ranked cue constraint decides that the intervocalic EP tap is mapped onto the Mandarin alveolar lateral.

In contrast to the intervocalic tap, which is composed of both CV and VC formant transitions, the syllable-final tap lacks CV formant transition[68], which may exacerbate auditory cue reliability in coda. Consequently, the highest ranked CV cue constraint, which is decisive in parsing the intervocalic tap is not activated, because of the absence of CV transition from the input, and the relatively lower ranked orthographic constraint comes in, regulating the perceived SF as /ɹ/, as shown in tableau (21).

[67] After the operation of splitting one constraint into two, the novel two constraints should maintain the same function as the original constraint, i.e. the candidates it disfavours and its ranking with respect to other constraints.

[68] One may argue that the syllable-final tap in EP may come with a supporting vowel, providing somewhat CV transition; however, acoustic evidence has suggested that the supporting vowel only occurs 29.88% and 12.59% of the time in word-internal and in word-final coda, respectively. Therefore, the acoustic cue available in coda is indeed less than in intervocalic position.

(21) EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as /ɹ/ (both auditory and orthographic input presented)

[2542 Hz]$_{\text{VC}}$ <r>	[2542 Hz]$_{\text{CV}}$ not /ɹ/	<r> /ɹ/	[2542 Hz]$_{\text{VC}}$ not /ɹ/	[2542 Hz]$_{\text{VC}}$ not /l/	[2542 Hz]$_{\text{CV}}$ not /l/
☞ /ɹ/			*		
/l/		*!		*	

As illustrated in (20) and (21), the same constraint ranking (a single grammar) yields different outputs across syllable contexts, due to the difference in terms of cue availability between positions. If L1-Mandarin learners store the onset tap as an alveolar lateral and the coda tap as a Mandarin approximant in their L2 Portuguese lexicon, the restricted use of /ɹ/ syllable-finally in L2 production comes as no surprise.

Another possible explanation for the prosodically-conditioned orthographic influence is based on the assumption that orthographic forms and underlying forms may be co-activated in L2 speech production (see Veivo et al., 2018 for experimental evidence). Before presenting the formalisation of L2 speech production, it is necessary to figure out how L2 UFs would look like across prosodic context.

The difference in cue reliability between prosodic positions may give rise to divergent categorization patterns even without being conditioned by orthography. Specifically, as shown in tableau (22), in intervocalic onset position, whereby the auditory cue is reliable, represented by the presence of both CV and VC formant transitions, the ranking between the orthographic constraint and two cue constraints expressing VC transition is irrelevant (thus can be unranked with respect to each other). This is due to the fact that, as long as they are all ranked below the first CV cue constraint, an auditory event [F3: 2542] will not be perceived as a Mandarin rhotic.

(22) EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as /l/ (both auditory and orthographic input presented)

[2542 Hz]$_{VC}$ [2542 Hz]$_{CV}$ <r>	[2542 Hz]$_{CV}$ not /ɹ/	<r> /ɹ/	[2542 Hz]$_{VC}$ not /ɹ/	[2542 Hz]$_{VC}$ not /l/	[2542 Hz]$_{CV}$ not /l/
/ɹ/	*!		*		
☞/l/		*		*	*

By contrast, the same constraint ranking cannot decide which SF is the best fit for syllable-final [ɾ]$_{Aud}$, due to the lack of CV transition cue (thus the highest CV cue constraint is inapplicable), so that either /l/ or /ɹ/ is an optimal SF for the auditory input. This is illustrated in tableau (23). As a consequence, L1-Mandarin learners might store an underspecified UF for the coda tap, which is compatible with both /l/ and /ɹ/.

(23) EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as (both auditory and orthographic input presented)

[2542 Hz]$_{VC}$ <r>	[2542 Hz]$_{CV}$ not /ɹ/	<r> /ɹ/	[2542 Hz]$_{VC}$ not /ɹ/	[2542 Hz]$_{VC}$ not /l/	[2542 Hz]$_{CV}$ not /l/
☞ /ɹ/			*(!)		
☞/l/		*(!)		*(!)	

As we see in tableaux (22) and (23), orthography does not play a role in the construction of L2 UFs neither in onset nor in coda position. The construction of the phonological representation in the L2 lexicon hinges on phonological categorization. As a result, the onset tap is stored as |l|[69], whereas the coda tap as |@|[70].

In the following part of this section, we argue that the observed prosodically-conditioned use of /ɹ/ might be triggered by the activation of orthographic representation in L2 production.

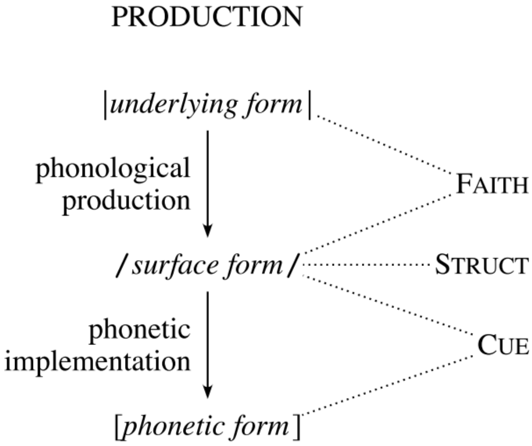


Figure 4.9: Speech production model in BiPhon (Boersma & Hamann, 2009).

In (L2) speech production, a speaker will start with an UF retrieved from their lexicon and map it onto a SF, which is connected to the phonetic forms i.e. auditory and articulatory forms, see Figure 4.9. The mapping from the UF to the SF is formalised as an interaction between faithfulness constraints, which penalises any mismatch between the UF and the SF, and structural constraints, which evaluate whether the SF conforms to language-specific phonotactics. The faithfulness constraints used in this study are listed in (24).

(24) *Faithfullness constraints* (McCarthy & Prince, 1995)

[69] In Biphon, the SF is written between "/ /", the UF between "| |".

[70] "@" represents being underspecified.

a) IDENT (F): Assign a violation mark if the output contains a different feature value from the one in the input
b) MAX: Assign a violation mark if a segment present in the input is absent in the output (deletion).
c) DEP: Assign a violation mark if a segment absent in the input is present in the output (epenthesis).

Since one of the core assumptions of the BiPhon model is bidirectionality, i.e. the same set of constraints and constraint ranking are used both in perception and production, the orthographic constraints (Hamann & Colombo, 2017; Hamann, 2020) can be also employed in the production direction. They evaluate the mapping between the stored orthographic representation (along with UF in lexicon), interacting with faithfulness constraints in selecting an SF. How L2 production works under the orthographic influence is illustrated in tableaux (25) and (26).

(25) EP word "duro" is produced as du/l/o by L1-Mandarin learners

du|l|o <r>	IDENT (F)	<r> /ɹ/	*/l./ /t./
du/ɹ/o	*!		
☞ du/l/o		*	

(26) EP word "carta" is produced as ca/ɹ/ta by L1-Mandarin learners

ca|@|ta <r>	IDENT (F)	<r> /ɹ/	*/l./ /t./
☞ca/ɹ/ta			
car/l/ta		*!	*

Note: @ represents being underspecified

When a Portuguese word containing an intervocalic tap is intended, e.g. *duro* “hard”, the stored phonological representation du|l|o and stored orthographic form du<r>o (abbreviated as <r>) are co-activated as input to the production grammar (e.g. Veivo et al., 2018). The UF |l| (in onset) is faithfully realised as an alveolar lateral at the surface level, because the highest ranked fatefulness constraint IDENT (F) penalises feature mismatch between the UF and the SF. The orthographic constraint, although ranked as the second highest, is not involved in choosing the optimal SF in onset position, see tableau (25).

On the other hand, in coda position, when uttering a word with an internal coda tap, such as *carta* “letter”, the underspecified UF |@| will be activated together with the stored written form <r>, shown in (26). This time, the faithfulness constraint IDENT (F) cannot help choose an optimal SF, since both candidates /l/ and /ɹ/ do NOT mismatch the underspecified UF. The decision making now is left to the second highest ranked orthographic constraint, which will regulate the UF as /ɹ/, creating thus an asymmetry in the use of /ɹ/ across prosodic contexts.

Up to this point, we have shown that the observed prosodically-conditioned use of the Mandarin rhotic for the L2 tap can be formalised in BiPhon-OT either as the orthographic influence during UF construction or co-activation of stored orthographic form and phonological form (UF) in L2 production. Our formalisation of these two possible explanations are testable as the categorization hypothesis predicts that /ɹ/ is stored as (one of the) UF in the L2 lexicon, whereas the production hypothesis does not. Future tasks tapping into the L2 lexicon are needed to examine these two hypotheses.

4.4.2 Decreasing use of alveolar stops for the L2 tap

In this section, we formalise another difference between naïve phonological categorization and L2 production, in particular, the decreasing use of /t, t^h/ for the target EP tap. In an exploratory analysis conducted in chapter 2, a significant decrease in the use of stops as responses was found from the auditory condition to the orthographic one (compare the left and the right side in Figure 4.10 and in Figure 4.11). Instead, sonorant consonants (laterals and approximant) increased.

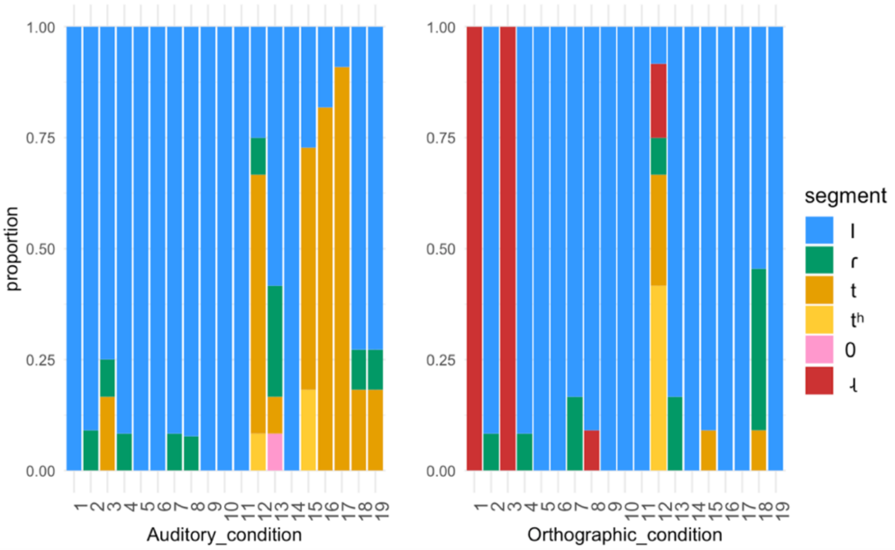


Figure 4.10: Categorization of the EP /ɾ/ in onset.

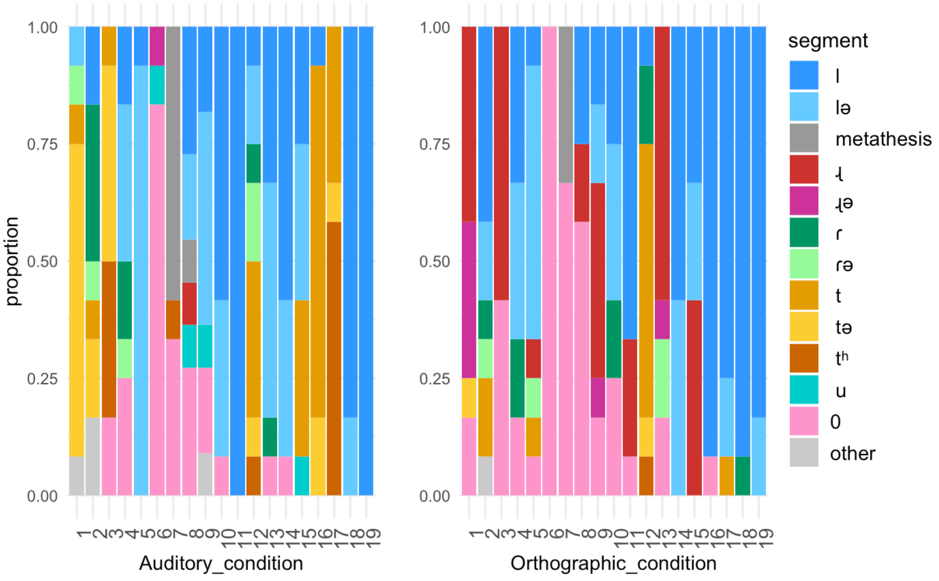


Figure 4.11: Categorization of the EP /ɾ/ in coda.

