L1 English / L2 Spanish: Orthography–phonology activation without contrasts

Christine Shea
University of Iowa, USA

Abstract
We consider how orthography activates sounds that are in a noncontrastive relationship in the second language (L2) and for which only one variant exists in the first language (L1). Participants were L1 English / L2 Spanish and native Spanish listeners. Intervocalically, Spanish graphemes 'b d g' correspond phonetically to stops and approximants (e.g. *lobo* 'wolf', lo[β]o), and in English they correspond only to stops. In Experiment 1, native and L2 Spanish listeners completed cross-modal (written–auditory) and within modal (auditory) priming tasks. Prime-target pairs were counterbalanced for phonetic variant. The results for L2 listeners in the cross-modal condition showed a significant interaction between variant and mode. Experiment 2 used long-term repetition priming to tap into longer-term representations and test whether L1 orthography is activated even when it is not strictly necessary to complete the task. Results for L2 speakers showed priming by both phonetic variants while for native listeners, only approximants showed a priming effect.

Keywords
adult second language acquisition, cross-modal priming, orthography, phonology, Spanish

I Introduction

There is considerable evidence that learning to read restructures phonological representations, strongly influenced by the close connections between orthography and phonology (Castles et al., 2011). Literate individuals have been shown to demonstrate sharper categorical perception boundaries than nonliterate individuals (Serniclaes et al., 2005), and literacy development (even in adulthood) has been shown to refine cortical organization (Dehaene et al., 2010). In children, phonological awareness improves in tandem

Corresponding author:
Christine Shea, Department of Spanish and Portuguese, University of Iowa, Iowa City, IA 52242, USA
Email: christine-shea@uiowa.edu

with increasing literacy (Alcock et al., 2010). In their seminal study, Tanenhaus et al. (1979) showed that auditory rime judgments for consistently spelled words (e.g. pie–die) were faster than for inconsistently spelled words (e.g. rye–tie), suggesting that spelling is activated even for tasks that do not explicitly draw upon orthographic representations. This orthographic consistency effect has been found across many different languages (French: Pattamadilok et al., 2011; Perre et al., 2009; English: Dich, 2011; Pattamadilok et al., 2010; Portuguese: Ventura et al., 2007, 2008).

While there is little debate regarding the existence of orthographic effects on phonological representations, researchers have questioned whether these effects are automatic or the result of strategic responses on the part of participants. Tasks such as phoneme detection, counting, elision, elimination (for discussion, see Bassetti, 2006: 99) and even auditory rime decision draw attention to the sound–spelling connection in a more or less explicit manner and rely on metaphonological knowledge for their execution. In priming tasks, the activation of a visual–word prime may lead to more positive responses on lexical decision tasks than on an auditory task. This would account for the finding that orthography–phonology consistency effects are found in auditory lexical decision tasks but not in shadowing tasks (Rastle et al., 2011), which does not involve meta-linguistic decision. Given this, it is possible that orthography is activated only when required to carry out specific tasks, reflecting strategic responses rather than obligatory activation under all contexts (Perre et al., 2011; Taft et al., 2008) and influences mainly meta-linguistic thinking about speech (Cutler and Davis, 2012).

These results have important repercussions for second language (L2) acquisition, since many L2 learners acquire the target language in classroom contexts where, from the first day of class, they are encouraged to read in their second language. It is common that target language reading and writing skills develop apart from and even prior to the establishment of L2 phonological categories (which may never achieve the same level of native-like proficiency), the opposite of what occurs with children becoming literate in their first language. Thus, many L2 learners may develop sound–spelling correspondences that are closer to their first language (L1) sound categories than their L2 categories, and this may be exacerbated when the L2 letters correspond to more than one sound in the L1 (i.e. when the sound–spelling correspondence is not 1:1 in either language). The question of how L1 orthography–phonology connections influence the establishment of L2 phonological categories – and whether these connections are automatic or task-dependent – is crucial to any investigation of L2 phonological development.

Recent work in the area of L2 orthography–phonology relationships has examined the role of L1 sound–spelling correspondences on nonce-word acquisition (Hayes-Harb et al., 2010) and L1–L2 correspondences where languages differ in writing systems (Bassetti, 2006; Showalter and Hayes-Harb, 2013). The evidence is somewhat mixed in terms of orthography universally helping or hindering L2 acquisition. Using a novel-lexicon-learning task, Escudero et al. (2008) showed that learners exposed to orthographic forms were more successful at acquiring a novel vowel contrast than those who were not, and orthography was most beneficial when acquired in tandem with the phonological form. However, there is also evidence that orthography can have mixed effects on the acquisition and development of L2 categories. Escudero and Wanrooij (2010) tested the perception of Dutch vowel contrasts by native Spanish speakers of different

Dutch proficiency. Their findings showed that for the most difficult contrast, orthography helped establish contrasts while for vowel contrasts that were less difficult, orthography hindered the perception of contrasts and suggest that when orthography is available, it can influence non-native (vowel) perception.

