

Article

Acquisition of the Tap-Trill Contrast by L1 Mandarin–L2 English–L3 Spanish Speakers

Matthew Patience 

Department of Spanish and Portuguese, University of Toronto, #208 - 91 Charles St. W., Toronto, ON M5S 1K7, Canada; matt.patience@utoronto.ca; Tel.: +1-416-813-4080

Received: 5 August 2018; Accepted: 7 November 2018; Published: 13 November 2018



Abstract: The goals of this study were to investigate the developmental patterns of acquisition of the Spanish tap and trill by L1 Mandarin–L2 English–L3 Spanish speakers, and to examine the extent to which the L1 and the L2 influenced the L3 productions. Twenty L1 Mandarin–L2 English–L3 Spanish speakers performed a reading task that elicited production of rhotics from the speakers’ L3 Spanish, L2 English, and L1 Mandarin, as well as the L2 English flap. The least proficient speakers produced a single substitution initially, generally [l]. The same non-target segment was produced for both rhotics, mirroring the results of previous studies investigating L1 English–L2 Spanish speakers, indicating that this may be a universal simplification strategy. In contrast to previous work on L1 English speakers, the L1 Mandarin–L2 English–L3 Spanish speakers who had acquired the tap did not tend to use it as the primary substitute for the trill. Overall, the L1 was a stronger source of cross-linguistic influence. Nonetheless, evidence of positive and negative L2 transfer was also found. The L2 flap allophone facilitated acquisition of the L3 tap, whereas non-target productions of the L2 /ɾ/ were also observed, revealing that both previously learned languages were possible sources of cross-linguistic influence.

Keywords: phonetic acquisition; L3 acquisition; Spanish; rhotics; Mandarin; phonetic acquisition; phonological acquisition; English; tap; trill

1. Introduction

The acquisition of the Spanish tap–trill contrast (the tap /ɾ/ and the trill /r/) has received considerable attention in recent years (e.g., [Waltmunson 2005](#); [Face 2006](#); [Colantoni and Steele 2008](#); [Johnson 2008](#); [Rose 2010](#); [Olsen 2012](#); [Kopečková 2014](#); [Amengual 2016](#); [Morales Reyes et al. 2017](#)). Findings from studies on the L2 production of the Spanish rhotics by L1 English speakers have revealed clear developmental paths. Specifically, such learners initially produce the English [ɹ] in place of both the tap and the trill ([Waltmunson 2005](#); [Olsen 2012](#)). The tap is the first rhotic mastered, and is used as a substitute for the trill until the latter is acquired ([Face 2006](#); [Johnson 2008](#)). This is interesting because the two rhotics are contrastive in intervocalic position, yet English speakers fail to reliably produce this contrast until the trill is acquired. The same tendency to substitute the tap for the trill was also observed in L1 German speakers learning Spanish ([Kopečková 2014](#)). Given that research to date has focused principally on either L1 English or L1 German learners of Spanish, it is unclear to what extent the patterns observed are general. Indeed, is this specific to native speakers of a Germanic language or is it a universal developmental pattern when acquiring the tap–trill contrast? One objective of the present study is thus to examine whether speakers of a typologically distinct L1, namely, Mandarin, demonstrate similar production patterns when acquiring the Spanish tap and trill. Previous work has demonstrated that L1 Mandarin speakers experience difficulty with both the perception ([Ortí Mateu 1990](#); [Chih 2013](#)) and production ([Ortí Mateu 1990](#)) of the /ɾ–l/ contrast in L2 Spanish, resulting in non-target lateral substitutions in place of the tap. L1 Mandarin speakers also have difficulty

producing the trill (Ortí Mateu 1990). Therefore, while some lateral substitutions are expected in the least proficient speakers when producing /r/ and /r/, it is unknown what developmental patterns these speakers might exhibit as they become more proficient.

While the participants of the present study were native speakers of Mandarin, they also spoke English as an L2. The second objective of the present study is to contribute to our understanding of L1 and L2 cross-linguistic influence (CLI) in L3 acquisition. One of the limitations in L3 phonetics and phonology (L3PP) is the small number of language combinations that have been researched, and also the limited variety of structures that have been analyzed (Cabrelli Amaro and Wrembel 2016). Consequently, to fully understand the role of the L1 and L2 in L3 production requires additional data from speakers of different language combinations and structures. The present experiment provides data from a unique language triad involving three typologically distinct languages (L1 Sino-Tibetan, L2 Germanic, L3 Romance), on segments that have received little attention (rhotics) in L3PP. Of particular interest here is the extent to which the L2 English plays a role in the acquisition of the L3 Spanish tap. Given the similarity of the English flap and the Spanish tap, and the fact that the English flap has been shown to facilitate the acquisition of the Spanish tap in L1 English–L2 Spanish speakers (Colantoni and Steele 2008; Olsen 2012), we might expect the same to be true for L1 Mandarin–L2 English–L3 Spanish speakers. This hypothesis was investigated by examining whether speakers who have acquired the L2 English flap produce the L3 Spanish tap with greater accuracy. The fact that L1 English–L2 Spanish speakers substitute the English [ɹ] for the two Spanish rhotics suggests that the [ɹ] may also be a likely substitute in the Spanish of L1 Mandarin–L2 English–L3 Spanish speakers, if they have indeed acquired the English [ɹ], and if the L2 is a more likely source of CLI, as two models of L3 acquisition (L2 Status Factor model; Bardel and Falk (2007, 2012); and the Typological Proximity Model, Rothman (2011, 2015)) would predict for the present study's learners. Therefore, in addition to investigating the potential positive transfer of the L2 English flap, the extent to which transfer of the L2 English [ɹ] is present will also be examined.

The goals of the present study were investigated by analyzing the intervocalic production of the Spanish tap–trill contrast by beginner to intermediate L1 Mandarin–L2 English–L3 Spanish speakers. The participants' production of the L2 English [ɹ] and [r] was also examined, in order to establish whether the speakers had the ability to produce and therefore transfer L2 sounds into their L3 Spanish.

The current work is organized as follows. In the remainder of the introduction, we will first discuss the relevant theoretical background on phonological development, in order to highlight the role that perception and perceptual categorization can play in shaping learners' production. A brief overview of L3 acquisition theory, as well as a summary of studies on the L3 acquisition of phonetics and phonology, is also presented. The phonetic and phonological characteristics of the Spanish, Mandarin, and English rhotics are then discussed in detail, in addition to other segments that may be relevant to the acquisition of the Spanish tap and trill. These include the English flap, due to its similarity to the Spanish tap, as well as /l/, given that Mandarin learners of Spanish experience difficulty with the /l–r/ contrast (Ortí Mateu 1990; Chih 2013). A summary of previous findings on the acquisition of the Spanish rhotics is subsequently presented, followed by specific predictions regarding the expected production patterns that L1 Mandarin–L2 English–L3 Spanish speakers will display when acquiring the Spanish rhotics. Following the introduction, the current study's methodology is summarized. The results are then presented, and the manuscript concludes with a discussion of the results' implications for L3 acquisition and development.

1.1. Theoretical Background on L2 and L3 Phonological Development

Although the present study focuses on production, perceptual categorization is expected to play an important role and can help explain and predict the production patterns of non-native speakers. The objective of this section is to briefly present Best and Tyler's (2007) Perceptual Assimilation Model (PAM-L2), which will be used to (1) motivate which L1 Mandarin and L2 English segments, including which characteristics of these segments, may play a role in the acquisition of the Spanish

tap–trill contrast (Section 1.2); (2) partially explain the similarities and differences observed in the production patterns of L1 Mandarin and L1 English learners of Spanish (Section 1.3); and (3) make predictions concerning L1 Mandarin–L2 English–L3 Spanish speakers’ production of the Spanish rhotics (Section 1.4). The PAM-L2 is expected to be the most appropriate model for the present study, given that it posits a role for phonetic, phonological, and orthographic similarity during the categorization of sounds. While the model was designed specifically for perception, the predictions it makes for the perceptual assimilation of sounds are expected to partially explain the production patterns observed in non-native speakers. Note that while the PAM-L2 was developed for L2 speech perception, one can potentially apply it to any non-native speech learning context, including L3 acquisition. Therefore, an additional goal of this section is to discuss how the PAM-L2 might be applied to L3 acquisition. While PAM-L2 has not been previously applied to an L3 learning context¹, the Speech Learning Model (Flege 1995), which is similar to the PAM-L2, has been both discussed and applied to L3 acquisition (Sypiańska 2016a; Lipińska 2017). However, one of the limitations of the SLM is that it cannot account for orthographic or phonological similarity, which have been shown to play a crucial role in the acquisition of rhotics (e.g., Face 2006; Johnson 2008). Given that the PAM-L2 considers not only phonetic but also phonological and orthographic similarity, it is expected to be a more suitable model for the acquisition of the Spanish tap and trill.

According to the PAM-L2 (Best and Tyler 2007), if two contrasting non-native sounds are very similar to an L1 sound, then acquisition of that contrast will be difficult, as the two non-native sounds may be assimilated to a single native sound. This has consequences for production: if learners perceptually assimilate a non-native sound to a native sound, then the non-native sound will be produced incorrectly using L1 gestural patterns. The PAM-L2 acknowledges that perceptual assimilation can happen at either the phonetic or phonological level. At the phonetic level, perceptual assimilation occurs if two sounds are similar acoustically, arising from similar articulatory gestures (e.g., manner of articulation, place of articulation, voicing). At the phonological level, perceptual assimilation can occur if sounds have similar phonotactics or even if they are represented by the same grapheme. The example given in Best and Tyler (2007) exemplifying phonological similarity is of French /ʁ/ and English /ɹ/, which tend to be perceptually assimilated, despite the fact that they are phonetically very different. However, both are represented by the same grapheme <r>, and behave similarly in terms of their syllable structure and phonotactics (e.g., both can occur syllable initially, finally, and in consonant clusters). These similarities (phonological and orthographical) cause English-speaking learners to perceptually categorize the two rhotics as a single phonological category, and substitute /ʁ/ with [ɹ] in French productions; similar substitutions would be expected for French-speaking learners of English.

To summarize, perceptual assimilation can be due to phonetic, phonological and/or orthographic similarity, all of which may lead to L1 influence in production. The example of the French and English rhotic is particularly relevant to this paper. As we will see in Section 1.2, while the two Spanish rhotics are phonetically distinct, they share some phonological as well as orthographic similarities with the English rhotic and, to a lesser extent, the Mandarin rhotic. Moreover, while the contrast of interest in this paper is the acquisition of the L3 Spanish tap and trill, the Spanish /l-r/ contrast is also relevant for L1 Mandarin learners of Spanish, because they tend to perceptually assimilate the Spanish /l/ and /r/ to Mandarin /l/ (Ortí Mateu 1990; Chih 2013).

While the PAM-L2 was developed to account for L2 acquisition, the model can be applied to L3 acquisition. The crucial difference is that both L1 and L2 sounds are potential sources for perceptual assimilation. Therefore, when considering the PAM-L2 in an L3 context, two possibilities arise: (1) The most similar sound from either the L1 or L2 will be the most likely source of perceptual assimilation

¹ While Lipińska (2017) investigates the L3 acquisition of German and summarizes the PAM-L2, there is no discussion on how the PAM-L2 might be adapted to L3PP.

for an L3 sound; and (2) sounds from one of the two previously learned languages will be more likely sources of perceptual assimilation. In order to determine which of the two proposed scenarios ((1) or (2)) is the most likely, we will consider three models of L3 acquisition, as well as findings from studies investigating the acquisition of L3PP.

Three models of L3/Ln acquisition that have been tested extensively are the L2 Status Factor model (L2SF) (Bardel and Falk 2007, 2012), the Typological Proximity Model (TPM) (Rothman 2011, 2015), and the Cumulative Enhancement Model (CEM) (Flynn et al. 2004). The L2SF predicts initial L2 transfer², regardless of typological similarity. This is due to the similar process by which post-puberty languages are acquired (i.e., using declarative memory), as opposed to the process by which languages are acquired before puberty (i.e., using procedural memory). In contrast, the TPM predicts that learners will initially transfer the (psycho)typologically³ most similar language, which is established by an L3 learner's linguistic parser. The linguistic parser (subconsciously) compares the L1 and L2 to the L3 during the initial stages of acquisition. Up to four linguistic domains are analyzed hierarchically: lexicon -> phonology -> functional morphology -> syntax. The analysis continues until sufficient similarity is encountered between the L3 and either the L1 or the L2 at one of the domains (for further details, see Rothman 2015). Therefore, if the lexicon of either the L1 or L2 were clearly more similar to the L3, the parser would stop the comparison at this point. However, if no clear similarities were found, the parser would continue the analysis and begin comparing the L1 and L2 phonologies to the L3 phonology, followed by the functional morphology and then the syntax, if necessary. Once the parser determines which language is the most similar, that language is transferred to the L3. The more similar language therefore functions as the initial state of the L3, analogous to the Full Transfer Full Access proposal of Schwartz and Sprouse (1996) in L2 acquisition. According to the TPM's predictions, in the present study's language combination, the linguistic parser would only need to compare the lexicons of the L1 Mandarin, L2 English, and L3 Spanish, due to the clear similarities of the L2 English and L3 Spanish lexicons. Such similarity is not present between the L1 and L3 lexicons, and thus we should assume that the parser would conclude that the L2 English is the more similar language. As a result, the L2 English phonological system would be the initial state of the learner's L3 Spanish. The third L3 model, the CEM (Flynn et al. 2004), does not assume that transfer originates from only one of previously learned languages. The model proposes that learning a language is a cumulative process, and that any previously learned language could influence the acquisition of a subsequent language. Crucially, however, the influence of previously learned languages is expected to be either facilitative or neutral, and therefore negative transfer is not expected. In the context of the present study, the CEM would assume that either the L1 Mandarin or L2 English could be a potential source of transfer, if facilitative in nature. However, as we will see in Sections 1.1–1.3, the only segment that is expected to result in positive transfer is the L2 English flap⁴. Therefore, while the CEM can potentially account for both L1 and L2 transfer, the only transfer it would predict for the present study's learners is L2 transfer of the English flap.

Note that the three L3 models are designed to determine the expected source of syntactic/morphosyntactic transfer, and in the case of the TPM, are specific to initial state learners. Consequently, the models may not apply to (1) phonetic and phonological transfer; or (2) experienced L3 speakers, and therefore are not necessarily applicable in the present study. Nevertheless, the three

² In the present paper, transfer is used to refer to the copying of an underlying structure or feature from one language to another, and is therefore representational in nature. In contrast, CLI is used as a more general term to refer to characteristics from previously learned languages that are present in the speech of L3 learners. While these characteristics may be the result of transfer, this is not necessarily the case. For example, persistent CLI from previously learned articulatory motor routines.

³ Psychotypological similarity refers to the perceived linguistic typological similarity between two languages (Wrembel 2015, p. 40).

⁴ Given that the CEM was not designed to predict transfer of phonetics and phonology, it is not clear at what level (segmental, sub-segmental) positive transfer should be expected to occur. For the purposes of the present study, it will be assumed that the CEM predicts transfer at the segmental level. However, it is plausible that positive transfer occurs at a sub-segmental level, as in Brown's (1998) feature-based L2 speech model (e.g., positive transfer of features such as a velar place feature).

models would predict L2 transfer, thus the L2 should be considered a likely source of CLI. Moreover, the predictions of the TPM demonstrate that despite English and Spanish being from typologically distinct language families, an L1 Mandarin–L2 English–L3 Spanish learner would likely consider their L2 English to be more similar to Spanish than their L1 Mandarin.