This substantial divergence between experimental conditions has led us to postulate that the written input altered listeners' cue weighting, similarly to what has been demonstrated by McGuire (2014). In particular, listeners who categorized /ɾ/ as a stop seem to give more weight to its brief closure cue than to its formant structure cue, otherwise a sonorant consonant, characterized by steady formants, would have been perceived. The simultaneous presentation of the orthographic form <r>, corresponding to a sonorant sound in Mandarin, seems to avert listeners' attention away from the closure cue. This finding, together with that in McGuire (2014), suggest that the auditory-orthographic cue competition and integration occur at a sub-phonemic level, in support for the view that acoustic information is mapped to phonological features in speech categorization (e.g. Lahiri & Reetz, 2010; Chládková et al., 2015; Monahan, 2018). In the following analysis, we formalise how orthography interferes the construction of L2 UF by dismissing stops which are preferred by the perception grammar.

The grammar of those listeners who weigh the closure cue above the formant cue, categorizing the EP tap predominantly as an alveolar stop, is replicated in tableau (27). Recall that their cue weighting preference is instantiated by the constraint ranking "[silence] not /+son/ or /+con/" > "[2542 Hz] not /-son/".

(27) *EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as an alveolar stop (silence cue > formant cue)*

[2542 Hz] ([ɾ]$_{Aud}$) [silence]	[2542 Hz] not /ɹ/	[silence] not /+son/ /+ con/	*[2542 Hz] not /-son/	[2542 Hz] not /l/
/ɹ/	*!	*		
/l/		*!		*
☞/t/			*	
☞/t^h/			*	

Since adult L2 learners are exposed to both auditory and written input from the very beginning when studying a foreign language, both forms are expected to serve as input for the construction of L2 lexical representations, as illustrated in (28).

(28) *EP [ɾ]$_{Aud}$ categorized by the Mandarin perception grammar as an alveolar lateral (silence cue > formant cue)*

[2542 Hz] ([ɾ]$_{Aud}$) [silence] <r>	[2542 Hz] not /ɹ/	*<r> /-son/	*<r> /+lateral/	[silence] not /+son/ /+ con/	*[2542 Hz] not /-son/	[2542 Hz] not /l/
/ɹ/	*!			*		
/t/		*!			*	
/t^{h}/		*!			*	
☞/l/			*	*		*

The orthographic constraint that we have used "<r> /ɹ/" can be split into two novel constraints which map the grapheme onto phonological features, instead of a phoneme, "* <r> /-son/" and "<r> /+lateral/". Such operation does not alter the orthographic constraint's function in the prior modelling, because the two novel constraints would still militate against SFs other than /ɹ/. In (27), the constraint "* <r> /+lateral/" is outranked by "* <r> /-son/", due to the fact that the feature /+sonorant/ has been considered as the most pertinent feature for cross-linguistic rhotics (Chabot, 2019; Natvig, 2020). Therefore, segments phonologically specified as /-sonorant/ are viewed as worse candidates for a rhotic, in comparison with laterals, which are phonologically [+sonorant].

A listener with the grammar in (27) weighs the closure cue above the formant cue, that is to say that they would perceive the EP tap as a stop without the presence of orthography; however, high ranked orthographic constraints favour a sonorant. Since the highest cue constraint does not allow the orthographic cue to override the auditory information completely, a compromise is made by choosing /l/ as the optimal output. Please note that, if orthography overrides auditory cues completely, the tap will be categorized as a Mandarin rhotic, as in (26). Listeners seem to have struggled to preserve some information from both auditory (high F3 values, /coronal/) and orthographic cue (/+son/), which gave rise to /l/. This suggests that L2 phonological categorization targets subphonemic units, i.e. distinctive features.

With more L2 experience, the SF generated by the L2 multimodal grammar (created when both auditory and orthographic input are given, i.e. formal learning in a classroom or self-learning with textbooks) will mediate the SF generated by the perception grammar (created when only the auditory input is given, e.g. listening to the radio), gradually updating the L2 UF by specifying it as [+sonorant]. We argue that this might account for the decreasing use of alveolar stops in L2 phonological acquisition of the EP tap by L1-Mandarin learners.

It is important to note that the attested orthographic influence on naïve phonological categorization challenges the widely cited L2 speech models, not only by showing that the L2 phonological categorization is multimodal, but also by contradicting that segment-size units are the (only) primitives in L2 speech perception, as assumed by the widely-cited SLM and PAM-L2. The BiPhon model, on the other hand, which comprises all relevant representational levels involved in speech, assuming bidirectionality (the same set of constraints and constraint ranking are used in perception and production), taking multimodal cue integration into account (Boersma, 2006 for visual cue and auditory cue; Hamann & Colombo, 2017 for orthographic cue and auditory cue), provides a more satisfactory account for multimodal L2 speech learning.

4.5 Formalising perception-production asymmetry in L2 phonological acquisition of EP /l/-/ɾ/

4.5.1 Confusability of EP /l/-/ɾ/ in L2 perception and production

Studies on L2 production of EP have shown that, in the intervocalic onset position, L1-Mandarin learners often replace the tap with [l], but never the reverse (*/l/ → [ɾ]) (Zhou 2017). In stark contrast, our identification task from chapter 3, together with Cao (2018) and Vale (2020), demonstrated that the segmental substitution in perceptual tasks takes place bidirectionally (([l] → /ɾ/ and [ɾ] → /l/), which leads to an apparent asymmetry between L2 perception and production.

A straightforward solution to this mismatch is to postulate that learners have developed distinct grammars for L2 perception and production (Ramus et al. 2010), as illustrated in Figure 4.12.

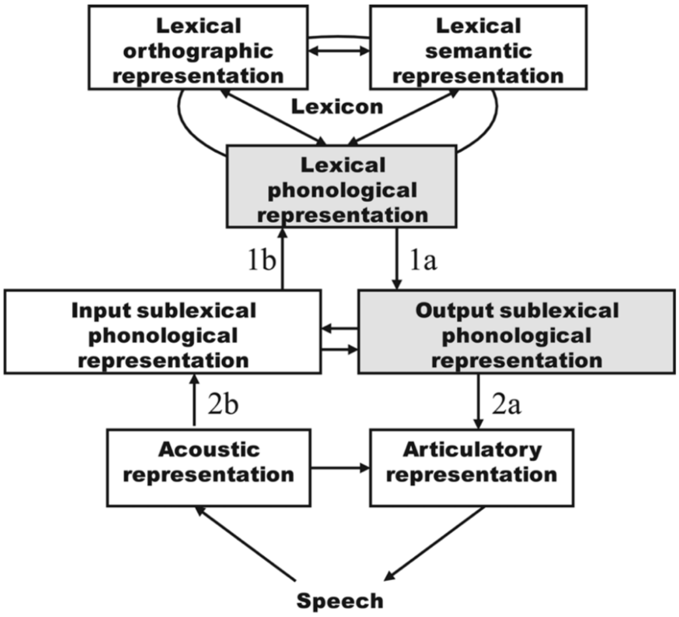

Lexical orthographic representation
Lexical semantic representation
Lexicon
Lexical phonological representation
1b
1a
Input sublexical phonological representation
Output sublexical phonological representation
2b
2a
Acoustic representation
Articulatory representation
Speech

Figure 4.12: Model for speech perception and production proposed by Ramus and colleagues (2010, p.313)

Nevertheless, such boxes-and-arrows model does not provide an explicit account of how perception and production could proceed, as pointed out by Boersma (2012) and a unified account without assuming separate grammars would be superior. In the current section, we thus argue against this distinct-grammar view by showing that the observed mismatch can emerge from an L2 phonological grammar which is identical in the two speech modalities.

We first build an L2 perception grammar, tableaux (29) and (30), where both auditory forms [l] and [ɾ] (simplified as F3 values) can be categorized as either /l/ or /ɾ/, in line with the perceptual results[71]. Please note that the L2 perception grammar represents the mapping of AudF onto SF in L2 phonology, which differs from the naïve Mandarin perception grammar that we have built in 4.3 and 4.4. In particular, the L2 perception grammar comprises a novel tap category and cue constraints expressing how an auditory event is mapped onto it, i.e. "[...] not /ɾ/", simplified as "*[...] /ɾ/".

In tableaux (29) and (30), the highest ranked faithfulness constraint IDENT (F), which evaluates the relationship between SF and UF is irrelevant in prelexical categorization, whereby no UF is involved. This is because, in a phoneme categorization task that is very often employed to investigate L2 perception, a listener is only asked to classify a given stimulus as one of the phonemes of his or her L1 or L2. Since the categorization task normally uses pseudowords or lexical minimal pairs (chapter 3 and Vale, 2020), in order to avoid lexical influence, only the mapping from auditory form to phonological surface form is activated and evaluated by cue constraints and structural constraints. In our modelling of L2 Portuguese perception, only cue constraints are relevant for the moment because the structural constraints, which express phonotactic restrictions, will not play a role in phoneme categorization in intervocalic onset position, where both the lateral and the tap are legitimate.

[71] Please note that we are interested in modelling the underlying mechanism which gives rise to the asymmetry between modalities, instead of simulating the mathematical tendency (e.g. 65% as tap, 35 as lateral), which can be achieved by calculating the possible value range of each constraint.

The reason why we include IDENT (F) in the L2 perception tableaux is to allow readers to be able to compare perception tableaux with production tableaux, emphasising the main point of our formalisation that the same set of constraints and constraint ranking are used both in perception and in production. The decision for choosing the perceived SF relies on the cue constraints pertaining to the same auditory event; for an auditory event [F3: 2692 Hz], only the ranking among cue constraint *[2692Hz] /A/, *[2692Hz] /B/, *[2692Hz] /C/ is relevant.

Since the ranking between the cue constraints of the same auditory event are still unfixed, segmental confusion was found to be bidirectional in perceptual experiments[72].

(29) EP [l]$_{Aud}$ categorized by L2 perception grammar as either /l/ or /ɾ/

[2692Hz] ([l]$_{\text{Aud}}$)	IDENT (F)	*[2542Hz] /l/	*[2542Hz] /ɾ/	*[2692Hz] /ɾ/	*[2692Hz] /l/
☞ /l/					*
☞ /ɾ/				*	

(30) EP [ɾ]$_{Aud}$ categorized by L2 perception grammar as either /l/ or /ɾ/

[2542 Hz] ([ɾ]$_{\text{Aud}}$)	IDENT (F)	*[2542Hz] /l/	*[2542Hz] /ɾ/	*[2692Hz] /ɾ/	*[2692Hz] /l/
☞ /l/		*			
☞ /ɾ/			*		

[72] What leads to the observed bidirectional perceptual confusion, in contrast to naïve phonological categorization where the distinction existed, was discussed in detail in chapter 3.

The production results, reminiscent of asymmetrical lexical access reported previously in the literature (e.g. Darcy et al 2013), suggest that the /l/-/ɾ/ distinction is somehow preserved, yet not target-like in the L2 lexicon (otherwise, the confusion would be bidirectional). We thus postulate that the L2 lateral, which bears no detectable difference from the L1 lateral, as evidenced by the naïve imitation results in chapter 2, is accurately represented in the lexicon, whereas the L2 tap seems to be underspecified, compatible both with /l/ and /ɾ/.

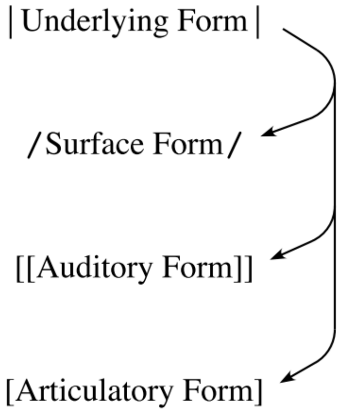


Figure 4.13: The phonological-phonetic production process, fully parallel edition (Boersma, 2011; p. 27)

In the production tableaux (31) and (32), where the underlying phonological forms retrieved from the learners' lexicon serve as input, the same set of constraints and constraint ranking are employed. The notation in the following production tableaux differs from the traditional notation in that the candidate cells contain paired representations, SF-AudF. This reflects the BiPhon model's assumption that the process of phonological-phonetic production is parallel, as illustrated in Figure 4.13. Such formalisation allows the "high-level" constraints, like faithfulness constraints, to interact with "low-level" constraints, such as cue and articulatory constraints (Boersma, 2008; 2009). In our case, the production being parallel implies that the mapping from UF to SF and the one from SF to AudF are evaluated at the same time.