Most of this previous work has examined the relationship between orthography and phonology, more specifically, at the levels of categories. An exception is Rafat (2015) who examined how naive speakers of a language produce a non-native sound that has differing acoustic phonetic properties but only one orthographic representation. She examined the production of assibilated rhotics in Spanish by naive and L2 Spanish–English speakers of varying proficiency levels under either auditory–auditory and auditory–orthography conditions. Rafat found a significant difference in the production patterns of the two groups and argued that orthography plays a key role in highlighting salient acoustic differences for L2 speakers. Evidence that orthographic input can even override acoustic input comes from a study by Hallé et al. (2000) who examined the perception of /b/ and [p] by French-speaking adults. The authors used a phoneme-monitoring task in which listeners had to detect the grapheme 'b' or 'p' in words such as *absurd*, which is realized phonetically as a[p]surd. Their results showed a higher detection of [b] in words where the sound corresponded to the grapheme 'b', in spite of the phonetic realization of the targets. These two studies provide important insights into how listeners produce and perceive subphonemic variants that correspond to the same orthographic symbol.

In the present study we extend this work and examine the activation of noncontrastive phonetic variants, or allophones. Crucial to our purpose here, allophones of the same phoneme typically share an orthographic symbol in alphabetic languages, which is also the name of the grapheme that matches it (e.g. all the 't' sounds in words such as 'water', 'teacher' and 'stop' are written with the letter named 't' and correspond to /t/ in the mind of the literate native speaker), and furthermore, by definition, allophones do not serve to contrast items in the lexicon of the native speaker. Together, these two points can make establishing the grapheme–allophone connection difficult for L2 learners.

The alternation examined here is the well-studied voiced stop–approximant allophonic distribution that occurs in Spanish, whereby stops tend to occur after pauses and nasals and approximants tend to occur intervocalically (see, amongst others, Carrasco et al., 2012) (e.g. a[β]uela 'grandmother', la[ɣ]o 'lake', la[ð]o 'side'). We use this allophonic variation to test whether L1 orthography–phonology connections are automatically activated during L2 lexical processing when the L2 phonetic variants share the same orthographic symbol, one that is also shared by the L1 phoneme. L1 English / L2 Spanish listeners carried out a masked cross-modal priming experiment in which the targets either followed the distributional probabilities of Spanish (approximants in intervocalic position) or followed the distributional probabilities of English (stops in intervocalic position). Furthermore, to verify that this effect is in fact driven by orthography and not only phonology, we also include a within-modal condition with auditory primes and targets. A group of native Mexican Spanish speakers was included as a control.

We predict that when L1 English / L2 Spanish listeners are exposed to Spanish orthographic input with intervocalic stops, learners will not activate the approximant variant and instead only activate the stop (Shea and Curtin, 2010, 2011). This prediction is based
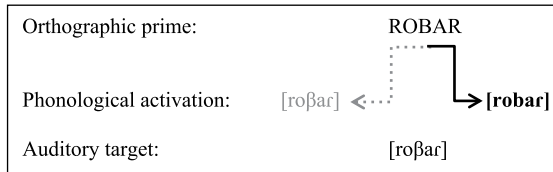
Orthographic prime: ROBAR

Phonological activation: [roβaɾ] ⟵⋯⋯   ⟶ **[robaɾ]**

Auditory target: [roβaɾ]

Figure 1. Interaction between written primes and L2 allophonic variants.
Note. If L2 listeners have both allophonic variants but lack the correspondence between the grapheme 'b' and the allophone [β], perceiving the letter 'b' is predicted to activate the phonological form [robaɾ] not the form [roβaɾ]. This leads to longer reaction times for auditory targets with the approximant variant.

on the idea that L1 orthography–phonology links will be activated when listeners perceive the written form, even though it is masked. If L1 orthographic connections play a determining role in the activation of one phonetic variant over the other, we predict that the approximant will show a greater priming effect in the within-modal (auditory–auditory) condition, in which the listeners are not exposed to any written forms. This is depicted in Figure 1. For the Spanish listeners, we predict an effect for phonetic variant independent of mode. Given that the stop variant is less probable in intervocalic position in Spanish (but not unacceptable) we predict longer reaction times for this condition than for the stop condition.