Given that the L3 models were developed to explain morphosyntactic transfer, it raises the question of whether they apply to the L3PP. Research investigating whether the L1 or L2 is a more likely source of CLI in L3PP has revealed inconsistent results. In contrast to the L2SF and the TPM, some recent findings suggest that the L1 (dominant) language may be a stronger source of CLI in learners. For example, [Kopečková \(2014\)](#) investigated the acquisition of the Spanish rhotics by L1 German–L2 English–L3 Spanish speakers (11- and 12-year-old children). With respect to the L3 trill production, the author found primarily either L1 transfer (German uvular fricative or approximant productions) or what the author refers to as interlanguage productions, that consisted of either L3 tap substitutions, or uvular trills (analyzed as a combination of an L1 and L3 segment). Likewise, regarding the L3 tap, the L1 proved to be the strongest source of transfer, with 33.3% of productions displaying clear L1 transfer (uvular fricative or approximant productions). However, the learners also produced target L3 taps approximately 50% of the time. It is possible that the target L3 tap productions were the result of positive transfer of the L2 English flap, but the study did not examine the L2 production of the flap. Consequently, it is not possible to know whether this was the case. Nevertheless, the results reported by [Kopečková \(2014\)](#) reveal that for her learners, the L1 was a stronger source of CLI. Similar findings were reported in [Pyun \(2005\)](#), [Llama and Cardoso \(2018\)](#), and [Llama and López-Morelos \(2016\)](#). [Pyun \(2005\)](#) investigated the acquisition of phonological processes (e.g., unreleased obstruents, consonant cluster simplification) in L1 Korean–L2 English–L3 Swedish speakers. Even though English is more similar to Swedish, CLI from the L1 Korean was more prevalent (although L2 English influence was also observed to a lesser extent). [Llama and Cardoso \(2018\)](#) examined VOT production in the L3 Spanish of L1 English–L2 French and L1 French–L2 English adults. For both groups, prevailing influence from the L1 was observed. However, for the L1 English–L2 French group, the L1 English influenced the L2 French productions, and it was not clear to what extent the L1, or both the L1 and L2, influenced the L3 Spanish VOT values. In a similar study, [Llama and López-Morelos \(2016\)](#) examined the VOT realization of voiceless stops by English dominant-Spanish heritage speakers acquiring L3 French. The L3 French VOT values patterned according to the speakers' English VOT values, and not the speakers' more similar Spanish VOT values. While the studies summarized here suggest that the dominant language (generally the L1) may have a privileged role, evidence has also been found for a combination of CLI from both the L1 and the L2 ([Blank and Zimmer 2009](#); [Wrembel 2014](#); [Sypiańska 2016b](#)), and from primarily the L2 (e.g., [Hammarberg and Hammarberg 2005](#); [Wrembel 2010](#); [Llama et al. 2010](#); [Chang 2015](#)). Some of the studies observing primarily L2 CLI involved L3s that were more similar to the L2 than the L1, and therefore support both the L2SF and the TPM. For example, [Wrembel \(2010\)](#) examined CLI in the L3 English of L1 Polish–L2 German speakers, and found that the L2 was initially a stronger source of CLI. Interestingly, the author also observed that as the learner became more proficient in the L3, the L2 influence diminished, whereas the L1 influence became more prevalent. This same trend was observed in [Hammarberg and Hammarberg \(2005\)](#), who investigated CLI in one L1 English–L2 German–L3 Swedish speaker. These two studies suggest that the source of influence may vary as learners become more proficient in the L3. While some previous research that observed primarily L2 transfer consisted of language groupings in which the L2 was more similar to the L3 than the L1, other research has observed stronger transfer from the L2, even when the L2 was clearly less similar to the L3 than the L1 (e.g., L1 French–L2 English–L3 Spanish; [Llama et al. 2010](#)), or when all three languages were from different typological families (L1 English–L2 French–L3 Japanese; [Tremblay 2007](#)).

Given the variable findings in research on L3PP, it remains unclear in what contexts L3 models might apply to the production of L3 segments, and whether the L1 or the L2 might be a stronger source of CLI. However, results from studies in L3PP have generally found that either the L1 or the

L2 is a more likely source of CLI, as opposed to both being equally likely sources. This suggests that perceptual assimilation is not determined on a segment-by-segment basis according to the most similar previously learned L1 or L2 sound, thus potentially ruling out scenario (1) discussed previously. Scenario (2) may be more likely, with perceptual assimilation being most probable from the segments of one of the two previously acquired languages, either the L1 or the L2. Specific predictions are laid out in Section 1.4.

In the following section, the phonetic, phonological, and orthographic characteristics of the two Spanish rhotics will be summarized. Other L1 Mandarin and L2 English segments that could play a role in L1 Mandarin–L2 English-speaking learners' perceptual categorization and, thus, production, will also be discussed.

1.2. Relevant Phonetic and Phonological Characteristics of Spanish, Mandarin, and English

In this section, similarities and differences between relevant English, Mandarin, and Spanish consonants are discussed. The goal is to provide a detailed summary of the knowledge that an L1 Mandarin–L2 English speaker has, and thus demonstrate what needs to be acquired to master the Spanish rhotics, and also to highlight what may influence (positively or negatively) acquisition. The characteristics of the Spanish rhotics will be discussed first, followed by the relevant English then Mandarin segments.

1.2.1. Spanish

Spanish has two rhotic consonants, the tap /ɾ/ and the trill /r/. The tap is a voiced segment produced with a rapid, brief contact between the tongue tip and alveolar ridge (Martínez Celdrán and Fernández Planas 2007), averaging only approximately 23 ms, and very rarely exceeding 30 ms (Blecua Falgueras 2001). The articulation of the trill similarly involves rapid contact of the tongue against the alveolar ridge. The primary difference is that the trill is produced with two or more rapid closures. The duration of the entire segment in intervocalic position is on average 82–88 ms (Quilis 1993). Like the tap, the trill is voiced (Navarro Tomás 1957), although voiceless realizations are also observed. For example, Lewis (2004) observed voiceless word initial trills with frequency of occurrence rates ranging from 5% to 40%, depending on the final segment of the preceding word. Waltmunson (2005) observed that word medial trills were sometimes partially devoiced (1.6/2, where 2 = fully voiced, 1 = partially voiced, 0 = voiceless).

Phonologically, both Spanish rhotics occur in intervocalic position where they are contrastive. There are approximately 30 minimal pairs in Spanish (e.g., pero /'pero/ 'but' versus perro /'pero/ 'dog'; caro /'caro/ 'expensive' versus carro /'caro/ 'car') (Willis and Bradley 2008). However, the two members of such pairs are generally from different lexical categories (e.g., conjunction versus noun and adjective versus noun in the examples above). Thus, a tap being produced in place of a trill, or vice versa, is unlikely to result in miscommunication. While the focus of the present paper is on intervocalic rhotics, it is important to note that the rhotics also appear in other positions in the word where they are not contrastive. The tap appears in word- and syllable-final positions (e.g., por /por/ 'for'; parte /'par.te/ 'part') as well as in syllable onset clusters (e.g., problema /pro.ble.ma/ 'problem'), whereas the trill is found word initially (e.g., rata /'ra.ta/ 'rat') and syllable initially (e.g., honra /'on.ra/ 'honour') after the consonants /l/ and /n/ (Hualde 2014). In intervocalic position the tap is represented orthographically as <r> whereas the trill is represented as <rr>. However, both rhotics are represented by a single <r> in non-contrastive word positions (e.g., parte /'par.te/ 'part' versus rata /'ra.ta/ 'rat') and thus share the same grapheme with English.

While the canonical realization of the /ɾ/ and /r/ phonemes is a tap and an alveolar trill, respectively, a significant amount of variation is present. In different regions of Latin America, rhotics may be produced with assibilation, as approximants, with a dorsal place of articulation, or as laterals (Hualde 2014). In certain regions of Spain, coda /ɾ/ is realized as a lateral, fricative, or is elided (Samper Pandilla 2011), whereas the trill is produced as a fricative, approximant, or with r-coloring (Henriksen

and Willis 2010). A significant amount of intra-speaker variation is also observed, due to stylistic differences and, in the case of the trill, due to articulatory complexity (Hualde 2014). For example, both rhotics are sometimes produced without a complete closure in the vocal tract, resulting in fricative productions (Blecua Falgueras 2001). Other variables, such as social factors, can also play a role, with males producing fewer occlusions when realizing trills compared to females (Henriksen 2014). Given that there is variability in rhotic production, we should assume that the learners of the present study have been exposed to multiple allophones of the rhotic phonemes. Nevertheless, the canonical tap–trill realizations are the most common variants, and are also the variants that are taught in educational settings.

In sum, while the two Spanish rhotics are phonetically quite different, they share some similarities in that both are produced with an alveolar place of articulation, appear intervocalically, and are represented with the same graphemes. These similarities may cause learners to perceptually categorize the tap and trill as a single segment, which is what occurs in L1 English–L2 Spanish speakers (at least in production; see Section 1.3 for details). In the following sections, we will see that while the English and Mandarin rhotics share some similarities themselves, they are both phonetically different segments compared to the Spanish rhotics. However, the English and Mandarin rhotics also share some characteristics with the tap and trill, such as their graphemic representations and, especially with respect to the English /ɹ/, their phonotactics. These shared characteristics can have important implications in terms of the substitutions that might surface in production.

1.2.2. Mandarin

Mandarin Chinese has one rhotic, a voiced apical post-alveolar approximant /ɻ/ (Lee 1999), with an average duration of 95 ms (Smith 2010). There is some debate as to the exact nature of the Mandarin rhotic. Duanmu (2000) states that some consider it to be a voiced retroflex fricative /ʐ/, but he presents two reasons why it should be considered an approximant. First, the consonant has very little frication. Second, Mandarin has no other voiced obstruents, so it would be ‘phonologically odd’ to have a single voiced obstruent. Lee (1999), on the other hand, by means of an articulatory and acoustic study, determined that the rhotic has a post-alveolar articulation (not retroflex). He also found that it is produced with no frication, thus it is clearly an approximant. Note, however, that Lee’s study was based on just four speakers, and only involved word initial /ɻ/ produced in limited contexts (the number of different stimuli used was not stated). Consequently, the study may not capture all of the variability in production of the Mandarin rhotic. Finally, Cerini (2013) also argues that the rhotic is an approximant, because it is the most frequent realization, but admits that a fricative variant exists (in addition to a hybrid of both). It will therefore be assumed that speakers may produce either variant.

Phonologically, Mandarin /ɻ/ occurs in syllable onset position, and, similar to the Spanish rhotics, can occur in intervocalic position (lǎorén /lauɻən/ ‘old people’). It can also occur syllable finally in affixed forms as a syllabic consonant (e.g., gao-gaor-de [kau-kaɻ-də], ‘rather tall’; (Duanmu 2007). The Mandarin rhotic does not occur in consonant clusters, which are lacking in general from the language. Mandarin /ɻ/ is transcribed orthographically as <r> in Pinyin (Duanmu 2007), which is the phonetically transparent, Romanized writing system of Chinese. Mandarin speakers learn the Pinyin system when they begin elementary school. Children become very competent reading in Pinyin, and every time they learn a new character in school, it is accompanied by the corresponding pinyin representation so that they know how to pronounce the new character (Hanley 2005). Pinyin is also frequently used to type on computers and smartphones by adults. While the Chinese characters are the dominant script, and the writing system with which Chinese speakers will have the greatest amount of exposure, the fact that they also use a Romanized system to some extent indicates that orthographic transfer in the production of Spanish should be considered a possibility in L1 Mandarin speakers. However, we might also expect that orthographic transfer is less likely in L1 Mandarin compared to L1 English speakers, given that they have had less exposure to a Roman writing system.

The characteristics of the Mandarin rhotic reveal that it does not share many characteristics with the Spanish rhotics. Phonetically, they are very different; phonologically, they are only somewhat similar; specifically, the Mandarin rhotic also occurs in initial and medial positions, but does not occur in clusters, or codas. Finally, while the Mandarin and Spanish rhotics are both represented orthographically with <r>, the extent to which orthography will play a role in perceptual categorization is expected to be less in Mandarin compared to English learners of Spanish. In view of these differences, the likelihood of any perceptual categorization between Mandarin /ɻ/ and Spanish /r/ or /r/ is moderate, and it is conceivable that L1 Mandarin beginner learners of Spanish will substitute some other segment for either or both of the rhotics. The most likely of these with respect to the tap is the Mandarin dento-alveolar lateral. This hypothesis is supported by [Ortí Mateu \(1990\)](#), who observed that L1 Mandarin–L2 Spanish speakers tend to produce the lateral in place of /r/. The lateral /l/ has a similar place of articulation (dental) with the tap (alveolar), and is often confused perceptually with /r/ by Mandarin speakers ([Ortí Mateu 1990](#); [Chih 2013](#)). The voiced stop [d], which is an allophone of /t/ that can surface in unstressed syllables ([Duanmu 2007](#)), is also a possible candidate for perceptual assimilation, as it shares some similarities with the Spanish tap (place of articulation and voicing).

Given that the speakers of this study spoke English as an L2, English sounds also present possible sources of influence. The most likely candidates for substitutions will now be discussed.

1.2.3. English

The speakers of the present study were living in Eastern Canada, which was the dialect to which they had the greatest amount of exposure. This section will therefore focus on characteristics specific to the Canadian dialect.

English has one rhotic, a voiced retroflex or bunched-tongue approximant /ɻ/ ([Ladefoged and Maddieson 1996](#)) with an average duration of 95 ms ([Smith 2010](#)). It occurs in word initial (e.g., run /ɹʌn/), medial (e.g., merry /mɛ.ɹi/), and final positions (e.g., poor /pʊɹ/) as well as in stop-rhotic clusters (e.g., tree /tri/). The English /ɻ/ is also represented orthographically by <r> or <rr> (e.g., <correct>). Thus, while English /ɻ/ is phonetically quite different from the Spanish rhotics, both native and non-native speakers of English learning Spanish may recognize that /ɻ/ appears in similar positions, and is represented by the same grapheme. Consequently, perceptual categorization with the English and Spanish rhotics is a possibility.

In addition to /ɻ/, English has a flap allophone [ɾ], which surfaces in place of intervocalic /t/ or /d/ after a stressed vowel ([Ladefoged and Maddieson 1996](#)), such as in water /wɒtəɹ/ [ˈwɒ.ɾə]. A flap is considered to be nearly identical acoustically to a tap, but differs slightly in articulation. The English flap is a very brief (10–40 ms), voiced segment ([Zue and Laferriere 1979](#)). While there are slight articulatory differences between a tap and a flap⁵, four variants in English have been reported for the flap allophone, one of which is a tap ([Derrick and Gick 2011](#)). This variation in realizations is only partially dependent on context, as speakers tend to use more than one variant even when producing the same word. These findings suggest that English speakers may already have experience producing taps⁶. Given the acoustic similarity of the tap and the flap, the possibility exists that they will be perceptually categorized as the same sound; moreover, given the articulatory similarity of the tap and flap, if speakers can produce a flap, then this should facilitate the production of the tap, because learners will already have acquired the principal articulatory gestures required for the latter's production.

⁵ Articulatorily, a flap involves retraction of the tongue to a position behind the alveolar ridge. The tongue tip then briefly contacts the alveolar ridge in a forward movement; in contrast, when a tap is produced, the tongue tip makes a direct movement upward to the alveolar ridge, and does not involve a retraction of the tongue ([Ladefoged and Maddieson 1996](#)).

⁶ The flap allophone will continue to be referred to as a flap, in order to differentiate it from the Spanish tap, even though it is often produced as a tap.

As with the Spanish rhotics, the flap and the English /ɹ/ are produced with some variation across dialects. For example, while intervocalic, post-tonic /t/ and /d/ are flapped in most Canadian and American varieties, they are produced as [t] and [d] in Standard Southern British English, (Yavas 2016), or as [ʔ] (Wells 1982), whereas /ɹ/ is produced with an alveolar place of articulation, or can be elided in certain positions (Yavas 2016). In other UK dialects, /ɹ/ may be realized as [ɻ] or [ɹ]; the [ɹ] realization is also common in South African English (McMahon 2002). While the goal here is not to describe all of the variation present in English, it is important to point out that the learners in the present study have almost certainly experienced some variability in English production, and have potentially been taught different ways of pronouncing the English /ɹ/ and post-tonic intervocalic /t/ and /d/.

A summary of the Spanish, Mandarin, and English rhotics is displayed in Table 1. Three main characteristics should be highlighted. First, while English /ɹ/ does not resemble the Spanish rhotics from an acoustic or articulatory perspective, phonotactically the two languages' rhotics are very similar. This, along with the fact that the English and Spanish rhotics are represented with the graphemes <r> and <rr>, means that perceptual categorization of the tap and trill with /ɹ/ by L1 Mandarin–L2 English–L3 Spanish speakers is a possibility, especially in the more proficient English speakers. The Spanish rhotics could also be perceptually categorized with Mandarin /ɹ/, but this segment's phonotactics are less similar, thus perceptual categorization may be less likely to occur with Mandarin /ɹ/ compared to the English /ɹ/. Second, the English flap and Spanish tap are acoustically and articulatorily nearly identical. As a result, perceptual categorization between these two segments is likely if speakers have enough experience with English. Moreover, if speakers have the ability to produce a flap, they should also be able to produce a tap, given the articulatory similarities. Third, speakers may also perceive the tap to be more similar to /d/ or, for less proficient speakers of English, /l/. As a result, these two segments are additional candidates for perceptual assimilation, and thus may surface in production. Indeed, the presence of /l/ productions in L1 Mandarin–L2 Spanish speakers has been observed, which will now be discussed in detail. This will then be followed by a summary of the literature on L1 English–L2 Spanish speakers.