In particular, as shown in tableau (31), when an underlying form containing the lateral is activated (e.g. |mala|, "suitcase"), the highest-ranked constraint IDENT (F) rules out candidates with a surface tap (/.ma.ɾɐ./). The first and second candidates both with a surface /l/ are then submitted to the evaluation of cue constraints. Since the cue constraints militate against the auditory form [2542Hz] (a prototypical F3 value of the tap), the second candidate with [ɾ]$_{\text{Aud}}$ turns out to be less preferable than the first candidate with [l]$_{\text{Aud}}$. Consequently, the underlying lateral will be realized with [l]$_{\text{Aud}}$, which has F3 values of 2692 Hz.

(31) EP |l| is realised by L2 production grammar as [l]$_{Aud}$

|mala|	IDENT (F)	* /l/ [2542Hz]	* /ɾ/ [2542Hz]	* /ɾ/ [2692Hz]	* /l/ [2692Hz]
☞ /.ma.lɐ./ [2692Hz] ([l]$_{\text{Aud}}$)					*
/.ma.lɐ./ [2542Hz] ([ɾ]$_{\text{Aud}}$)		*!			
/.ma.ɾɐ./ [2542Hz] ([ɾ]$_{\text{Aud}}$)	*!		*		
/.ma.ɾɐ./ [2692Hz] ([l]$_{\text{Aud}}$)	*!			*	

(32) EP |@| (tap) is realised by L2 production grammar as [l]$_{Aud}$

|ka@a|	IDENT (F)	* /l/ [2542Hz]	* /ɾ/ [2542Hz]	* /ɾ/ [2692Hz]	* /l/ [2692Hz]
☞ /.ka.lɐ./ [2692Hz] ([l]$_{\text{Aud}}$)					*
/.ka.ɾɐ./ [2542Hz] ([ɾ]$_{\text{Aud}}$)			*!		
☞ /.ka.ɾɐ./ [2692Hz]([l]$_{\text{Aud}}$)				*	
/.ka.lɐ./ [2542Hz] ([ɾ]$_{\text{Aud}}$)		*!			

Note: @ represents being underspecified

On the other hand, when a word is intended with a the underlying rhotic, which is stored as the underspecified |@| (e.g. |ka@a| "face"), the decision for choosing an output hinges on cue constraints, because the high ranked IDENT cannot regulate an underspecified representation (no feature mismatch occurs). Therefore, he underlying rhotic is produced as $[l]_{Aud}$, see tableau (32).

Please note that what underlies the mismatch between perceptual (output: /l/ and /ɾ/) and production results (output: [l]) is the bidirectional use of cue constraints. In particular, irrespective of the UF and SF, the cue constraint ranking in tableaux (31) and (32) always favours [2692] Hz, the auditory form of the lateral. This seems to suggest that the L2 lexicon does not play a role in decision making. However, this is not the case as we will further illustrate.

For now, our simulated L2 learners cannot distinguish the EP lateral-tap contrast in perception and only use a lateral in production. Nevertheless, the perceptual data shows that the identification accuracy is above chance level and learners do sometimes produce a target tap.

Recall that the bidirectional perceptual confusion is attributed to the fact that the cue constraints for the same auditory event are unranked with respect to each other, see tableaux (29) and (30). According to the stochastic OT adopted by the BiPhon model, being unranked means that the value ranges of these constraints coincide with each other completely. This can be viewed as the very initial state of L2 speech learning. With being more exposed to the target language, L1-Mandarin learners are expected to develop more target-like cue knowledge. As a result, the constraint value ranges of those initially unranked constraints will start to move apart, as in Figure 4.14.

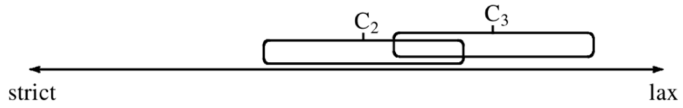


Figure 4.14: Two partially overlapped constraint value ranges

Boersma & Hayes (2001, p.48)

In stochastic OT, the constraint ranking value can be temporarily altered by a small random positive or negative value (noise) chosen from a pre-set range at each evaluation time (Boersma, 1998; Boersma & Hayes, 2001). Provided that a ranking value of C2 was chosen from the rightmost part of the range and the value of C3 from the leftmost, then C3 would outrank C2. This probabilistic re-ranking mechanism of stochastic OT predicts that the two cue constraints whose value ranges overlap partially can yield two different rankings, e.g. *[2542Hz] /ɾ/ > *[2542Hz] /l/ or *[2542Hz] /l/ > *[2542Hz] /ɾ/. Consequently, this allows the L1-Mandarin learners to perceive the lateral and the tap as separate categories, to some extent, as shown in tableaux (33) and (34).

(33) EP [l]$_{Aud}$ categorized by L2 perception grammar as /l/

[2692Hz] ([l]$_{Aud}$)	IDENT	*[2542Hz] /l/	*[2692Hz] /ɾ/	*[2542Hz] /ɾ/	*[2692Hz] /l/
☞ /l/					*
/ɾ/			*!		

(34) EP [ɾ]$_{Aud}$ categorized by L2 perception grammar as /ɾ/

[2542Hz] ([ɾ]$_{Aud}$)	IDENT	*[2542Hz] /l/	*[2692Hz] /ɾ/	*[2542Hz] /ɾ/	*[2692Hz] /l/
/l/		*!			
☞ /ɾ/				*	

Note: @ represents being underspecified

The probabilistic constraint re-ranking between the one in tableaux (29) - (30) and the one in tableaux (33) - (34) will thus successfully stimulate learners' bidirectional confusability between [l]-[ɾ] in perception, as well as their above-chance accuracy in an identification task, corroborated by the experimental results obtained in chapter 3 and in previous perceptual studies (Cao, 2018; Vale, 2020).

In the production direction, this probabilistic L2 grammar yields unidirectional segmental confusion, conforming to the production data (Zhou, 2017). This is illustrated in (35) and (36).

In the production of an EP words with intervocalic lateral, the highest ranked faithfulness constraint IDET (F) is violated by candidates with the SF /ɾ/, which will be no longer in the running, as a consequence. Since the cue constraints favour the auditory form of a lateral [2692 Hz], the UF |l| is realised faithfully as [l]$_{\text{Aud}}$, see (35).

(35) EP |l| is realised by more target-like L2 production grammar as [l]$_{\text{Aud}}$

|mala|	IDENT (F)	* /l/ [2542Hz]	* /ɾ/ [2692Hz]	* /ɾ/ [2542Hz]	* /l/ [2692Hz]
☞ /.ma.lɐ./ [2692Hz] ([l]$_{\text{Aud}}$)					*
/.ma.lɐ./ [2542Hz] ([ɾ]$_{\text{Aud}}$)		*!			
/.ma.ɾɐ./ [2542Hz] ([ɾ]$_{\text{Aud}}$)	*!			*	
/.ma.ɾɐ./ [2692Hz] ([l]$_{\text{Aud}}$)	*!		*		

When the underspecified UF of the tap is retrieved from the L2 lexicon, serving as the input in production, the decision making is contingent completely on cue constraints alone, because the underspecified UF is compatible with both SF/l/

and /ɾ/. Since the cue knowledge represented by the cue constraint ranking in tableau (36) is still not target-like, the production of the rhotic will vary between the lateral and the tap.

(36) EP |@| (tap) is realised by L2 production grammar as either [l]$_{Aud}$ or [ɾ]$_{Aud}$

|ka@a|	IDENT	* /l/ [2542Hz]	* /ɾ/ [2692Hz]	* /ɾ/ [2542Hz]	* /l/ [2692Hz]
☞ /.ka.lɐ./ [2692Hz] ([l]$_{Aud}$)					*
☞ /.ka.ɾɐ./ [2542Hz] ([ɾ]$_{Aud}$)				*	
/.ka.ɾɐ./ [2692Hz]([l]$_{Aud}$)		*!			
/.ka.lɐ./ [2542Hz] ([ɾ]$_{Aud}$)			*!		

Note: @ represents being underspecified

Recall that in the previous production tableaux (29) and (30), the bidirectional use of cue constraint underlies the unidirectional confusion in L2 production and L2 lexicon does not seem to play a role. However, as illustrated in tableaux (35) and (36), the lexical influence is important for decision making, otherwise the cue constraint alone will produce either a lateral or a tap, as in tableaux (36).

In sum, our simulated learner with the probabilistic L2 grammar shows bidirectional confusability between [l]-[ɾ] in a perception experiment, while still scoring above chance; in L2 production, the intended rhotic will be produced either as [l] or [ɾ], whereas the underlying lateral will be faithfully realised. Our formalisation shows that the mismatch between L2 perceptual and production may not be due to two separate phonological grammars (constraint rankings), but to the fact that the two paralinguistic processes targeted by perception and production studies involve different mappings: in the perception experiment only the mapping from auditory to phonological surface form is triggered, while production also involves mapping the lexical form onto the phonological surface form (and thus employs bidirectional use of cue constraints and the L2 lexicon influences decision making).

One of the major shortcomings of our proposal is that the production process is modelled with assumed underlying representations |l| and |@| (underspecified tap). This problem of assuming that discrete units have already taken place is shared by the majority of the existing computational models (Dupoux, 2018), especially for those based on OT. However, our proposal provides a conceivable and detailed enough model of L2 perception-production asymmetry and it can be falsified by experimental studies. For instance, a lexical decision task can be performed to test the underlying forms assumed in the modelling. In particular, if |l| is accurate, learners should find it easy to refute mispronunciations with [ɾ]-alternation (e.g. ma[ɾ]a is not ma|l|a); on the other hand, if the underlying tap is underspecified, both correct and mispronounced forms will be accepted by learners (e.g. both ca[l]a and ca[ɾ]a are ca|@|a). Nontheless, our modelling within BiPhon-OT is more promising, in comparison with other existing L2 speech models with respect to this issue, because the Neural Network edition of BiPhon, which derives from BiPhon-OT, has successfully modelled the emergence of discrete phonological categories from continuous auditory input, without pre-assumed abstraction (Seinhorst et al., 2019; Boersma et al., 2020). In other words, BiPhon-OT can be regarded as a fairly adequate transition from the formal phonological theory to the computational modelling of human speech processing (Escudero & Boersma, 2004; Boersma & Escudero, 2008).

4.5.2 Structural modifications of EP /ɾ/ in L2 perception and production

Another asymmetry between L2 perception and production reported in chapter 3 concerns the structural repairs of EP syllable-final tap by L1-Mandarin learners. In particular, the epenthesis has been shown to be perceptually driven whereas the segmental deletion is restricted to production. As we have already shown in 4.3, an illusory vowel may emerge during the construction of L2 UF, as in tableau (37).

(37) Emergence of an illusory vowel in naïve categorization

[2542 Hz] ([ɾ]$_{\text{Aud}}$) [silence]	[2542 Hz] not / /	[2542 Hz] not /ɻ/	[2542 Hz] not /-son/	[silence] not /+son/ /+ con/	*/l./ /t./	[ ] not /ə/	[2542 Hz] not /l/
/pa.fa./	*!						*
/paɻ.fa./		*!		*			
/pat.fa./			*!		*		
/pat^{h}.fa./			*!		*		
/pa.tə.fa./			*!				*
/pal.fa./				*(!)	*!		*
☞/pa.lə.fa./				*(!)		*	*

The perceived SF with an epenthetic schwa may be passed onto the underlying level and stored as UF |paləfa|. Consequently, the epenthetic schwa in UF will be realised in production.

Regarding the deletion of coda tap, which is restricted to L2 production (chapter 3). a straightforward solution to the production-specific L2 syllable-final tap deletion is to posit that L2 perception and production have distinct grammars (Ramus et al., 2010). We will show in the following part of this section that assuming separated grammars across modalities is not necessary, because the observed coda tap deletion in L2 production can stem from the same grammar (constraint ranking) used in perception.

The first question that needs to be addressed is what kind of UF underlies the deleted tap. Since our formalisation does not predict any empty category (i.e. no content at all for the tap) in the lexicon, on the basis of the categorization results in chapter 2 and 3, we speculate that the UF behinds the deleted tap could be |l|, |t|, |t^h| or underspecified. This depends on a learner's perception grammar.

The next question is whether the tap deletion occurs at the phonological part or the phonetic part of the L2 production grammar. The following formalisation shows that both are possible.

