Zhou, Chao & Hamann, Silke. 2024. Modelling the acquisition of the Portuguese tap by L1-Mandarin learners: A BiPhon-HG account for individual differences, syllable-position effects and orthographic influences in L2 speech. *Glossa: a journal of general linguistics* 9(1). pp. 1–39. DOI: https://doi.org/10.16995/glossa.9692

Modelling the acquisition of the Portuguese tap by L1-Mandarin learners: A BiPhon-HG account for individual differences, syllable-position effects and orthographic influences in L2 speech

Chao Zhou, School of Arts and Humanities, Center of Linguistics, University of Lisbon, PT, zhouchao@edu.ulisboa.pt

Silke Hamann, Amsterdam Center for Language and Communication, University of Amsterdam, NL, s.r.hamann@uva.nl

The present study provides a formal account for three types of experimental findings recurrently reported in the literature, yet not explicitly formalised in current L2 speech theories, namely individual variability, syllable-position effects and orthographic influences. By analysing examples from L2 acquisition of Portuguese, we show that all these L2 speech patterns can be formalised within a single generative linguistic model, the Harmonic-Grammar version of the *Bidirectional Phonology and Phonetics Model,* which was originally proposed for L1 speech perception and production. Without resorting to any mechanism specific to L2 acquisition, our formal modelling suggests that L2 speech data can be addressed in a broader context, benefiting from well-developed formal phonological theories.

1 Introduction

Speech models of second language (henceforth: L2) acquisition seek to account for phenomena observed over the course of acquiring a non-native sound system. Decades of studies have shown that, when acquiring an L2, learners tend to rely on the knowledge of their first language(s) (henceforth: L1). It is not surprising then that many L2 speech theories have been developed to model the interaction between the learners' L1 and L2, which is also referred to as cross-linguistic influence (CLI). Three such theories are most prominent in the literature: the *Speech Learning Model* (SLM; Flege 1995; Flege & Bohn 2021), the *Perceptual Assimilation Model-L2* (PAM-L2, Best & Tyler 2007) and the *Second Language Linguistic Perception* model (L2LP, Escudero & Boersma 2004; Escudero 2005). All three converge on the idea that the learner's L1 acts as a perceptual sieve, modulating how the L2 speech input is parsed, and all three formulate CLI as a function of L2-to-L1 sound category mapping. CLI is predicted, for instance, if an L2 sound is essentially different from any existing segment in the learner's L1, yet similar enough (in terms of phonetic and/or phonological characteristics) to be considered a good token of its closest L1 counterpart (the "similar" scenario in the SLM; the "single-category assimilation" pattern in the PAM-L2; the "new" scenario in the L2LP).

The adequacy of formalising CLI as perceptually driven L2-to-L1 category assimilation has been evidenced in an extensive amount of studies examining different L1-L2 pairs (see, e.g., Colantoni et al. 2015; Bohn 2017, for overviews). Yet not all observations in L2 speech can be accounted for by CLI. For instance, Colantoni & Steele (2008) reported that in L2 acquisition of French and Spanish rhotics, L1 speakers of English target the intervocalic onset before the word-internal coda position, which can be explained by phonetic complexity (in word-internal coda, it is necessary to master consonant-consonant coarticulation) but not by CLI. Another piece of evidence comes from the acquisition of the English /i/-/ɪ/ contrast. It has been shown that learners with different L1 backgrounds such as Spanish, Portuguese, Russian and Mandarin rely mainly on duration to perceptually distinguish the two categories (see Bohn 2020 for review), despite the fact the learners' L1s do not use duration to contrast vowels and that the L2 English contrast /i/-/ɪ/ is only marginally cued by vowel quantity. Their reliance can thus hardly be ascribed to CLI. These are merely two instances of many L2 speech phenomena that the models focussing on CLI cannot account for.

In addition, category assimilations have been shown to be not only driven by category-internal factors. They are also influenced by L1-specific restrictions on the position of the category in the syllable and the word, and its possible cooccurrences with other categories, i.e., language-specific phonotactics. The documentation of L1 phonotactic influence on L2 speech perception can be at least dated back to Polivanov (1931) and has been recurrently reported in empirical studies (see e.g. Dupoux et al. 1999; Kabak & Idsardi 2007; Cardoso 2011; Zhou & Rato 2023). It is thus of great importance to incorporate such cross-linguistic influences beyond the segmental level into theorisation to better understand and predict L2 speech patterns.

In this article, we aim to narrow the gap between L2 speech theories and experimental evidence by formalising three patterns attested in the acquisition of the European Portuguese (henceforth: Portuguese) tap by Mandarin L1 speakers (Zhou 2017; Liu 2018; Zhou & Hamann 2020). These are i) between- and within-speaker variability, ii) a syllable-position effect and iii) an orthographic influence. Those L2 speech phenomena were chosen because they have been reported in a growing number of studies (see Colantoni et al. 2015 for review) and, although accumulated empirical evidence of this kind has called for an expansion of existing L2 models, such attempts are still rare, to the best of our knowledge.

For our formalisation, we employ the *Bidirectional Phonology and Phonetics Model* (henceforth, BiPhon; Boersma 2007, 2011; Boersma & Hamann 2009a) with its associated reading grammar (Hamann & Colombo 2017). We will demonstrate that, as a comprehensive linguistic model, BiPhon constitutes a promising tool for handling a wide range of phenomena in L2 speech learning.

The article is structured as follows. Section 2 summarises the relevant data on the Mandarin L1 speakers' acquisition of Portuguese /ɾ/ across syllable contexts. In Section 3, we introduce the BiPhon model that is adopted in the current study. Section 4 presents the formalisation of the three phenomena observed in the L2 acquisition of Portuguese /ɾ/. In Section 5, we offer some conclusions and point out directions for future research.

2 The data: L2 speech learning of Portuguese /ɾ/ by L1-Mandarin learners

In this section, we summarise previous experimental findings on the acquisition of Portuguese /ɾ/ by Mandarin L1 speakers. We start with a brief introduction to the phonetics and phonology of the apical rhotics in the two languages that are involved.

The Portuguese rhotic /ɾ/ can occur in all contexts except in word-initial position (Mateus et al. 2005; Zhou & Jesus 2022). Its most frequent realisation is a tap, though other variants can be also found, depending on the adjacent segment and the syllabic position of the rhotic. For instance, in coda position followed by a stop in the onset of the following syllable, the rhotic is mostly realised as a tap with a following vocoid, whereas in the same position preceding a fricative, a fricative realisation is found (Silva 2014). Word-finally, /ɾ/ is often produced as a voiceless fricative (Jesus & Shadle 2005) and can even be omitted, especially when the following word starts with a consonant (Mateus & Rodrigues 2003; Rodrigues 2003).

The Mandarin rhotic partially resembles the distribution of its Portuguese counterpart by occurring both in syllable onset and coda (but not in onset clusters), and its syllable-dependent phonetic realisation. While in onset position the Mandarin rhotic can vary between an approximant and a fricative (Zhu 2007; Chen & Mok 2019), in syllable-final position it is always an approximant (Chen & Mok 2019; Jiang et al. 2019). At the underlying phonological level, the

Mandarin rhotic is assumed to be an approximant /ɻ/ in most studies (e.g., Chao 1968; Duanmu 2005; Lin 2007), but has also been analysed as an underlying obstruent by some researchers (e.g., Shi 2004; Duanmu 2007) due to its allophonic variation as a fricative in syllable onset. In the present article, we assume the former for two reasons (following Duanmu 2007 and Hall & Hamann 2010): the frication that the Mandarin rhotic bears is weaker than that of a canonical fricative (Zhu 2007), and a fricative analysis would make it the only voiced obstruent in the Mandarin phonological inventory.

When acquiring the Portuguese tap, L1-Mandarin learners very often replace it with an alveolar lateral /l/, which is arguably one of the most perceptible characteristics of Chinese-accented Portuguese (e.g. Batalha 1995; Martins 2008). Recent experimental studies have suggested that this notorious L2 speech learning difficulty goes beyond the confusability between /ɾ/ and /l/ and is constrained by the syllabic position of the tap. Zhou (2017) observed that in a picture-naming task, 14 L1-Mandarin learners with relatively homogenous experience of Portuguese (2-year formal instruction plus 3-month immersion in Portugal) produced more target-like instantiations of /ɾ/ in coda (M = 0.69) than in onset (M = 0.39). When failing to produce [ɾ], learners used [l] exclusively in onset, whereas in coda they deleted the segment, inserted a schwa (and thus created a new onset), or replaced the segment with [l] or [ɻ]. This onset-coda asymmetry was replicated in Liu (2018), using both a picture-naming and a text reading task. It is worth noting that, in addition to [l] or [ɻ], the 15 participants in Liu (2018), who were all first-year college students majoring in Portuguese, also replaced the syllable-final /ɾ/ with a coronal stop [d, t, tʰ] or a nasal [n].

In order to examine whether the above syllable-dependent repair strategies are prompted by CLI (the perceptually driven L2-to-L1 category assimilation), as predicted by the major L2 speech theories introduced in Section 1, Zhou & Hamann (2020) performed a categorisation (delayed-imitation) task with 19 naïve Mandarin listeners that did not have any knowledge of Portuguese. It was reasoned that naïve imitators would produce the L1 category to which they assimilate the L2 input. A further point tested in the experiment was the influence of orthography. Half of the test items were therefore presented auditorily only (auditory condition), the other half with an accompanying written form (orthographic condition, which followed the auditory condition). This was done because Zhou & Hamann speculated that one of the deviant productions was unlikely to be perceptually motivated, namely the use of Mandarin [ɻ] (with low third formants; cf. Smith 2010) for Portuguese /ɾ/ (with high third formants, cf. Rodrigues 2015). As both sounds are represented with the letter <r>[1], orthography might be responsible for this substitution. The

[1] In Mandarin, <r> is used in the alphabetic script Pinyin, a romanization system for Standard Mandarin Chinese. Pinyin is generally used to help children and non-native speakers associate pronunciation with Chinese characters. Apart from being a learning tool, Pinyin also plays an important role in Mandarin speakers' daily communication, since it has become the dominant method for entering Chinese text into computers and smartphones.

prediction on L2-to-L1 category assimilation was borne out as the responses by the Mandarin naïve imitators resembled those by L2 learners of Portuguese: the imitators employed the same position-dependent repairs for the Portuguese tap, namely [l] for /ɾ/ in syllable onset, and both segmental replacement ([l], [d/t/tʰ] and [ɻ]) and structural modification (epenthesis and deletion) in coda. In spite of this relatively good predictive power of CLI, there were several findings in Zhou & Hamann's study that cannot be ascribed to perceptually driven L2-to-L1 category assimilation.