Masked priming provides a window onto how listeners process stimuli when they are unaware that the stimuli have been presented (Forster and Davis, 1984) and reveals prime-target connections that are automatic. In a series of experiments using Dutch and French bilinguals, Brysbaert and colleagues used cross-linguistic masked homophone priming and showed that the magnitude of the homophonic priming effect was equally large in L2 as in L1 (Brysbaert et al., 1999; Van Wijnendaele and Brysbaert, 2002), even for participants who were not balanced bilinguals, suggesting that the activation of phonological information from a visually presented word is equally strong in both languages. Thus, the grapheme–sound correspondence rules from the language not in use are still applied to the processing of target stimuli, reflecting nonselective access to both languages. Veivo and Järvikivi (2013) carried out three cross-modal masked priming with lexical decision to investigate orthographic and phonological processing by L1 Finnish / L2 French speakers of low and high proficiency levels. For the present purposes, Veivo and Järvikivi's Experiment 1 is the most relevant, since it examines cross-modal priming in participants' L2. In this experiment (a partial replication of Grainger et al., 2003[1]), learners were exposed to three conditions: priming with orthographic equivalents to the auditory target (e.g. [staʒ] *stage* 'course') nonword pseudohomophones that followed French grapheme–phoneme correspondence rules (e.g. [staʒ] *staje*) and, finally, nonword controls with no form overlap with the target. Previous research had shown that native French speakers show priming for both the orthographic equivalent forms and the pseudohomophone forms and, thus, Veivo and Järvikivi predicted that if orthography facilitated lexical recognition, the orthographic equivalents should show the greatest degree of priming, and if orthography also plays a role in pseudohomophone activation, then the L2 listeners should show effects similar to the native French speakers. The results showed that, indeed, facilitation was observed for both the orthographically

equivalent prime-target pairs and also for the pseudohomophone pairs. The priming effect for the equivalent condition was slightly higher. In the present study, we give a different perspective on the L2 orthography–phonology connections than Veivo and Järvikivi by examining how the correspondence of one L1–L2 shared grapheme activates two phones in the L2 but only one phone in the L1.

Current models of bilingual lexical activation do not tend to take into account how variability may affect lexical activation in bilingual spoken (or written) language processing, whether variability originating in the context of interaction, individual shifting proficiency levels or task demands that may lead a bilingual to favor one language over another at certain moments. The Bilingual Language Interaction Network for Comprehension of Speech (BLINCS; Shook and Marian, 2013) is an exception to this and uses self-organizing maps to map information to nodes that are similar. Furthermore, BLINCS also explicitly encodes orthographic information, connected to positionally-faithful syllable slots that can accommodate variability such as that examined here, specifically referring to positional variants. This is more fully addressed in Section IV.

In Experiment 1 we use cross-modal and within-modal priming to test whether L1 sublexical allophonic variants are activated by matching orthographic symbols in the L2 (i.e. one allophone in the L1, two allophones in the L2, both of which correspond to the same orthographic symbol). Our goal in using these two methods is to test whether L1 orthography plays a role in activating subphonemic variants in second language learners. The prime in the cross-modal masked priming experiment is, by definition, not available for conscious processing while in the auditory priming condition it is. Opting for masked priming in this case allows us to test for the subconscious activation of orthography, or activation that is not under the conscious control of the listener. In the auditory prime condition, orthographic activation would be similarly subconscious if it were to occur. Thus, by using cross-modal priming and auditory priming we can test for the activation of L1 orthography–phonology connections under equal conditions.

The activation of the incorrect allophonic variant by L1 English / L2 Spanish listeners may be due to two closely related but ultimately distinct causes, one acoustic and one orthographic–phonological. First, it is possible that the L2 learners only have the stop variant in their phonetic inventory and the orthography–phonology correspondence rule for the subphonemic approximant variant cannot be established; the only category available for activation is the stop. If L2 listeners have not established the two phonetic categories corresponding to the phonological category of voiced stops in Spanish, we predict that for both the cross-modal and within modal conditions, the stop targets will show greater priming effect than the approximant targets. Alternatively, it is possible that both phonetic categories exist in the listener's inventory but the probabilistic relationship between orthography and phonology has not yet been established. In this situation, seeing the orthographic symbol activates the L2 category closest to the L1 category that corresponds to the grapheme, given that L1 orthography–phonology associations override L2 associations. However, for the within-modal priming condition, a priming effect for the approximant variant may emerge. The final possibility mirrors the predictions for the native Spanish speakers, that is, the existence of two phonetic categories and strong L2 orthography–phonology connections. In this situation, we predict that the stop variant will exhibit less priming and there will be very little, if any difference across the two priming modes.