(38) *Coda deleted by the phonological part of the L2 grammar (UF specified)*

|kalta|	IDENT (F)	*/l./ /t./	DEP	MAX
/kaɻ.ta./	*!			
/kat.ta./	*!	*		
/ka.tə.ta./	*!		*	
/kal.ta./		*!		
/ka.lə.ta./			*!	
☞ /ka.ta./				*

As shown in tableau (38), in the phonological part of the production grammar, where UF is mapped to SF, provided that the UF for L2 rhotic is specified (either as a lateral or an alveolar stop), the highest ranked faithfulness constraint IDENT (F) is violated by all SFs with segmental change, keeping those with the same feature specification as in UF and the form in which the liquid is deleted in the running (the last three candidates). The relative ranking between the structural constraint * /l./; /t./ and the constraint DEP, which prohibits segmental insertion, is irrelevant, as long as they outrank the faithfulness constraint MAX, which militates against deletion, because the structural constraint and DEP rule out all candidates, except the SF with deleted coda.

In the case where the underlying tap is underspecified, the current constraint ranking still yields the same result, as shown in tableau (39). Since the UF for the L2 rhotic is underspecified, the faithfulness constraint IDENT (F) cannot rule out any SF, as no mismatch between UF and SF occurs. The candidate SF with the Mandarin rhotic can be in principle kicked out by an L2 structural constraint */ɻ/[73], which expresses the Portuguese phonotactic knowledge that /ɻ/ does not exist in the Portuguese segmental inventory. The SF with the omission of syllable-final tap is chosen as the most harmonic candidate because MAX is outranked by the L1 structural constraint * /l./; /t./ and the faithfulness constraint DEP.

(39) *Coda deleted by the phonological part of the L2 grammar (UF underspecified)*

|ka@tə|	IDENT (F)	*/ɻ/	*/l./ /t./	DEP	MAX
/kaɻ.ta./		*!			
/kat.ta./			*!		
/ka.tə.ta./				*!	
/kal.ta./			*!		
/ka.lə.ta./				*!	
☞ /ka.ta./					*

Note: @ represents being underspecified

Please note that the difference between L2 perception and production is not due to different constraint rankings between perception and production, but to different types of constraints involved in L2 production (UF to SF) and phonological categorization (AudF to SF).

[73] The inclusion of this L2 structural constraint will not affect the decision making in tableau (37).

The second possibility is that the coda tap deletion is driven by the phonetic part of the grammar, namely the articulatory constraint. As reviewed in 1.2.2, deviations in L2 speech may be invoked by articulatory difficulties, irrespective of phonological representation. Therefore, the omission of syllable-final tap might be a result of articulatory imprecision. The Portuguese tap imposes great articulatory complexity since it stipulates a ballistic movement of the tongue tip and a constriction towards to the pharynx (Berti, 2010; Barberena et al., 2014; Barberena et al., 2019)[74], which is entirely novel for L1-Mandarin learners. Accordingly, it is very likely that learners sometimes fail to realise the complex and novel gestures required for the Portuguese tap, especially in word-internal coda position, where consonant-to-consonant co-articulation increases the articulatory difficulty.

The articulatory difficulty can be expressed by an articulatory constraint *[[CC]][75], which militates against the articulation of two adjacent consonants. This articulatory constraint should be ranked very low in EP, since the articulatory gestures for the tap, followed by another consonantal gesture, can be realised with ease by native Portuguese speakers. By contrast, this constraint is expected to be highly ranked in L2 production grammar, before Mandarin speakers master this gestural coordination.

As shown in tableau (40), if the articulatory constraint ranks higher than the faithfulness (e.g. IDENT), structural (e.g. */.../) and cue constraints, segmental omission will be expected, irrespective of the specification of phonological representation (/ɾ/, /l/ or any other segments), or the mapping between SF and AudF. The constraint ranking among faithfulness, structural and cue constraints used in L2 perception can also be used in L2 production as

[74] To the best of our knowledge, all studies pertaining to the articulatory characteristics of the Portuguese tap investigated the Brazilian variety, and no comparable studies exist for EP currently. However, the potential articulatory differences between the EP tap and the Brazilian Portuguese one do not invalidate our argument that the articulation of this segment is challenging for Mandarin native speakers.

[75] The articulatory forms are written within [[ ]].

the segmental omission is triggered by the involvement of ArtF and articulatory constraint, which are irrelevant in speech perception.

(40) *Coda deleted by the phonetic part of the L2 grammar*

/paɾ.fa./	*[[CC]]	IDENT	*/…/	/…/ […]
[ɾf]$_{\text{Aud}}$ [[ɾf]]$_{\text{Art}}$	*!			
[lf]$_{\text{Aud}}$ [[ɾf]]$_{\text{Art}}$	*!			
[ɾf]$_{\text{Aud}}$ [[lf]]$_{\text{Art}}$	*!			
[lf]$_{\text{Aud}}$ [[lf]]$_{\text{Art}}$	*!			
☞ [lf]$_{\text{Aud}}$ [[f]]$_{\text{Art}}$				
☞ [ɾf]$_{\text{Aud}}$ [[f]]$_{\text{Art}}$				

Future studies examining different phonological levels (surface and underlying) and combing both perception and production data from the same group of L1-Mandarin learners of EP are needed in order to test the hypothesis put forward by our modelling.

In sum, in this chapter, by adopting the BiPhon model, we managed to formalise several intriguing experimental findings that cannot be explained by the current L2 speech learning models, in particular, variations in L2 phonological categorization, interaction between phonology and orthography in category creation and the asymmetry between L2 perception and production. Our formalisation not only bridges some gaps between the L2 empirical evidence and formal phonological theory, but also puts forward testable predictions for future research.

Chapter 5: Conclusion

Research has shown that mastering the EP /l/ and /ɾ/ contrast can be challenging for L1-Mandarin learners. This thesis has investigated what constrains the development of these L2 phonological categories across different prosodic positions and how different modalities interact during this L2 speech learning process. To achieve this aim, we employed both laboratory experiments and theoretical modelling. The following sections of this chapter summarize the main experimental findings of this project and discuss directions for further research.

5.1 Summary of the main experimental findings and contributions

The first study of this thesis (chapter 2) explored the role of cross-linguistic influence as well as of orthography in L2 category formation. Firstly, in order to attest cross-linguistic influence directly, a delayed-imitation task with Mandarin speaking natives without any knowledge of Portuguese was conducted. This task assessed how Mandarin phonology influences the parsing of the EP input ([l], [ɾ]) in intervocalic onset position and in word-internal coda position. Secondly, the orthographic effect was examined by manipulating the input types that were given (auditory input alone vs. auditory + written input). Results of the experimental condition where the participants received both types of input replicated the previously reported L2 prosodic effects (i.e., position-dependent repair strategies; Zhou, 2017; Liu, 2018), providing evidence for the cross-linguistic interaction between phonological categorization and orthography at the onset of L2 phonological categories' development. This study highlighted the multimodal nature of adult L2 speech learning, which should be incorporated in L2 speech models, in order to achieve a better understanding of the underlying mechanisms behind non-native phonological acquisition.

In a follow-up study (chapter 3), we further examined the interaction between speech perception and production in L2 speech learning and the perceptual plasticity in the learning of difficult L2 categories, by examining whether the L2 deviant productions stem from misperception and whether L2 phonological categories remain malleable at a mid-late stage of L2 speech learning. To answer the first question, two perceptual experiments (an AXB discrimination task and a forced-choice identification task) were conducted to test L1-Mandarin learners' discrimination ability between the target Portuguese form and the deviant form that they often employ in production. I reasoned that a deviant form would have a perceptual basis if learners failed to reliably discriminate it from the target form. Expanding on prior research (Cao, 2018; Vale, 2020), I investigated the potential perceptual confusability across syllable constituency and took both segmental replacement as well as structural modifications into account. Results show that, although L1-Mandarin learners perceptually confused the target [ɫ] and [ɾ] with the deviant forms they tend to produce (e.g., [w] for the velarised lateral; [l] and [ɾə] for the tap), some imprecise productions cannot be attributed to misperception (deletion of syllable-final tap) and production may even precede perception (confusability between /l/ ↔ /ɾ/ in perception; however, no confusability in production, /ɾ/ → [l], never /l/ → [ɾ]). The correspondence as well as discrepancy between modalities signal a complex relationship between L2 speech perception and production, which has not appropriately been addressed in L2 speech models.

To investigate L2 phonological categories' plasticity, two groups of L1-Mandarin learners were recruited to participate in two perceptual tasks. They differed substantially in terms of L2 experience (Intermediate Group: 2-year formal instruction vs. Advanced Group: 4-year formal instruction with immersion experience in Portugal, ranging from 1 to 4 years). No main effect of L2 experience was found in the generalized mixed effect models. This null effect of L2 experience corroborates the prediction by the Perceptual Assimilation Model-L2 (Best & Tyler, 2007): at a mid-late stage of L2 learning, learners' attention might be driven away from the critical acoustic differences between two difficult categories.

In order to contribute to bridging the gap between experimental research and L2 speech theory, in the fourth chapter of this thesis, we formalized the aforementioned findings that cannot be accounted by current L2 models by adopting the Bidirectional Phonology and Phonetics Model (BiPhon; Boersma 2011) with an additional reading grammar (Hamann & Colombo, 2017):

1) The between-subject variation in L2 phonological categorization can be attributed to individual cue-weighting strategies (e.g. formant cue vs. silence cue). The L1-Mandarin listeners who predominantly categorize the EP tap as a lateral weigh formant cue over silence cue (formant cue > silence cue), while those whose locus of attention is on the silence cue (silence cue > formant cue) identify the tap most often as a stop. This was formalised as different rankings of cue constraints, which map the auditory input onto the surface phonological form.

2) The within-subject variation in L2 speech learning, on the other hand, was formalised as a re-ranking of the cue constraints that express a listener's cue weighting. The occurrence of probabilistic constraint re-ranking might be due to the fact that the ranking between relevant cue constraints for categorizing the EP tap is not decisive in the learner's L1 Mandarin (irrelevant for L1 cue-to-mapping) and that the relative ranking between these cue constraints is unstable.

3) The variation in the categorization of /ɾ/ as a function of prosodic context can be regarded as an instance of the interaction between cue constraints and structural constraints. The latter reflects phonotactic constraints from the learners' L1 Mandarin, which only allows nasals and an approximant syllable-finally. Therefore, structural modifications, either

epenthesis or deletion, are employed in coda position, in order to accommodate the illegal sequences.

4) The interaction between phonological categorization and orthography during L2 category construction was expressed by means of cue constraints and orthographic constraints that simulate learners' L1 grapheme-phoneme conversion, e.g. mapping the written form onto L1 phonological categories. Since in the BiPhon-OT model all constraints are used both in perception and production, the observed L2 orthographic effect can stem either from phonological categorization (orthographic constraints ranked higher than cue constraints) or from co-activation of the orthographic representation in production (orthographic constraints ranked higher than faithfulness constraints). It is even conceivable to formalise the prosodically-conditioned orthographic effect (the Mandarin rhotic [ɻ] is only employed in coda but not in onset by L2 learners), if we assume that specification of the L2 tap differs across syllable constituents.

5) The perception-production asymmetry observed in L2 acquisition of /l/ and /ɾ/ was formalised within a single grammar, showing that the mismatch between evidence from L2 perception and production is not due to two separate phonological grammars (constraint rankings), but to the fact that the two paralinguistic processes targeted by perception and production studies involve different mappings: in the perception experiment only the mapping from auditory to phonological surface form is triggered, while the production task also involves mapping of the lexical form onto the articulatory form.

5.2 Future research

The research line of the current thesis can be further pursed in three directions, namely experimental research, modelling and applied linguistics.

On the basis of the experimental findings, we formalised the prosodically-conditioned orthographic effect and the asymmetry between L2 perception and production in BiPhon, assuming the (under)specification of the L2 underlying representations; however, we did not test the L2 lexical-phonological representations directly. Future experimental research should target the L2 phonological representations at the lexical level, making use of a lexical task, for instance, a mispronunciation detection paradigm combined with pupillometry. Pupillometry is a novel and sensitive method, which does not require participants to provide explicit responses (e.g. only passive listening is needed), but also allows us to observe the gradual development of L2 lexical-phonological representations (task-evoked pupillary response is continuous, in contrast to the binary response elicited in a classical lexical decision task).

The modelling in the current thesis was carried out within BiPhon, making use of OT-like constraints. One major incompatibility between the OT and the emergentist approach to phonological acquisition lies in the fact that discrete phonological categories cannot emerge from an auditory input distribution via distributional learning (Boersma et al., 2003; Boersma et al., 2018). For instance, we had to assume the specification of L2 phonological representations in our modelling, which we did on the basis of experimental data. Future research can consider the neural network version of the Bidirectional Phonology and Phonetics Model (BiPhon-NN; Boersma et al., 2020), which has successfully modelled the emergence of discrete phonological features from an auditory continuum (Seinhorst et al., 2019).