First, a considerable variability was attested in the L2 categorisation of the Portuguese tap. **Figure 1** gives an overview of the variation: each bar represents the categorisation responses of one subject. On the basis of these patterns, the 19 Mandarin naïve listeners can be classified as two types: Type I systematically identifies the EP tap as /l/ (listeners 1, 2, 4, 5, 6, 7, 8, 9, 10, 11 and 14), while Type II maps [ɾ] onto /l/ or an alveolar stop (/t/ or /tʰ/) (listeners 3, 12, 13, 15, 16, 17, 18 and 19).[2]

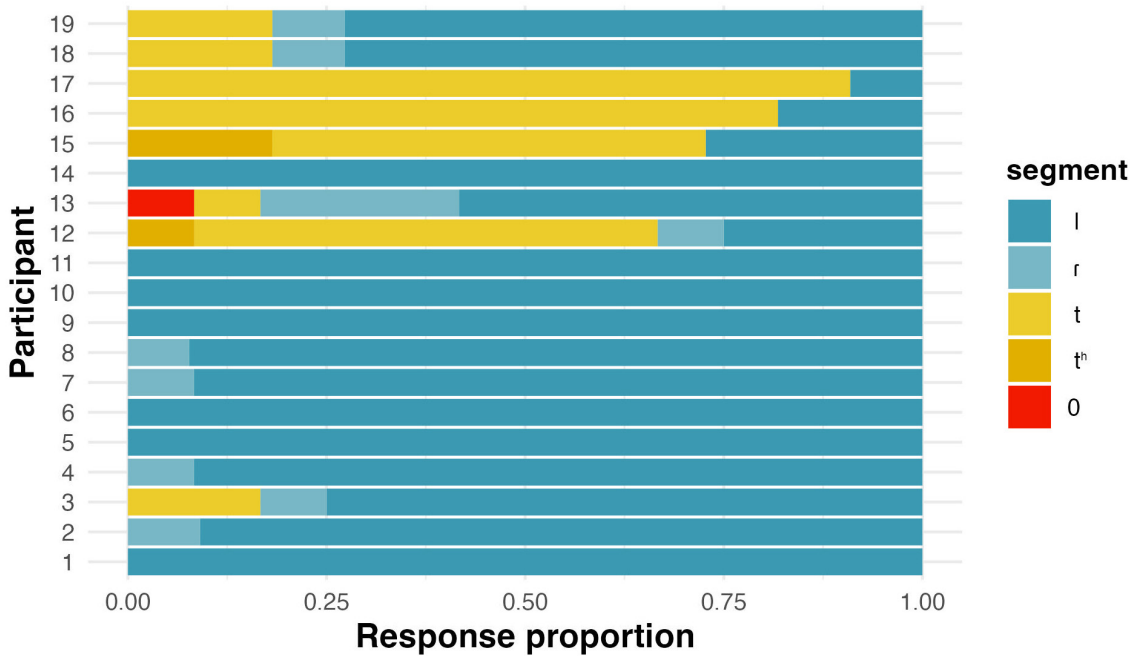

Figure 1: Categorisation of the Portuguese tap in syllable onset position by 19 naïve Mandarin speakers (based on Zhou & Hamann 2020: 4488), 0 = deletion.

Mapping an L2 sound onto multiple L1 categories has been mentioned by Escudero and Boersma (2002) as Multiple Category Assimilation (and has been part of L2LP since then), while in PAM-L2 it is the "uncategorised" category assimilation. It has been reported in prior research,

[2] We ignore here the occasional correct categorisations as tap (38 productions in total, i.e. 4.24%).

e.g., on Australian English vowels categorised by Egyptian Arabic speakers (Faris et al. 2016). However, it remains unclear what underlies the learners' choice of more than one L1 category and how such intra- and inter-speaker variability can be accounted for.

Second, apart from the segmental replacement, changes involving the syllabic structure of the input such as epenthesis and deletion were also employed to accommodate the syllable-final /ɾ/. This interaction between segmental material and syllable position has been long attested both in L2 speech perception and production studies (e.g. Dupoux et al. 1999; Cardoso 2011; Cabrelli et al. 2019), yet has not been incorporated into L2 speech models.

Finally, the use of Mandarin [ɻ] was restricted to the condition when written input was presented (though this orthographic effect on L2 phonological categorisation was restricted to some listeners). Interactions of phonology and orthography in L2 speech have attracted increasing attention from L2 researchers (see Bassetti et al. 2015 and Hayes-Harb & Barrios 2021 for reviews). And although it has been demonstrated that orthography can aid, hinder, or have no effect on L2 phonological acquisition, this type of cross-modal interaction has not been incorporated into existing L2 speech theories.

In the following section, we introduce a grammar model that is able to account for all three aforementioned aspects.

3 The model: BiPhon-HG and its associated reading grammar

The BiPhon model, proposed by Boersma (2007, 2011), adopts a modular view of phonetics and phonology.[3] The phonological module of BiPhon is composed of two discrete phonological representations, an Underlying Form (UF; the phonological form stored together with the meaning in the lexicon; given in pipes) and a Surface Form (SF; the prosodically detailed representation containing features, segments, syllables and larger prosodic constituents; given in slashes). The phonetic module consists of two phonetic representations, an Auditory Form (AudF; a continuous representation of speech sound consisting of auditory events such as noise, pitch, spectrum and duration; given in single square brackets) and an Articulatory Form (ArtF; a continuous representation of the articulatory gestures, e.g. tongue and lip movements, jaw depression; given in double square brackets). **Figure 2** illustrates how these representations are organised in BiPhon.

[3] For the advantages of BiPhon over a non-modular model in which phonetic and phonological representations are commensurable, interested readers are referred to Hamann (2011).

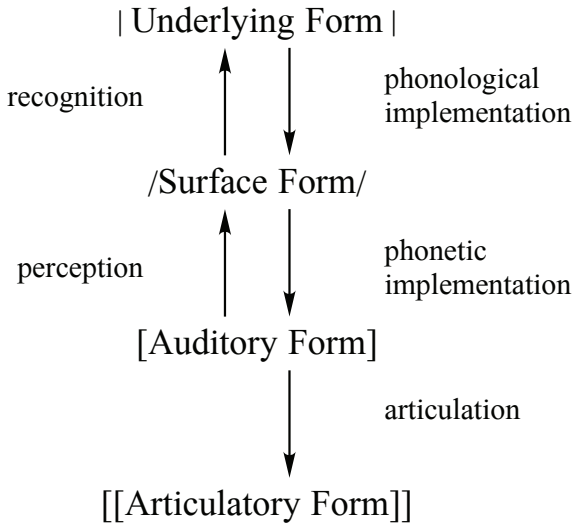

Figure 2: The BiPhon Model (based on Boersma 2007, 2011).

In line with the literature in psycholinguistics (McQueen & Cutler 1997; Ramus et al. 2010), L1 phonological acquisition (Fikkert 2010) and L2 speech learning (Escudero & Boersma 2004; Darcy et al. 2013; Flege & Bohn 2021), BiPhon advocates that speech comprehension is not a single mapping but consists of at least two processes. Starting with AudF, listeners map the speech-relevant auditory information (e.g. pitch, spectra, silences, transitions and durations) onto discrete phonological units, i.e., the SF.[4] This mapping is designated in the literature as prelexical perception (McQueen & Cutler 1997) or phonological categorisation (Escudero & Boersma 2004); these two terms will be used interchangeably henceforward. With the perceived SF, listeners access the intended lexical meaning by matching it to the phonological representation stored in the long-term memory, the UF, which is linked to the semantic representation. The mapping between the two phonological forms is referred to as word/lexical recognition. Note that these two processes do not need to proceed sequentially but can apply in parallel (see Boersma 2011 for a discussion).

In the production direction (the right side of **Figure 2**), BiPhon displays an analogous modularity, which is compatible with psycholinguistic models of speech production (e.g. Levelt

[4] 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). BiPhon assumes that SF refers to a hierarchically organised tree-like structure of abstract phonological elements such as features, segments, syllables, and other prosodic constituents. In principle, all these units can be targeted in speech perception, which presumably hinges on the specific experimental task, see Samuel (2020) for a discussion.

1989) and traditional generative models of phonology (Chomsky & Halle 1968). In order to produce a meaningful word, speakers retrieve the UF of an intended lexical entry from the mental lexicon and translate it into a prosodically detailed SF. This fully specified phonological representation is then converted into a phonetic form. In BiPhon, the SF is assumed to map onto an AudF, which is transformed by sensorimotor knowledge to an ArtF.

In contrast to many psycholinguistic models of speech comprehension and/or production that consist of box-and-arrow graphs (see e.g. Levelt 1989; McQueen & Cutler 1997; Ramus et al. 2010; Darcy et al. 2013), BiPhon is a linguistic model that provides a formalisation of the knowledge that listeners and speakers have, making explicit language-specific mappings between different representational levels, and providing a learning algorithm for the acquisition of these mappings.

In the rest of this section, we will present several crucial features of BiPhon, on the basis of which we will argue that it is the most suitable model for formalising the L2 speech phenomena addressed by the current study.

3.1 The grammatical computation in BiPhon is probabilistic

In BiPhon, the mappings between representations can be modelled with ranked constraints in Optimality Theory (Prince & Smolensky 1993; henceforth: OT), with weighted constraints in Harmonic Grammar (Legendre et al. 1990; henceforth: HG) or with weighted connections in Neural Networks (Boersma et al. 2020; henceforth: NN). In the current study, we adopt the HG version of BiPhon (henceforth: BiPhon-HG).

HG represents a phonological grammar by a set of violable constraints, just like OT. However, instead of being ranked with respect to each other, constraints in HG are assigned with numerical weights reflecting their relative strength. At evaluation time, the HG grammar yields a Harmony score (H) for each candidate, which is the sum of all constraint violations or satisfactions multiplied with the corresponding constraint weight. The winning candidate is the one that receives the highest H.

Many HG accounts employ constraint violations, just like OT, where violations are penalised, resulting in negative harmony scores with a maximum of 0 (see, e.g., Jesney & Tessier 2007: 4 for elaboration). In the present study, we follow Boersma & Pater (2007) and Kimper (2016) in assigning candidates positive scores for constraint satisfaction, resulting in positive H scores. This departs drastically from OT's decision mechanism by negative exclusion, which requires a large number of negative constraints when e.g. the mapping of gradual phonetic realisations onto phonological SF categories (or vice versa) is considered (cf. Cue constraints in Section 3.2; for an illustration of their working in BiPhon-OT, see e.g. Boersma & Hamann 2008). Apart from this formal consideration, constraint satisfaction seems also to be more in line with the almost

exclusively positive evidence that language learners encounter in the acquisition process, while constraint violation either has to assume that all negative constraints are innate (Gnanadesikan 2004) or requires an extra mechanism that allows negative constraints to be acquired through positive evidence (Ingram 1995). These two advantages of positively formulated constraints that can be satisfied rather than violated plus the ganging-up effect to be elaborated below are the reasons why we employ the HG version of BiPhon in the present article.

Note that BiPhon assumes bidirectionality (Smolensky 1996), i.e., the same set of constraints and constraint rankings/weights are used in perception and production. In this paper, we focus on the perception direction, though a short illustration is given in Section 5.2 of how the constraints that were introduced for perception are used in production. For further bidirectional use of constraints in BiPhon, interested readers are referred to Boersma and Hamann (2008, 2009b).

The workings of a BiPhon-HG tableau are illustrated in (1). Here and in the following tableaux, constraint weights are shown above the constraint names, and the harmony H for each candidate in the last column. Note that we indicate satisfaction of the positively formulated constraints by "+1", in contrast to the common notation of "1" that indicates violation of usually negatively formulated constraints (both in OT and HG).

