Cross-linguistic interaction between phonological categorization and orthography predicts prosodic effects in the acquisition of Portuguese liquids by L1-Mandarin learners

Chao Zhou[1], Silke Hamann[2]

[1]University of Lisbon, Portugal
[2]University of Amsterdam, The Netherlands
zhouchao@edu.ulisboa.pt, silke.hamann@uva.nl

Abstract

Prior research has revealed that L1-Mandarin learners employed position-dependent repair strategies for European Portuguese /l/ and /ɾ/. In this study we examined whether this L2 prosodic effect can be attributed to a cross-linguistic influence and whether the replacement of the Portuguese rhotic by the Mandarin [ɻ] is due to perception or orthography. We performed a delayed imitation task with naïve Mandarin listeners and manipulated the presented input types (auditory form alone or a combination of auditory and written forms). Results showed that naïve responses were reminiscent of L1-Mandarin learners' behaviour, and that [ɻ] was used almost exclusively in the presence of written input, suggesting that the prosodic effect attested in L2 acquisition of European Portuguese /l/ and /ɾ/ stems from cross-linguistic interaction between phonological categorization and orthography.

Index Terms: second-language acquisition, Mandarin, European Portuguese, liquids, orthography

1. Introduction

Recent research [1, 2] has shown that second-language (L2) acquisition of the European Portuguese (EP) lateral /l/ and rhotic /ɾ/ by L1-Mandarin learners is conditioned by prosodic position: learners almost never mispronounce EP /l/ in onset, but in coda frequently vocalize it as [w]. EP /ɾ/ is replaced by the learners with [l] in onset, while in coda they delete the segment, insert a schwa (and thus create an onset), or substitute it either with [l], a coronal stop or the Mandarin rhotic [ɻ].

Decades of studies on L2 speech lead researchers to converge on the idea that most divergence between the learners' output and the target form can be attributed to cross-linguistic influence (CLI), i.e. an interaction between the learner's mother tongue (L1) and the target language (L2), see [3] for a review. In the present study, we explored whether CLI can also explain the observed prosodically-conditioned repair strategies by Mandarin learners 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/. If CLI were responsible for the position-dependent treatment of EP liquids, then we expected to observe similar results for naïve Mandarin speakers.

For the previously-reported replacement of EP /ɾ/ by the Mandarin rhotic [ɻ] in coda position, which is similar to the use of [ɻ] for the Spanish tap by L1-Mandarin learners (observed in [4]), we furthermore tested whether this stems from CLI via phonological categorization or via orthography: From the point of categorization, EP /ɾ/ and Mandarin /ɻ/ could be argued to share acoustic-perceptual cues, such as first (F1) and second formant frequencies (F2) and trajectories (see [5]), or phonological features [6], which would result in EP /ɾ/ being perceived as Mandarin /ɻ/. Evidence for this comes from a non-native perception experiment [7] where various types of rhotics, i.e. [r, ɻ, ʀ], were treated as one class. Alternatively, the replacement with [ɻ] could be driven by orthography, since both Mandarin /ɻ/ (in Pinyin) and Portuguese /ɾ/ are represented by the letter <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. Apical lateral and rhotic consonants in EP and Mandarin

Both Standard EP and Mandarin manifest a contrast between an apical lateral and a rhotic, whereas they differ in distribution and phonetic realization of the two segments.

2.1. European Portuguese

EP /l/ is traditionally described as exhibiting two allophonic variants, namely an alveolar lateral [l] in onset and a velarized [ɫ] in coda [8]. Acoustic and articulatory studies have challenged this allophonic alternation as they found that EP /l/ always bears a certain degree of velarization, regardless of position and adjacent context [9, 10, 11]. A recent study by Rodrigues et al. [12] reported that /l/ has indeed consistently low F2 values in EP, due to velarization across positions, however, its third formant (F3) values, another acoustic correlate of degree of velarization, are higher syllable-finally, justifying the assumption of two distinct allophones of /l/.

The Standard EP rhotic /ɾ/ can occur in all prosodic contexts apart from word-initial position [13]. Note that we are not concerned with dialectal trill realizations in the present article. Its most frequent realization is a tap, though other variants can also be found, depending on the adjacent segment and position [14]. In coda position, followed by a stop in the onset of the following syllable, it is mostly realized as tap plus epenthetic vowel, whereas in the same position with a following fricative, a realization as fricative is more prevalent. In word-final position, /ɾ/ is often produced as a voiceless fricative [15].