Table 1. Comparison of Spanish, Mandarin, and English rhotics.

Language	Segment	Manner	Place	Voicing	Phonotactics	Orth.	Duration
Spanish	/ɾ/	Tap	Alveolar	Voiced	M, F, C	<r>	23 ms
	/r/	Trill	Alveolar	Voiced	I, M, C	<r>, <rr>	85 ms
English	/ɹ/	Approximant	Retroflex	Voiced	I, M, F, C	<r>, <rr>	95 ms
Mandarin	/ɹ̥, ʐ/	Fricative/	Post-alveolar/ Retroflex	Voiced	I, M, F	<r>	95 ms
		Approximant					

Notes. Orth. = orthography. Phonotactics: I = initial; M = medial (i.e., intervocalic); F = final (i.e., coda); C = clusters.

1.3. Previous Studies on the Acquisition of the Rhotics

1.3.1. The Acquisition of Rhotics by Mandarin Speakers

Various recent papers focused on teaching methods and evaluation of L2 Spanish learners' pronunciation have identified that L1 Mandarin speakers experience difficulty with the Spanish /l–ɾ/ contrast, both in perception and production, as well as with the production of /r/ (Bertola de Urgorri 2009; Cortés Moreno 2014; Igarreta Fernández 2015). However, few experimental perception or production studies have examined these patterns in detail.

Ortí Mateu (1990) investigated the perception and production of several Spanish minimal pairs by 120 early L1 Mandarin–L2 Spanish learners, who were studying Spanish at a university in China. To determine the learners' ability to perceptually discriminate the /l–ɾ/, /l–r/, and /ɾ–r/ contrasts, participants performed a forced choice word identification task. When presented with an /l–ɾ/ minimal pair (e.g., /pala/ 'stick'–/para/ 'for'), they only identified the correct item 59% of the time, essentially at chance. Note that it was not specified if the errors occurred when presented with the /ɾ/ token, the /l/ token, or both. Learners had no difficulty with the /r–l/ and /r–ɾ/ minimal pairs, as they

selected the correct word more than 90% of the time. [Ortí Mateu \(1990\)](#) also tested the participants' production ability via a repetition/word reading task. Participants heard a series of minimal pairs in addition to being presented with the written forms of each pair. They were then required to produce each word of the minimal pair (e.g., /pelo/ 'hair'–/pero/ 'but'; /coro/ 'choir'–/coro/ 'I run'). In the production tasks, the /r-l/ contrast was also difficult with non-target production occurring 33% of the time. An error-type analysis was not presented, therefore it is not clear whether participants experienced difficulty with the production of /l/, /r/ or both. However, Ortí Mateu does state in a summary that [l] was generally produced in place of /r/, so presumably the errors surfaced when attempting to produce the words with /r/. Moreover, /l/ is a phoneme in Mandarin, so it is unlikely that it resulted in difficulty in production. The author also found that the /r-r/ contrast was difficult in production (51% error rate). Again, it is not clear which of the two sounds in this contrast was (more) difficult, but the author also stated in the production summary that participants were often unable to produce the trill. We can thus infer that production of the trill was difficult for these speakers.

Another experimental study that analyzed the perception of rhotic and lateral contrasts by L1 Mandarin learners of Spanish is [Chih \(2013\)](#), who investigated the perception of eight minimal pairs in Spanish including /l-r/ and /r-r/ by third- and fourth-year L2 Spanish students in Taiwan using a forced-choice written-word identification task. Participants were given a sheet of paper with the written forms of 10 minimal pairs per contrast. They then heard one of the two pairs and had to circle which of the pairs they heard. Some difficulty was observed with /l-r/ (26% error rate); the /r-r/ contrast proved to be easier (12% error rate). The results are comparable to those reported in [Ortí Mateu \(1990\)](#), and taken together, these studies corroborate the observations discussed in the pedagogical literature (that Mandarin speakers have difficulty with the /l-r/ contrast in Spanish). Overall, while it is evident that Mandarin speakers of Spanish have difficulty with the perception and production of the /r-l/ contrast, it is unclear at what point during acquisition this contrast is acquired; moreover, our understanding of trill production by Mandarin speakers is negligible. Previous work has revealed that the trill is a difficult segment for Mandarin speakers; however, it is unknown what learners may produce when attempting a trill, and whether they make a contrast between the tap and trill in production. Consequently, a more controlled experimental study is crucial for determining difficulty and developmental patterns of these sounds.

While very little previous work has investigated the acquisition of the Spanish tap–trill contrast by native Mandarin speakers, [Falahati \(2015\)](#) examined the production of the Persian rhotic by L1 Mandarin speakers. When summarizing the previous literature, the author notes that native Persian speakers have been shown to produce several different variants for the rhotic, including taps, approximants, fricatives, trills. It is therefore difficult to know what the target is for learners of the Persian rhotic. Nevertheless, the author found that all seven of the L1 Mandarin speakers produced some trills, although they primarily produced either taps or approximants. Two of the learners also produced some laterals. The learners in this study were L3 Persian speakers, and spoke English as an L2. However, the role of the L2 was not examined.

In the next section, previous research on L1 English–L2 Spanish speakers will be presented, which will reveal recurrent developmental patterns reported for the L2 acquisition of the Spanish rhotic contrast, as well as demonstrate the facilitative role of the English flap when acquiring the tap. These findings help to shed light on what developmental patterns L1 Mandarin–L2 English–L3 Spanish speakers may follow, because they illustrate what role English could play in the acquisition of Spanish. A summary of previous work investigating the acquisition of the Spanish rhotics by speakers of L1s other than English and Mandarin will also be discussed.

1.3.2. The Acquisition of Spanish Rhotics by Speakers of Other Native Languages

A significant amount of previous research reveals clear, consistent acquisition patterns in the L2 acquisition of the Spanish rhotics by native English speakers. Regarding the tap, [Walton \(2005\)](#) and [Face \(2006\)](#) report very similar patterns of development. L2 learners first substitute the English [ɹ].

With greater proficiency, they tend to produce a mix of the English [ɾ] and the tap. Finally, at more advanced levels, learners master the tap. Similar patterns have been observed with the acquisition of the trill. The English [ɾ] is by far the most commonly produced segment by early learners. As the speakers become more advanced, those who have acquired the tap tend to produce it in place of the trill, before eventually acquiring the trill, which is particularly difficult (Waltmunson 2005; Face 2006; Johnson 2008). For example, Johnson (2008) found that Spanish students who had taken four semesters of Spanish were no more accurate than first-semester students, and it was not until the third or fourth year of learning that speakers began producing trills consistently. The most advanced speakers in the study (Spanish majors and graduate students) trilled at frequencies similar to those of native speakers, but even with L2 speakers of this level of proficiency, a small proportion of the speakers could not produce a trill (~5%; estimated from a histogram, exact numbers were not provided). Similarly, in Waltmunson (2005), one of 11 Spanish instructors with over 10 years of Spanish speaking experience was still unable to produce the trill. Thus, while some similarities are observed in the acquisition of the two Spanish rhotics, the trill is acquired much later than the tap; until the trill is acquired, it is usually substituted with the tap (or [ɾ] in less proficient learners). The fact that L1 English–L2 Spanish speakers produce [ɾ] in place of both rhotics is likely due to some form of perceptual categorization, as a result of similar orthographies and phonotactics.

A variable that has received relatively little attention but that merits consideration is whether the acquisition of the Spanish tap is conditioned by stress (i.e., whether the target rhotic is in a stressed or unstressed syllable). As was mentioned in Section 1.2, the English flap is an allophone of /t/ or /d/ that surfaces after a stressed syllable. The relative success that L1 speakers have when learning Spanish has been argued to be partially due to the flap allophone (Colantoni and Steele 2008; Olsen 2012). If this is the case, tap production in post-tonic syllables should be easier (and thus more accurate) for English speakers due to L1 influence; this is what was found in Olsen (2012). The L1 English–L2 Spanish speakers in his study were more accurate producing the tap in post-tonic (62% accuracy) compared to tonic position (45%). We might expect the same asymmetry in accuracy to be present in L1 Mandarin–L2 English–L3 Spanish speakers, if mastery of the L2 English flap facilitates acquisition of the L3 Spanish tap and such learners have acquired the former segment.

While the acquisition of the Spanish tap–trill contrast by L1 English speakers has been well-researched, less work has investigated the acquisition of the tap–trill contrast by speakers of other L1s. As was mentioned in Section 1.1, Kopečková investigated the production of the L3 Spanish tap and trill by L1 German–L2 English children, aged 11–12. She found that the tap was used as a substitute for the trill. Moreover, she found that the L1 was a stronger source of transfer. In another study with a language pairing similar to the present study, Morales Reyes et al. (2017) investigated the acquisition of intervocalic taps and trills by six L1 Korean–L2 English–L3 Spanish, 19 L1 English–L2 Spanish, and nine L1 Spanish control children (ages = 4–8). With respect to the tap, the L3 Spanish learners produced primarily taps or tap-like realizations (80.0% of the time), as well as a small proportion of other segments, such as alveolar approximants (8.5%), and trill-like segments consisting of a tap followed by frication (5.7%). No [ɾ] substitutes were produced, by either the L2 or L3 Spanish learners. With respect to the trill, the L3 Spanish speakers tended to produce either taps followed by a fricative portion (similar to a trill, with frication instead of a clear closure; 40.0% occurrence), and fricatives (36.6%). Only one trill was produced. Interestingly, both groups of learners produced almost no tap substitutions, which is what we should expect, given the pattern observed in the older children in Kopečková's (2014) study, and the results from studies on L1 English–L2 Spanish adults (e.g., Face 2006; Johnson 2008). Overall, the Morales Reyes et al. (2017) study revealed very little transfer from the L1 or L2 in the L3 speakers. These results contrast with those from other previous studies on the acquisition of rhotics. However, the differences observed should be expected to some extent, given the very different learner profiles (i.e., young children compared to older children and adults). Moreover, the lack of transfer observed in the L2 and L3 Spanish learners in the Morales Reyes et al. (2017) study may have been partly due to the fact that the young children would have had relatively little exposure

to orthography. As will be discussed in the next section, orthography plays a key role in the acquisition of non-native segments.

1.3.3. Orthographic Transfer in L2 Acquisition

Previous work on the L2 acquisition of the Spanish rhotics has revealed that orthographic transfer is a likely contributor to the English [ɹ] substitutes observed in production. This is supported by the literature on orthographic transfer. Numerous studies have found that orthography can have either a facilitative effect (e.g., Escudero et al. 2008; Showalter and Hayes-Harb 2013) or a non-facilitative effect (e.g., Bassetti and Atkinson 2015; Rafat 2015, 2016; Hayes-Harb and Cheng 2016) on the acquisition of non-native segments. For example, Rafat (2016) investigated the production of Spanish segments by naïve English speakers of Spanish. The naïve Spanish speakers were presented with pictures accompanied by a corresponding audio form of five grapheme-phoneme pairs that are different in English and Spanish, such as <v>–/b/ and <d>–/ð/. Some speakers were also presented with the corresponding orthographic form either when learning the sounds, when performing the production task, or both, whereas one group of speakers was never presented with the orthographic form. The latter group (audio-only condition) demonstrated a transfer rate of only 8% overall, with just two of the five consonant pairs revealing any transfer. In contrast, the three groups that were presented with the orthographic forms demonstrated overall transfer rates that ranged from 43% to 54%, depending on the condition, and transfer was observed for all five of the consonant pairs. These results clearly demonstrate that orthographic transfer must be considered as a variable that can influence non-native production. Nevertheless, one of the questions that arises is whether speakers of a language who are less dependent on a Roman writing system (i.e., Mandarin speakers) will still be influenced by orthographic transfer when learning a language that uses a Romanized writing system (e.g., English, Spanish). Very few studies have examined this, but the answer appears to be affirmative. Detey (2009) investigated the role of orthographic influence in 120 L1 Japanese–L2 French speakers who had been studying French for 1–2 years in university. Japanese speakers primarily use a non-Romanized writing system (a morphophonemic system called *kanji*, and two moraic systems called the *katakana* and the *hiragana*; Detey and Nespoulous 2008). However, like Mandarin, they learn and use to a lesser extent a Romanized script called *romaji*. In *romaji*, the Japanese /ɹ/ is transcribed as <r>. Participants performed a forced choice perception task. They listened to the auditory form of a French non-word (e.g., *ladeko*), and had to select what they heard from two options presented to them in orthographic form (<radeko> versus <ladeko>). Confusion rates were similar (23.7% for /l/, 23.5% for /ɹ/), but the authors argue that they should have been much higher for /l/, given its greater similarity to the Japanese /ɹ/ (which also has an /l/ allophone) than the French /ɹ/, which is not acoustically similar to /ɹ/. The authors conclude that orthography was an important factor. In terms of orthography, Japanese is similar to Mandarin. Therefore, the fact that Japanese speakers are influenced by orthography when learning a French suggests that Mandarin speakers may be susceptible to orthographic transfer when learning Spanish.

To recapitulate, previous research has revealed that L1 Mandarin speakers have difficulty with the /l-ɹ/ contrast in Spanish, which is likely at least partly due to difficulty perceiving the contrast. They also have difficulty producing the trill, as do L1 English speakers, which can be attributed to the precise aerodynamic conditions required for its production. Moreover, both L1 groups tend to initially substitute the same segment for the tap and trill. While Mandarin speakers may be less susceptible to orthographic transfer than native English speakers, it must be considered a possibility in Mandarin speakers as well, given their use of the Pinyin system and the fact that rhotics in that system are represented orthographically with <r>.

1.4. The Present Study: Research Questions and Predictions

The present paper has two goals. First, given that previous research on the acquisition of the Spanish rhotics has focused primarily on L1 English–L2 Spanish speakers, the goal here is to investigate

the production of the Spanish rhotics by speakers of a typologically distinct L1. Second, our current knowledge of L1 and L2 CLI in L3 acquisition is limited, due to the complexity of L3 acquisition. More research is needed from different language combinations and/or different structures, in order to obtain a comprehensive understanding of L1 and L2 CLI in L3 speech. Therefore, the second goal is to examine the extent to which the L1 and the L2 play a role in the production of the L3 Spanish rhotics. Of particular interest is whether the L2 flap allophone facilitates the acquisition of the L3 Spanish tap; and whether L1 or L2 segments are more likely substitutes for the L3 Spanish rhotics. To achieve these goals, the study was designed to answer the following questions, which are presented below with their respective hypotheses.

RQ1. What developmental patterns do L1 Mandarin–L2 English–L3 Spanish speakers exhibit when acquiring the Spanish tap?

RQ2. What developmental patterns do L1 Mandarin–L2 English–L3 Spanish speakers exhibit when acquiring the Spanish trill?

RQ3. Does the production accuracy of the Spanish tap and trill increase with higher L3 Spanish oral proficiency?

RQ4. Do L1 Mandarin–L2 English–L3 Spanish speakers transfer L1 or L2 segments?

RQ5. Does the ability to articulate an L2 English flap facilitate acquisition of the L3 Spanish tap?

Hypotheses

Hypotheses 1 (H1). *Based on patterns observed in L2 Spanish tap acquisition (e.g., Ortí Mateu 1990; Waltmunson 2005; Face 2006), learners are expected to initially produce a single non-native substitution (see Hypothesis 4 for expected substitutions). As learners begin to acquire the tap, the taps are expected to gradually replace the substitutions, until primarily taps are produced.*

Hypotheses 2 (H2). *As observed in L1 English–L2 Spanish trill production (Waltmunson 2005; Face 2006; Johnson 2008), speakers are expected to initially produce the same substitute for the trill as they do for the tap, until the tap is acquired; tap substitutions are then expected, until the trill is acquired.*

Hypotheses 3 (H3). *Based on the finding that L2 Spanish rhotic production accuracy is generally higher in more proficient speakers (Face 2006), the frequency with which learners produce target taps and trills is expected to increase with higher L3 Spanish oral proficiency.*

Hypotheses 4a (H4a). *If the TPM and L2SF are applicable to L3 phonology, we should expect primarily L2 English [ɹ] substitutions.*

Hypotheses 4b (H4b). *If the L1 or dominant language has a privileged status, we should expect primarily [l] substitutions.*

Hypotheses 5 (H5). *As predicted by the CEM, L1 Mandarin speakers who can produce an L2 English flap are expected to produce the L3 Spanish tap with higher levels of accuracy, especially when producing the tap in post-tonic position (i.e., where positive transfer from their L2 English is most direct).*

2. Materials and Methods

2.1. Participants

2.1.1. Language Profile

Twenty L1 Mandarin–L2 English–L3 Spanish speakers and 10 L1 Spanish–L2 English controls participated in the study. Participants were recruited via posters and advertisements posted on Spanish course websites at a major Canadian university and were paid CDN\$10 for their participation. The inclusion criteria for participation were as follows: (1) Participants had to be native speakers of Mandarin, having grown up in a city where Mandarin was the official language, and they had to have

spoken Mandarin at home with their parents; (2) Participants had to have learned English as a second language in school, in China (speakers who had moved to an English-speaking country as children (and that could be considered heritage speakers) were not accepted); (3) Participants had to be enrolled in a university-level Spanish course, or had to have completed (recently) a minimum level of Spanish equivalent to first-year-university Spanish.