(1)

	1.0	0.7	*H*
Input	Constraint A	Constraint B	
☞ Candidate 1	+1		1.0
Candidate 2		+1	0.7

In Tableau (1), Candidate 1 is selected by the HG grammar as optimal, because it receives a higher harmony than its opponent, Candidate 2, which satisfies a constraint with lower weight.

This decision-making mechanism, i.e., adding up all scores across constraints, gives rise to another major divergence between HG and OT: the ganging-up effect, illustrated in Tableau (2). If a candidate (Candidate 1 in Tableau 2) satisfies two constraints (here B and C) that both have lower weight than a third constraint (A), then these two constraint satisfactions together result in a higher harmony than the harmony of a candidate (Candidate 2) satisfying only the constraint with the highest weight (Constraint A). We thus see that two constraints with lower weight may gang up to overcome a third one with higher weight. This is not possible in OT with its strict ranking, where the violation of the highest ranked constraint is decisive.[5]

[5] The constraint scores chosen in this paper are arbitrary, since it is the relative weighting between constraints that determines the winning candidate. To calculate the relative constraint weights in a given grammar that allows a certain candidate to be selected as the optimum, one may use error vectors, see Boersma and Pater (2016) for discussion.

(2)

	1.0	0.7	0.5	*H*
Input	Constraint A	Constraint B	Constraint C	
☞ Candidate 1		+1	+1	1.2
Candidate 2	+1			1

Patterns found in L2 speech (learning) normally show a considerable degree of variability (e.g., Hamann 2009; Cardoso 2011; Zhou & Hamann 2020). To account for this variability and for L2 speech phenomena in a more realistic way, compared to other existing L2 speech models, we apply here stochastic BiPhon-HG (Noisy HG; Boersma & Pater 2016). Much like Stochastic OT (Boersma 1998), Noisy HG is made probabilistic by assuming that a random noise value (e.g., transmission/background noise; drawn from a normal distribution with a mean of 0 and a standard deviation of 1) is temporarily added to each constraint weight at each evaluation time. Accordingly, the computation of harmony is influenced by the noise values added to each weight, leading to the selection of different winners across instances of evaluation. This variation prompted by noisy evaluation is illustrated in (3) and (4). The basic weights of Constraint A and B are 1.0 and 0.7 respectively, given in brackets. The weights after the addition of noise are shown below those values, according to which either Candidate 1 (in Tableau 3) or Candidate 2 (in Tableau 4) wins. Such a change in winning candidates is only possible when the involved constraints have weights that are close.

(3)

	(1.0) 0.87	(0.7) 0.89	*H*
Input	Constraint A	Constraint B	
☞ Candidate 1		+1	0.89
Candidate 2	+1		0.87

(4)

	(1.0) 1.04	(0.7) 0.81	*H*
Input	Constraint A	Constraint B	
Candidate 1		+1	0.81
☞ Candidate 2	+1		1.04

3.2 Phonological categorisation as interaction between cue knowledge and structural restrictions

As reviewed in Section 1, converging lines of evidence have suggested that CLI in L2 speech is very often ascribed to a novel sound being perceptually categorised as an L1 category. In other

words, the locus of CLI lies in phonological categorisation, i.e., the mapping from AudF (L2 sound) to SF (L1 category) in **Figure 1**. This mapping is formalised in BiPhon with Cue and Structural constraints. Cue constraints (Escudero & Boersma 2004; Boersma 2009) represent the knowledge of how a certain auditory event is mapped onto abstract phonological categories in each language. In BiPhon-HG, Cue constraints are formulated as follows.

(5) *Cue constraints*[6]
"A value *x* on the auditory continuum [*y*] perceived as the phonological category /z/ receives a score of *n*."

For an example of relevance to the current study, consider how the Portuguese sounds [ɾ] and [l] are categorised by L1 Portuguese listeners. These two Portuguese liquids differ from each other in terms of both spectral (formant values and formant transitions) and durational dimensions (Rodrigues 2015). For simplicity, the auditory events used as input here are restricted to the F3 formant values and to durational values, which are typically 2542 Hz and 33 ms for /ɾ/, and 2692 Hz and 92 ms for /l/ (cf. Rodrigues 2015). The Cue constraints for perceiving the F3 values of 2542 Hz are given in (6a), those for perceiving F3 values of 2692 Hz in (6b), those for perceiving durational values of 33 ms in (6c) and those for perceiving durational values of 92 ms in (6d).

(6)

a)	[2542 Hz] /ɾ/	This constraint is satisfied if an F3 value of [2542 Hz] is mapped onto the phonological category /ɾ/.
	[2542 Hz] /l/	This constraint is satisfied if an F3 value of [2542 Hz] is mapped onto the phonological category /l/.
b)	[2692 Hz] /ɾ/	This constraint is satisfied if an F3 value of [2692 Hz] is mapped onto the phonological category /ɾ/.
	[2692 Hz] /l/	This constraint is satisfied if an F3 value of [2692 Hz] is mapped onto the phonological category /l/.
c)	[33ms] /ɾ/	This constraint is satisfied if a durational value of [33ms] is mapped onto the phonological category /ɾ/.
	[33ms] /l/	This constraint is satisfied if a durational value of [33ms] is mapped onto the phonological category /l/.
d)	[92ms] /ɾ/	This constraint is satisfied if a durational value of [92ms] is mapped onto the phonological category /ɾ/.
	[92ms] /l/	This constraint is satisfied if a durational value of [92ms] is mapped onto the phonological category /l/.

In a more realistic situation, Cue constraints as in (6) exist for every possible occurring value of *x*, making sure that a given phonological category /z/ connects to all occurring auditory events. Constraint weighing determines the likelihood of mapping a given *x* value to the category /z/.

[6] Due to cue integration (the use of more than a single auditory continuum), Cue constraints have to be negatively formulated in BiPhon-OT, see Boersma & Escudero (2008: 296–7) for a detailed discussion and simulation. Due to the ganging-up effect, a negative formulation is not necessary in HG.

To give an example, the Cue constraint responsible for mapping the prototypical value of a given category will outweigh the constraint targeting a peripheral value.

Considering the constraints in (6a), since an F3 of 2542 Hz is a typical formant value of Portuguese /ɾ/, the Cue constraint that links it to the SF /ɾ/ has greater weight than the one that links it to /l/, see Tableaux (7) and (8). The inverse situation is true for the Cue constraints relevant for 2692 Hz in (6b), which is a typical F3 value for Portuguese /l/. The same logic applies to the Cue constraints in (6c) and (6d).

For reasons of space, only Cue constraints relevant for categorising the Portuguese liquids in question are used here, and only prototypical values in Portuguese and Mandarin on the relevant AudF are considered. These Cue constraints in (6) are sufficient to model the correct perception of prototypical F3 values of Portuguese /ɾ/ and /l/:

(7) *Portuguese* $[ɾ]_{Aud}$ *categorised by the Portuguese perception grammar as* /ɾ/

	1.0	1.0	1.0	1.0	0.6	0.6	0.5	0.5	*H*
[2542Hz, 33ms]	[2692Hz] /l/	[2542Hz] /ɾ/	[33ms] /ɾ/	[92ms] /l/	[2542Hz] /l/	[33ms] /l/	[2692Hz] /ɾ/	[92ms] /ɾ/	
/l/					+1	+1			1.2
☞ /ɾ/		+1	+1						2.0

(8) *Portuguese* $[l]_{Aud}$ *categorised by the Portuguese perception grammar as* /l/

	1.0	1.0	1.0	1.0	0.6	0.6	0.5	0.5	*H*
[2692Hz, 92ms]	[2692Hz] /l/	[2542Hz] /ɾ/	[33ms] /ɾ/	[92ms] /l/	[2542Hz] /l/	[33ms] /l/	[2692Hz] /ɾ/	[92ms] /ɾ/	
/l/	+1			+1					2.0
☞ /ɾ/							+1	+1	1.0

In the tableaux above, we see how the two Portuguese alveolar liquids with prototypical acoustic values are categorised by L1-Portuguese listeners. This kind of optimal listening scenario, however, does not always occur. Due to co-articulation, speakers' physiological differences, speech style and many other factors, speech input shows notable variability. For instance, it is possible that a Portuguese speaker will produce an F3 value for the tap that overlaps with another speaker's F3 for the lateral, let us say, 2617 Hz. Still, most of the time, a Portuguese listener will be able to correctly identify the intended segment in these speakers' productions,

because the other cue, i.e. segmental duration, will also inform their perception.[7] The fact that a listener may make advantage of multiple acoustic cues for a given phonological contrast has been long known to the speech perception literature as cue integration, and BiPhon formalises it with Cue constraints straightforwardly (e.g. Boersma 2009). As illustrated in (9), where only constraints relevant for perceiving the input auditory events are shown, the two Cue constraints relevant for perceiving F3 values cannot determine the winner (both have the same constraint weight), as an F3 value of 2617 Hz (an intended /ɾ/ by the speaker) is compatible with both Portuguese liquid categories. The accurate categorisation of /ɾ/ then relies on the Cue constraints targeting the segmental duration. It is important to note that all Cue constraints used in (7), (8) and (9) make part of a single Portuguese perception grammar. For the sake of space, each tableau only includes the relevant Cue constraints for categorising the auditory input and can thus be viewed as a fraction of the whole perception grammar.

(9) *Portuguese* $[ɾ]_{\text{Aud}}$ *with a peripheral formant value but prototypical duration value categorised by the Portuguese perception grammar as /ɾ/*

	1.0	0.5	0.2	0.2	*H*
[2617Hz, 33ms]	[33ms] / ɾ /	[33ms] /l/	[2617Hz] /l/	[2617Hz] /ɾ/	
/l/		+1	+1		0.7
☞ /ɾ/	+1			+1	1.2

Apart from Cue constraints, speech perception is also prone to structural restrictions (i.e., phonotactics), which are formalised in BiPhon-HG as Structural constraints, cf. (10).

(10) *Structural constraints*
"The satisfaction of the structural requirement */x/* receives a score of *n.*"

An example for a Structural constraint of Portuguese is given in (11), where "#" denotes a word boundary. This constraint reflects the fact that only the fricative rhotic /ʁ/ but not the tap /ɾ/ can occur in word-initial position:

(11) */#ʁ/* This constraint is satisfied if a fricative rhotic is occurring in word-initial position.

[7] Real listeners of course perform speaker normalisation when they encounter productions of different speakers (for an overview, see Johnson & Sjerps 2021), which reduces the problem of overlap in the realisation of contrasting categories across speakers. We assume that some kind of normalisation has been performed before the process of speech perception that we model here (see e.g. Escudero & Bion 2007 for a formalisation thereof).

In the perception of L1 speech, Cue and Structural constraints are not conflicting (they are not favouring different output candidates). A conflict is, however, possible for non-native input as is the case in naïve L2 perception.

An instance of such naïve L2 perception by Portuguese listeners reflected in loanword adaptation is the treatment of the English approximant [ɹ]. Even though the English approximant is acoustically closer to the Portuguese tap /ɾ/ (both have clear formant structure), an English word-initial [ɹ] is nativized as a posterior fricative /ʁ/ (with weak formant structure but clear frication noise), conforming to the Portuguese Structural constraint in (11). This process is illustrated in (12), where the input is the auditory form of English [ɹ], strongly simplified here as [formant], which is in line with the Cue constraint "[formant] /ɾ/", as both the English approximant and the Portuguese tap manifest clear formant structures.