2.2. Mandarin

The Mandarin lateral /l/ is alveolar and its occurrence is restricted to single onset positions [16]. The Mandarin retroflex rhotic /ɻ/ can occur both in onset or coda. While in onset position it can vary between approximant and fricative realizations [17], syllable-finally it is always an approximant and resembles the rhotic in English, in as far as it varies between bunched and retroflex articulations [18].

3. Delayed imitation task

To explore whether CLI is responsible for the prosodic effect on L2 production of EP /l/ and /ɾ/, we tested how the EP liquids are categorized by naïve Mandarin listeners across prosodic contexts 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 [19], and L2 imitation was shown to be strongly related to L2 phonological categorization [20]. Accordingly, we reason that naïve imitators will only produce what they perceive and 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 e.g. in a graded rating perceptual similarity task.

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 the target segments in differing syllable positions. 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> → /ɻ/. For the two conditions we used, we reasoned that if the replacement of EP /ɾ/ with [ɻ] occurred in the auditory condition, it would provide evidence for perception-based CLI, while orthography-based CLI would be supported if the replacement with [ɻ] only appeared in the orthographic condition.

3.1. Speech material and recording procedures

Test materials comprised the Portuguese liquids /l/ and /ɾ/ in intervocalic onset and word-internal coda position. 16 pseudo-words were created, where target /l/ and /ɾ/ were always in a stressed syllable. Intervocalically, the liquid appeared between two /a/ (e.g., *pa/l/afa*, *pa/ɾ/afa*); syllable-finally, the target liquid followed the vowel /a/ and preceded either the voiceless stop /p/ or the voiceless 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. 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 [21].

3.2. Participants

23 L1-Mandarin listeners participated in 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). 19 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.

3.3. Procedure

A delayed imitation task was constructed using Microsoft PowerPoint. In the first part of the experiment, the 48 test items were presented 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 visual stimulus 请重复 'Please repeat' on a computer screen. The temporal interval between offset of the sound stimulus and onset of the visual cue was set to 1200 ms, with the purpose of encouraging phonological categorization rather than merely acoustic mimicry [22]. In the second part, the written form of each test item was presented on the screen along with its auditory form. The order of the two parts of the experiment was not counterbalanced between 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 words.

All auditory stimuli were presented to subjects via headphones at a comfortable listening level. Participants' imitations were recorded individually in a quiet room.

3.4. 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], and [ɻ] by having a low F3. 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.

All coding was performed by the first author, a Mandarin native speaker with near-native proficiency in EP, and checked by two trained Portuguese phoneticians. The annotations were then extracted and used to calculate the occurrence frequency of each segmental realization. Note that for three participants, the imitation results 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 5.

4. Results

4.1. Lateral

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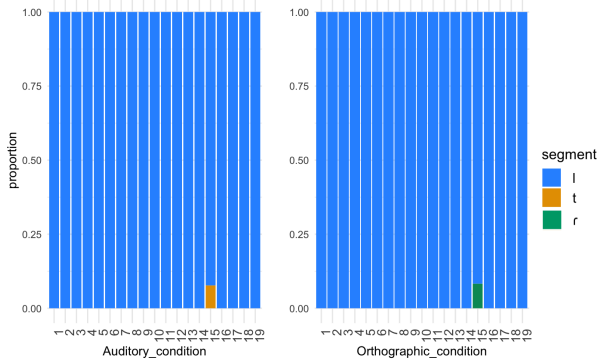
Figure 1: *Categorization of EP /l/ in onset position split by participants*

Figure 2 shows that in coda, the EP /l/ was most often identified as vocalized (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 by a non-velarized lateral (/l/ or /lə/), replacement by other segments (/t, s, x/, coded *other*), or deletion (coded as *0*).
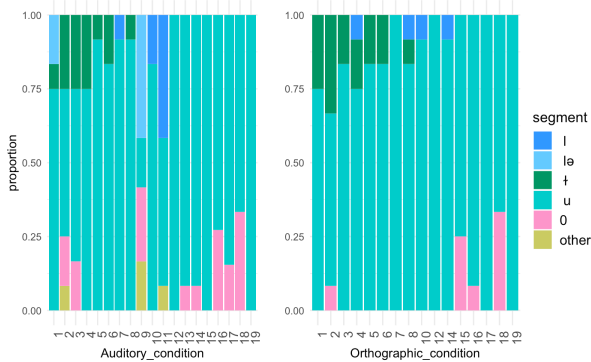

Figure 2: *Categorization of the EP /l/ in coda position split by participants.*

4.2. 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 /tʰ/ (M=0.19) in the auditory condition, cf. Figure 3 left. In the orthographic condition, the use of /l/ for target /ɾ/ was still prevailing (M=0.79), but here it was followed by /ɻ/ (M=0.12) and a coronal stop (/t/ or /tʰ/, M=0.05), cf. Figure 3 right.