Participants completed a language background questionnaire which ensured that they met the inclusion criteria. The majority of participants were completing either first or second-year Spanish although six upper-year students who had completed at least six semesters of Spanish courses also participated. Ideally, more third- and fourth-year students would have been included in order to obtain a clearer picture of the full developmental path of acquisition, but participants that matched the other requirements for participation were sparse in upper-year Spanish courses. The participants were on average 20.0 years old ($SD = 1.32$).

Regarding L2 English experience, most speakers had begun learning English around the age of seven, but were not immersed in English until much later, around the age of 15 when they moved to Canada. Thus, while the L2 English oral proficiency varied to some extent between participants (see Section 2.1.2 for further details), their type of education and experience with English was similar. Note also that all 20 participants continued to use their L1 Mandarin on a daily basis, and indicated that they were more comfortable communicating in Mandarin than English. The speakers are therefore considered to be L1 Mandarin-dominant. A summary of the participant characteristics is displayed in Table 2, whereas a detailed description of individual L1 Mandarin–L2 English–L3 Spanish participant profiles is provided in Appendix A.

Table 2. Summary of L1 Mandarin–L2 English–L3 Spanish speaker participant profiles.

L2 English			L3 Spanish		
AoA	LoRE	Eng-Man use	AoA	LoRS	SPA
7.20	4.25	50.8	17.9	0	3.70

Notes. AoA = age of onset of acquisition (age of first exposure to the language, in years); LoRE = length of residence in an English speaking country (in years); LoRS = length of residence in a Spanish speaking country; Eng-Man use = how much L2 English is used compared to L1 Mandarin at work, home, school, and in social situations (expressed as a %); SPA = semesters of Spanish courses completed.

Regarding the 10 L1 Spanish–L2 English controls that participated, all were born and had spent at least the first 12 years of their lives in a Spanish-speaking country (Spain, Cuba, Argentina, and Mexico). When growing up in their home country, they spoke only Spanish at home, and were educated in Spanish. They were living in Canada at the time of testing, but stated that they spoke Spanish on a daily basis with friends and family. All speakers spoke dialects that produced the alveolar tap and trill. They were on average 32.1 years old ($SD = 6.11$).

2.1.2. Oral Proficiency

Spanish and English oral proficiency were determined via accent ratings. Accentness ratings specifically target oral proficiency (as opposed to proficiency in other domains, such as the lexicon or morphosyntax), thus are one of the more relevant measures for determining oral proficiency in a non-native language (Colantoni et al. 2015, p. 89). For this reason, they have often been used to establish oral proficiency in previous work (e.g., Colantoni and Steele 2007, 2008; Kopečková 2016; Lloyd-Smith et al. 2017). All L1 Mandarin–L2 English–L3 Spanish participants were required to read ‘The North Wind and the Sun’ in English as well as its Spanish equivalent. The Spanish recordings were presented in random order to native ($N = 10$) Spanish speakers, who listened to and rated on a

scale of 1–5 how strong they felt each speaker’s accent was⁷. The scores from all judges were averaged, and the resulting values were used as measures of each speaker’s overall oral proficiency. The English recordings were also rated by 10 native English speakers using the same procedure. Cronbach’s alpha was calculated for both sets of ratings in order to determine interrater reliability. Results were $\alpha = 0.892$ and $\alpha = 0.938$ for the Spanish and English accent ratings, respectively, both of which indicate a high degree of interrater agreement. A dependent samples two-tailed *t*-test comparing the mean L3 Spanish accent rating ($M = 1.75$, $SE = 0.15$) to the L2 English one ($M = 2.89$, $SE = 0.25$) revealed that the participants’ L2 English proficiency was higher overall than their L3 Spanish proficiency ($t = -8.603$, $df = 19$, $p < 0.001$). For individual oral proficiency ratings, see Appendix A.

2.2. Experimental Tasks

The L3 Spanish participants performed tasks in three languages—Spanish, English, and Mandarin, whereas the L1 Spanish controls only performed the same Spanish task. These tasks will be discussed in the same order in which the tasks were presented to the participants.

A word-reading task was used to elicit the Spanish rhotics. Participants were presented with the orthographic form of each target word (e.g., *perro*), one after another, on a computer screen. Each slide was advanced at a steady (untimed) pace by the experimenter. The participants were required to produce the word when they saw it appear on the screen. One of the motivations for presenting the speakers with the orthographic form was that it ensured the speakers would attempt to produce the segment of interest (e.g., tap or trill), given that they are represented by <r> and <rr>, respectively. While a more complex task would have been preferred, the least experienced speakers had very low Spanish oral proficiency and little experience speaking Spanish. Pilot testing revealed that reading the passage ‘The North Wind and the Sun’ in addition to producing a series of sentences was tiring for some speakers. Pilot testing was also used to assess the feasibility of a picture naming task; however, many speakers had difficulty identifying the pictures they were presented with. Therefore, in order to avoid articulatory fatigue and to be able to include even the least proficient speakers in the study, a word reading task was considered to be the most viable option.

The motivation for analyzing the L2 English production of the participants was to examine whether they had acquired, and could therefore transfer, the English flap and the English /ɹ/, both of which were expected to be sources of L2 transfer. The primary task used to elicit the two English segments was also a word-reading task. However, the orthographic stimuli were accompanied by their corresponding audio form, for two reasons. First, the flap is written with <t> or <d> (e.g., <water> /wɑtəɹ/ [ˈwɑ.ɹə]; <ladder> /ˈlædəɹ/ [ˈlæ.ɹə]), which could result in orthographic influence and therefore the production of [t] or [d] in place of [ɹ]. Second, participants may have been exposed to English varieties that do not flap (e.g., British English). It is thus possible that the participants have acquired the flap, but do not always use it when speaking English. The audio form ensured that the participants attempted to produce a flap, and not some other segment. Note, however, that one of the limitations of a repetition task is that it may not indicate the extent to which a speaker uses the target segment in context. Consequently, successful production of [ɹ] or [ɹ̠] in the word-reading task may not reveal conclusively whether the speaker had actually acquired the two L2 English segments. Therefore, the production of the English passage ‘The North Wind and the Sun’ that participants read before the word-reading task (used to determine L2 English oral proficiency) was also examined. The passage elicited the production of four flaps (disputing [dɪs.ˈpju.ɹɪŋ], immediately [ə.ˈmi.ɹ.jət.li], succeeded [sək.ˈsi.ɹəd], considered [kən.ˈsɪ.ɹəd]) and numerous /ɹ/ tokens. The production of the flaps and four of the English /ɹ/ (in various positions: agreed [ə.ˈɡɪd], wrapped [ɹæpt], around [ə.ˈraʊnd], more [mɔ.ɹ]) were analyzed to determine whether speakers who successfully realized [ɹ]

⁷ The following scale was provided to judges: 1 = ‘clearly non-native, very strong foreign accent’; 2 = ‘strong foreign accent’; 3 = ‘noticeable foreign accent, but not too strong’; 4 = ‘almost no accent’; 5 = ‘no accent (native speaker)’.

and [ɹ] in the word-reading task also produced these segments in the reading passage. Speakers who consistently produced the targets in the word-reading task and in the reading passage were assumed to have acquired the segment. In contrast, the speakers who produced the targets accurately in the word-reading task, but not in the reading passage, were excluded from the analyses that specifically examined the extent to which L2 transfer occurred. For these speakers, it would not be possible to establish without doubt whether or not they had acquired the segments of interest.

The third task that participants performed was a word reading task in Mandarin, which was designed to elicit Mandarin intervocalic /ɹ/. The primary reason for eliciting the Mandarin rhotic was to be aware of the characteristics present in each individual's L1 rhotic productions, in the case that the same characteristics were observed in the learners' L3 Spanish rhotic productions. While transfer of the Mandarin rhotic was expected to be minimal, it had to be considered as a possible candidate, given that it is a rhotic, that it is represented by <r> in Pinyin, and that the dominant language should be considered a potential source of transfer.

2.3. Stimuli

The Spanish word-reading task involved 24 target stimuli in total: six stimuli per stress condition × two stress conditions (tonic versus post-tonic) × 2 rhotics (e.g., *corrí* /ko.'ri/ 'I ran', *carro* /'ka.ro/ 'car', *veré* /be.'re/ 'I saw', *pero* /'pe.ro/ 'but'). Care was taken to select frequent words that are learned in first-year Spanish, based on the first-year Spanish textbook that was being used at the university where the students were studying and a Spanish frequency dictionary (Davies 2006). Thirty-eight distractor stimuli were also included. The word list was read twice by each speaker, resulting in 48 rhotic and 76 distractor tokens per speaker.

A total of 60 stimuli were included in the English word-reading task. There were 10 stimuli for the flap in post-tonic position (e.g., *water* /wɑtəɹ/ ['wɑ.rə]), 10 stimuli for the English /ɹ/ in both post-tonic (*berry* /'bɛ.ɹi/) and tonic positions (*direct* /dɪ.'ɹɛkt/), and 30 distractors. The English stimuli were accompanied by their audio form. To create the English stimuli, a 24-year-old female native speaker of Canadian English was recorded reading the stimuli in isolation. She was instructed to read the list of stimuli at a normal rate, as clearly as possible. The list was read three times, and the best realization of each stimulus was chosen.

Twenty stimuli were included in the Mandarin task, which consisted of 10 target stimuli with the Mandarin rhotic /ɹ/ in intervocalic position (牛肉 *niú.ròu* 'beef'), as well as 10 distractors. The entire stimuli list for all three languages can be found in Appendix B.

2.4. Testing Protocol

Each participant was recorded individually in a quiet lab. Participants were first asked several questions about their language background, in order to acquire relevant data (e.g., age of acquisition, use) of previously learned languages. They were subsequently recorded reading 'The North Wind and the Sun' passage in Spanish. Participants were given time to read the passage once on their own before being recorded, in order to become familiar with the content. After the reading passage, participants performed the Spanish word-reading task. They were asked to produce each word that they saw on the computer screen. Participants were recorded in Spanish first because it was their weakest language, and the most likely to be affected by articulatory fatigue. Following the word reading task in Spanish, participants had a short break, and were then recorded reading 'The North Wind and the Sun' passage in English. As with the Spanish passage, they had the opportunity to read and become familiar with the English passage before being recorded. Participants subsequently performed the English word-reading task. They were given the same instructions as the Spanish word-reading task; however, they were instructed that they would also hear a recording of the word they were supposed to produce. They were asked to listen to the recorded word first, before reading the word on the screen. The English word-reading task was followed by a short break. The session concluded with participants reading the target stimuli in Mandarin Chinese, which followed the same procedure as

the Spanish word-reading task. All productions were recorded using a Marantz PMD561 recorder and a unidirectional condenser microphone.

2.5. Data Preparation and Analysis

All words containing the target segments were extracted and examined acoustically in Praat (Boersma and Weenink 2017). The segmental boundaries were marked, which were determined based on a decrease in intensity (for all segments) in the waveform and the spectrogram (Figure 1), and, for the English and Mandarin rhotics, a decrease in F3. Some Mandarin /ɹ/'s were produced as fricatives. For these segments, boundaries were marked according to the changes in periodicity. After marking segment boundaries, the type of segment produced (e.g., [r], [r], [l]) was identified through a visual analysis of acoustic cues present in the waveform and spectrograms (e.g., changes in intensity and formants, the presence/absence of noise), combined with an auditory analysis. A total of 720 L3 tap (30 speakers × 24 tokens), 720 L3 trill, (30 speakers × 24 tokens), 280 L2 flap (20 speakers × 14 per tokens), 480 L2 /ɹ/ (24 per speaker), and 200 L1 /ɹ/ tokens (10 per speaker) were analyzed. Segments were considered taps/flaps if there was a brief closure and trills if two or more rapid closures were observed. Regarding the English and Mandarin rhotics, segments were coded as /ɹ/ or /ɹ/ (respectively) if a noticeable drop in F3 was visible. The Mandarin rhotics produced as fricatives did not have a decrease in F3, and were identified through the presence of noise. These productions were coded as [ʒ]. Examples are displayed in Figure 1. All coding was performed by the author, a native English, near-native Spanish speaker trained in acoustic analysis.

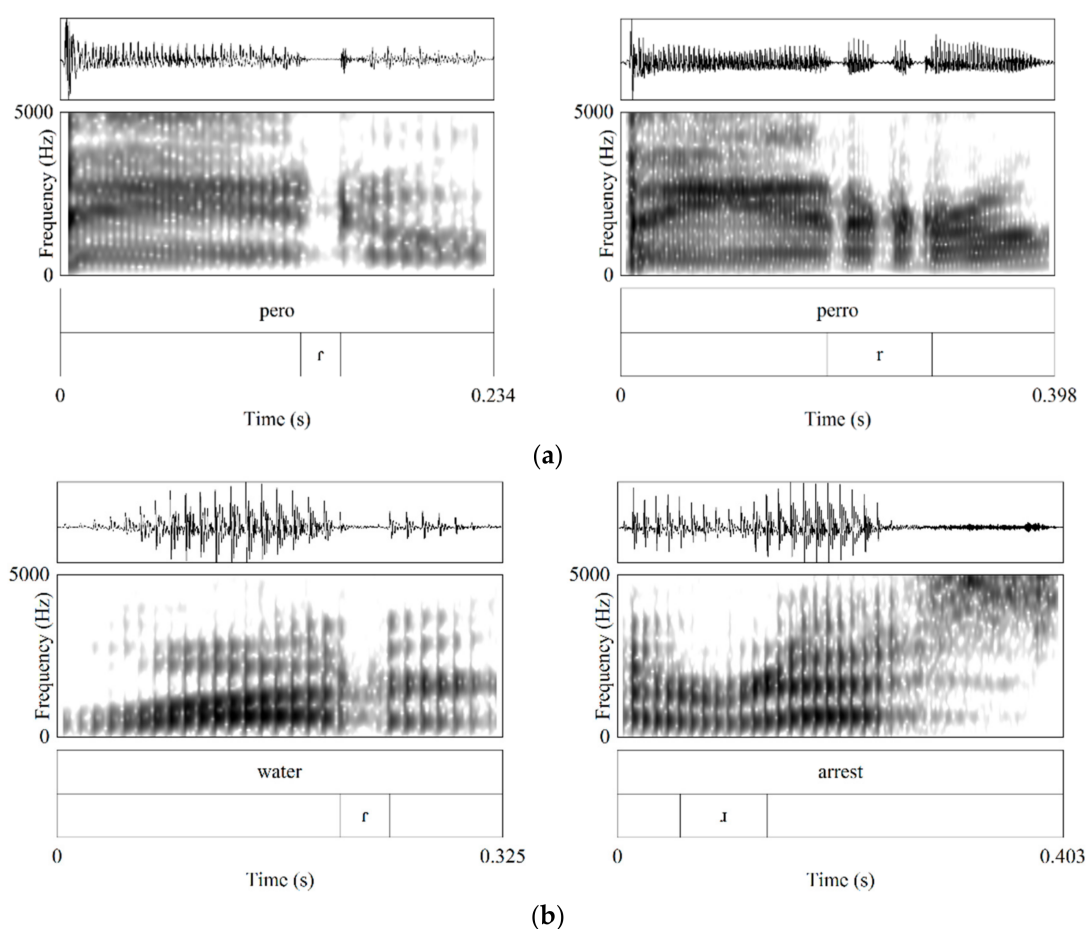


Figure 1. Cont.