In chapter 3, the null effect of L2 experience suggests that L1-Mandarin learners' difficulties with the Portuguese /l/ and /ɾ/ cannot be overcome by merely being exposed to more L2 input. Future studies on Laboratory trainings, such as High Variability Perceptual Training, may be promising as they are designed to direct learners' attention to the critical acoustic differences between difficult L2 phonological categories and have been shown to improve L2 perceptual performance only after a few training sections (Wong, 2013; Rato, 2014; Oliveira 2020).

References

Aliaga-Garcia, C. (2010). Measuring perceptual cue weighting after training: A comparison of auditory vs. articulatory training methods. In K. Dziubalska- Kołaczyk, M. Wrembel, & M. Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech*, pp. 2-7.

Altenberg, E., and R. M. Vago. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning* 33(4): 427–447.

Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, *37*(5), 1221–1251. https://doi.org/10.1017/S0142716415000557

Amorim, C. (2014). *Padrão de aquisição de contrastes do pE: a interação entre traços, segmentos e sílabas*. (Unpublished doctoral dissertation), University of Porto, Porto, Portugal.

Amorim, C. & J. Veloso (2018). O estatuto fonológico do rótico dorsal à luz dos dados de aquisição (pp. 131-150) in C. Lazzaroto-Volcão & M. J. Freitas (orgs.) *Estudos em Fonética e em Fonologia, Coletânea em Homenagem a Carmen Matzenauer*. Editora CRV. Curitiba. Brasil.

Andrade, E. (1977). *Aspects de la phonologie (générative) du Portugais*, Lisboa, INIC.

Andrade, A. (1999). On /l/ velarization in European Portuguese. *International Congress of Phonetic Sciences (ICPhS)*, San Francisco, 543-546.

Andruski, J. E., Blumstein, S. E., and Burton, M. (1994). The effect of sub-phonetic differences on lexical access, *Cognition* 52, 163–187. Baker,

Antoniou, M., Ettlinger, M., & Wong, P. C. M. (2016). Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning. *PLoS ONE, 11*(7), Article e0158812. https://doi.org/10.1371/journal.pone.0158812

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6

Apoussidou, D. (2007). *The Learnability of Metrical Phonology*. Ph.D. dissertation, University of Amsterdam.

Archibald, J. & Young-Scholten, M. (2003). The Second Language Segment Revisited. *Second Language Research*, 19 (3), 163-167.

Azevedo, R. Q. (2016). *Formalização fonético-fonológica da interação de restrições na produção e na percepção da epêntese no português brasileiro e no português europeu*, (Unpublished PhD thesis), Universidade Católica de Pelotas, Brazil.

Barberena LS, Keske-Soares M, Berti LC. (2014) Descrição dos gestos articulatórios envolvidos na produção dos sons /r/ e /l/. *Audiol Commun Res*. 19(4):338-44.

Barberena, Luciana da Silva, Uberti, Letícia Bitencourt, Rosado, Isadora Mayer, Moraes, Denis Altieri de Oliveira, Mancopes, Renata, Berti, Larissa Cristina, & Keske-Soares, Márcia. (2019). Comparison of articulatory gestures between men and women in the production of sounds /r/, /l/ and /j/. *Audiology - Communication Research*, *24*, e2059. Epub September 16, 2019. https://dx.doi.org/10.1590/2317-6431-2018-2059

Barrios, S. L., & Hayes-Harb, R. (2020). Second language learning of phonological alternations with and without orthographic input: Evidence from the acquisition of a German-like voicing alternation. *Applied Psycholinguistics*, *41*(3), 547–577. https://doi.org/10.1017/S0142716420000077

Bassetti, B., Escudero, P., & Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, *36*(1), 1–6. https://doi.org/10.1017/S0142716414000393

Batalha, G. N. (1995). *O Português falado e escrito pelos Chineses de Macau*. Instituto Cultural de Macau.

Bates, D., Mächler, M., Bolker, B.N. & Walker, S. C. (2015). Fitting Linear mixed-effects models using lme4, *Journal of Statistical Software*, 67, pp. 1-48.

Bent, T., Bradlow, A. R., and Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech: all's well that begins well, in *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. eds M. J. Munro and O.-S. Bohn (Amsterdam: John Benjamins), 331–347.

Bernhardt, B. & J. Stemberger. (1998). *Handbook of phonological development: From the perspective of constraint-based non-linear phonology*. San Diego, CA: Academic Press.

Berti, L. C. (2010). Investigação da produção de fala a partir da ultrassonografia do movimento de língua. In: *Anais do 18º Congresso Brasileiro de Fonoaudiologia*; 2010 Set 22-25; Curitiba. São Paulo: Sociedade Brasileira de Fonoaudiologia. 1-5.

Best, C. & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of phonetics*, 20: 305–330

Best, C. T. (1995). A direct realist per*spective on cross-language speech perception. In W. Strange (Ed.), Speech perception and linguistic experience: Issues in cross-language research*, pp. 167–200, Timonium, MD: New York Press.

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

Bion, R., Escudero, P., Rauber, A., & Baptista, B. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. *Interspeech 2006 – ICSLP*, 17-21.

Boersma, Paul (1998). *Functional phonology: formalizing the interaction between articulatory and perceptual drives*. Ph.D. dissertation, University of Amsterdam

Boersma, P. (2006). Prototypicality judgments as inverted perception. *Gradience in Grammar*, 167–184. Retrieved from http://www.fon.hum.uva.nl/paul/papers/Prototypicality.pdf

Boersma, P. (2008). Emergent ranking of faithfulness explains markedness and licensing by cue. *Rutgers Optimality Archive 954*, (May 2007), 1–30.

Boersma, P. (2009). Cue constraints and their interactions in phonological perception and production. *Phonology in Perception*, (July), 55–110.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. *Bidirectional Optimality Theory*, (180), 33–72.

Boersma, P. (2012). Modeling phonological category learning. *Phonological Architecture: Empirical, Theoretical and Conceptual Issues*, (August), 299–312.

Boersma, P. and Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32:45–86.

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences Barcelona*, 1–4. https://doi.org/10.1250/ast.23.213

Boersma, P., & Escudero, P. (2008). Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account. *Rutgers Optimality Archive*, *684*, 45–86. Retrieved from http://roa.rutgers.edu/files/684-0904/684-0904-0-0.PDF

Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, *25*(2), 217–270. https://doi.org/10.1017/S0952675708001474

Boersma, P., & Hamann, S. (2009a). Introduction: models of phonology in perception. *Phonology in Perception*, (January), 1–24.

Boersma, P., & Hamann, S. (2009b). Loanword adaptation as first-language phonological perception. *Loanword Phonology*, (July), 1–41.

Boersma, P., & van Leussen, J.-W. (2017). Efficient Evaluation and Learning in Multilevel Parallel Constraint Grammars. *Linguistic Inquiry*, *48*(3), 349–388. https://doi.org/10.1162/ling_a_00247

Boersma, P., Benders, T., & Seinhorst, K. (2020). Neural network models for phonology and phonetics. *Journal of Language Modelling*, *8*(1), 103–177. https://doi.org/10.15398/jlm.v8i1.224

Boersma, P., & Weenink, D. (2019). *Praat: doing phonetics by computer* [Computer program], retrieved from http://www.praat.org/.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research,* pp. 275–300, Timonium, MD: New York Press.

Bohn, O.-S. (2017). Cross-language and second language speech perception. In E. M. Fernandéz & H. S. Cairns (Eds.), *The handbook of psycholinguistics,* pp. 213–239, Hoboken, NJ: John Wiley & Sons.

Bohn, O.-S., & J. E. Flege, (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. A*pplied Psycholinguistics*, 11, pp. 303-328.

Bonet, E. & J. Mascaró (1997). On the representation of contrasting rhotics In F. Martínez- Gil & A. Morales-Front (eds.) *Issues in the Phonology and Morphology of the Major Iberian Languages*. Washington: Georgetown University Press.

Briére, E. J. (1968). *A psycholinguistic study of phonological interference*. The Hague: Mouton.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology*, *64*(1), 74–95. https://doi.org/10.1080/17470218.2010.499174

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, *27*(7–8), 1205–1224. https://doi.org/10.1080/01690965.2012.660170

Broselow, E., Chen, S., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, 20(2), 261-280. doi:10.1017/S0272263198002071

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, 14(2), 136–193. https://doi.org/10.1191/026765898669508401

Browman, C. P., & Goldstein, L. (1995). *Dynamics and articulatory phonology.* In R. F. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (p. 175–193). The MIT Press.

Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M., & Perkell, J. (2011). The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation. *Journal of Speech*, Language, and Hearing Research, 54(3), 727–73

Buchwald, A., & Miozzo, M. (2011). Finding levels of abstraction in speech production: Evidence from sound production impairment. *Psychological Science, 22(9),* 1113-1119.

Buchwald, A., & Miozzo, M. (2012). Phonological and motor errors in individuals with acquired impairment. *Journal of Speech, Language and Hearing Research, 55(5),* 1573-1586.

Cao, Q. (2018). *Perceção das Consoantes Líquidas por Aprendentes Chineses do Português Língua Estrangeira.* (Unpublished MA thesis), University of Aveiro, Aveiro, Portugal.

Cabrelli, J., Luque, A., & Finestrat-Martínez, I. (2019). Influence of L2 English phonotactics in L1 Brazilian Portuguese illusory vowel perception. *Journal of Phonetics*, *73*, 55–69. https://doi.org/10.1016/j.wocn.2018.10.006

Cardoso, W. (2011). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research*, 27(4), 433–465. https://doi.org/10.1177/0267658311413540

Carlisle, R. S. (2001). Syllable structure universal and second language acquisition. *International Journal of English Studies*, *1*(1), 1-19.

Carvalho, J. B. (2006). Markedness gradient in the Portuguese verb: How morphology and phonology interact. In Ivan Fónagy, Yuji Kawaguchi, Tsunekazu Moriguchi (eds.). *Prosody and Syntax*. Amsterdam: Benjamins, p. 157-174.

Chabot, A. (2019). What's wrong with being a rhotic? *Glossa: A Journal of General Linguistics*, *4*(1), 1–24. https://doi.org/10.5334/gjgl.618

Cerni, T., Bassetti, B., & Masterson, J. (2019). Effects of Orthographic Forms on the Acquisition of Novel Spoken Words in a Second Language. *Frontiers in Communication*, *4*(July). https://doi.org/10.3389/fcomm.2019.00031

Cheng, B., & Zhang, Y. (2015). Syllable structure universals and native language interference in second language perception and production: Positional asymmetry and perceptual links to accentedness. *Frontiers in Psychology*, *6*(NOV), 1–17. https://doi.org/10.3389/fpsyg.2015.01801

Cheung, C., Hamiton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *ELife*, *5* (MARCH2016), 1–19. https://doi.org/10.7554/eLife.12577

Chládková, K., Boersma, P., & Benders, T. (2015). The perceptual basis of the feature vowel height. *Proceedings of the 18th International Congress of Phonetic Sciences*.

Church, K. W. (1987). Phonological parsing and lexical retrieval. *Cognition*, 25(1–2), 53–69.

Cichocki, W., House, A. B., Kinloch, A. M., & Lister, A. C. (1999). Cantonese speakers and the acquisition of French consonants. *Language Learning, 49*, 95–121.

Clements, G. N. & Hume, E. (1995). Internal organization of speech sounds, in: John Goldsmith (ed.), *The Handbook of Phonological Theory*, Cambridge, Mass., Basil Blackwell, 245–306.

Cohen, E. G. (2013). Crazy little thing called /r/: Unlocking the mysteries of the Hebrew rhotic. Paper presented at the *Fourth International Symposium on Rhotics*, Autrans (Grenoble), France.

Cohen, E.G. (2015). Phoneme complexity and frequency in the acquisition of Hebrew rhotics. *Journal of Child Language Acquisition and Development* 3(1). 1–11.

Colantoni, L., & Steele, J. (2007). Acquiring /R/ in context. *Studies in Second Language Acquisition*, 27, 381–406.

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*(3), 489–534. https://doi.org/10.1017/S0142716408080223

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice.* Cambridge: Cambridge University Press.

Collins, B., & I. M. Mees. (1984). *The Sounds of English and Dutch*. Leiden: Leiden University Press.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route model and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology*, *7*, 1–17. https://doi.org/10.3389/fpsyg.2016.01345

Costa, T. (2010). *The Acquisition of the Consonantal System in European Portuguese: Focus on Place and Manner Features*. (Unpublished doctoral dissertation), University of Lisbon, Lisbon, Portugal.