(12) *English* $[ɹ]_{\text{Aud}}$ *nativized by the Portuguese perception grammar as word-initial /#ʁ/*

	1.0	0.8	0.6	*H*
[formant]	/#ʁ/	[frication] /ʁ/	[formant] /ɾ/	
☞ /#ʁ/	+1			1.0
/#ɾ/			+1	0.6

Since the Structural constraint representing the Portuguese phonotactic restriction is weighted higher than the relevant Cue constraints, an English word like *rugby* with word-initial [ɹ] is adapted as *râguebi* with initial /ʁ/ rather than /ɾ/. For word-medial English [ɹ], the phonotactic restriction obviously does not apply, hence an adaptation with a tap is predicted.

3.3 Orthographic influence on phonological categorisation as interaction between perception and reading grammar

Another important feature that makes BiPhon a good model to account for L2 speech patterns is its inclusion of the orthographic form (Hamann & Colombo 2017, Hamann 2020). Although the mutual influence between phonology and orthography has been long acknowledged in experimental research (including L2 research, see Bassetti et al. 2015 and Hayes-Harb & Barrios 2021 for overviews), orthography has been underexplored in formal phonology and in L2 speech models. BiPhon fills this gap by proposing a reading grammar, which interacts with the speech perception and production grammar as illustrated in **Figure 3**.

The reading grammar (which can also be used to model writing due to the bidirectionality of the model) consists of Orthographic constraints that represent the phoneme-grapheme conversion

in each language, and the already introduced Structural constraints. The general formulation of Orth(ographic) constraints is given in (13).

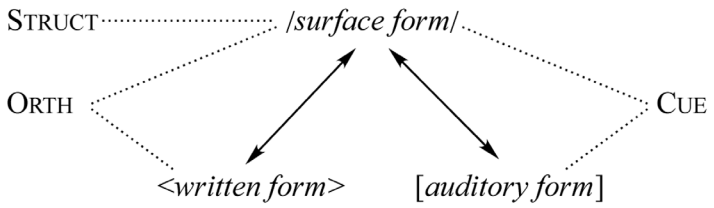

Figure 3: The BiPhon perception/phonetic implementation part (on the right) with the extension of the reading grammar (on the left; Hamann & Colombo 2017: 701).

(13) *Orth(ographic) constraints*
"A letter <x> that is mapped onto an SF /y/ receives a score of *n*."

A concrete example for an Orth constraint in Portuguese is given in (14), accounting for the fact that the letter <h> is silent in the Portuguese orthography.

(14) <h> / / This constraint is satisfied if a letter <h> (not proceeded by another consonantal letter)[8] is mapped onto nothing in the phonological surface form.

Orth constraints are acting together with Cue and Structural constraints in selecting an SF in phonological categorisation when both a written and an auditory form are provided at the same time. An example for this interaction can be found again in loanword phonology. In the adaptation of English words into Portuguese, the English word-initial glottal fricative [h] is systematically deleted, as e.g., in the English word *hardware* which is realised in Portuguese without [h]. Given that [h] is a dialectal variant of /ʁ/ in Portuguese (Pereira 2020) and it is the most prevailing form of the word-initial rhotic in Brazilian Portuguese (Rennicke 2015), it seems untenable to assume that its deletion is perceptually driven (see Sousa 2017 for preliminary perceptual results). If that were the case, L1 listeners of Portuguese would experience difficulties in detecting the /ʁ/ in the speech of many lusophones. We therefore assume that the deletion of /h/ stems from orthographic influence, as formalised in (15).

[8] Together with a preceding consonantal letter, as is the cases in <nh>, <lh> and <ch>, the <h> represents a palatal place of articulation of the consonant in Portuguese, that is /ɲ/, /ʎ/ and /ʃ/, respectively, and would be covered in a detailed orthographic account by separate Orth constraints referring to digraphs (such as <nh> /ɲ/, <lh> /ʎ/, and <ch> /ʃ/) with higher weights than the constraint in (14) (see e.g. Hamann 2020 for an example how Orth constraints referring to more than one grapheme interact with those referring to a single grapheme). As the present analysis cannot provide a complete formalisation of Portuguese orthography, the Orth constraint in (14) has the addition in brackets to account for these special cases.

(15) *Word-initial English* [h]$_{\text{Aud}}$ *deleted in Portuguese loanword adaptation due to the orthographic constraint* (14)

	0.8	1.0	*H*
[frication] <h…>	[frication] /ʁ/	<h> / /	
/ʁ/	+1		0.8
☞ / /		+1	1.0

The input, the English glottal fricative, consists of both an auditory form [frication] and the orthographic form <h>. Although the Cue constraint clearly favours /ʁ/, the phonological categorisation is determined by the high-weighted Orthographic constraint.

The characteristics of BiPhon-HG that were elaborated in the present section are shown in the following section to enable a formalisation that can account for the three L2 speech phenomena introduced in Section 2.

4 A formal account of the data

In this section, we present a BiPhon-HG formalisation of the three L2 speech patterns reported in the acquisition of the Portuguese tap by L1-Mandarin learners, namely between and within-subject variation (Section 4.2), position-dependent repair strategies (Section 4.3), and influence of orthography (Section 4.4). Section 4.5 then provides the complete grammar, i.e., all subset of constraints and their weights employed in the formalisations. Before we provide these accounts, we first elaborate on the initial stage of acquisition in Section 4.1.

4.1 The initial L2 grammar

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. During the subsequent L2 acquisition process, this perception grammar is adapted based on the input the learners receive, by changing the weights of the Cue and Structural constraints but also by creating new categories and new Cue constraints involving these new categories. We will come back to this development in Section 5.2.

To model the initial stage of L2 perception for the present case at hand, we must construct phonological categories and Cue constraints for Mandarin, which are then used to parse the Portuguese sounds. Mandarin has the liquid categories /l/ and /ɻ/, with the respective prototypical F3 values of 2643 Hz, and 2118 Hz (Smith 2010). The mapping

of these values onto the corresponding Mandarin categories is shown in Tableaux (16) and (17).

(16) *Mandarin* [l]$_{\text{Aud}}$ *categorised by the Mandarin perception grammar as* /l/

	1.0	1.0	0.2	0.2	*H*
[2643 Hz]	[2118Hz] /ɻ/	[2643 Hz] /l/	[2118Hz] /l/	[2643Hz] /ɻ/	
☞ /l/		+1			1.0
/ɻ/				+1	0.2

(17) *Mandarin* [ɻ]$_{\text{Aud}}$ *categorised by the Mandarin perception grammar as* /ɻ/

	1.0	1.0	0.2	0.2	*H*
[2118Hz]	[2118Hz] /ɻ/	[2643Hz] /l/	[2118Hz] /l/	[2643Hz] /ɻ/	
/l/			+1		0.2
☞ /ɻ/	+1				1.0

The liquid perception grammar in (16) and (17) is only able to deal with prototypical F3 values. But how would a Mandarin listener categorise an input F3 of 2542 Hz that corresponds to the Portuguese tap? 2542 Hz is in close proximity to the 2643 Hz value that is prototypical for the Mandarin category /l/. It can therefore be assumed that the Mandarin lateral is often enough also realised with this less-prototypical value, and hence that the corresponding Cue constraint "[2542 Hz] /l/" has higher weights than the constraint "[2542 Hz] /ɻ/". An input [2542 Hz] would therefore be perceived as /l/, shown in Tableau (18):

(18) *Portuguese* [ɾ]$_{\text{Aud}}$ *categorised by the Mandarin perception grammar as* /l/

	1.0	1.0	0.7	0.3	0.2	0.2	*H*
[2542Hz]	[2118Hz] /ɻ/	[2643Hz] /l/	[2542Hz] /l/	[2542Hz] /ɻ/	[2643Hz] /ɻ/	[2118Hz] /l/	
☞ /l/			+1				0.7
/ɻ/				+1			0.3

Note that at this stage, naïve listeners have no Portuguese tap category, which means Cue constraints referring to /ɾ/ do not exist (yet) (but see Section 5.2 below).

Up to now we only considered the perception of single sounds, where Structural constraints do not play any role, but we will integrate them in Section 4.3. Since we will be only looking at the perception of Portuguese sounds in the following, we do not further include Cue constraints referring to the prototypical Mandarin sound realisations (i.e., the first, second, fifth and sixth constraint in Tableau (18)) in the following three sub-sections, but will come back to how all these constraints combine in one single Mandarin naïve perception grammar in Section 4.5.

4.2 Between and within-subject variations in L2 phonological categorisation

As described in Section 2, the experimental results by Zhou & Hamann (2020) regarding the phonological categorisation of the Portuguese tap by Mandarin naïve listeners showed considerable variation, and two types of listeners emerged: Type I systematically identified the Portuguese tap as /l/ and Type II as /l/ or alveolar stop /t/ or /t^h/.

Tableau (18) in the previous section can successfully account already for Type I naïve listeners, as it categorises Portuguese [ɾ] as /l/. However, that tableau only considered one cue, which is simplistic, as phonological distinctions are often signalled by multiple acoustic cues. A stop voicing contrast, for instance, is usually implemented by Voice Onset Time, adjacent vowel duration, closure duration and F0 in the following vowel. The use of such an abundance of cues makes phonological categorisation stable even in adverse conditions (e.g., if some cues are masked by background noise). In the case of the Portuguese tap, not only formant structures but also a brief silence caused by tongue tip closure is a common cue (Silva 2014).

Prior research has suggested that, when presented with multiple cues whose cooccurrence is entirely novel, non-native listeners may manifest individual differences in cue use and weighting (e.g. Chandrasekaran et al. 2010; Wanrooij et al. 2013; Schertz et al. 2015; Kim et al. 2017). Therefore, when being presented with both spectral and closure cues, some Mandarin L1 listeners may have higher perceptual weights for the spectral cue than for the closure one, while others may have the reverse cue-weighting. If the closure cue is regarded more important than the formant cue by a Mandarin listener, this is implemented in their perception grammar by the closure Cue constraint outweighing the spectral one, and will lead to the Portuguese tap being categorised as an alveolar stop, as shown in (19), where the input consists now of two auditory events, an F3 of [2542 Hz] and a [closure]. The high-weighted Cue constraint "[closure] /t, t^h/" expresses the Mandarin cue knowledge that the presence of a closure in the auditory input implies the existence of a stop. Since this Cue constraint has higher weight than the two spectral Cue constraints, the Portuguese auditory input $[ɾ]_{Aud}$ is categorised as a stop (either plain or aspirated) by this Mandarin perception grammar.