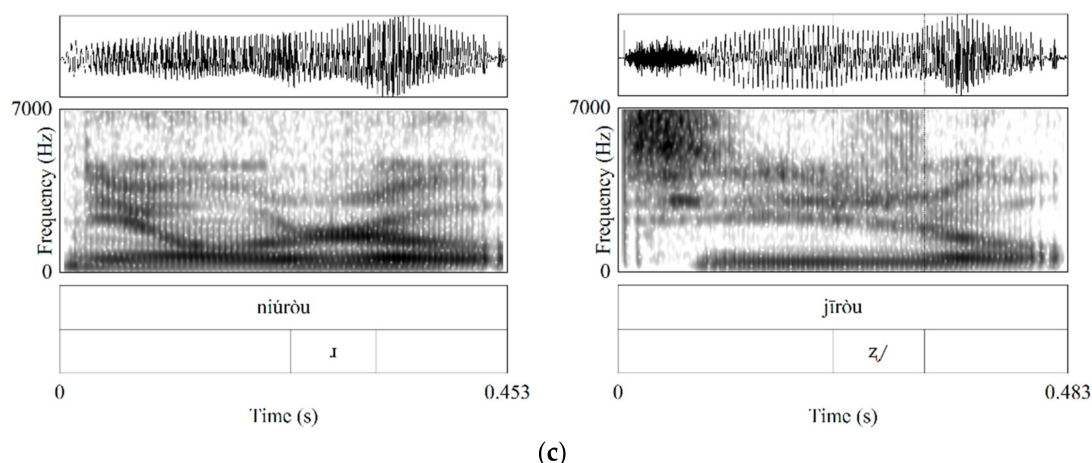


Figure 1. (a) Example of a typical tap (left) in Spanish *pero* /'pero/ 'but' and a typical trill with three closures (right) in *perro* /'pero/ 'dog' as produced by a native Spanish speaker (S01); (b) Example of a typical L2 English [ɾ] (left) in *water* /'wɔtəɹ/ and English /ɹ/ (right) in *arrest* /ə'ɹɛst/ as produced by an L1 Mandarin–L2 English–L3 Spanish speaker (M05); (c) Examples of the Mandarin rhotics. A typical approximant production [ɹ] (left) in Mandarin *niú ròu* /nju.rou/ 'beef' and a typical fricative production [ʒ] (right) in Mandarin *jī ròu* /ti.rou/ 'chicken meat', as produced by an L1 Mandarin–L2 English–L3 Spanish speaker (M01).

After segmenting and coding the data, the identified segmental realizations were extracted and used to calculate the frequency of production of each target and non-target segment. To determine the role of predictors on the outcome of target versus non-target tap/trill production, a series of mixed effects binomial logistic regression models were run. Each model included a random intercept for 'participant'. Categorical predictors were coded using treatment coding. All statistics were run in SPSS v. 23 (IBM, Armonk, NY, USA), with a significance level of $p = 0.05$.

3. Results

The results of the experiment are presented in this section, starting with the production results for the L3 tap. This is followed by a comparison of the accuracy rates of the L2 English flap and the L3 Spanish tap, in order to determine whether the L2 flap facilitates the acquisition of the L3 tap. The section concludes with the presentation of the results for the trill.

Regarding the production of the tap, recall that primarily either [l] or [ɹ] were expected initially. Results revealed that, at the group level, a [ɹ] was produced 42.9% of the time by the L3 Spanish speakers. Tap targets were also frequently produced as laterals [l], making up 27.6% of all productions. The remaining 29.5% consisted of English [ɹ] approximant productions (11.8%), stop-liquid clusters [dɹ]⁸ (6.2%), stops [d] (6.2%), and sporadic productions of a variety of other non-target segments (5.3%) (approximants [ʃ], fricatives, [ð], trills [r] and deletions). No productions of L1 Mandarin [ɹ] were observed. In contrast to the L3 Spanish learners, native Spanish speakers produced taps 98.7% of the time, in addition to stops in very few instances (1.3%). A mixed effects binomial logistic regression comparing the target tap production of the learners versus the controls revealed that the learners were significantly less accurate ($\beta = -5.262$; $SE = 1.132$; $t = 4.648$; $p < 0.001$).

The segments produced by the learners, arranged in increasing order of L3 tap accuracy, are displayed in Figure 2. Each bar represents the percentage realization of taps and other non-target

⁸ An example of a non-target [dɹ] cluster in place of the tap is [pedro] instead of /pero/ 'but'. The [dɹ] clusters were also produced in place of the trill by some speakers.

segments, for each participant, among each individual's productions. The final bar represents the percentage realization by the native Spanish speakers (averaged together).

While the native speakers produced almost exclusively taps, and there was an increase in the proportion of taps with oral proficiency among the L3 learners, a significant amount of variation was observed in the non-native productions. Four of the learners were unable to produce a single tap (M13, M12, M16, M08), while an additional three (M10, M06, M14) only produced taps less than 15% of the time. These learners also had seven of the nine lowest L3 oral proficiency ratings. Five speakers (M20, M05, M04, M01, M11,) were able to produce the L3 tap with very high accuracy rates (80% or more). The rest of the speakers produced the L3 tap with accuracy rates ranging from 29.1% to 75.0%. The most frequently produced substitute was [l], making up 48.3% of non-native realizations, followed by English [ɹ] (23.0% of non-native realizations), [d] (12.1%), and [dr] (12.1%). A mixed effects binomial logistics regression was run on the learner data to analyze whether the L3 oral proficiency was a predictor of accurate tap production. The results revealed that it was, with target tap production increasing as the L3 oral proficiency increased ($\beta = 3.378$; $SE = 0.734$; $t = 4.000$; $p < 0.001$).

The production results indicate what appear to be three developmental stages. At the first stage, speakers almost categorically substituted a single segment for the L3 tap—usually [l] but in some cases English [ɹ] or the sequence [dr]. At the second stage, speakers began to acquire the tap but also produced a variety of other non-target segments (e.g., stops, approximants). Finally, at the third stage (M05, M01, M04, M11, M03), speakers produced primarily taps (at least 90% of the time).

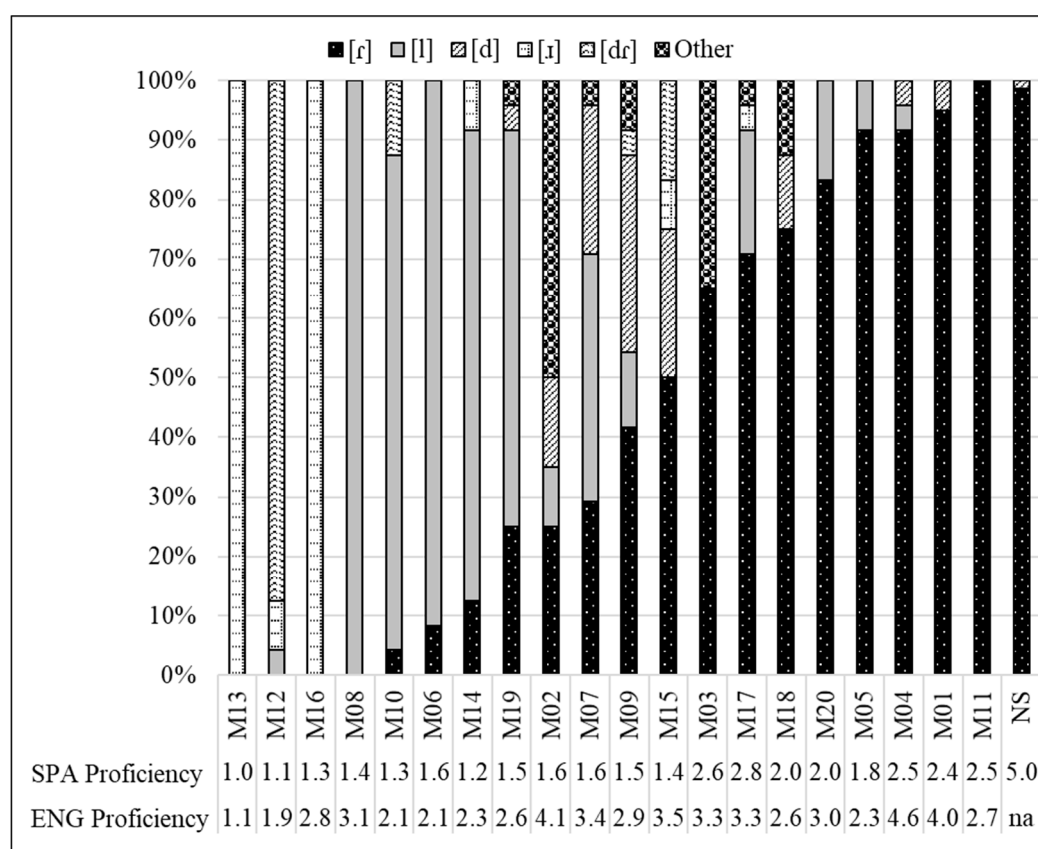


Figure 2. Percentage realization of each segment for the Spanish tap target, as produced by L1 Mandarin–L2 English–L3 Spanish speakers (M01–20) and native Spanish speakers (NS; average results presented in the final column). Results are ordered by percent accuracy of L3 tap production, from lowest to highest. ‘Other’ refers to [ð], [ð], [r], or deletions.

The low occurrence of L2 English [ɹ] productions raises the question of whether speakers did not substitute [ɹ] with more frequency because they had not yet acquired the segment, or because most

learners did not consider it to be a valid substitute. An analysis of the L2 English productions revealed no difficulty with /ɹ/. All speakers produced a target [ɹ] at least 80% of the time in the word-reading task. In the passage-reading task, all speakers produced a target [ɹ] 100% of the time, except one speaker (M13) who produced [ɹ] 75% of the time. These results indicate that all speakers had acquired /ɹ/, and used it consistently when speaking English. Therefore, while early L1 English–L2 Spanish speakers tend to transfer the English [ɹ] (e.g., Waltmunson 2005), this was generally not the case for the L1 Mandarin–L2 English–L3 Spanish speakers (other than M13 and M16). Interestingly, one of the two speakers who consistently produced [ɹ] was the least proficient L2 English speaker (1.1 oral proficiency), whereas some of the more proficient L2 English speakers with the lowest levels of Spanish proficiency (e.g., M07, M08) substituted primarily [l] and not [ɹ]. This finding suggests that the use of [ɹ] did not correlate with L2 English proficiency. None of the other non-native segments used at each stage appear to depend on L2 English proficiency, except for perhaps [d]. The [d] substitute occurred with the greatest frequency in the productions of four of the five most proficient L2 English speakers (M02, M07, M09, M15), although it was also produced to some extent by one speaker with a somewhat lower proficiency (M18). To determine whether any correlations were present between the type of non-native substitutions and L2 English oral proficiency, a series of Spearman’s rho correlations were run (Table 3), comparing each speaker’s L2 English oral proficiency score with the proportion of non-native substitutions produced. A strong significant correlation between [d] productions and L2 English proficiency was observed ($\rho = 0.53$; $p = 0.016$), confirming that [d] is more likely to be used by more advanced L2 English speakers. No other significant correlations were found.

Table 3. Spearman’s rho correlations comparing L1 Mandarin–L2 English–L3 Spanish speakers’ L2 English oral proficiency with the proportion of the five most frequently produced non-native substitutions for the L3 Spanish tap target.

	Segment				
	[ɹ]	[l]	[d]	[ʔ]	[dr]
ρ	−0.29	−0.18	0.53	0.19	−0.20
<i>p</i> -value	0.213	0.440	0.016 *	0.428	0.404

Notes. * = $p < 0.05$. Values in bold are those that reached significance.

In order to determine whether the ability to produce the L2 flap might play a factor in the acquisition of the tap, the percent accuracy of L2 flap production from the word-reading task and the passage-reading task was examined. The proportion of target-like L2 flaps realized by speaker and task is displayed in Table 4. In general, the most successful flap producers in the word-reading task also produced the flap in the passage-reading task, which indicates that these speakers were not only capable of producing L2 flaps, but that they produced them regularly when speaking English. Three speakers (M03, M06, M10) produced some flaps in the word-reading task, but none in the passage-reading task. These speakers were therefore not included in the analysis of whether the ability to produce an L2 flap might facilitate the acquisition of the L3 tap, given that it is not conclusive whether these speakers produce flaps in English or not.

Figure 3 displays the percent accuracy of L2 flap production from the word-reading task (horizontal axis) compared with the percent accuracy of the L3 tap production (vertical axis). If the production of the L2 flap facilitates acquisition of the L3 tap, we would expect L2 flap accuracy rates to be a significant predictor of L3 tap accuracy.

The scatterplot indicates that the five speakers with L2 flap accuracy rates that were lower than 50% (i.e., those who had not acquired the flap) also had low accuracy rates producing the L3 tap. Of the twelve speakers with at least a 50% L2 flap accuracy production, eight produced the L3 tap with a 50% or higher accuracy rate (and six of them with a much higher accuracy rate). The results suggest that L2 flap accuracy does indeed correlate with L3 tap accuracy to some extent, an observation that was confirmed with a mixed-effects binomial logistic regression. L2 flap accuracy was found to be

a significant predictor of target L3 tap production ($\beta = -4.954$; $SE = 2.069$; $t = -2.395$; $p = 0.017$)⁹. Note that the predictor ‘stress’ (tonic-post-tonic) was not significant ($\beta = -1.257$; $SE = 0.760$; $t = -1.654$; $p = 0.099$), nor was there any interaction between ‘flap accuracy’ and ‘stress’ ($\beta = 1.726$; $SE = 1.064$; $t = 1.623$; $p = 0.105$). These results reveal that the more successful L2 flap producers were more likely to produce an accurate L3 tap overall, but that it was no different in tonic versus post-tonic position. These data support the hypothesis that the ability to produce the L2 flap facilitates the production of the L3 tap. However, it was not the case that all speakers who had acquired the L2 flap were necessarily able to produce the L3 tap, as four speakers had high accuracy rates producing the L2 flap, but low accuracy rates producing the L3 tap.

Table 4. Percent accuracy realization of the L2 English flap in the word-reading task and the passage-reading task.

Participant	% Accuracy	
	Word Reading	Passage Reading
M13	0.00	0.00
M16	0.00	0.00
M12	0.10	0.25
M09	0.20	0.25
M10	0.20	0.00
M06	0.30	0.00
M14	0.30	0.25
M11	0.50	0.25
M18	0.50	0.50
M20	0.50	0.50
M03	0.60	0.00
M15	0.60	0.50
M04	0.70	0.75
M05	0.80	0.75
M07	0.80	1.00
M08	0.80	0.50
M17	0.90	0.75
M01	1.00	0.75
M02	1.00	0.75
M19	1.00	0.75

We will now turn to the results for trill production. Recall that the trill production was expected to be particularly difficult, and that learners were expected to substitute the same segments they substitute for the tap, until the trill is acquired. Overall, the trill was, as expected, much more difficult than the tap for the L3 speakers. While the control group produced trills 89.2% of the time, only 14.1% of productions by the learners were trills. A mixed effects binomial logistic regression comparing target productions by each group revealed that this difference was highly significant ($\beta = 6.975$; $SE = 1.245$; $t = 5.602$; $p < 0.001$).

Trill targets were produced most frequently as laterals [l] (22.9%) and taps [ɾ] (22.4%), followed by English [ɹ] (13.7%), [dɹ] clusters (13.2%), approximants [ɻ] (4.5%), fricatives [ð]/[ʃ] (4.3%), and stops [d] (3.6%). The remaining 1.3% consisted of infrequent productions of [w], [dl], [ðɹ]. The approximants were generally similar to the approximants substituted for the tap (very open dental approximants [ɻ]), although in some cases they were produced with a particularly long duration (>100 ms), which was not the case with the tap target. The fricatives that were produced varied, ranging from short

⁹ Given that three speakers were omitted from the analysis, the binomial logistic regression was run on a smaller data set, totaling 400 observations (across 17 speakers). The standard error for the predictor L2 flap accuracy was slightly more than 2, revealing a moderately high degree of variability. Consequently, these results should be considered with caution.

dental or alveolar fricatives as observed in the tap substitutions, to long, failed alveolar trills. They did not resemble the L1 Mandarin fricatives [ʐ].

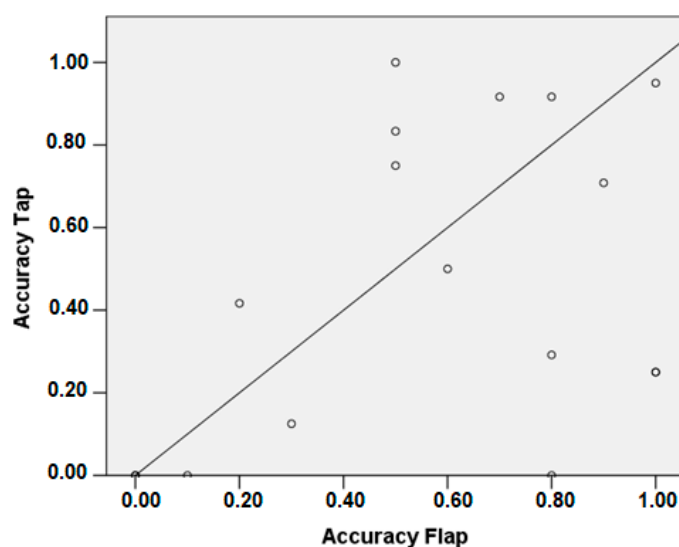


Figure 3. Comparison of the production accuracy rate of the L3 Spanish tap (Y-axis) with the production accuracy rate of the L2 English flap (X-axis), produced by L1 Mandarin–L2 English–L3 Spanish speakers ($n = 17$).