Cuetos, F., Hallé, P. A., Domínguez, A., & Segui, J. (2011). Perception of prothetic /e/ in #sC utterances: Gating data. In W. S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, pp. 540–543. Hong Kong: City University of Hong Kong.

Cunillera, T., Càmara, E., Laine, M., & Rodríguez-Fornells, A. (2010). Speech segmentation is facilitated by visual cues. *Quarterly Journal of Experimental Psychology*, *63*(2), 260–274. https://doi.org/10.1080/17470210902888809

Curtin, S., Goad, H., & Pater, J. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, 14, 389–405.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, 36(1), 115–128. https://doi.org/10.1017/S0142716414000459

Darcy, I., Dekydtspotter, L., Sprouse, R.., Glover, J., Kaden, C., McGuire, M.,& Scott, J. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research*, 28, 5-40.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. https://doi.org/10.1075/ml.8.3.06dar

Darcy, I., & Thomas, T. (2019). When blue is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism*. https://doi.org/10.1017/S1366728918001050

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics, 19,* 619–633.

Davidson, L. (2006). Phonotactics and articulatory coordination interact in phonology: evidence from nonnative production. *Cognitive Science*, 30, 837–862. doi: 10.1207/s15516709cog0000_73

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics*, 40(2), 234–248. https://doi.org/10.1016/J.WOCN.2011.11.005

De Jong, K.J., Silbert, N.H. and Park, H. (2009), Generalization Across Segments in Second Language Consonant Identification. *Language Learning*, 59: 1-31. https://doi.org/10.1111/j.1467-9922.2009.00499.x

Dollmann, J., Kogan, I., & Weißmann, M. (2020). Speaking accent-free in L2 beyond the critical period: The compensatory role of individual abilities and opportunity structures. *Applied Linguistics*, *41*(5), 787–809. https://doi.org/10.1093/applin/amz029

Duanmu, S. (2005). Chinese (Mandarin): phonology. *Encyclopedia of Language and Linguistics, 2nd Edition*, ed. by Keith Brown, 351-355. Oxford, UK: Elsevier Publishing House.

Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed.). Oxford: Oxford University Press.

Dupoux, E. (2018). Cognitive Science in the era of Artificial Intelligence: A roadmap for reverse-engineering the infant language-learner. *Cognition*, 173, 34-59.

Dupoux, E., Pallier, C., Sebastian-Gallés, N., & Mehler, J. (1997). A destressing deafness in French? *Journal of Memory and Languag*e, 36(3), 406–421. https://doi.org/10.1006/jmla.1996.2500

Dupoux E, Sebastián-Gallés N, Navarrete E, Peperkamp S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, 106(2): 682–706. 10.1016/j.cognition.2007.04.001

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from?*Journal of Memory and Language*, 64(3), 199–210. https://doi.org/10.1016/J.JML.2010.12.004

Durvasula, K., Huang, H.-H., Uehara, S., Luo, Q., & Lin, Y.-H. (2018).

Phonology modulates the illusory vowels in perceptual illusions: Evidence from Mandarin and English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1). https://doi.org/10.5334/labphon.57

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27(2), 315–330. https://doi.org/10.1111/j.1467-1770.1977.tb00124.x

Eckman, F. R. (1984). Universals, typologies and interlanguages. In W.E. Rutherford (ed.), *Language universals and second language acquisition*, pp. 79-105. Amsterdam: Benjamins.

Eckman, F. (2004). From phonemic differences to constraint rankings. *Studies in Second Language Acquisition*, 26, 513–549.

Eisner, F., & Mcqueen, J. (2018). *Speech perception.* (pp. 1–46). In Stevens' *Handbook of Experimental Psychology and Cognitive Neuroscience*, Fourth Edition, edited by John T. Wixted. https://doi.org/10.1002/9781119170174.epcn301

Elvin J., & Escudero P. (2019) Cross-Linguistic Influence in Second Language Speech: Implications for Learning and Teaching. In: Gutierrez-Mangado M., Martínez-Adrián M., Gallardo-del-Puerto F. (eds) *Cross-Linguistic Influence: From Empirical Evidence to Classroom Practice. Second Language Learning and Teaching*. Springer, Cham. https://doi.org/10.1007/978-3-030-22066-2_1

Engstrand, O., Frid, J., & Lindblom, B. (2007). A perceptual bridge between coronal and dorsal /r/. In: Solé, M.J., Beddor, P.S., Ohala, M. (Eds.), *Experimental Approaches to Phonology*. OUP, Oxford, GB, pp. 175–191.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition. LOT, The Netherlands.*

Escudero, P. (2007). Second language phonology: The role of perception. In *Phonology in Context*. New York: Palgrave Macmillan.

Escudero, P. (2009). The Linguistic Perception of similar L2 sounds. In P. Boersma & S. Hamann (Eds.), *Phonology in Perception*, pp. 152–190.

Escudero, P., & Boersma, P. (2004). Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Studies in Second*

Language Acquisition, *26*(04), 551–585. https://doi.org/10.1017/S0272263104040021

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360. https://doi.org/10.1016/j.wocn.2007.11.002

Escudero, P., & Wanrooij, K. (2010). The Effect of L1 Orthography on Non-native Vowel Perception. *Language and Speech*, *53*(3), 343–365. https://doi.org/10.1177/0023830910371447

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, *17*(02), 384–395. https://doi.org/10.1017/S1366728913000436

Espadinha, M. A., & Silva, R. (2009). O Português de Macau. Paper presented at the *II Simpósio Mundial de Estudos em Língua Portuguesa (SIMELP)*, University of Evora, Portugal.

Fadiga, Luciano, Laila Craighero, Giovanni Buccino and Giacomo Rizzolatti (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience* 15: 399-402.

Faris, M. M., Best, C. T., & Tyler, M. D. (2016). An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized. *The Journal of the Acoustical Society of America*, 139(1), EL1–EL5. https://doi.org/10.1121/1.4939608

Fikkert, P. (1994). *On the acquisition of prosodic structure*. PhD Dissertation, HIL dissertations 6, Leiden University. The Hague: Holland Academic Graphics.

Fikkert, P. (2007). Developing word representations in the lexicon: evidence from perception and production. *Development*, *134*(4), 635–646.

Fikkert, P., & De Hoop, H. (2009). Language acquisition in optimality theory. *Linguistics*, *47*(2), 311–357. https://doi.org/10.1063/1.3033202

Flege, J. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, 86, 1684-1697.

Flege, J. (1995). Second Language Speech Learning: Theory, Findings and Problems. In Strange, W. (Ed), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 233-277). Timonium, MD: New York Press.

Flege, J. (2021). New methods for second-language (L2) speech research, Manuscript, available at https://www.researchgate.net/publication/348266082_New_methods_for_second-language_L2_speech_research

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

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470. https://doi.org/10.1006/JPHO.1997.0052

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23(4), 527-552.

Flege, J., & MacKay, I. (2004) Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34. doi: 10.1017/50272263104261010

Flege, J., & Bohn, O. (2021). The revised Speech Learning Model, Manuscript, available at https://www.researchgate.net/publication/348265785_The_revised_Speech_Learning_Model_SLM-r

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

Freitas, M. J. (1997). *Aquisição da estrutura silábica do Português Europeu.* (Unpublished PhD Dissertation). University of Lisbon, Lisbon, Portugal.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123, 71–99

Frota, S. (2014). The intonational phonology of European Portuguese. In S. A. Jun (Ed.) *Prosodic Typology II*. Oxford: Oxford University Press, pp. 6-42.

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

Funatsu, S., & Fujimoto, M. (2012). Mechanisms of vowel epenthesis in consonant clusters: an EMA study. *Proceedings of Acoustics, Nantes Conference*, (August), 341–346. Retrieved from http://hal.archives-ouvertes.fr/hal-00810613/

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performanc*e, 6(1), 110–125. https://doi.org/10.1037//0096- 1523.6.1.110

Gay, T., Lindblom, B. & Lubker, J. (1981). Production of bite block vowels: Acoustic equivalence by selective compensation. Journal of Acoustic Society of America, 69, pp. 802-810.

Gick, B., Wilson, I., Koch, K., Cook, C., 2004. Language-specific articulatory settings: Evidence from inter-utterance rest position. *Phonetica* 61, 220–233.

Gordon, Gor, K. (2015). Phonology and morphology in lexical processing. In *The Cambridge Handbook of Bilingual Processing* (pp. 173–199).

Goswami. U, Ziegler. JC, & Richardson, U. (2005). The effects of spelling consistency on phonological awareness: a comparison of English and German. *J Exp Child Psychol* 92(4):345-65.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, 9, 317–323. doi:10. 1016/0028-3932(71)90027-3

Greenberg, J. (1978). Some generalizations concerning initial and final consonant clusters. In J. Greenberg (ed.), *Universals of human language, II: Phonology*. Stanford University

Guan, Q. (2019). Emerging modes of temporal coordination: Mandarin and non-native consonant clusters, (Unpublished PhD thesis), Université Sorbonne Paris Cité.

Gulian, M., Escudero, P., & Boersma, P. (2007). Supervision hampers distributional learning of vowel contrasts. *Proceedings of the international congress of phonetic sciences*, pp. 1893–1896, Saarbrücken, Germany

Guirao, M., & García Jurado, M. A. (1991). Los perfiles acústicos y la identificación de /l/ y /r/[Acoustic profiles and the identification of /l/ and /r/]. Revista Argentina de Lingüística, 7,21–42.

Gundel, j., Houlihan, k., & Sanders, G. (1986). Markedness distribution in phonology and syntax. In F. Eckman, E., Moravcsik, & J. Wirth (eds.), *Markedness*, New York: Plenum Press, pp. 107-138.

Hale, M. & and Kissock, M. (2007). The phonetics-phonology interface and the acquisition of perseverant underspecification. In: Gillian Ramchand and Charles Reiss (eds.) *The Oxford Handbook of Linguistic Interfaces*. Oxford: Oxford University Press, 81-102.

Hamann, S. (2009a) Variation in the perception of an L2 contrast: A combined phonetic and phonological account. In: F. Kügler, C. Fery & R. v. d. Vijver (eds.) *Variation and Gradience in Phonetics and Phonology*. Berlin: Mouton de Gruyter, 79–105.

Hamann, S. (2009b) The learner of a perception grammar as a source of sound change. In: P. Boersma & S. Hamann (eds.) *Phonology in Perception*. Berlin: Mouton de Gruyter, 111–149.

Hamann, S. (2011). The phonetics-phonology interface. In *Continuum Companion to Phonology* (pp. 202–224).

Hamann, S. (2020). One phonotactic restriction for speaking, listening and reading: The case of the *no geminate* constraint in German. In: Martin Evertz-Rittich & Frank Kirchhoff (eds.) *Geschriebene und gesprochene Sprache als Modalitäten eines Sprachsystems - Written and spoken language as modalities of one language system*. Berlin: de Gruyter, 57–78. doi: 10.1515/9783110710809-004

Hamann, S., & Colombo, I. E. (2017). *A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. Natural Language and Linguistic Theory* (Vol. 35). https://doi.org/10.1007/s11049-017-9362-3

Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5–33. https://doi.org/10.1177/0267658307082980

Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech*

Communication, *52*(11–12), 996–1009. https://doi.org/10.1016/j.specom.2010.05.003

He, Y. (2015). Production of English Syllable Final /l/ by Mandarin Chinese Speakers. *Journal of Language Teaching and Research*, 5(4), pp. 742–750. https://doi.org/10.4304/jltr.5.4.742-750

Heselwood, B. (2009). Rhoticity without F3: lowpass filtering, F1-F2 relations and the perception of rhoticity in the ‘north-force,’ ‘start’ and ‘nurse’ words. Leeds Work. *Paper. Linguist. Phon.* Vol. 14 pp. 49-64.

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.

Howson, P. (2018a). Rhotics and palatalization: an acoustic examination of upper and lower Sorbian. *Phonetica* 75 (2), 132–150 .

Howson, P. (2018b). *A Phonetic Examination of Rhotics: Gestural Representation Accounts for Phonological Behaviour*. (Unpublished Ph.D. Thesis). University of Toronto, Toronto, Canada .

Howson, P. J., & Monahan, P. J. (2019). Perceptual motivation for rhotics as a class. *Speech Communication*, *115*(July), 15–28. https://doi.org/10.1016/j.specom.2019.10.002

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. 共 2003 兲. “A perceptual interference account of acquisition difficulties for non-native phonemes,” Cognition 87, B47 – B57.