Figure 3 also reveals that there was notable between-subject variation. In the auditory condition, some listeners constantly identified [ɾ] as 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 /ɻ/. Listeners 2 and 5, for instance, categorized [ɾ] solely as /ɻ/. As this only happened in the orthographic condition, we can conclude that it was only the orthographic cue that triggered their categorization as /ɻ/, and that participants 2 and 5 disregarded the auditory information in this condition completely.
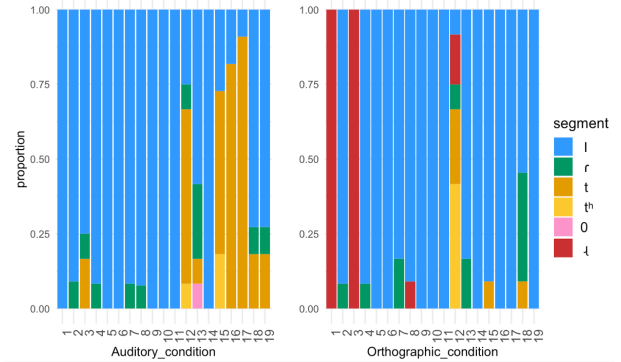

Figure 3: *Categorization of the EP /ɾ/ in onset.*

Syllable-final [ɾ] in the auditory condition, cf. Figure 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 coronal stop (/t/, /tə/ or /tʰ/, M=0.25) or some other segment (e.g. /s, ʂ/, coded as *other*, M < 0.04 each). In the orthographic condition on the left of Figure 4, post-vocalic [ɾ] was again 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).
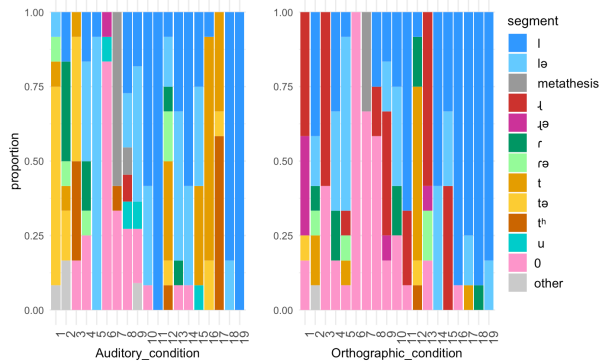

Figure 4: *Categorization of the EP /ɾ/ in coda.*

In Figure 4 we can see that categorization of syllable-final EP [ɾ] by naïve Mandarin listeners manifested 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). 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).

4.3. 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 [23] in R [24] on the imitation results by the listeners who produced [ɻ] in either auditory or orthographic condition (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 Mandarin rhotic is due to orthographic influence.

In addition, a visual inspection of Figure 3 and Figure 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 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, due to the fact that the naïve Mandarin listeners categorized EP [ɾ] as coronal stop to a lesser extent in the orthographic condition than in the auditory condition.

5. 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 [1, 2] could be accounted for by CLI. The results of our delayed imitation task with naïve Mandarin listeners largely replicated the modifications employed by L1-Mandarin learners, as shown in Table 1.

Table 1: *Imitation results by Mandarin naïve speakers and repair strategies by L1-Mandarin learners of EP*
(C[ə] = schwa epenthesis, ∅ = deletion)

	/l/$_\text{VCV}$	/l/$_\text{VC}$	/ɾ/$_\text{VCV}$	/ɾ/$_\text{VC}$
Naïve	[l]	[w]	[l]	[l], [t,tʰ], [ɻ], C[ə], ∅
Learners	[l]	[w]	[l]	[l], [t,d,tʰ], [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 [25].

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. 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.

In response to the second research question, 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 [26]. Adherents of traditional phonological accounts usually dismiss orthographic explanations by criticising that only some of the observed changes can be accounted for by orthography. 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 3 and 4). Learners thus may manifest a different weighting of auditory vs. orthographic cues [27]. 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 [28], presumably because syllable-final EP /ɾ/ displays larger allophonic variability [14] 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 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 [29]. In particular, listeners who categorized /ɾ/ as 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. If the orthographic form <r>, corresponding to a sonorant sound in Mandarin, is simultaneously presented, this seems to "hinder" the perception of an obstruent.

Since all our participants shared 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/, it is therefore 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 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, [7] showed that the mastery of [ɾ] in L2 English does not aid its acquisition in an L3.

A methodological limitation of using an imitation task to access L2-to-L1 category assimilation is that it measures production rather than the perception output one would like to directly tap into. 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.