The individual results are presented in Figure 4, ordered by the percent accuracy of targetlike trills, from the least to most accurate. Only five of the twenty participants were able to produce trills, and of those five, only two (M05, M01) produced trills with native-like accuracy levels. The least proficient learners tended to substitute a single segment, either [l], [ɹ], or [dɹ]. As speakers became more proficient, a variety of non-target segments were produced in place of the trill, including taps. Native speakers produced primarily trills (89.2%), as well as some fricatives (10.8%). In contrast to the tap results, a mixed effects binomial logistic regression revealed that the L3 Spanish oral proficiency of the learners was not a predictor of trill accuracy ($\beta = 1.781$; $SE = 1.625$; $t = 1.096$; $p = 0.274$).

The production patterns of the trill were similar to those of the tap, but there appear to be four stages of acquisition as opposed to three. At Stage 1 (M13 to M15), speakers almost categorically produced a substitution; [l] was produced most frequently, but [ɹ] and [dɹ] were also observed. This was followed by a stage where speakers produced a variety of non-target segments (M03 to M18), including taps, but also a number of other segments (e.g., [ð], [d], [ɹ]). The third stage (M20, M11, M19) involved some trill productions, but with low frequency. The final stage involved trill production with frequencies similar to those of native speakers. This stage was only achieved by two speakers (M05, M01).

When discussing the previous literature in Section 1, it was identified that L1 English–L2 Spanish speakers often do not make a contrast between the tap and the trill (i.e., speakers produce the same segment for the tap and the trill, beginning first with the English /ɹ/, followed by the tap once it is acquired). A question of interest in the present paper is whether this may be a universal strategy or whether it is a strategy specific to L1 English–L2 Spanish speakers. In order to determine whether the L1 Mandarin speakers tended to produce the same segment for the tap and trill, the productions for each speaker were analyzed and compared. Table 5 displays the most frequent productions by speaker, ordered by L3 Spanish oral proficiency, for each target (tap and trill). The shaded sections indicate which speakers produced the same segment for both targets (i.e., did not make a contrast), whereas the unshaded sections indicate the speakers that produced different segments (i.e., made a contrast). Note that speakers who could produce the trill (M01, M05, M11, M19, M20) were not included, given that they had already acquired the ability to produce a contrast. The results revealed that the six least proficient speakers produced the same segment for both rhotic targets, whereas

five of the six most proficient speakers produced different segments. This pattern differs from the pattern observed in L1 English–L2 Spanish speakers. Specifically, in contrast to L1 English–L2 Spanish speakers, L1 Mandarin–L2 English–L3 Spanish speakers begin producing a contrast as they become more proficient, and do not tend to use the tap in place of the trill. A possible explanation for the different patterns is discussed in detail in Section 4.2.

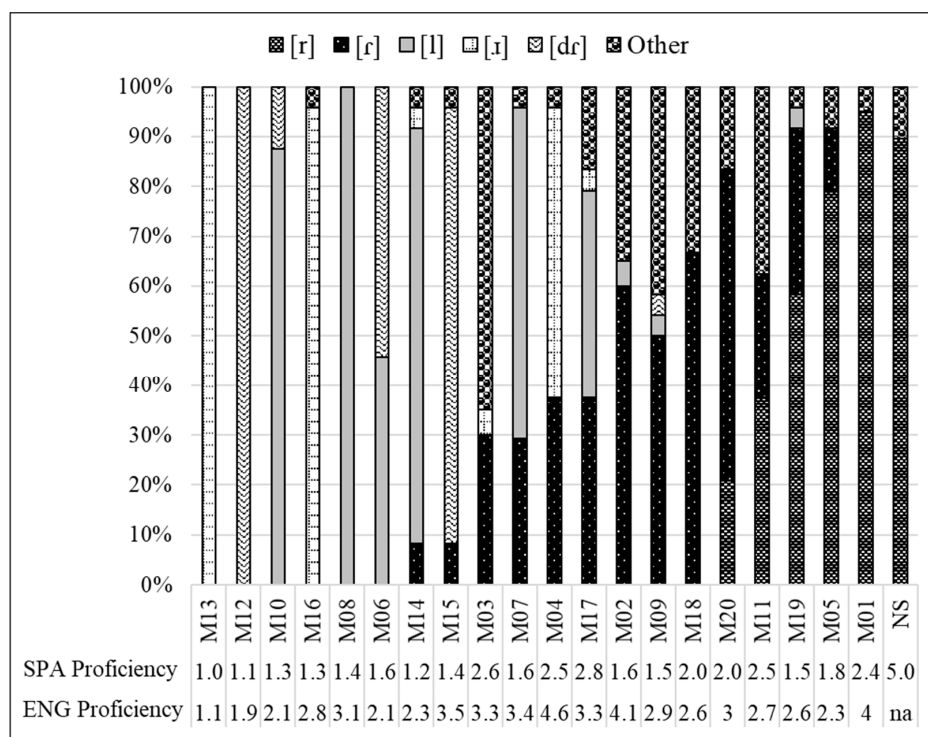


Figure 4. Percentage realization of each segment for the Spanish trill target, as produced by Mandarin L3 learners of Spanish (M01–M20) and native Spanish speakers (NS; average results presented in the final column). Results are ordered by percent accuracy of trill production, from lowest to highest. ‘Other’ refers to [ʔ], [ð], [d], [w], [dl], or [ðr].

Table 5. Comparison of the most frequently produced segment for tap and trill targets, by L1 Mandarin–L2 English–L3 Spanish speakers who could not yet produce the trill ($n = 15$), ordered by Spanish oral proficiency. Shaded cells indicate which speakers produced the same substitution for both targets.

Speaker	Spanish Proficiency	Target	
		Tap	Trill
M13	1.0	[ɹ]	[ɹ]
M12	1.1	[dr]	[dr]
M14	1.2	[l]	[l]
M10	1.3	[l]	[l]
M16	1.3	[ɹ]	[ɹ]
M08	1.4	[l]	[l]
M15	1.4	[r]	[dr]
M09	1.5	[r]	[r]
M07	1.6	[l]	[l]
M02	1.6	[ʔ]	[r]
M06	1.6	[l]	[dr]
M18	2.0	[r]	[r]
M04	2.5	[r]	[ɹ]
M03	2.6	[r]	[ʔ]
M17	2.8	[r]	[l]

4. Discussion

4.1. Summary and Discussion of Results

In the present study, L1 Mandarin–L2 English–L3 Spanish speakers performed a word reading task that was designed to elicit the production of the L3 Spanish tap and trill in intervocalic position. Learners displayed clear developmental patterns as they acquired the two rhotics. These patterns are summarized in Figure 5.

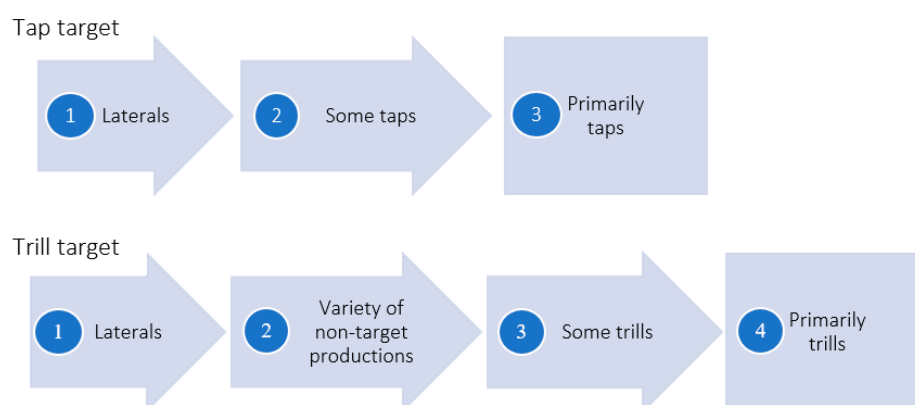


Figure 5. Developmental patterns of acquisition of the Spanish rhotics by L1 Mandarin–L2 English–L3 Spanish speakers.

Regarding the tap, the results suggest that, as predicted in H1, learners initially produce a single substitution (generally [l]). Once learners begin producing taps, in contrast to what was predicted (primarily taps and either [l] or [ɾ]), they also produce a variety of other non-target segments. Learners then begin to produce taps frequently, until the tap becomes the primary segment produced. Finally, some learners produce taps more than 90% of the time, which is similar to the frequency of tap production observed in native Spanish speakers. L3 Spanish oral proficiency was found to be a strong predictor of tap accuracy. Specifically, accuracy increased with proficiency, confirming the prediction put forward in H3.

Some variability was observed in the development of the tap, and L2 English was found to be a factor. Specifically, [d] substitutions, while not common overall, were produced primarily by the more advanced L2 English speakers. Moreover, the L2 English flap was found to be a facilitating factor in the L3 Spanish tap production accuracy. In general, participants who were unable to produce the L2 flap, were also unable to produce the L3 tap, whereas speakers who were able to produce the L2 flap were also able to produce the L3 tap. Nevertheless, four speakers who clearly had acquired the L2 flap still could not consistently produce target L3 taps, which indicates that speakers who have acquired the L2 flap do not ‘automatically’ acquire the L3 tap. This is not fully consistent with H5, which assumed that all learners who could produce the L2 flap would transfer it to their L3 Spanish. One possible explanation for this pattern is the allophonic status of the L2 flap. While contrasting phonemes result in semantic differences, allophones do not, which influences the awareness a speaker has of phonemes compared to allophones (Tobin 1997). Moreover, each phoneme is represented by a specific grapheme, whereas two allophones of the same phoneme are represented by the same grapheme. Native English speakers are therefore more likely to think they are producing [t] for the /t/ in water and not [ɾ]. We should expect the same to be true for the present study’s speakers as well. Consequently, if the L1 Mandarin–L2 English–L3 Spanish speakers are unaware of the fact that they produce the L2 flap allophone, they may not realize that they are already able to produce a tap-like sound, until they become more proficient and are able to re-categorize the L2 flap allophone as an L3 Spanish phoneme. The data support this proposal. The four speakers who could produce L2 flaps, but did not consistently produce L3 taps, were some of the least proficient Spanish speakers (oral proficiency ≤ 1.6 ; Scale: 1 =

‘clearly non-native, very strong foreign accent’; 2 = ‘strong foreign accent’). Note that similar findings are observed with L1 English–L2 Spanish speakers. At first, they produce [ɹ] in place of the L2 tap, even though they have the articulatory ability to produce a tap-like sound. It is not until they become more advanced that they recategorize the L1 flap allophone as an L2 tap phoneme. In sum, the results suggest that L2 allophones do facilitate the acquisition of L3 phonemes, just as L1 allophones facilitate acquisition of L2 phonemes. However, the re-categorization is not immediate.

Regarding the trill, the results revealed that initially, consistent with H2, learners almost categorically produce a single non-target segment. As learners become more proficient, a variety of non-target segments are produced. While some of these are taps, many learners who have acquired the tap do not use it as a substitute for the trill, in contrast to the prediction put forward in H2. The production of the trill eventually emerges. However, at this stage learners produce trills infrequently, and also produce several other non-trill segments. Finally, some speakers are able to produce trills frequently (at least 75% of the time). In contrast to the tap results, the L3 Spanish oral proficiency was not found to be a predictor of trill production accuracy, contrary to what was predicted in H3. While the least proficient Spanish speakers were not able to produce trills, some of the most proficient Spanish speakers were also unable to produce trills. This is in line with previous work, which has found that even learners with several years of experience speaking Spanish are not always able to produce trills (Waltmunson 2005; Johnson 2008). This is likely due to the high degree of articulatory complexity required to produce a trill, which highlights the importance of production difficulty in non-native speech learning.

The observed developmental trends of the tap and the trill diverge from the predictions put forward in Section 1.4 in three ways. First, while the tap was acquired before the trill (as predicted), most of the learners who had acquired the tap did not tend to use it as the primary substitute for the trill, contrary to what L1 English–L2 Spanish speakers do. This divergent pattern will be discussed in detail below when the results of the present study are compared with previous findings. Second, once learners began attempting to produce the tap and trill (i.e., the stage after categorical non-target substitutions), they also produced a large variety of non-target segments, more so than the least proficient learners who tended to substitute a single (non-target) segment. This is in contrast to what was predicted in H1 and H2, and will be examined in detail in Section 4.2, where developmental patterns of non-native speech are discussed. Third, while most speakers substituted what is presumed to be an L1 segment (e.g., [l]), some speakers substituted their L2 ([ɹ]). The results are therefore more consistent with H4a as opposed to H4b, although neither hypothesis was accurate in all cases. The role of cross-linguistic influence from previously learned languages will be discussed in more detail in Section 4.3.

The results of the present study are consistent with previous work, which found that L1 Mandarin–L2 Spanish speakers have difficulty with the /l-r/ contrast (Ortí Mateu 1990; Chih 2013). However, the learners in the present study were able to overcome the initial difficulty as they became more proficient. Previous findings on L1 English–L2 Spanish learners also share many similarities with the learners of the present study, especially with respect to the tap. In both cases, learners begin with substitutions, which is followed by some tap realizations, and subsequently by native-like tap production. The primary difference is that L1 English speakers tend to substitute the English [ɹ], whereas L1 Mandarin speakers tend to substitute [l]. This is noteworthy, because the L1 Mandarin speakers of the present study had acquired the L2 English [ɹ], but unlike L1 English speakers, they rarely used it as a substitute for the Spanish rhotics, despite the fact that many speakers had high L2 English proficiency and had acquired the L2 [ɹ]. Why do the L1 Mandarin–L2 English–L3 Spanish speakers substitute a lateral and not [ɹ]? This is likely attributable to perceptual assimilation and perceived similarity of /l-r/. Ortí Mateu (1990) and Chih (2013) both found that L1 Mandarin–L2 Spanish speakers have difficulty perceiving the /r-l/ contrast. While perception was not tested in the current study, it can be assumed from the production patterns and previous research on perception, that the less proficient learners of Spanish have difficulty perceiving the difference between /ɹ/ and

/l/ due to perceptual assimilation. Consequently, these learners produce /l/ in place of /r/, until they acquire (perceptually) the contrast, and also the ability to produce both sounds. /ɹ/ is less similar acoustically to /r/, thus perceptual assimilation of /ɹ/ and /r/ is less likely, and only happens in L1 English–L2 Spanish speakers because they do not have difficulty perceiving a difference between /l/ and /r/; the L1 English–L2 Spanish speakers thus consider /ɹ/ to be the most similar sound (most likely due to orthographic influence and phonotactic similarity), which is why it is used as a substitute for the Spanish tap.

The lateral is also often used as a substitute for the trill, and the motivation for this is less transparent. /l/ and /r/ are not similar either phonetically or orthographically, so why do the least proficient L1 Mandarin–L2 English–L3 Spanish speakers substitute [l] for /r/? As was demonstrated in Table 5, the less proficient learners tend to produce the same segment for both the tap and the trill, which is the same strategy used by L1 English–L2 Spanish speakers. This finding suggests that learners of Spanish follow a universal strategy when acquiring the rhotics. Specifically, the tap and trill are perceptually categorized with a single non-target segment in initial stages, probably based on perceived similarity between the tap and a previously acquired segment. For L1 Mandarin–L2 English–L3 Spanish speakers, this segment is /l/, as it is for L1 Mandarin–L2 Spanish speakers. One other possibility, proposed in Colantoni et al. (2015), is that Mandarin speakers may link laterals and rhotics as belonging to the same liquid class. This claim is made in regards to data in Steele (2002), who found that some L1 Mandarin–L2 English–L3 French speakers produced laterals in place of the French /ʁ/ in stop-liquid clusters. Colantoni et al. (2015) propose that the link may be partly conditioned by the similar orthographies in the L2 English and L3 French (in both languages, the laterals are represented by <l>, and the rhotics by <r>). L1 Mandarin–L2 English speakers tend to perceptually assimilate the L2 English rhotic to the L2 English lateral (both approximants with similar phonotactics), suggesting that these speakers consider the two segments to belong to the same category. Thus, when these L1 Mandarin–L2 English speakers learn L3 French, their knowledge of L2 English could influence how they perceive the L3 French /l/ and /ʁ/, and can explain why they may categorize the two L3 liquids as members of the same phonological class. It is therefore conceivable that L1 Mandarin–L2 English–L3 Spanish speakers do, at some abstract level, categorize the trill as belonging to the same liquid class as laterals, and as a result, substitute a lateral when first learning the trill. Note that three speakers in this study substituted English [ɹ] for /r/, and their L2 English proficiencies ranged from low to high. It is difficult to know the motivation for why some speakers substitute [ɹ] as opposed to [l], especially if L2 English proficiency is not a factor. However, the fact that not all speakers substitute the same segment suggests that perceived similarity is variable across speakers.