(19) *Portuguese* [ɾ]$_{\text{Aud}}$ *(with closure and spectral cues) categorised by the Mandarin perception grammar as an alveolar stop*

	0.8	0.7	0.3	*H*
[2542Hz, closure]	[closure] /t, tʰ/	[2542Hz] /l/	[2542Hz] /ɻ/	
/l/		+1		0.7
/ɻ/			+1	0.3
☞ /t/	+1			0.8
☞ /tʰ/	+1			0.8

If an L1-Mandarin listener gives more weigh to the spectral cues, on the other hand, the Cue constraint "[closure] /t, tʰ/" in the perception grammar is outweighed by "[2542Hz] /l/", and Portuguese [ɾ]$_{\text{Aud}}$ is parsed as /l/, resulting in the behaviour of Type I listeners again, recall Tableau (18), but now with a more realistic model that involves two cues, as shown in (20):

(20) *Portuguese* [ɾ]$_{\text{Aud}}$ *(with closure and spectral cues) categorised by the Mandarin perception grammar as lateral*

	0.7	0.5	0.3	*H*
[2542Hz, closure]	[2542Hz] /l/	[closure] /t, tʰ/	[2542Hz] /ɻ/	
☞ /l/	+1			0.7
/ɻ/			+1	0.3
/t/		+1		0.5
/tʰ/		+1		0.5

Tableau (19) does not fully account for Type II listeners, yet, as these listeners were not consistent in their categorisation of the Portuguese tap and alternated between /l/ and an alveolar stop in their responses. This within-subject variation can be captured by the probabilistic grammar put forward in Section 3.1. The random noise value temporarily added to each constraint weight at evaluation time can result in the selection of different winners across several instances of evaluation, if the two decisive constraints have weights that are fairly close to each other. This is the case with the two constraints "[closure] /t, tʰ/" and "[2542Hz] /l/" in (19). The two following tableaux illustrate such possible variation, where the original weights are given in brackets in the first row, and the resulting weights with added evaluation noise are shown underneath. In (21), the added noise does not change the winner with respect to the outcome without added noise in (19), while in (22) it does.

(21) *Portuguese* [ɾ]$_{\text{Aud}}$ *(with closure and spectral cues) categorised by the Mandarin noisy perception grammar as an alveolar stop*

	(0.8) 0.87	(0.7) 0.72	(0.3) 0.34	*H*
[2542Hz, closure]	[closure] /t, tʰ/	[2542Hz] /l/	[2542Hz] /ɻ/	
/l/		+1		0.72
/ɻ/			+1	0.34
☞ /t/	+1			0.87
☞ /tʰ/	+1			0.87

(22) *Portuguese* [ɾ]$_{\text{Aud}}$ *(with closure and spectral cues) categorised by the Mandarin noisy perception grammar as* /l/

	(0.8) 0.74	(0.7) 0.79	(0.3) 0.25	*H*
[2542Hz, closure]	[closure] /t, tʰ/	[2542Hz] /l/	[2542Hz] /ɻ/	
☞ /l/		+1		0.79
/ɻ/			+1	0.25
/t/	+1			0.74
/tʰ/	+1			0.74

It is noteworthy that such a change in winning candidates does not occur when the involved constraints have weights that are far apart. The third constraint "[2542Hz] /ɻ/" in Tableaux (21) and (22), for instance, has a much lower weight than the other two, and the addition of evaluation noise to it does not influence the outcome of evaluation.

In this section, we have shown that the between and within-subject variation attested in L2 phonological categorisation can be accounted for with a model that makes explicit the mapping between several available auditory cues onto the involved phonological categories, and listener-specific preferences in this mapping. The Cue constraints of the BiPhon-HG model formalise such individual cue weighting strategies, and the stochasticity of the model with added noise at evaluation time allows for within-listener variation in the choice between different categories.

In the appendix we provide the details of a computer simulation of a Type II learner with an OTGrammar object in Praat (Boersma & Weenink 2023) set to a Noisy-HarmonicGrammar decision strategy with constraint satisfaction (Appendix 1a), in which all initial constraint

weights are 100. This grammar is fed 100,000 pairs with auditory forms of [2542Hz, closure] and surface forms that are 70% lateral, 15% plain stop and 15% aspirated stop realisations (see Appendix 1b for the input-output distributions). The weights of the constraints are then adjusted stepwise with the Gradual Learning Algorithm (GLA; Boersma 1998; Boersma & Pater 2016 for HG). The resulting acquired constraint weights reproduce the distribution of output forms in the perception of [2542Hz, closure] (see Appendix 1c for a resulting grammar with percentage of occurring output forms).

4.3 Position-dependent repair strategies in L2 phonological categorisation

As reviewed in Section 2, there is an interaction between segmental and syllabic information in the categorisation of the Portuguese tap by L1-Mandarin learners. In syllable onset, Mandarin speakers mainly rely on segmental replacement (/ɾ/ as [l] or [t, t^h]). In coda position, production studies (Zhou 2017; Liu 2018) have reported also structural modifications such as epenthesis ([lə] or [tə]) and deletion of the tap. A recent perceptual study by Zhou & Rato (2023)[9] confirms that L1-Mandarin learners may hear an illusory vowel after the tap in coda position, but they do not delete the tap perceptually. Therefore, in this section, we will consider only the perceptual epenthesis in coda position (the production-specific repair by deletion is discussed in Section 5.1).

The dependence of tap categorisation on syllable position can be attributed to Mandarin phonotactics, which only allows /ɻ/ and nasals (/n/ and /ŋ/) in syllable-final position (Duanmu 2005; Lin 2007). Other segments that are in coda in the original language are therefore perceived as occupying an onset position.

As introduced in Section 3.2, phonotactic knowledge is captured in BiPhon by language-specific Structural constraints, and we will employ in the following the general constraint "ManPhono" to represent Mandarin phonotactic well-formedness:

(23) ManPhono This constraint is satisfied if a candidate is in line with the structural requirements of Mandarin.

Since adherence to phonotactics is an important requirement, this constraint has a high weight. How the Mandarin perception grammar with this Structural constraint parses the Portuguese tap in syllable onset and in coda position is shown in Tableaux (24) and (25), respectively. The tableaux have a weighting of Cue constraints that favors the lateral, i.e., they model a Type I listener from Section 4.2. The input to these tableaux consists of whole words (the pseudowords used in Zhou & Hamann 2020), where phonetic cues are only provided for the tap (in curly brackets), whereas the rest of the word is given in phonetic transcription.

[9] The perception study by Zhou & Rato (2023) employed natural stimuli, produced by a male L1 speaker of Portuguese.

(24) *Portuguese tap in* [pɐɾafɐ] *(with closure and spectral cues) categorised by the Mandarin perception grammar as /l/ in onset*

	0.9	0.7	0.5	0.3	*H*
[pɐ{closure, 2542Hz}afɐ]	ManPhono	[2542Hz] /l/	[closure] /t, tʰ/	[2542Hz] /ɻ/	
☞ /pa.la.fa/	+1	+1			1.6
/pa.ɻa.fa/	+1			+1	1.2
/pa.ta.fa/	+1		+1		1.4

In (24), all candidates are rewarded by the high-weighted Structural constraint as they satisfy the Mandarin phonotactics. The decision between them is due to Cue constraints, which select the candidate with an onset /l/ as the optimum.

The decisive role of the Structural constraint is revealed in (25), where the Portuguese tap occurs in syllable-final position in the input. This time, the Structural constraint rewards the candidates that conform to the Mandarin phonotactic requirement, namely those with syllable-final /ɻ/ and those without a coda. Given that the candidate /pa.lə.fa/ is not only rewarded by the Structural constraint, but also by the highest-weighted Cue constraint, it is selected as winner by the perception grammar.

(25) *Portuguese tap in* [paɾfɐ] *(with closure and spectral cues) categorised by the Mandarin perception grammar as /lə/ in coda*

	0.9	0.7	0.5	0.3	*H*
[pa{closure, 2542Hz}fɐ]	ManPhono	[2542Hz] /l/	[closure] /t, tʰ/	[2542Hz] /ɻ/	
/pal.fa/		+1			0.7
☞ /pa.lə.fa/	+1	+1			1.6
/paɻ.fa/	+1			+1	1.2
/paɻə.fa/	+1			+1	1.2
/pat.fa/			+1		0.4
/patə.fa/	+1		+1		1.4

Candidates with other epenthetic vowels than schwa (e.g. /pa.la.fa/ or /pa.li.fa/) were not included here for the sake of space. Like /pa.lə.fa/, they would also satisfy the Mandarin phonotactic restrictions. In contrast to /pa.lə.fa/, such candidates would not satisfy an additional Cue constraint that allows the insertion of schwa if this surface schwa is breaking up the illicit

consonant cluster of lateral and any following consonant lC. This Cue constraint could be formalised as "[ ] /(.l)ə(.C)/", where the empty auditory form indicates absence of cues, which is then interpreted as a schwa in the surface form; parentheses denote the surface environment (see Boersma & Hamann 2009b for a similar account of the epenthetic vowel [ɨ] in Korean loanword adaptation). Such a Cue constraint is based on the fact that Mandarin has a process in production whereby neutral-tone vowels are often reduced to schwa (Yip 2002) and deleted (Cheng 1973; Weinberger 1996). Mandarin listeners are therefore accustomed to insert a schwa in auditory forms that contain consonant clusters like [lC]. Durvasula et al. (2018) provide evidence from L2 perception experiments for this process.

The formalisation that we presented in this subsection demonstrates that BiPhon-HG is capable of including the L1 phonotactic influence on L2 speech perception we have observed in the L1-Mandarin listeners of Portuguese. The subset of constraints that we employed here will be integrated with those in the previous section in 4.5.

4.4 The interaction between phonology and orthography in L2 speech learning

In the acquisition of L2 Portuguese, L1-Mandarin speakers replaced the Portuguese tap with their L1 rhotic /ɻ/ only when written input was given (Zhou & Hamann 2020). These findings indicate an orthographic influence on L2 phonological categorisation. In this section, we formalise this attested interaction. For this, we extend the perception grammar from the previous sections to a multimodal grammar that integrates the Mandarin grapheme-phoneme conversion knowledge as the Orth constraint "<r> /ɻ/", which rewards a candidate that maps written input <r> to the SF /ɻ/.

(26) *Portuguese* [ɾ] *(only auditory input presented) categorised by the Mandarin multimodal grammar as* /l/

	0.9	0.7	0.5	0.3	0.2	*H*
[2542Hz, closure]	ManPhono	[2542Hz] /l/	[closure] /t, t^h/	[2542Hz] /ɻ/	<r> /ɻ/	
☞ /l/		+1				0.7
/ɻ/				+1		0.3
/t/			+1			0.5

As illustrated in Tableau (26), when only the auditory input is given, the multimodal grammar functions exactly like the perception grammar, because the newly added Orth constraint "<r> /ɻ/" lacks an orthographic input to evaluate. When both auditory and written input serve as input, a possible influence of orthography depends on the weight of the Orth constraint, as shown in Tableaux (27) and (28).

(27) *Portuguese* [ɾ] *(both auditory and orthographic input presented) categorised by the Mandarin multimodal grammar as* /l/

	0.9	0.7	0.5	0.3	0.2	*H*
[2542Hz, closure] <r>	ManPhono	[2542Hz] /l/	[closure] /t, tʰ/	[2542Hz] /ɻ/	<r> /ɻ/	
☞ /l/		+1				0.7
/ɻ/				+1	+1	0.5
/t/			+1			0.5

In (27), the Orth constraint has a relative low weight, and therefore the high-weighted Cue constraint "[2542 Hz] /l/" determines the selected winner. In this case, the presence of an orthographic cue, in addition to the AudF, does not affect the output of L2 phonological categorisation, accounting for the listeners whose categorisation was not influenced by the presence of orthography in Zhou & Hamann (2020), introduced in Section 2. However, if a Mandarin listener displays more reliance on the orthographic information, formalised as a higher weight of the Orth constraint, orthography will change the results of phonological categorisation, as illustrated in (28).