Jesus, Luis M. T. & Christine H. Shadle. 2005. Acoustic analysis of European Portuguese uvular [χ, ʁ] and voiceless tapped alveolar [ɾ̥] fricatives. Journal of the International Phonetic Association 35(1). 27–44. DOI: https://doi.org/10.1017/S0025100305001866

Jiang, S., Y-C. Chang and F.-f. Hsieh (2019). An EMA study of er- suffixation in Northeastern Mandarin monopthongs, *Proceedings of ICPhS,* 2019, Melbourne, 2149–2153.

John, P., & Cardoso, W. (2017). On syllable structure and phonological variation: The case of I-epenthesis by Brazilian Portuguese learners of English. *Ilha Do Desterro*, *70*(3), 169–184. https://doi.org/10.5007/2175-8026.2017v70n3p169

Jusczyk, Peter W. (1997). *The Discovery of Spoken Language*. Cambridge (Massachusetts): MIT Press.

Kabak, B., & Idsardi, W. (2003). Syllabically Conditioned Perceptual Epenthesis. *Annual Meeting of the Berkeley Linguistics Society*, *29*(1), 233. https://doi.org/10.3765/bls.v29i1.1018

Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech*, 50(1), 23–52. DOI: https://doi.org/10.1177/00238309070500010201

Kirchner, R. (1998). Lenition in Phonetically-Based Optimality Theory, (Unpublished PhD thesis), UCLA, USA.

Kojima, C., & Darcy, I. (2014). Learners' Proficiency and Lexical Encoding of the Geminate / Non-geminate Contrast in Japanese, In *Selected Proceedings of the 2012 Second Language Research Forum*, ed. Ryan T. Miller et al., 30-38. Somerville, MA: Cascadilla Proceedings Project.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 121– 154, Timonium, MD: New York Press.

Ladefoged, P. & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 637–676). Berlin, Germany: Mouton de Gruyter.

Lahiri, A. & R., Henning. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*. 38. 44-59. 10.1016/j.wocn.2010.01.002.

Legendre, G., M., Yoshiro & P., Smolensky. (1990). Can Connectionism Contribute to Syntax? Harmonic Grammar, with an Application, *Computer Science Technical Reports*. 467.

Lehiste I. (1964). Acoustical characteristics of selected English consonants. *Indiana University Research Center in Anthropology, Folklore and Linguistics*, 34 (1964), pp. 10-50.Lenneberg, E.H. (1967). *Biological Foundations of Language*. Wiley. ISBN 978-0-89874-700-3.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Li, M., Zhang, J. (2017). Perceptual distinctiveness between dental and palatal sibilants in different vowel contexts and its implications for phonological contrasts. *Laboratory Phonology*, 8(1).

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 65(4), 497–516.

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

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, *30*(2), 133–143. https://doi.org/10.3758/BF03204471

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

Lin, H. (2001). *A grammar of Mandarin Chinese*. Muenchen: Lincom Europa.

Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Lindau, M. 1985. The story of /r/. In Victoria Fromkin (ed.), *Phonetic linguistics: Essays in honour of Peter Ladefoged*, 157–168. Orlando: Academic Press. Magnuson,

Lindblom, B., J. Lubker and T. Gay (1979). Formant frequencies of some fixed mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics* 7: 147-161

Liu, W. (2018). *Aquisição da Vibrante Simples [r] pelos Alunos Chineses Aprendentes de Português como Língua Estrangeira*. (Unpublished MA thesis), University of Macau, Macau, China.

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

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/and /l/: III. Long-term retention of new phonetic categories. Journal of the Acoustical Society of America, 96, 2076-2087.

Llompart, M. & Reinisch, E. (2019). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts," *Language and Speech*, 62(3), pp. 594–622.

Lousada, M. L. (2006). *Estudo da produção de oclusivas do português europeu*. Unpublished MA dissertation, University of Aveiro, Portugal.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, I. (2013). *Lateral Consonants*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Mah, J., Goad, H., & Steinhauer, K. (2016). Using Event-Related Brain Potentials to Assess Perceptibility: The Case of French Speakers and English [h]. *Frontiers in Psychology*, 7(OCT), 1–14. https://doi.org/10.3389/fpsyg.2016.01469

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect discrimination. *Cognition* 82, B101–B111. doi: 10.1016/S0010-0277(01)00157-3

Major, R. (2008). Transfer in second language phonology: A review. In J. G. Hasen Edwards and M. L. Zampini (Eds), *Phonology and Second Language Acquisition*, pp. 63–94. Amsterdam: John Benjamins.

Mann, V. A. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English "l" and "r". *Cognition*, 24(3), 169–196.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 28(3),213–228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal ofthe Acoustical Society of America*, 69(2), 548–558.

Marques, I. (2010). *A Variação Fonética da Lateral Alveolar no Português Europeu*. (Unpublished Master dissertation). University of Aveiro.

Martins, M. M. (2008). *O português dos chineses em Portugal – O caso dos imigrantes da área do comércio e restauração em Águeda*. (Unpublished MA thesis), University of Aveiro, Aveiro, Portugal

Martins P, Oliveira C, Silva A. Articulatory Characteristics of European Portuguese Laterals: a 2D & 3D MRI Study. (2010). *VI Jornadas en Tecnología del Habla and II Iberian SLTech Workshop*. pp. 33–6.

Mateus, M. H. M. & Rodrigues, C. (2003). A vibrante em coda no português. *Teoria Lingüística. Fonologia e outros temas*, 181-199.

Mateus, M. H. M., Falé, I. & Freitas, M. J. (2005). *Fonética e Fonologia do Português*. Lisboa: Universidade Aberta.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314-324. doi:10.3758/s13428-011-0168-7

Matthews, J., & Brown, C. (2004). When intake exceeds input: Language specific perceptual illusions induced by L1 prosodic constraints. *International Journal of Bilingualism*, 8, 5–27. doi:10.1177/13670069040080010201

Melinger, A., Branigan, H. P., & Pickering, M. J. (2014). Parallel processing in language production. *Language, Cognition and Neuroscience, 29*, 663-683.

Melnik, G. A. (2019). *Issues in L2 phonological processing*, (Unpublished PhD thesis), Ecole Normale Supérieure de Paris, Paris. France.

Melnik, G. A., & Peperkamp, S. (2019). Perceptual deletion and asymmetric lexical access in second language learners. *The Journal of the Acoustical Society of America*, 145(1), EL13–EL18. https://doi.org/10.1121/1.5085648

McCandliss BD, Fiez JA, Protopapas A, Conway M, McClelland JL. (2002) Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogn Affect Behav Neurosci*, 2(2):89-108. doi:10.3758/cabn.2.2.89

McCarthy, J. & A. Prince (1995): Faithfulness and reduplicative identity. In J.Beckman, L. W. Dickey & S. Urbanczyk (eds.), *Papers in Optimality*

Theory. *University of Massachusetts Occasional Papers* 18. Amherst, Mass.: Graduate Linguistic Student Association. pp. 249–384.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86

McGuire, G. (2014). Orthographic effects on phonetic cue weighting. *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, pp. 201–204, 2014.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264:746–748.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), 33–42. https://doi.org/10.1016/S0010-0277(02)00157-9

McQueen, J. M. and Cutler, A. (1997). Cognitive processes in speech perception. In Hardcastle, W. J. and Laver, J., editors, *The Handbook of Phonetic Sciences*, pages 566–585. Blackwell, Oxford.

Miatto, V., S. Hamann & P. Boersma (2019). Self-reported L2 input predicts phonetic variation in the adaptation of English final consonants into Italian. In: S., Calhoun, P., Escudero, M., Tabain & P., Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*. Canberra: Australasian Speech Science and Technology Association Inc; 949–953.

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction with- out phonemes in speech perception. *Cognition*, 129(2), 356–361. doi:10.1016/j.cognition. 2013.07.011 Morillon,

Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language*, *98*, pp. 77–92. https://doi.org/10.1016/j.jml.2017.09.005

Monahan, P. J., Takahashi, E., Nakao, C., & Idsardi, W. J. (2009). Not All Epenthetic Contexts are Equal : Differential Effects in Japanese Illusory. In *Proceedings of the 17th Annual Japanese/Korean Lingustics Conference* (pp. 391–405).

Monahan, P. J. (2018). Phonological Knowledge and Speech Comprehension. *Annual Review of Linguistics*, *4*(1), 21–47. https://doi.org/10.1146/annurev-linguistics-011817-045537

Montrul, S. (2014) Interlanguage, transfer and fossilization. Beyond second language acquisition. In Han, Z., & Tarone, E. (2014). *Interlanguage: Forty years later,* pp.75-104. Amsterdam: John Benjamins Publishing Company.

Natvig, D. (2020). Rhotic underspecification: Deriving variability and arbitrariness through phonological representations. *Glossa: A Journal of General Linguistics, 5*(1), 1–28.

Navarra, Jordi & Soto-Faraco, Salvador. (2007). Hearing lips in a second language: Visual articulatory information enables the perception of second language sounds. *Psychological research*. 71. 4-12. 10.1007/s00426-005-0031-5.

Nixon, J. S. (2020). Of mice and men: Speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition*, *197*(January), 104081. https://doi.org/10.1016/j.cognition.2019.104081

Nobre-Oliveira, D. (2007). *The Effect of Perceptual Training on the Learning of English Vowels by Brazilian Portuguese Speakers* (Unpulisbied Doctoral dissertation). Retrieved from http://repositorio.ufsc.br/

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–325.

Norris, D., Mcqueen, J., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Ohala, J.J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P.F. (ed.), *The Production of Speech*. New York: Springer-Verlag, 189- 216.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues.*The Journal of Acoustic Society of Amercia*, 99, pp.1718-1725.

Ohala, J.J., (2011). Accommodation to the Aerodynamic Voicing Constraint and its Phonological Relevance. In *ICPhS 2011*, pp. 64-67.

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

Oliveira, C. Martins, P., Teixeira, A., Marques, I., Sá Couto, P. (2011). An articulatory and acoustic study of the European Portuguese /l/. *17th International Congress of Phonetic Sciences (ICPhS)*, Hong Kong.

Oliveira, D. (2016). *Perceção e Produção de Sons Consonânticos do Português Europeu por Aprendentes Chineses*. (Unpublished MA disseration), University of Minho, Braga, Portugal.

Oliveira, D. (2020). *Auditory selective attention and performance in high variability phonetic training: The perception of Portuguese stops by Chinese L2 learners*. (Unpublished PhD thesis), University of Minho, Braga, Portugal.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.

Pape, D., & Jesus, L. M. T. (2014). Cue-weighting in the perception of intervocalic stop voicing in European Portuguese. *The Journal of the Acoustical Society of America*, 136(3), 1334–1343. https://doi.org/10.1121/1.4890639

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

Paradis, C., & Prunet, J.-F. (2000). Nasal Vowels as Two Segments: Evidence from Borrowings. *Language*. 76. 324-357. 10.1353/lan.2000.0117.

Paradis, C., & LaCharité, D. (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry*, *36*(2), 223–258.

Park, H., & de Jong, K. J. (2017). Perceptual category mapping between English and Korean obstruents in non-CV positions: Prosodic location effects in second language identification skills. *Journal of Phonetics*, *62*, 12–33. https://doi.org/10.1016/J.WOCN.2017.01.005

Perkell, J. S. (1969). Physiology of speech production: Results and implications of a quantitative cineradiographic study. *M.I.T. research monograph, No 53*.

Pereira, R. (2020). *O R-forte em Português Europeu: análise fonológica de dados dialetais*. (Unpublished MA thesis), University of Lisbon, Lisbon, Portugal.

Perre, L. & Ziegler, J.C. (2008). On-line activation of orthography in spoken word recognition, Brain Research, Volume 1188, pp. 132-138, ISSN 0006-8993, https://doi.org/10.1016/j.brainres.2007.10.084.

Polivanov, E. D. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague*, 4: 79–96. [English translation: The subjective nature of the perceptions of language sounds. In E.D. Polivanov (1974): *Selected Works: Articles on general linguistic*s, pages 223–237. Mouton, The Hague: Mouton].