While L1 Mandarin–L2 English–L3 Spanish and L1 English–L2 Spanish speakers follow a similar strategy initially, as they become more advanced a significant difference is observed between the two groups. L1 English speakers tend to produce the tap in place of the trill, thus they continue to avoid a contrast between the two rhotics. This was not the case for L1 Mandarin–L2 English–L3 Spanish speakers, who produced many other non-target segments besides the tap (e.g., [l], [ð], [dr], [ɹ]), and the segment most frequently realized for the trill target was generally a different segment than what they produced for the tap target (Table 5). In other words, L1 Mandarin–L2 English–L3 Spanish speakers begin producing a contrast once they become more proficient. It is unclear why the two groups differ in this respect, but the findings suggest that the strength of the perceptual categorization of the two rhotics is stronger for L1 English speakers. This difference in perceptual categorization could be due to orthographic influence, which is more likely to be a factor in L1 English speakers, who use a Roman alphabet. The L1 Mandarin speakers are also familiar with a Roman alphabet through Pinyin and through their L2 English, but would not have the same amount of experience with it. The orthography could influence the likelihood that the two Spanish rhotics be perceptually categorized with a single segment, which would explain why no contrast is made in production by L1 English–L2 Spanish speakers.

4.2. Developmental Patterns in the Acquisition of Non-Native Segments

A notable finding regarding the observed production patterns as speakers acquired the Spanish tap–trill contrast was that less proficient speakers followed a substitution strategy at first. The speakers primarily substituted a single segment for the target rhotic; as they became more advanced, they began to produce a wide variety of non-target segments, resulting in a significant amount of variability, especially when acquiring the trill. For example, in the L3 trill production, eight different non-target realizations were observed in the seven speakers who were no longer producing primarily a single substitute and that still could not produce trills (M03 to M18, Figure 4). Each of these speakers produced at least three different non-target segments. This raises the question as to why the variability increases as learners become more proficient, and more importantly, what can this tell us about phonetic and phonological development? The results suggest that upon first learning a new language, new sounds that are articulatorily complex (e.g., the trill) are initially produced with a substitution of a previously learned and mastered segment (e.g., [l]). At this stage, the learners are less likely to attempt to produce the target segment, because the cognitive load is very high—learners may be more focused on trying to communicate by combining familiar sounds, recalling vocabulary, and forming logical sentences. As learners become more advanced in the target language, linguistic tasks such as recalling vocabulary and forming sentences become automatic. Learners can therefore dedicate more explicit attention to attempts to produce difficult sounds. It is at this stage that a significant amount of variability is observed, as the learners are either unable to articulate the target sound (as is often the case with the trill) or they can only do so with a low accuracy rate. The errors that result in missed attempts are what lead to the variability, and the amount of variability is likely to be higher with segments that are articulatorily complex. This is arguably why more variability was observed in trill compared to tap production. It is difficult to determine whether this was also observed in the studies investigating L1 English–L2 Spanish speakers, as individual manner of articulation results were not provided, other than in [Waltmunson \(2005\)](#), who did observe some variability (in terms of the number of non-target manners produced). Of the seven participants in his study who had difficulty with the trill (between 10–50% accuracy), five produced at least two types of non-target segments other than taps and English [ɹ]. Three other participants with less than 10% trill accuracy produced at least four different types of segments. Similarly, [Johnson \(2008\)](#) also commented that the intermediate speakers in his study demonstrated more variability in production than the beginner speakers, but only in word-initial position. Thus, the speakers in [Waltmunson \(2005\)](#) and [Johnson \(2008\)](#) appear to pattern in a similar way to the speakers of the present study. Whereas nearly categorical substitutions are observed in less proficient speakers, as speakers become more proficient they produce a larger variety of non-target variants. Future work should examine the extent to which this developmental pattern is observed in the acquisition of non-native segments, and what factors might influence the degree of variability.

4.3. Implications for Models of L3 Acquisition

Given that the speakers in the present study spoke three languages, they had the ability to transfer previously learned segments from their L1 or their L2. Overall, more negative transfer was observed from the L1 Mandarin (in the form of [l])¹⁰, especially in the least proficient speakers, regardless of L2 English proficiency. These findings are not fully consistent with the three L3 models discussed in Section 1.1 (CEM; [Flynn et al. 2004](#); L2SF; [Bardel and Falk 2007, 2012](#); TPM; [Rothman 2011, 2015](#)). Recall that both the TPM and the L2SF would predict L2 English transfer in L1 Mandarin–L2 English–L3 Spanish speakers, whereas no negative transfer should be observed according

¹⁰ Note that the laterals observed could potentially have been transferred from L2 English [l], which is very similar to the Mandarin [l], but this was not tested explicitly. However, L1 Mandarin–L2 Spanish learners also transfer [l], whereas L1 English–L2 Spanish speakers do not. Given that the speakers in this study behaved in a similar manner to L1 Mandarin speakers, (and in many cases were not highly proficient English speakers), the [l] is assumed to be an L1 transfer. A future study would have to test this in more detail via acoustic analysis.

to the CEM. Indeed, the results of the present study are more consistent with studies observing that the L1 or dominant language is a more likely source of transfer (Pyun 2005; Kopečková 2014; Llama and López-Morelos 2016; Llama and Cardoso 2018). Nevertheless, some clear evidence of L2 transfer was found; the L2 flap facilitated acquisition of the L3 tap (but not in the least proficient L3 speakers), and some productions of L2 [ɾ] were also observed. While the CEM correctly predicted the observed positive transfer of the L2 flap, some of the less proficient L3 Spanish learners did not transfer their L2 flap, which is not a pattern that the CEM would predict. The results of the present study therefore reveal that none of the L3 models can adequately account for the data, which suggests that a model that is specific to phonetics and phonology may be necessary (as discussed in Wrembel 2015, Kopečková et al. 2016¹¹, and Cabrelli Amaro and Wrembel 2016). Such a model must be able to account for CLI from both previously learned languages (positive and negative). The fact that speakers transfer from the L1 and/or L2 suggests that L2 speech models may be better able to account for the data, and could serve as a point of departure for an L3 model. For example, Flege (1995) suggests that L1 and L2 sounds share a common phonological space. If this is the case, we would then expect that upon acquisition of an L3, segments of all three languages would share a common phonological space. Indeed, Sypiańska (2016a) found some support for this proposal by analyzing the production of L1, L2, and L3 vowels by L3 speakers. We might also expect that perceptual assimilation of native and non-native segments would work in a similar way, with the only difference being that L3 segments could potentially be perceptually assimilated with either L1 or L2 segments, depending on which previously learned segment the L3 segment is most similar to. The perceived similarity of L1, L2, and L3 segments may differ across speakers, which could explain the assortment of L1 and L2 segments that were substituted for the two L3 rhotic targets. Exactly how well a model adapted from Flege (1995) and/or Best and Tyler (2007) could work to predict and explain L3 CLI would have to be corroborated through more perception-based studies, and the literature on CLI in L3 speech is primarily based on production (but see Enomoto 1994; Gogoi 2010; Kopečková 2015; Onishi 2016; Qin and Jongman 2016). Previous studies on perceptual assimilation in L2 perception have focused on determining which L1 segments a speaker perceives to be the most similar to L2 segments (e.g., Strange et al. 2004; Rose 2010). A similar study testing the perceived similarity of L1 and L2 to L3 segments, combined with a production component, could reveal whether speakers are transferring what they perceive to be the most similar segment. The results of such a study could be used to determine whether speakers do indeed transfer on a segment by segment basis, according to perceived similarity.

While the L2 speech models are possible starting points for models of L3 speech, the limitation of the models is that they do not consider the role of articulatory complexity, or dominant articulatory routines. The difficulty of acquiring the trill is clearly related to its complexity, and the variability observed in the production of the trill is likely to be driven by articulatory complexity. Nevertheless, the SLM and the PAM-L2 do not consider difficulty in production. Nor do they consider the role of dominant articulatory routines, which have generally been shown to influence L3 production (Hammarberg and Hammarberg 2005; Llama and López-Morelos 2016; Lloyd-Smith et al. 2017). Accordingly, any L3 model should be able to account for factors that influence production difficulty, while also considering the role of perceptual assimilation.

5. Conclusions

The present study investigated the production of the Spanish rhotics by L1 Mandarin–L2 English–L3 Spanish speakers, and had two objectives. The first objective was to establish the developmental stages of acquisition of the Spanish tap and trill by native speakers of a non-Germanic

¹¹ (Kopečková et al. 2016, p. 439) propose modeling of L3 phonetics and phonology through a Dynamic Systems Theory (e.g., de Bot 2012) approach, arguing that it would provide significant explanatory potential. However, no formal L3 phonetics and phonology model was proposed that could be tested explicitly.

language (i.e., L1 Mandarin), and to compare these stages to those reported for L1 English–L2 Spanish speakers. The second objective was to examine which of previously learned L1 and L2 segments were more likely sources of transfer in L3 production. Three main findings were revealed. First, speakers at initial stages tended to produce the same non-target segment for both rhotics, which mirrors the findings for L1 English–L2 Spanish speakers. This suggests that it is a universal simplification strategy. Second, while non-native substitutions were observed initially, speakers produced a variety of non-target segments as they became more proficient, suggesting that the speakers were attempting to produce the target, but were unsuccessful. These findings indicate that at first, learners rely on a single substitution (most likely to decrease complexity); as they become more advanced, and can thus focus additional attention on production, they are more likely to attempt target articulations. The result is an increase in the amount of variability produced by each speaker. Third, CLI surfaced from both the L1 Mandarin and L2 English, varying by speaker. The results suggest that transfer may be determined on a segment by segment basis, according to perceived similarity of the target and previously acquired segments. The perceived similarity of the segments is not the same for each speaker; consequently, different substitutes are observed.

As already highlighted in the discussion, future research in phonetic and phonological development should investigate what factors might lead speakers to follow a strategy involving categorical substitutions initially (of the target segment), followed by attempts at realizing the target as speakers become more proficient. With regards to L3 speech research, there is a need to determine in greater detail under what conditions L1 and L2 sounds are available to an L3 learner, and any model of L3PP should be able to account for the possibility of speakers substituting both L1 and L2 sounds.

Funding: This research was funded by the Social Sciences and Humanities Research Council of Canada grant number [767-2014-2571].

Acknowledgments: I would like to thank Jeffrey Steele, Laura Colantoni, and two anonymous reviewers for their valuable feedback, which greatly improved the manuscript.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. L1 Mandarin–L2 English–L3 Spanish participant profiles, ordered by L3 Spanish oral proficiency ratings, from lowest to highest.

Participant	Age	L2 English				L3 Spanish			
		AoA	AoUse	LoRE	Oral Prof.	Average Eng-Man Use	AoA	Semesters of SPA	Oral Prof.
M13	19	9	15	4	1.6	31	19	1.5	1
M12	21	6	17	4	2.1	50	19	1.5	1.1
M14	19	10	16	3	2.4	53	18	1.5	1.2
M10	19	6	17	2	2.6	35	18	1.5	1.3
M16	20	6	15	5	2.8	43	19	2	1.3
M15	19	6	16	3	3.3	83	16	1.5	1.4
M08	19	5	15	4	3.4	37	18	1.5	1.4
M19	21	6	14	7	2.6	40	15	10	1.5
M09	17	7	16	1	3.1	38	18	1.5	1.5
M06	20	7	17	3	2.3	44	19	1.5	1.6
M07	19	10	15	4	3.3	35	18	1.5	1.6
M02	22	7	15	7	4.1	50	19	8.5	1.6
M05	22	7	18	4	2.5	64	19	1.5	1.8
M18	20	6	10	10	2.6	75	14	8	2.0
M20	20	5	15	5	3.0	42	16	8	2.0
M01	21	6	16	5	3.9	42	18	6	2.4
M11	21	13	13	8	2.7	70	20	1.5	2.5
M04	19	5	5	1	4.6	60	18	1.5	2.5
M03	19	10	18	1	3.1	48	19	3.5	2.6
M17	22	7	16	4	3.3	75	18	10	2.8
Average	20.0	7.20	15.0	4.25	2.97	50.8	17.9	3.70	1.76

Notes. AoA = age of onset of acquisition (age of first exposure to the language, in years); AoUse = age of onset of use (age, in years, when use of the target language began in an immersion setting); LoRE = length of residence in an English-speaking country (in years); Average Eng-Man use = how much L2 English is used compared to L1 Mandarin at work, home, school, and at work (expressed as a %); Prof = Proficiency. The following scale was provided to judges when rating the L2 English and L3 Spanish oral proficiency: 1 = ‘clearly non-native, very strong foreign accent’; 2 = ‘strong foreign accent’; 3 = ‘noticeable foreign accent, but not too strong’; 4 = ‘almost no accent’; 5 = ‘no accent (native speaker)’.

Appendix B

Table A2. L1 Mandarin, L2 English, and L3 Spanish target stimuli.

Spanish Stimuli					
Tap /ɾ/					
Tonic			Post-tonic		
morí	/mo'ri/	'I died'	caro	/'karo/	'expensive'
veré	/be're/	'I will see'	pero	/'pero/	'but'
lloré	/o're/	'I cried'	hora	/'ora/	'hour'
herido	/e'rido/	'hurt'	cero	/'θero/	'zero'
verano	/be'rano/	'summer'	era	/'era/	'was'
marido	/ma'rido/	'husband'	pare	/'pare/	'stop'
Trill /r/					
Tonic			Post-tonic		
corrí	/ko'ri/	'I ran'	carro	/'karo/	'car'
cerré	/θe're/	'I closed'	perro	/'pero/	'dog'
borré	/bo're/	'I erased'	corra	/'kora/	'run'
derrito	/de'rito/	'I melt'	cierro	/'θjero/	'I close'
cerrado	/θe'rado/	'closed'	guerra	/'gera/	'war'
arriba	/a'riba/	'up'	barre	/'bare/	'sweep'
English stimuli					
Rhotic /ɹ/			Flap [ɾ]		
Tonic		Post-tonic			
arrive	/ə'ɹajv/	Mary	/'mɹi/	later	/'lejtəɹ/
parade	/pə'ɹejd/	porridge	/'pɹɔɪd/	butter	/'bʌtəɹ/
maroon	/mə'ɹun/	parrot	/'pɹət/	letter	/'letəɹ/
direct	/dɪ'ɹɛkt/	berry	/'bɹi/	attic	/'ætɪk/
correct	/kə'ɹɛkt/	prairie	/'pɹɛɹi/	water	/'wɔtəɹ/
arise	/ə'ɹajz/	teary	/'tiɹi/	kitty	/'kɪti/
arrange	/ə'ɹejnd/	sparrow	/'spɛɹɔw/	fatty	/'fæti/
erase	/ə'ɹeɪs/	carrot	/'kɛɹət/	better	/'betəɹ/
derive	/dɪ'ɹajv/	narrow	/'nɹɔw/	lettuce	/'letəs/
arrest	/ə'ɹɛst/	Tara	/'tɹə/	pretty	/'pɹɪti/
Mandarin stimuli					
/ʅ/					
牛肉 niúròu		/njuʅou/		'beef'	
鸡肉 jīròu		/tʃiʅou/		'chicken meat'	
渔人 yúrén		/yuʅən/		'fisherman'	
华人 huárén		/xuɑʅən/		'Chinese people'	
老人 lǎorén		/lauʅən/		'old people'	
昨日 zuórì		/tsuoʅi/		'yesterday'	
每日 měirì		/meiʅi/		'daily'	
体弱 tǐruò		/tiʅuo/		'debility'	
譬如 pìrú		/piʅu/		'for example'	
偶然 ǒurán		/ouʅan/		'randomly'	