(28) *Portuguese* [ɾ] *(both auditory and orthographic input presented) categorised by the Mandarin multimodal grammar as* /ɻ/

	0.9	0.7	0.5	0.5	0.3	*H*
[2542Hz, closure] <r>	ManPhono	[2542Hz] /l/	[closure] /t, tʰ/	<r> /ɻ/	[2542Hz] /ɻ/	
/l/		+1				0.7
☞ /ɻ/				+1	+1	0.8
/t/			+1			0.5

In (28), the Orth constraint "<r>/ɻ/" is still outweighed by two Cue constraints "[2542 Hz] /l/" and "[closure] /t, tʰ/". However, due to the decision-making mechanism of HG, the Orth constraint together with the lowest-weighed Cue constraint "[2542 Hz] /ɻ/", both of which reward the SF /ɻ/, may gang up to overcome the two Cue constraints with higher weight. In this case, the multimodal grammar selects as optimum the candidate /ɻ/, which is least favoured by the perception grammar (Cue constraints), simulating an orthographic influence on L2

phonological categorisation. Tableau (28) therefore accounts for the listeners that categorised the Portuguese tap as an L1 /ɻ/ only when the written form was given along with the auditory form.

In principle, the candidate /ɻ/ would still win even if we assigned a really high weight to the Orth constraint, let us say, 1.0. In this case, the relative weighting of the Cue constraints is not decisive anymore for evaluation, because the highest-weighed Orth constraint will determine the winner, much like how decision-making is performed in strict OT. Although the same candidate will be chosen as winner by the multimodal grammar, regardless of whether it is due to a gang-up effect or overriding orthographic effect, there is at least one theoretical consequence that differs between these two formal treatments. In an account where the Orth constraint had such a high weight that it determined the output of evaluation regardless of cue knowledge, one would expect the candidate /ɻ/ to win even if the auditory input contained cues for a stop or a fricative. This can hardly be the case, as previous studies have suggested that the L2 orthographic effect is likely to be triggered only when the L2 sound and the L1 category represented by the grapheme share some acoustic similarity (e.g., Rafat 2015). The role of such acoustic similarity in inducing L1-based orthographic influence is better accounted for by the ganging-up effect in HG, suggesting that BiPhon-HG outperforms its OT counterpart under certain circumstances.

4.5 The complete grammar – an overview of all constraints and their weights

In this section, we bring together all subsets of constraints introduced in the previous sections to show that they are compatible with each other and form one grammar that accounts for the observed syllable-position effects and orthographic influences. For the individual differences, the studies reported in Section 2 showed that there are two types of listeners: Type I, who categorises the Portuguese tap as lateral all the time, and Type II, who categorises the tap either as lateral or as stop. We implemented this difference in our BiPhon-HG account in 4.3 by giving a different weight to the Cue constraint that maps the tap closure cue onto a stop category ("[closure] /t, t^h/"): for Type I listeners this constraint has lower weight (0.5) than for Type II learners (0.8). The complete grammars with all constraints for both types of learners are provided in (29) and (30). As can be seen, the two differ only in the weight given to Cue constraint "[closure] /t, t^h/" (boldfaced):

(29) *Mandarin naïve listener Type I: Portuguese* [ɾ] *is categorised as lateral*

1.0	1.0	0.9	0.7	**0.5**	0.5	0.3	0.2	0.2
[2118Hz] /ɻ/	[2643Hz] /l/	Man Phono	[2542Hz] /l/	**[closure] /t, t^h/**	<r> /ɻ/	[2542Hz] /ɻ/	[2643Hz] /ɻ/	[2118Hz] /l/

(30) *Mandarin naïve listener Type II: Portuguese* [ɾ] *is categorised either as stop or lateral (within-subject variation)*

1.0	1.0	0.9	**0.8**	0.7	0.5	0.3	0.2	0.2
[2118Hz] /ɻ/	[2643Hz] /l/	Man Phono	**[closure] /t, tʰ/**	[2542Hz] /l/	<r> /ɻ/	[2542Hz] /ɻ/	[2643Hz] /ɻ/	[2118Hz] /l/

The data overview in Section 2 also indicated that some listeners relied less on the influence of orthography than others. For these listeners we proposed a lower weight (of 0.2) for the Orthographic constraint <r> /ɻ/ in Section 4.4, again with all other constraints and weights staying the same.

Note that none of the winning candidates in the tableaux of the previous sections formalising the Mandarin naïve perception of the Portuguese tap would change if the complete set of constraints were used in the evaluation process rather than the smaller set we employed for each case.

5 Future directions

5.1 Development towards a target-like grammar

In this paper, we formalised three intriguing L2 speech patterns by feeding the Portuguese input to a naïve Mandarin perception grammar. The next step is to determine how learners develop from this to a more Portuguese-like grammar that allows them to eventually perceive (and produce) an L2 tap.

As briefly discussed in 4.1, we assume that learners simply use a copy of their L1 grammar at the onset of L2 speech learning. Based on subsequent L2 input, learners will then adapt their interlanguage grammar by creating new categories and new constraints involving these new categories. In the present case of the Portuguese tap, advanced learners create a new surface phonological category of a tap, connections from auditory cues onto this new category (via Cue constraints) and restrictions on the distribution of this tap (via Structural constraints).

Although L1-Mandarin learners tend to assimilate the L2 tap to their L1 lateral category, a new category can still be constructed on the basis of sufficient distributional information (see Boersma et al. 2003 and Escudero & Boersma 2004 for other cases). L2 category creation can further be assisted by orthographic knowledge, as empirical evidence has shown (e.g., Escudero et al. 2008). For Portuguese, the fact that the letters <r> and <l> represent two different sounds could support the acquisition of a tap category.

After creating the new SF tap, learners will build UFs for words containing the tap in their lexicon, and optimise the connections between these two representational levels via Faithfulness constraints (short: Faith).

An example for a target-like Portuguese L2 grammar that can accurately perceive the L2 tap (i.e., can distinguish it from other segments in Portuguese, such as the lateral) as a SF and UF tap category, is given in (31). In this tableau, the Portuguese structural restrictions are summarised as PortPhono, analogous to the constraint ManPhono defined in (23), with a high weight of 0.9. Similarly high weights are also given to Faith and the Cue constraint for perceiving the most prototypical F3 vale [2542Hz] as tap, resulting in target-like mapping between AudF, SF and UF.

(31) *L2 Portuguese recognition of* [ca{2542Hz}ta] *as* |caɾta|

	0.9	0.9	0.9	0.4	*H*
[ca{2542Hz}ta]	Faith	Port Phono	[2542Hz] /ɾ/	[2542Hz] /l/	
☞ /caɾ.ta/ |caɾta|	+1	+1	+1		2.7
/caɾ.ta/ |calta|		+1	+1		1.8
/cal.ta/ |caɾta|		+1		+1	1.3
/cal.ta/ |calta|	+1	+1		+1	2.2

Tableau (31) illustrates that, upon hearing the auditory form of the word *carta* “letter” with a prototypical F3 value of [2542Hz] for the tap, this learner can recognise the intended UF (and subsequently, its meaning, though this is not incorporated in our formalisation), by selecting the correct SF and UF pair from all candidate pairs. The candidate pairs with a mismatch between SF and UF (candidates two and three) will always be less optimal, since they do not satisfy Faith. All four given candidate pairs have a SF that is in line with the Portuguese phonotactics. Candidates three and four map the prototypical tap value onto a surface /l/ instead /ɾ/, satisfying the Cue constraint “[2542Hz] /l/”, with a lower weight than “[2542Hz] /ɾ/”.

For the learning/optimisation between all levels of representation in the advanced L2 perception grammar, the GLA is assumed (as introduced at the end of Section 4.2 for the computer simulation). Grammatical learning through the GLA has been implemented many times in prior research (see especially Escudero & Boersma 2004; Boersma & Escudero 2008). In particular, each time the learner detects a mismatch between the input and their own production, the GLA will be triggered, adjusting constraint weights for achieving a more target-like grammar. The supervising mechanism of GLA corroborates nicely previous empirical studies showing that supervision (either via explicit corrective feedback or via implicit lexical guidance) plays a crucial role in L2 sound learning (see Felker et al. 2021 for a recent review).

5.2 Perception-production asymmetry

The modelling presented so far in this paper was only concerned with the perception direction. As the use of the same constraints and constraint weights in both processing directions is one of the most striking features of BiPhon, this section will give a brief illustration of how L2 production is handled in BiPhon, targeting the perception-production asymmetry noted at the beginning of Section 4.3: Recall that in L2 Portuguese production, L1-Mandarin learners resort to both epenthesis and deletion as repair strategies for the Portuguese tap in coda position (Zhou, 2017; Liu, 2018), while in perception, they always hear the tap, very often accompanied by a following illusory vowel (Zhou & Rato 2023). Therefore, epenthesis may have a perceptual basis, but tap deletion does not.

Although a straightforward account for the production-specific repair strategy is to posit that L2 perception and production have distinct grammars (Ramus et al. 2010), we argue that this is not necessary. Instead, we assume that the coda tap deletion is driven by articulatory restrictions that only apply in the production direction of the BiPhon model (cf. **Figure 2**), which therefore can account for such a perception-production asymmetry.

Unlike perception, speech production in BiPhon involves an additional representational level, the ArtF, and is thus subject to the influences of articulatory constraints, which represent the degree of articulatory difficulty. The articulation of the Portuguese tap implies considerable complexity, demanding a ballistic movement of the tongue tip and a constriction towards the pharynx (Berti 2010; Barberena et al. 2014; 2019). Before mastering such complex gestures, L1-Mandarin learners may very likely fail to realise the Portuguese tap, especially in word-internal coda position, where consonant-to-consonant co-articulation further increases articulatory difficulty. The production of a voiceless plosive after a vowel in coda position, as in /at/, on the other hand, is easy for L1-Mandarin learners because it occurs in their L1. This is expressed in the Articulatory constraint in (32a). Articulations with an intervening lateral or tap, which are more difficult because they are non-occurring in Mandarin, are expressed in the Articulatory constraints (32b) and (32c), respectively.

(32) a) [[at]] This constraint is satisfied if jaw lowering for a low vowel followed by a full alveolar closure for a voiceless stop is produced.
b) [[alt]] This constraint is satisfied if jaw lowering for a low vowel followed by a lateral alveolar closure followed by a full alveolar closure is produced.
c) [[aɾt]] This constraint is satisfied if jaw lowering for a low vowel followed by a ballistic movement for an alveolar tap followed by a full alveolar closure is produced.

The different degrees of articulatory difficulty for L2 speakers are expressed by decreasing weight given to the three constraints. The influence of these three Articulatory constraints on production are shown in Tableau (33). This tableau formalises the production of a word-internal coda tap in

the word *carta*, “letter”, i.e., it is the production counterpart of the recognition tableau in (31). It employs the same Faith and PortPhono and Cue constraints and weights as the former tableau, only now the Cue constraints are read in the opposite direction, e.g. “[2542Hz] /ɾ/” as “This constraint is satisfied if the phonological category /ɾ/ is mapped onto an F3 value of [2542 Hz]”. Input is the UF retrieved from the lexicon, and candidates are now triplets of SF, AudF and ArtF. Only SF which are faithful to the UF are considered, as a different SF will always be less optimal.