6. Acknowledgements

7. References

[1] C. Zhou, *Contributo para o estudo da aquisição das consoantes líquidas do português europeu por aprendentes chineses*, MA dissertation, University of Lisbon, 2017.

[2] W. Liu, *Aquisição da Vibrante Simples [r] pelos Alunos Chineses Aprendentes de Português como Língua Estrangeira*, MA dissertation, University of Macau, 2018.

[3] R. C. Major, "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, 2008.

[4] M. Patience, "Acquisition of the tap-trill contrast by L1 Mandarin–L2 English–L3 Spanish Speakers," *Languages*, 3(4), 42, 2018.

[5] P. Howson, "Rhotics and palatalization: An acoustic examination of Upper and Lower Sorbian," *Phonetica*, 75, pp. 132–150, 2018.

[6] T. A. Hall, *The Phonology of Coronals*, Amsterdam: John Benjamins, 1997.

[7] P. Howson and P. Monahan, "Perceptual motivation for rhotics as a class," *Speech Communication*, 115, pp.15–28, 2019.

[8] M. H. M. Mateus and E. Andrade, *The Phonology of Portuguese*. Oxford: Oxford University Press, 2000.

[9] A. Andrade, "On /l/ velarization in European Portuguese," *Proceedings of ICPhS,* 1999, San Francisco, 543–546.

[10] P. Martins, C. Oliveira, A. Silva and A. Teixeira, "Articulatory characteristics of European Portuguese laterals: a 2D & 3D MRI study," *Proceedings of FALA* Vigo, Spain, pp. 33–36, 2010.

[11] C. Oliveira, P. Martins, A. Teixeira, I. Marques and P. Sá-Couto, "An articulatory and acoustic study of the European Portuguese /l/," *Proceedings of ICPhS*, 2011, Hong Kong, pp. 1538–1541.

[12] S. Rodrigues, F. Martins, S. Silva and L. Jesus, "/l/ velarisation as a continuum," *PLoS ONE*, 14(3), pp. 1–22, 2019.

[13] M. H. M. Mateus, I. Falé and M. J. Freitas, *Fonética e Fonologia do Português*. Lisboa: Universidade Aberta, 2005.

[14] A. Silva, *Análise Acústica da Vibrante Simples do Português Europeu*. MA dissertation, University of Aveiro, 2014.

[15] L. Jesus and C. Shadle, "Acoustic analysis of European Portuguese uvular [χ, ʁ] and voiceless tapped alveolar [ɾ̥] fricatives," *Journal of the International Phonetic Association*, 35(1), pp. 27–44, 2005.

[16] S. Duanmu, *The Phonology of Standard Chinese (2nd Edition)*. Oxford: Oxford University Press, 2007.

[17] X. Zhu, "近音—附论普通话日母," [About approximant], *Fangyan*, 1, pp. 2–9, 2007.

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

[19] M.E.H. Schouten, "Imitation of synthetic vowels by bilinguals," *Journal of Phonetics*, 5, pp. 273–283, 1977.

[20] M. Llompart and E. Reinisch, "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, 2019.

[21] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer. Version 6.1.14.*, 2020. Retrieved from www.praat.org

[22] P. Escudero, T. Benders and S.C. Lipski, "Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners," *Journal of Phonetics*, 37(4), pp. 452–465, 2009.

[23] D. Bates, M. Mächler, B. M. Bolker and S. C. Walker, "Fitting Linear mixed-effects models using lme4," *Journal of Statistical Software*, 67, pp. 1-48, 2015.

[24] R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

[25] Y. He, "Production of English syllable final /l/ by Mandarin Chinese speakers," *Journal of Language Teaching and Research*, 5(4), pp. 742–750, 2014.

[26] C. Paradis and D. LaCharité, "Category preservation and proximity versus phonetic approximation in loanword adaptation," *Linguistic Inquiry*, 36(2), pp. 223–258, 2005.

[27] A. Giannakopoulou, M. Uther and S. Ylinen, "Phonetic and orthographic cues are weighted in speech sound perception by second language speakers: Evidence from Greek speakers of English," *Proceedings of the Acoustical Society of America*, 20 (060011), 2016.

[28] C. Best and M. Tyler, "Nonnative and second language speech perception: Commonalities and complementarities," In O. Bohn, and M. Munro (Eds), *Language Experience in Second Language Speech Learning – In Honor of James Emil Flege*, pp.13–34, Amsterdam/Philadelphia: John Benjamins Publishing Company, 2007.

[29] G. McGuire, "Orthographic effects on phonetic cue weighting", *Proceedings of the 15th Australasian International Conference on Speech Science and Technology*, pp. 201–204, 2014.