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39(4), 467–478. https://doi.org/10.1016/j.wocn.2010.08.007

Prince, A. and Smolensky, P. (1993). Optimality Theory: Constraint interaction in generative grammar. Technical Report 2, Rutgers University Center for Cognitive Science.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rafat, Y. (2016). Orthography-induced transfer in the production of English-speaking learners of Spanish. *Language Learning Journal,* 44(2), 197–213. https://doi.org/10.1080/09571736.2013.784346

Rafat, Y., & Stevenson, R. A. (2018). Auditory-orthographic integration at the onset of L2 speech acquisition. *Language and Speech*. https://doi.org/10.1177/0023830918777537

Ramanarayanan, V., L. Goldstein, & S. S. Narayanan. (2013). Spatio-temporal articulatory movement primitives during speech production: Extraction, interpretation, and validation. *The Journal of the Acoustical Society of America*, 134(2), 1378–1394.

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., & Dupoux, E. (2010). A psycholinguistic perspective on the acquisition of phonology. In *Laboratory Phonology 10*. Berlin, Boston: De Gruyter Mouton.

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. https://doi.org/10.1016/j.jml.2007.04.001

Rasmussen, S., & Bohn, O. (2019). Vowel Context Affects Danish L2 Chinese Learners' Identification of Postalveolar Sibilants, In Sasha Calhoun, Paola

Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia 2019. Canberra, Australia: Australasian Speech Science and Technology Association Inc.

Rato, A. (2014). Cross-language perception and production of English vowels by Portuguese learners: The effects of perceptual training, (Unpublished PhD dissertation), University of Minho, Braga, Portugal.

Rauber, A. S., Rato, A., Silva, A. (2010). Percepção e produção de vogais anteriores do inglês por falantes nativos de mandarim. *Diacrítica,* 24(1), 5-23.

Recasens, D. (1996). An Articulatory-Perceptual Account of Vocalization and Elision of Dark /l/ in the Romance Languages. *Language and Speech, 39(1), 63–89.* doi:10.1177/002383099603900104

Recasens, D. & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear /l/ and dark /l/: evidence from two Catalan dialects. *Journal of the International Phonetic Association* 35/1, pp. 1–25.

Recasens, D. & Espinosa, A. (2010). A perceptual analysis of the articulatory and acoustic factors triggering dark /l/ vocalization. In: *Experimental Phonetics and Sound Change*, eds. Daniel Recasens, Fernando Sánchez Miret & Kenneth Wireback. Lincom Europa, München, pp.71–82.

Rennicke, I. & Martins, P. (2013). As realizações fonéticas de /ʀ/ em português europeu: análise de um corpus dialetal e implicações no sistema fonológico. In F. Silva, I. Falé & I. Pereira (eds.), *Textos Selecionados do XXVIII Encontro Nacional da Associação Portuguesa de Linguística*. Coimbra: Associação Portuguesa de Linguística, 509–523.

Rodrigues, C. (2003). *Lisboa e Braga: Fonologia e Variação*. (Unpublished Ph.D. dissertation), University of Lisbon, Lisbon, Portugal.

Rodrigues, C., & da Hora, D. (2016). Main Current Processes of Phonological Variation. *The Handbook of Portuguese Linguistics*, 504–525. https://doi.org/10.1002/9781118791844.ch28

Rodrigues, Susana. (2015). *Caracterização acústica das consoantes líquidas do Português Europeu.* (Unpublished PhD Dissertation). University of Lisbon, Lisbon, Portugal.

Rodrigues, S., Martins, F., Silva, S., & Jesus, L. M. T. (2019). /l/ velarisation as a continuum. *PLoS ONE, 14*(3), 1–22.

Rogers, C. L., & Dalby, J. (2005). Forced-choice analysis of segmental production by Chinese-accented English speakers. *Journal of speech, language, and hearing research : JSLHR, 48*(2), 306–322. https://doi.org/10.1044/1092-4388(2005/021)

Saito, K., Sun, H., & Tierney, A. (2020). Domain-general auditory processing determines success in second language pronunciation learning in adulthood: A longitudinal study. *Applied Psycholinguistics, 41*(5), 1083-1112. doi:10.1017/S0142716420000491

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187-224. doi:10.1017/S0142716417000418

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language*, *111*(November). https://doi.org/10.1016/j.jml.2019.104070

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. https://doi.org/10.1016/j.wocn.2015.07.003

Schouten, M.E.H. (1977). Imitation of synthetic vowels by bilinguals,*Journal of Phonetics*, 5, pp. 273–283.

Sebregts, K. (2014). *The sociophonetics and phonology of Dutch r*. Utrecht: Utrecht University dissertation.

Seinhorst, K., P., Boersma & S., Hamann (2019). Iterated distributional and lexicon-driven learning in a symmetric neural network explains the emergence of features and dispersion. In: Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.) *Proceedings of the 19th International Congress of Phonetic Sciences*. Canberra: Australasian Speech Science and Technology Association Inc; 1134–1138.

Shiller, D. M., Rvachew, S., & Brosseau-Lapré, F. (2010). Importance of the auditory perceptual target to the achievement of speech production accuracy. *Canadian Journal of Speech-Language Pathology and Audiology, 34*(3), 181–192.

Smith, J. G. (2010). *Acoustic Properties of English /l/ and /ɹ/ Produced by Mandarin Chinese Speakers*, (Unpublished MA disseration), University of Toronto, Toronto, Canada.

Silva, A. (2014). *Análise Acústica da Vibrante Simples do Português Europeu.* (Unpublished MA dissertation). University of Aveiro, Aveiro, Portugal.

Simonchyk, A. (2017). *The Relationships between Perception, Production, Lexical Coding and Orthography in the Acquistion of Palatalization in L2 Russian*, (Unpublished PhD thesis), Indiana University, USA.

Sproat, R. & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21:291-311.

Stilp, C. (2020). Acoustic context effects in speech perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(1), 1–18. https://doi.org/10.1002/wcs.1517

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, 39(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

Steriade, D. (1999). *The phonology of perceptibility effects: The P-map and its consequences for constraint organization.* Unpublished manuscript, University of California, Los Angeles.

Święciński, R. (2013). An EMA study of articulatory settings in Polish speakers of English.W: Waniek-Klimczak, Ewa; Shockey, Linda R. (Eds.) *Teaching and Researching EnglishAccents in Native and Non-native Speakers*. Heidelberg: Springer Publications. 73-82.

Taft, M. (2006). Orthographically influenced abstract phonological representation: Evidence from non-rhotic speakers. *Journal of Psycholinguistic Research*, *35*(1), 67–78. https://doi.org/10.1007/s10936-005-9004-5

Tartter VC, Kat D, Samuel AG, Repp BH. (1983). Perception of intervocalic stop consonants: the contributions of closure duration and formant transitions. The Journal of the Acoustical Society of America, 74(3):715-725.

Teixeira, António., Martins, Paula., Oliveira, Catarina., & Silva, Augusto. (2012). Production and Modeling of the European Portuguese Palatal Lateral. In H. Caseli, A. Villavicencio, A. Teixeira, and F. Perdigão (Eds.), *Computational Processing of the Portuguese Language*. Berlin: Springer –Verlag, pp. 318-328

Trubetzkoy, N. S. (1977). *Grundzüge der Phonologie*, 6th Edn. Göttingen: Van den Hoeck & Ruprecht.

Tyler, Vale, A. (2020). *Perceção das Consoantes líquidas / ɾ / e / l / do Português Europeu sob Influência do Mandarim L1*. (Unpublished MA thesis), University of Minho, Braga, Portugal.

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Frontiers in Psychology*, *6*(August), 1–12. https://doi.org/10.3389/fpsyg.2015.01000

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1269–1291.

Vale, A. (2020). *Perceção das Consoantes líquidas /ɾ/ e /l/ do Português Europeu sob Influência do Mandarim L1*. (Unpublished MA thesis), University of Minho, Braga, Portugal.

Veivo, O., Porreta, V., Hyönä, J., & Järvikivi, J. (2018). Spoken second language words activate native language orthographic information in late second language learners. *Applied Psycholinguistics*, *39*, 1–22. https://doi.org/10.1017/S0142716418000103

Veloso, J. (1997). Vozemento, duração e tensão nas oposições de sonoridade das oclusivas orais do português. *Línguas e Literaturas*, XIV, 59–80.

Veloso, J. (2015). The English R Coming! The never ending story of Portuguese rhotics. *OSLa. Oslo Studies in Language*. 7(1): 323-336.

Veloso, J. (2019). Complex Segments in Portuguese: The Unbearable Heaviness of Being Palatal. In I. Zendoia, O. Nazabal (eds.). *Bihotz ahots. M. L. Oñederra irakaslearen omenez*. Bilbao: Universidad del País Vasco, Euskal Herriko Unibertsitatea, pp. 513- 526

Vigário, M. (2003). *The Prosodic Word in European Portuguese*. (Interface Explorations Series, 6). Berlin/ New York: Mouton de Gruyter.

Vigário, M. (2019) Phonetics and phonology of Portuguese. In Gabriel, Christoph, Gess, Randall & Meisenburg, Trudel (eds) *Manual of Romance phonetics and phonology*. Berlin/New York: De Gruyter.

Vogele, A. C. E. (2020). O Modelo BiPhon-OT e a formalização dos Transtornos dos Sons da Fala Using the BiPhon-OT to Model the Speech Sound Disorders. *Brazilian Journal of Health Review,* 18014–18053. https://doi.org/10.34119/bjhrv3n6-205

Waltmunson, J. (2005). *The relative degree of difficulty of Spanish /t, d/, trill and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy*. Seattle: University of Washington. (Unpublished Doctoral dissertation).

Wang, X. (2008). *Perceptual Training for Learning English Vowels – Perception, Production, and Long-Term Retention*. Saarbrücken: VDM Verlag Dr. Müller.

Watkins, Kate E., A.P. Strafella and Tomáš Paus (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia* 41: 989-994.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Wetzels, L. (2000). Consoantes palatais como geminadas fonológicas no Português Brasileiro. *Revista de Estudos da Linguagem*, Vol.9,2: 5-15.

Wiese, R. (2004). How to optimize orthography. *Written Language and Literacy* 7: 305–331. Zingarelli,

Wiese, R. (2011). The representation of rhotics. In Marc van Oostendorp, Elizabeth Ewen, Colin J. Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 711– 729. Oxford: Blackwell. DOI: https://doi.org/10.1002/9781444335262.wbctp0030

Wilson, I., & Gick, B. (2014). Bilinguals use language-specific articulatory settings. *Journal of Speech, Language, and Hearing Research*, *23*, 361–373.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience* 7:701–702. doi: 10.1038/nn1263

Wong, J.W.S. (2012). Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, Sydney, Australia, 3-6.

Wong, J.W.S. (2013). The effects of perceptual and or productive training on the perception and production of English vowels /I/ and /i:/ by Cantonese ESL learners. In F. Bimbot et al. (Eds.), *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, pp. 2113–2117. Lyon, France: ISCA.

Wong, J.W.S. (2014). The Effects of High and Low Variability Phonetic Training on the Perception and Production of English Vowels /e/-/æ / by Cantonese ESL Learners with High and Low L2 Proficiency Levels. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014)*, ed. Haizhou Li, Helen Meng, Bin Ma, Eng Siong Chng, Lei Xie, 524- 528. Singapore. ISCA.

Xing. K. (2019). Onset /R/-ticulation in Mandarin. Paper presented at *'R-atics6*, Paris, France.

Zhou, C. (2017). *Contributo para o estudo da aquisição das consoantes líquidas do português europeu por aprendentes chineses*. (Unpublished MA thesis), University of Lisbon, Lisbon, Portugal.

Zhu, X. (2007). Jinyin-fulun Putonghua rimu [About approximant]. *Fangyan* 1: 2-9.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review*, *5*(4), 683–689. https://doi.org/10.3758/BF03208845

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, 12, 379–384

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Zimmer, M., & Alves, U. (2012). Uma visão dinâmica da produção da fala em L2: o caso da Dessonorização Terminal. *Revista da Abralin*, 11 (1), 221-272.

Appendix I

语言背景调查

Language Background Questionnaire

姓名 **Name** 性别 **Gender** 出生日期 **Date of Birth**

出生地 **Place of Birth**

1. 你的中文方言

Your mother tongue (dialect)

2. 从几岁开始说普通话?

When did you start to learn (use) Mandarin?

3. 学习葡语的时长以及学习的方式 (本科专业，自学)

Length of learning Portuguese (College Major or self-learning)

4. 除葡语外，掌握的其他外语以及学习时长

In addition to Portuguese, Foreign language(s) you speak and the respective length of study (year)

英语 English 法语 French 西班牙语 Spanish 意大利语 Italian 日语 Japanese 韩语 Korean

5. 你有任何听说读写方面的语言能力障碍吗?

Do you have any difficulties reading and writing/speaking and listening?

签名 **Signature** 日期 **Date**