(33) *L2 Portuguese production of* |carta| *as* [[cata]]

	0.9	0.9	0.9	0.4	1.0	0.5	0.1	*H*
|carta|	Faith	Port Phono	/ɾ/ [2542Hz]	/l/ [2542Hz]	[[at]]	[[alt]]	[[art]]	
/car.ta/ [carta] [[carta]]	+1	+1	+1				+1	2.8
/car.ta/ [carta] [[calta]]	+1	+1	+1			+1		3.2
☞ /car.ta/ [carta] [[cata]]	+1	+1	+1		+1			3.7

Due to the non-native-like lower weights of the Articulatory constraints involving the tap (the last two constraints in the tableau), the L2 learner represented with this grammar will produce a form without a tap. Once this learner can master the more complex articulations with tap, the weights increase and a tap production becomes possible.[10]

6 Conclusions

In this article we showed that three L2 speech patterns, widely observed in empirical studies but largely ignored by major L2 theoretical models, can be accounted for within a comprehensive generative linguistic model, namely BiPhon-HG.

We formalised the between and within-subject variability in L2 phonological categorisation as learners’ individual cue-weighting strategies and the stochasticity of the learners’ perception grammar. This was done with the help of Cue constraints, which express the mapping from input auditory events to listeners’ abstract phonological categories.

[10] Note that the mapping between AudF and ArtF is regulated via Sensorimotor constraints in BiPhon (Boersma 2009). Those were not included in the present formalisation to restrict it to the most relevant factors.

In addition to these Cue constraints, BiPhon also employs phonotactic (structural) and orthographic knowledge in the phonological categorisation. In our formalisation, we have shown that the competition between Cue and Structural constraints gives rise to position-dependent repair strategies in L2 speech learning: in the present case, it caused the use of perceptual epenthesis to break illicit consonant clusters. The L2 orthographic influence on phonological categorisation, on the other hand, was accounted for by the interaction between Cue and Orth constraints.

This paper constitutes the first attempt to formalise all these three recurrent observations in L2 speech research within a single model, responding to the needs of a more comprehensive model rather than one that mainly concerns the mapping between acoustic input and phonetic/phonological categories, see Yazawa et al. (2020) for a similar point.

Apart from accounting for the existing data, our modelling further makes some predictions that can be falsified by future experiments. For instance, it predicts that in L2 phonological acquisition, the orthographic effect is triggered only if the L2 sound and the L1 category represented by the grapheme share some acoustic similarity, as discussed in 4.4. If the auditory input for instance contains only vowel-like formants but the written form represents a stop, our model predicts that learners would ignore the orthographic information and rely solely on auditory cues to construct a novel phonological category for that sound (thus no orthographic effect would be expected).

The OT version of BiPhon has been previously applied to account for individual variation in L2 speech (Hamann 2009) and L1 phonotactic restriction on L2 phonological categorisation (Boersma 2009). Due to the exclusion mechanism of OT, constraints had to be negatively formulated in these studies. This is not necessary in our formalisation employing HG. Although both constraint satisfaction and violation can be implemented in BiPhon-HG, from an emergentist perspective, positive constraints are more in line with the type of evidence that a learner normally encounters in the input, and these positive constraints might be more intuitive for (L2) researchers working in different frameworks. The application of BiPhon-HG therefore allows L2 speech patterns to be addressed in a broader context. Furthermore, the formal HG grammar is compatible with learning algorithms developed for neural and statistical approaches to language learning (e.g., Boersma & Escudero 2008; Pater 2009). And lastly, L2 speech phenomena can be handled within the same model originally proposed for L1 speech perception and production. Accordingly, there is no need to resort to any mechanism specific to L2 acquisition, and modelling L2 speech patterns can benefit from well-developed formal phonological theories.

Appendix 1a. OTMulti grammar

```
“ooTextFile”
“OTGrammar 2”
<HarmonicGrammar>
0 ! leak
3 ! constraints
     “[closure]/t, th/”    100 100 1
     “[2542Hz]/l/”         100 100 1
     “[2542Hz]/ɻ/”         100 100 1
0 ! fixed rankings
1 ! number of accepted inputs
“[2542Hz, closure]” 4  ! input form with number of output candidates
     “/l/”  0 -1 0     ! output candidate with constraint satisfactions
     “/ɻ/”  0 0 -1
     “/t/”  -1 0 0
     “/tʰ/” -1 0 0
```

Appendix 1b. Input-output distribution

```
“ooTextFile”
“PairDistribution”
3 ! pairs
“[2542Hz, closure]” “/l/”    70  ! number of occurrences
“[2542Hz, closure]” “/t/”    15
“[2542Hz, closure]” “/tʰ/”   15
```

Appendix 1c. Learned grammar and occurrence of candidates

	104.43	102.96	92.61	*H*
[2542Hz, closure]	[2542Hz] /l/	[closure] /t, tʰ/	[2542Hz] /ɻ/	
70.07% ☞ /l/	+1			104
/ɻ/			+1	93
14.83% ☞ /t/		+1		102
15.09% ☞ /tʰ/		+1		102

Funding information

This research was supported by the Doctoral Degree Scholarship Program 2017, Language Sciences by University of Lisbon, and also by grant UIDB/00214/2020 from the Portuguese Foundation of Science and Technology (FCT).

Acknowledgements

We would like to thank Paul Boersma for his feedback on the computational simulations. Thanks also to the Associate Editor Sara Finley and three anonymous reviewers for their insightful comments and suggestions, which helped us to improve the article greatly.

Competing interests

The authors have no competing interests to declare.

References

Barberena, Luciana da Silva & Keske-Soares, Márcia & Berti, Larissa Cristina. 2014. Descrição dos gestos articulatórios envolvidos na produção dos sons /r/ e /l/. *Audiology – Communication Research* 19(4). 338–44. DOI: https://doi.org/10.1590/S2317-64312014000400135

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. DOI: https://doi.org/10.1590/2317-6431-2018-2059

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

Batalha, Graciete Nogueria. 1995. *O Português falado e escrito pelos Chineses de Macau*. Instituto Cultural de Macau.

Berti, Larissa Cristina. 2010. Investigação da produção de fala a partir da ultrassonografia do movimento de língua. In *Anais do 18º Congresso Brasileiro de Fonoaudiologia*. São Paulo: Sociedade Brasileira de Fonoaudiologia.

Best, Catherine & Tyler, Michael. 2007. Nonnative and second language speech perception: Commonalities and complementarities. In Bohn, Ocke-Schwen & Munro, Murray J. (eds.), *Language experience in second language speech learning – In honor of James Emil Flege*, 13–34. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.17.07bes

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

Boersma, Paul. 2007. Some listener-oriented accounts of h-aspiré in French. *Lingua* 117(12). 1989–2054. DOI: https://doi.org/10.1016/j.lingua.2006.11.004

Boersma, Paul. 2009. Cue constraints and their interactions in phonological perception and production. Boersma, Paul & Hamann, Silke (eds.), *Phonology in Perception*, 55–110. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110219234.55

Boersma, Paul. 2011. A programme for bidirectional phonology and phonetics and their acquisition and evolution. In Benz, Anton & Mattausch, Jason (eds.), *Bidirectional Optimality Theory*, Vol. (180), 33–72. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.180.02boe

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

Boersma, Paul & Escudero, Paola. 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, Paul & Escudero, Paola & Hayes, Rachel. 2003. Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In Solé, Maria Josep & Recasens, Daniel & Romero Joaquín (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013–1016.

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

Boersma, Paul & Hamann, Silke. 2009a. Introduction: models of phonology in perception. In Boersma, Paul & Hamann, Silke (eds.), *Phonology in Perception*, 1–24. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110219234.1

Boersma, Paul & Hamann, Silke. 2009b. Loanword adaptation as first-language phonological perception. In Calabrese, Andrea & Wetzels, Leo (eds.), *Loanword Phonology*, 11–58. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.307.02boe

Boersma, Paul & Pater, Joe. 2007. Constructing constraints from language case of Canadian English diphthongs. Paper presented at *North East Linguistic S (NELS)* 38. Ottawa. Handout retrieved from https://www.fon.hum.uva.nl/paul/presentations/CanadianRaisingNELS2007.pdf

Boersma, Paul & Pater, Joe. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. In McCarthy, John & Pater, Joe (eds.), *Harmonic Serialism and Harmonic Grammar*, 389–434. Sheffield: Equinox.

Boersma, Paul & Weenink, David. 2023. Praat: doing phonetics by computer [Computer program]. Version 6.3.10, retrieved 3 May 2023 from http://www.praat.org/

Bohn, Ocke-Schwen. 2017. Cross-language and second language speech perception. In Fernandéz, Eva M. & Cairns, Helen S. (eds.), *The handbook of psycholinguistics*, 213–239. Hoboken, NJ: John Wiley & Sons. DOI: https://doi.org/10.1002/9781118829516.ch10

Bohn, Ocke-Schwen. 2020. Cross-language phonetic relationships account for most, but not all L2 speech learning problems: The role of universal phonetic biases and generalized sensitivities. In Wrembel, Magdalena & Kiełkiewicz-Janowiak, Agnieszka & Gąsiorowski, Piotr (eds.), *Approaches to the Study of Sound Structure and Speech: Interdisciplinary Work in Honour of Katarzyna Dziubalska-Kołaczyk*, 171–184. Routledge. DOI: https://doi.org/10.4324/9780429321757-13

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

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

Chandrasekaran, Bharath & Sampath, Padma & Wong, Patrick. 2010. Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America* 128. 456–465. DOI: https://doi.org/10.1121/1.3445785

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, Shuwen & Mok, Peggy. 2019. Speech production of rhotics in highly proficient bilinguals: acoustic and articulatory measures. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 1818–1822.

Cheng, Chin-Chuan. 1973. *A Synchronic Phonology of Mandarin Chinese*. The Hague: Mouton. DOI: https://doi.org/10.1515/9783110866407

Chomsky, Noam & Halle, Morris. 1968. *The sound pattern of English*. New York: Harper & Row.

Church, Kenneth. 1987. Phonological parsing and lexical retrieval. *Cognition* 25(1–2). 53–69. DOI: https://doi.org/10.1016/0010-0277(87)90004-7

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

Colantoni, Laura & Steele, Jeffrey & Escudero, Paola. 2015. *Second language speech: Theory and practice*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139087636

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

Duanmu, San. 2005. Chinese (Mandarin): phonology. *Encyclopedia of Language and Linguistics, 2nd Edition*, ed. by Keith Brown, 351–355. Oxford, UK: Elsevier Publishing House. DOI: https://doi.org/10.1016/B0-08-044854-2/00096-1

Duanmu, San. 2007. *The Phonology of Standard Chinese* (2nd ed.). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780199215782.001.0001

Dupoux, Emmanuel & Kakehi, Kazuhiko & Hirose, Yuki & Pallier, Christophe C. & Mehler, Jacques. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6). 1568–1578. DOI: https://doi.org/10.1037/0096-1523.25.6.1568

Durvasula, Karthik & Huang, Ho-Hsin & Uehara, Sayako & Luo, Qian & Lin, Yen-Hwei. 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). DOI: https://doi.org/10.5334/labphon.57

Escudero, Paola. 2005. *Linguistic Perception and Second Language Acquisition*. Utrecht: University of Utrecht dissertation.