References

- Amengual, Mark. 2016. Acoustic correlates of the Spanish tap-trill contrast: Heritage and L2 Spanish speakers. *Heritage Language Journal* 13: 88–112.
- Bardel, Camilla, and Ylva Falk. 2007. The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research* 23: 459–84. [[CrossRef](#)]

- Bardel, Camilla, and Ylva Falk. 2012. Procedural distinction. In *Third Language Acquisition in Adulthood*. Edited by Jennifer Cabrelli Amaro, Suzanne Flynn and Jason Rothman. Amsterdam: John Benjamins, pp. 61–78.
- Bassetti, Bene, and Nathan Atkinson. 2015. Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics* 36: 67–91. [CrossRef]
- Bertola de Urgorri, Mercedes. 2009. La diferencia cultural y la pronunciación: Dos aspectos a tener en cuenta en la enseñanza de español a alumnos chinos [Cultural difference and pronunciation: Two aspects to keep in mind in the teaching of Spanish to Chinese students]. *Signos Universitarios Virtual* 8: 1–8.
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins, pp. 13–34.
- Blank, Cintia A., and Marcía C. Zimmer. 2009. A transferência fonético-fonológica L2 (francês)—L3 (inglês): Um estudo de caso [Phonetic-phonological transfer in L2 (French)—L3 (English): A case study]. *Revista de Estudos da Linguagem* 17: 207–33. [CrossRef]
- Blecua Falgueras, Beatriz. 2001. Las vibrantes del español: Manifestaciones acústicas y procesos fonéticos [Rhotics in Spanish: Acoustic Manifestations and Phonetic Processes]. Doctoral dissertation, Universitat Autònoma de Barcelona, Barcelona, Spain.
- Boersma, Paul, and David Weenink. 2017. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.29. Available online: <http://www.praat.org/> (accessed on 24 May 2017).
- Brown, Cynthia A. 1998. The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14: 136–93. [CrossRef]
- Cabrelli Amaro, Jennifer, and Magdalena Wrembel. 2016. Investigating the acquisition of phonology in a third language—A state of the science and an outlook for the future. *International Journal of Multilingualism* 13: 395–409. [CrossRef]
- Cerini, Marco. 2013. The pronunciation of Mandarin Chinese According to the ^{can}IPA Natural Phonetics Tonetics Method. Doctoral dissertation, Ca' Foscari University of Venice, Venice, Italy.
- Chang, Seung-Eun. 2015. Degree and direction of foreign accent in L2 and L3 Korean speech. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Edited by The Scottish Consortium for ICPhS 2015. Glasgow: The University of Glasgow.
- Chih, Martín T. C. 2013. E/LE en Taiwán: Problemas de apreciación fonética en estudiantes universitarios de grado [Spanish as a Foreign Language in Taiwan: Problems with the phonetic assessment of undergraduate students]. *SinoELE* 9: 17–32.
- Colantoni, Laura, and Jeffrey Steele. 2007. Acquiring /ʁ/ in context. *Studies in Second Language Acquisition* 29: 381–406. [CrossRef]
- Colantoni, Laura, and Jeffrey Steele. 2008. Integrating Articulatory Constraints into Models of Second Language Phonological Acquisition. *Applied Psycholinguistics* 29: 489–534. [CrossRef]
- Colantoni, Laura, Jeffrey Steele, and Paula Escudero. 2015. *Second Language Speech*. Cambridge: Cambridge University Press.
- Cortés Moreno, Maximiano. 2014. Dificultades lingüísticas del español para los estudiantes sinohablantes y búsqueda de soluciones motivadoras [Linguistic difficulties of Spanish for Chinese speaking students and a search for motivating solutions]. *Monográficos SinoELE* 10: 173–208.
- Davies, Mark. 2006. *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York: Routledge.
- de Bot, Kees. 2012. Rethinking multilingual processing: From a static to a dynamic approach. In *Third Language Acquisition in Adulthood*. Edited by Jennifer Cabrelli Amaro, Suzanne Flynn and Jason Rothman. Amsterdam: John Benjamins, pp. 77–93.
- Derrick, Donald, and Bryan Gick. 2011. Individual variation in English flaps and taps: A case of categorical phonetics. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 56: 307–19. [CrossRef]
- Detey, Sylvain. 2009. Phonetic input, phonological categories and orthographic representations: A psycholinguistic perspective on why language education needs oral corpora. The case of French-Japanese interphonology development. In *Corpus Analysis and Variation in Linguistics*. Edited by Yuji Kawaguchi, Makoto Minegishi and Jacques Durand. Amsterdam: John Benjamins, pp. 179–200.
- Detey, Sylvain, and Jean-Luc Nespoulous. 2008. Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua* 118: 66–81. [CrossRef]
- Duanmu, San. 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

- Duanmu, San. 2007. *The Phonology of Standard Chinese*, 2nd ed. Oxford: Oxford University Press.
- Enomoto, Kayoko. 1994. L2 Perceptual Acquisition: The Effect of Multilingual Linguistic Experience on the Perception of a 'Less Novel' Contrast. *Edinburgh Working Papers in Applied Linguistics* 5: 15–29.
- Escudero, Paola, Rachel Hayes-Harb, and Holger Mitterer. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60. [\[CrossRef\]](#)
- Face, Timothy L. 2006. Intervocalic rhotic pronunciation by adult learners of Spanish as a second language. In *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*. Edited by Carol A. Klee and Timothy L. Face. Somerville: Cascadia Press, pp. 47–58.
- Falahati, Reza. 2015. The production of Persian rhotics by native Mandarin speakers. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Edited by The Scottish Consortium for ICPhS 2015. Glasgow: The University of Glasgow.
- Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Edited by Winnifred Strange. Baltimore: York Press, pp. 233–77.
- Flynn, Suzanne, Claire Foley, and Inna Vinnitskaya. 2004. The cumulative-enhancement model for language acquisition: Comparing adults' and children's patterns of development in L1, L2 and L3 acquisition of relative clauses. *The International Journal of Multilingualism* 1: 3–16. [\[CrossRef\]](#)
- Gogoi, Divya V. 2010. Acquisition of Novel Perceptual Categories in a Third Language: The Role of Metalinguistic Awareness and Feature Generalization. Doctoral dissertation, University of Florida, Gainesville, FL, USA.
- Hammarberg, Björn, and Britta Hammarberg. 2005. Re-setting the basis of articulation in the acquisition of new languages: A third language case study. In *Processes in Third Language Acquisition*. Edited by Björn Hammarberg. Edinburgh: Edinburgh University Press, pp. 74–85.
- Hanley, J. Richard. 2005. Learning to read Chinese. In *The Science of Reading: A Handbook*. Edited by Margaret J. Snowling and Charles Hulme. Oxford: Blackwell Publishing, pp. 316–35.
- Hayes-Harb, Rachel, and Hui-Wen Cheng. 2016. The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word forms by native English speakers. *Frontiers in Psychology* 7: 785. [\[CrossRef\]](#) [\[PubMed\]](#)
- Henriksen, Nicholas C. 2014. Sociophonetic analysis of phonemic trill variation in two sub-varieties of Peninsular Spanish. *Journal of Linguistic Geography* 2: 4–24. [\[CrossRef\]](#)
- Henriksen, Nicholas C., and Erik W. Willis. 2010. *Selected Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*. Edited by Marta Ortega-Llebaria. Somerville: Cascadia Proceedings Project, pp. 115–27.
- Hualde, Jose I. 2014. *Los Sonidos del español [The Sounds of Spanish]*. Cambridge: Cambridge University Press.
- Igarreta Fernández, Alba. 2015. La corrección de la pronunciación de los estudiantes sinohablantes en el aula de E/LE [The correction of Chinese speakers' pronunciation in the Spanish as a Foreign Language classroom]. *Foro de Profesores de E/LE* 11: 189–96.
- Johnson, Keith E. 2008. Second Language Acquisition of the Spanish Multiple Vibrant Consonant. Doctoral dissertation, University of Arizona, Tucson, AZ, USA.
- Kopečková, Romana. 2014. Crosslinguistic influence in Instructed L3 child phonological acquisition. In *Essential Topics in Applied Linguistics and Multilingualism*. Edited by Mirosław Pawlak and Larissa Aronin. Cham: Springer International Publishing, pp. 205–24.
- Kopečková, Romana. 2015. Differences in the perception of English vowel sounds by child L2 and L3 learners. In *Universal or Diverse Paths to English Phonology*. Edited by Ulrike Gut, Robert Fuchs and Eva-Maria Wunder. Berlin: De Gruyter, pp. 71–90.
- Kopečková, Romana. 2016. The bilingual advantage in L3 learning: A developmental study of rhotic sounds. *International Journal of Multilingualism* 13: 410–25. [\[CrossRef\]](#)
- Kopečková, Romana, Marta Marecka, Magdalena Wrembel, and Ulrike Gut. 2016. Interactions between three phonological subsystems of young multilinguals: The influence of language status. *International Journal of Multilingualism* 13: 426–43. [\[CrossRef\]](#)
- Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell Publishing.
- Lee, Wai-Sum. 1999. An articulatory and acoustical analysis of the syllable-initial sibilants and approximants in Beijing Mandarin. In *Proceedings of the International Congress on Phonetic Sciences 1999*. Edited by John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville and Ashlee C. Bailey. San Francisco: The Regents of the University of California, pp. 413–16.

- Lewis, Anthony M. 2004. Coarticulatory effects on Spanish trill production. In *Proceedings of the 2003 Texas Linguistics Society Conference*. Edited by Augustine Agwuele, Willis Warren and Sang-Hoon Park. Somerville: Cascadia Proceedings Project, pp. 116–27.
- Lipińska, Dorota. 2017. The Influence of L2 Status on L3 Pronunciation. *English Insights* 1: 69–86.
- Llama, Raquel, and Walcir Cardoso. 2018. Revisiting (Non-)Native Influence in VOT Production: Insights from Advanced L3 Spanish. *Languages* 3: 30. [\[CrossRef\]](#)
- Llama, Raquel, and Luz P. López-Morelos. 2016. VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism* 13: 444–58. [\[CrossRef\]](#)
- Llama, Raquel, Walcir Cardoso, and Laura Collins. 2010. The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism* 7: 39–57. [\[CrossRef\]](#)
- Lloyd-Smith, Anika, Henrik Gyllstad, and Tanja Kupisch. 2017. Transfer into L3 English. *Linguistic Approaches to Bilingualism* 7: 131–62. [\[CrossRef\]](#)
- Martínez Celdrán, Eugenio, and Ana María Fernández Planas. 2007. *Manual de fonética española: Articulaciones y sonidos del español* [Manual of Spanish Phonetics: Articulations and Sounds of Spanish]. Barcelona: Editorial Ariel.
- McMahon, April. 2002. *An Introduction to English Phonology*. Edinburgh: Edinburgh University Press.
- Morales Reyes, Alexandra, Begoña Arechabaleta-Regulez, and Silvina Montrul. 2017. The acquisition of rhotics by child L2 and L3 learners. *Journal of Second Language Pronunciation* 3: 242–66. [\[CrossRef\]](#)
- Navarro Tomás, Tomás. 1957. *Manual de pronunciación española* [Spanish Pronunciation Manual]. New York: Hafner Publishing Company.
- Olsen, Michael K. 2012. The L2 acquisition of Spanish rhotics by L1 English speakers: The effect of L1 articulatory routines and phonetic context for allophonic variation. *Hispania* 95: 65–82.
- Onishi, Hiromi. 2016. The effects of L2 experience on L3 perception. *International Journal of Multilingualism* 13: 459–75. [\[CrossRef\]](#)
- Ortí Mateu, Rosa. 1990. Comparación fonética, diagnóstico y tratamiento de las dificultades de los estudiantes chinos para aprender español [Phonetic Comparison, Assessment, and Treatment of Chinese Students' Difficulties Learning Spanish]. Doctoral dissertation, University of the Philippines, Quezón City, PH, USA.
- Pyun, Kwang-Soo. 2005. A model of interlanguage analysis—The case of Swedish by Korean speakers. In *Introductory Readings in L3*. Edited by Britta Hufeisen B and Robert J. Fouser. Tübingen: Stauffenberg, pp. 55–70.
- Qin, Zhen, and Allard Jongman. 2016. Does second language experience modulate perception of tones in a third language? *Language and Speech* 59: 318–38. [\[CrossRef\]](#) [\[PubMed\]](#)
- Quilis, Antonio. 1993. *Tratado de Fonética y Fonología españolas* [Treatise of Spanish Phonetics and Phonology]. Madrid: Gredos.
- Rafat, Yasaman. 2015. The interaction of acoustic and orthographic input in the acquisition of Spanish assimilated/fricative rhotics. *Applied Psycholinguistics* 36: 43–66. [\[CrossRef\]](#)
- Rafat, Yasaman. 2016. Orthography-induced transfer in the production of English-speaking learners of Spanish. *The Language Learning Journal* 44: 197–213. [\[CrossRef\]](#)
- Rose, Marda. 2010. Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In *Selected Proceedings of the 2008 Second Language Research Forum*. Edited by In Matthew T. Prior, Yukiko Watanabe and Sang-Ki Lee. Somerville: Cascadia Proceedings Project, pp. 181–96.
- Rothman, Jason. 2011. L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research* 27: 107–27. [\[CrossRef\]](#)
- Rothman, Jason. 2015. Linguistic and cognitive motivations for the typological primacy model of third language (L3) transfer: Considering the role of timing of acquisition and proficiency in the previous languages. *Bilingualism: Language and Cognition* 18: 179–90. [\[CrossRef\]](#)
- Samper Pandilla, José A. 2011. Socio-phonological variation and change in Spain. In *The Handbook of Hispanic Sociolinguistics*. Edited by Manuel Díaz-Campos. Oxford: Blackwell Publishing, pp. 98–120.
- Schwartz, Bonnie D., and Rex A. Sprouse. 1996. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12: 40–72. [\[CrossRef\]](#)
- Showalter, Catherine E., and Rachel Hayes-Harb. 2013. Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research* 29: 185–200. [\[CrossRef\]](#)
- Smith, James G. 2010. Acoustic Properties of English /l/ and /r/ Produced by Mandarin Chinese Speakers. Master's thesis, University of Toronto, Toronto, CA, USA.

- Steele, Jeffrey. 2002. Representation and Phonological Licensing in the L2 Acquisition of Prosodic Structure. Doctoral dissertation, University of McGill, Montreal, CA, USA.
- Strange, Winifred, Ocke-Schwen Bohn, Sonja A. Trent, and Kanae Nishi. 2004. Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America* 115: 1791–807. [[CrossRef](#)] [[PubMed](#)]
- Sypiańska, Jolanta. 2016a. L1 vowels of multilinguals: The applicability of SLM in multilingualism. *Research in Language* 14: 79–94. [[CrossRef](#)]
- Sypiańska, Jolanta. 2016b. Multilingual acquisition of vowels in L1 Polish, L2 Danish and L3 English. *International Journal of Multilingualism* 13: 476–95. [[CrossRef](#)]
- Tobin, Yishai. 1997. *Phonology as Human Behavior: Theoretical Implications and Clinical Applications*. Durham: Duke University Press.
- Tremblay, Marie-Claude. 2007. L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals. Paper presented at the Satellite Workshop of ICPhS XVI, Freiburg, Germany, August 3–4.
- Waltmunson, Jeremy C. 2005. The relative degree of difficulty of L2 Spanish /d, t/, trill, and tap by L1 English speakers: Auditory and acoustic methods of defining pronunciation accuracy. Doctoral dissertation, University of Washington, Seattle, WA, USA.
- Wells, John C. 1982. *Accents of English*. Cambridge: Cambridge University Press.
- Willis, Erik W., and Travis G. Bradley. 2008. Contrast maintenance of taps and trills in Dominican Spanish: Data and analysis. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Edited by Laura Colantoni and Jeffrey Steele. Somerville: Cascadilla Proceedings Project, pp. 87–100.
- Wrembel, Magdalena. 2010. L2-accented speech in L3 production. *International Journal of Multilingualism* 7: 75–90. [[CrossRef](#)]
- Wrembel, Magdalena. 2014. VOT patterns in the acquisition of third language phonology. *Concordia Working Papers in Applied Linguistics* 5: 750–70.
- Wrembel, Magdalena. 2015. *In Search of a New Perspective: Cross-Linguistic Influence in the Acquisition of Third Language Phonology*. Poznań: Wydawnictwo Naukowe UAM.
- Yavas, Mehmet. 2016. *Applied English Phonology*. West Sussex: John Wiley & Sons.
- Zue, Victor W., and Martha Laferriere. 1979. Acoustic study of medial /t, d/ in American English. *The Journal of the Acoustical Society of America* 66: 1039–50. [[CrossRef](#)]



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).