Escudero, Paola & Bion, Ricardo. 2007. Modeling vowel normalization and sound perception as sequential processes. In Trouvain, Jürgen & Barry, William John (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*, 1413–1416.

Escudero, Paola & Boersma, Paul. 2002. Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Arunachalam, Sudha & Kaiser, Elsi & Williams, Alexander (eds.), *Penn Working Papers in Linguistics* 8(1). 71–85.

Escudero, Paola & Boersma, Paul. 2004. Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition* 26(4). 551–585. DOI: https://doi.org/10.1017/S0272263104040021

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

Faris, Mona & Best, Catherine & Tyler, Michael 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. DOI: https://doi.org/10.1121/1.4939608

Felker, Emily & Broersma, Mirjam & Ernestus, Mirjam. 2021. The role of corrective feedback and lexical guidance in perceptual learning of a novel L2 accent in dialogue. *Applied Psycholinguistics* 42(4). 1029–1055. DOI: https://doi.org/10.1017/S0142716421000205

Fikkert, Paula. 2010. Developing representations and the emergence of phonology: Evidence from perception and production. *Laboratory Phonology* 10(4). 227–255. DOI: https://doi.org/10.1515/9783110224917.3.227

Flege, James. 1995. Second Language Speech Learning: Theory, Findings and Problems. In Strange, Winifred (ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research,* 233–277. Timonium, MD: New York Press.

Flege, James & Bohn, Ocke-Schwen. 2021. The Revised Speech Learning Model (SLM-r). In Wayland, Ratree (ed.), *Second Language Speech Learning: Theoretical and Empirical Progress,* 3–83. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781108886901.002

Fowler, Carol. 1986. An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14. 3–28. DOI: https://doi.org/10.1016/S0095-4470(19)30607-2

Gnanadesikan, Amalia. 2004. Markedness and faithfulness constraints in child phonology. In Kager, René & Zonneveld, Wim & Pater, Jospeh (eds.), *Fixing priorities: Constraints in phonological acquisition*, 73–108. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486418.004

Hall, T. A. & Hamann, Silke. 2010. On the cross-linguistic avoidance of rhotic plus high front vocoid sequences. *Lingua* 120(7). 1821–1844. DOI: https://doi.org/10.1016/j.lingua.2009.11.004

Hamann, Silke. 2009. Variation in the perception of an L2 contrast: A combined phonetic and phonological account. In Kügler, Frank & Fery, Caroline & van de Vijver, Ruben (eds.), *Variation and Gradience in Phonetics and Phonology*. Berlin: Mouton de Gruyter, 79–105. DOI: https://doi.org/10.1515/9783110219326.71

Hamann, Silke. 2011. The phonetics-phonology interface. In Kula, Nancy & Botma, Bert & Nasukawa, Kuniya (eds.), *Continuum Companion to Phonology,* 202–224. London: Continuum.

Hamann, Silke. 2020. One phonotactic restriction for speaking, listening and reading: The case of the *no geminate* constraint in German. In Evertz-Rittich, Martin & Kirchhoff, Frank (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: https://doi.org/10.1515/9783110710809-004

Hamann, Silke & Colombo, Ilaria. 2017. A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. *Natural Language and Linguistic Theory* 35. 683–714. DOI: https://doi.org/10.1007/s11049-017-9362-3

Hayes-Harb, Rachel & Barrios, Shannon. 2021. The influence of orthography in second language phonological acquisition. *Language Teaching* 1–30. DOI: https://doi.org/10.1017/S0261444820000658

Ingram, David. 1995. The acquisition of negative constraints, the OCP, and underspecified representations. In Archibald, John (ed.), *Phonological acquisition and phonological theory*, 63–79. Psychology Press.

Jesney, Karen & Tessier, Anne-Michelle. 2007. Re-evaluating learning biases in Harmonic Grammar. In Becker, Michael (ed.), *University of Massachusetts Occasional Papers in Linguistics* 36*: Papers in theoretical and computational phonology*. GLSA.

Jesus, Luis M. T. & Shadle, Christine H. 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, Song & Chang, Yueh-chin & Hsieh, Feng-fan. 2019. An EMA study of er-suffixation in Northeastern Mandarin monophthongs. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 2149–2153.

Johnson, Keith & Sjerps, Matthias. 2021. Speaker normalization in speech perception. In Pardo, Jennifer & Nygaard, Lynne & Remez, Robert & Pisoni, David (eds.), *The handbook of speech perception*, 145–176. Hoboken: Wiley Blackwell. DOI: https://doi.org/10.1002/9781119184096.ch6

Kabak, Bariş & Idsardi, William 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

Kim, Donhyun & Clayards, Meghan & Goad, Heather. 2017. Individual differences in second language speech perception across tasks and contrasts: The case of English vowel contrasts by Korean learners. *Linguistics Vanguard* 3(1). DOI: https://doi.org/10.1515/lingvan-2016-0025

Kimper, Wendell A. 2016. Positive constraints and finite goodness in Harmonic Serialism. In Pater, Joe & McCarthy, John (eds.), *Harmonic Grammar and Harmonic Serialism,* 221–235. London: Equinox Press.

Lahiri, Aditi & Reetz, Henning. 2002. Underspecified recognition. In Gussenhoven, Carlos & Warner, Natasha (eds.), *Laboratory Phonology* Vol. 7, 637–676. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110197105.2.637

Legendre, Geraldine & Miyata, Yoshiro & Smolensky, Paul. 1990. Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the cognitive science society*, 388–395. Mahwah, NJ: Lawrence Erlbaum Associates.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lin, Yen-Hwei. 2007. *The Sounds of Chinese*. Cambridge: Cambridge University Press.

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

Martins, Marlene. 2008. *O português dos chineses em Portugal – O caso dos imigrantes da área do comércio e restauração em Águeda*. Aveiro: University of Aveiro dissertation.

Mateus, Maria Helena & Falé, Isabel & Freitas, Maria João. 2005. *Fonética e Fonologia do Português*. Lisboa: Universidade Aberta.

Mateus, Maria Helena & Rodrigues, Celeste. 2003. A vibrante em coda no português. *Teoria Lingüística. Fonologia e outros temas*, 181–199.

McClelland, James L. & Elman, Jeffrey L. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18(1). 1–86. DOI: https://doi.org/10.1016/0010-0285(86)90015-0

McQueen, James M. & Cutler, Anne. 1997. Cognitive processes in speech perception. In Hardcastle, William & Laver, John & Gibbon, Fiona (eds.), *The Handbook of Phonetic Sciences*, 566–585. Oxford: Blackwell.

Mitterer, Holger & Scharenborg, Odette & McQueen, James. 2013. Phonological abstraction without phonemes in speech perception. *Cognition* 129(2). 356–361. DOI: https://doi.org/10.1016/j.cognition.2013.07.011

Pater, Joe. 2009. Weighted constraints in generative linguistics. *Cognitive Science* 33. 999–1035. DOI: https://doi.org/10.1111/j.1551-6709.2009.01047.x

Pereira, Rodrigo. 2020. *O R-forte em Português Europeu: análise fonológica de dados dialetais.* Unpublished MA thesis, University of Lisbon, Lisbon, Portugal.

Polivanov, Evgenij 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 Polivanov, Evgenij D. 1974. *Selected Works: Articles on general linguistics*, 223–237. Mouton, The Hague: Mouton].

Prince, Alan & Smolensky, Paul. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar (Technical Report no. RuCCS-TR-2)*. New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Rafat, Yasaman. 2015. The interaction of acoustic and orthographic input in the acquisition of Spanish assibilated/fricative rhotics. *Applied Psycholinguistics* 36(1). 43–66. DOI: https://doi.org/10.1017/S0142716414000423

Ramus, Franck & Peperkamp, Sharon & Christophe, Anne & Jacquemot, Charlotte & Kouider, Sid & Dupoux, Emmanuel. 2010. A psycholinguistic perspective on the acquisition of phonology. In *Laboratory Phonology 10: Variation, Phonetic Detail and Phonological Representation*, 311–340. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110224917.3.311

Rennicke, Iiris. 2015. *Variation and Change in the Rhotics of Brazilian Portuguese.* Helsinki: University of Helsinki dissertation.

Rodrigues, Celeste. 2003. *Lisboa e Braga: Fonologia e Variação*. Lisbon: University of Lisbon dissertation.

Rodrigues, Susana. 2015. *Caracterização acústica das consoantes líquidas do Português Europeu.* Lisbon: University of Lisbon dissertation.

Samuel, Arthur G. 2020. Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language* 111. DOI: https://doi.org/10.1016/j.jml.2019.104070

Schertz, Jessamyn & Cho, Taehong & Lotto, Andrew & Warner, Natasha. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52. 183–204. DOI: https://doi.org/10.1016/j.wocn.2015.07.003

Schwartz, Bonnie D. & Sprouse, Rex A. 1996. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12(1). 40–72. DOI: https://doi.org/10.1177/026765839601200103

Shi, Dingxu. 2004. *Peking Mandarin*. Lincom, München.

Silva, Ana. 2014. *Análise Acústica da Vibrante Simples do Português Europeu*. Aveiro: University of Aveiro dissertation.

Smith, James G. 2010. *Acoustic Properties of English /l/ and /ɹ/ Produced by Mandarin Chinese Speakers*. Toronto: University of Toronto dissertation.

Smolensky, Paul. 1996. On the comprehension/production dilemma in child language. *Linguistic Inquiry* 27. 720–731.

Sousa, Elsa. 2017. *Production and perception of the English /h/: the case of native Portuguese speakers of English as a Foreign Language*. Braga: University of Minho dissertation.

Wanrooij, Karin & Escudero, Paola & Raijmakers, Maartje E. J. 2013. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *Journal of Phonetics* 41(5). 307–319. DOI: https://doi.org/10.1016/j.wocn.2013.03.005

Weinberger, Steven H. 1996. Minimal segments in second language phonology. In: James, Allan & Leather, Jonathan (eds.), *Second-language speech: structure and process*, 263–311. Berlin: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110882933.263

Yazawa, Kakeru & Whang, James & Kondo, Mariko & Escudero, Paola. 2020. Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research* 36(4). 557–581. DOI: https://doi.org/10.1177/0267658319832645

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Zhou, Chao. 2017. *Contributo para o estudo da aquisição das consoantes líquidas do português europeu por aprendentes chineses*. Lisbon: University of Lisbon dissertation.

Zhou, Chao & Hamann, Silke. 2020. Cross-linguistic interaction between phonological categorization and orthography predicts prosodic effects in the acquisition of Portuguese liquids by L1-Mandarin learners. *Proceedings of Interspeech 2020*. DOI: https://doi.org/10.21437/Interspeech.2020-2689

Zhou, Chao & Jesus, Alice. 2022. Portuguese has two underlying rhotics: Evidence from Lisbon and Carioca varieties. *Supplemental Proceedings of the 2021 Annual Meeting on Phonology*. DOI: https://doi.org/10.3765/amp.v9i0.5167

Zhou, Chao & Rato, Anabela. 2023. Syllable position effects in the perception of L2 Portuguese /l/ and /ɾ/ by L1-Mandarin learners. *Second Language Research*. DOI: https://doi.org/10.1177/02676583221137713

Zhu, Xiaonong. 2007. Jinyin-fulun Putonghua rimu [About approximant]. *Fangyan* 1. 2–9.