Additional support for an orthographically-mediated priming effect would be found if we could conclusively show that learners do in fact encode both forms in their lexicons and that the variant that corresponds to their L1 is primed in the cross-modal condition. In order to test this, Experiment 2 uses long-term repetition priming, in which participants are exposed to two blocks of words (see McLennan et al., 2003; Sumner and Samuel, 2009). Block 1 serves as the prime for Block 2. Generally, long-term repetition priming includes a high number of items in each block and when a break is added between blocks, considerable time may elapse between the presentation of the prime (Block 1) and the target (Block 2). Nonetheless, in spite of this lag, primed words are still identified more rapidly than unprimed words (Church and Schacter, 1994; Goldinger, 1996). In a study that looked specifically at long-term repetition priming and allophonic variation, Luce et al. (2003) used shadowing across two blocks of English stimuli that were counterbalanced for the presence of a flap or [t]/[d]. They found robust allophonic priming effects when items were matched across allophonic variants for high frequency words. Priming across two blocks would suggest that the items share a common underlying representation while a lack of priming would suggest that the two allophones do not.

II   Experiment 1: Cross-modal and within-modal forward masked priming with lexical decision

1   Method

a   Participants. Participants were 42 L1 English / L2 Spanish majors or minors studying at a public university in the American Midwest, tested in the USA, and 34 native Mexican Spanish speakers, tested in Mexico City. The English speakers received class credit for their participation and the Mexican Spanish speakers were compensated $20.00. All English participants completed Spanish vocabulary and grammar tests, and only individuals who scored within 1.5 standard deviations of the group mean on both tests were included in the study, leading to the elimination of 5 L1 English speakers.[2]

L1 English speakers had completed the Spanish-language requirement and were enrolled in Spanish major-level courses and had studied at least seven years of Spanish, between high school and college. None had studied abroad prior to participating in the study. The average age for the L1 English speakers was 20.4 years.

All L1 Spanish speakers were undergraduate students at a large public institution in Mexico City and none were English majors or studying undergraduate programs where English was a major component (e.g. tourism, English teacher, translation). None of the L1 Spanish participants had lived abroad nor did they have any family members or friends with whom they spoke English on a regular basis. Every effort was made to control for the level of bilingualism in the native Spanish speaker group, with the important caveat that it is almost impossible to find university students in Mexico who have not studied English at the post-secondary level. The average age of the L1 Spanish speakers was 22.3 years.

b   Stimuli and design. Lexical items were matched for frequency and length. Frequency data was taken from the online NIM database (Guasch et al., 2013) for Spanish, based on written

frequencies found in the *Léxico informatizado del Español* (LEXESP; Sebastián-Gallés et al., 2000). The corpus includes 5,629,279 Spanish tokens and 166,494 word types. The average length of the stimuli was 6.1 letters (range: 4–9) and the average log frequency was 1.73 (range: 1.33–2.6). Items that exhibited the contrast of interest had the grapheme 'b' or 'd' or 'g' in intervocalic position. Due to restrictions on item selection, 30% of the words had the target segment in the onset of the tonic syllable. The filler items were taken from the NIM database as well and reflected the same frequency and length range as the target items and ranged between 5–9 letters in length. None of the filler items had the target sounds in them. The nonword items were created by switching the penultimate or ultimate vowel in a real word. For example, instead of the word *mesa* 'table', the nonword became *mesu*, a word that does not exist in Spanish. Items included verbs, nouns and adjectives.

A female native speaker from Bogotá, Colombia (not related to the current study in any capacity), a trained linguist, recorded the stimuli. The Bogotá variety has been documented as being one of the few to maintain a clear distinction between the stop and approximant variants in both of the predicted contexts (Carrasco et al., 2012). The words were recorded in a soundproof booth using a Marantz PM 670 solid state recorder and a Sennheiser e835 microphone. The speakers read each word in the carrier phrase *Yo digo _____ una vez* ('I say _____ one time') at a comfortable pace. For the stop-medial items that did not follow the expected distributional pattern of Spanish, the speaker was asked to produce a stop. The word-list was read three times and the most representative tokens were selected for inclusion in the experiment by the author and a colleague who is a Colombian Spanish linguist (distinct from the speaker who recorded the stimuli) and is familiar with the process of stop–approximant alternation and produces it clearly in her speech. She was asked to select the best token of each type, for each item.

All auditory target items with either a stop or approximant in intervocalic position were subjected to an acoustic analysis, to ensure that the target segments were in fact realized as their target phonetic variants. Using the method outlined in Shea and Curtin (2011, based on Lavoie, 2001), we took the root mean square amplitude ratios for the target segment and the following vowel. Root mean square amplitude is a measure of the amplitude of a sound sample, in dB SPL (decibels in sound pressure level). For segments that are more approximant-like, the RMS (root mean square) amplitude will be higher than for sounds that are more stop-like. Thus, when comparing them to the following vowels, lower ratio values indicate more approximant-like segments while higher ratio values indicate more stop-like segments. For the stops, segment duration was measured from the start of stop closure to the onset of vowel formant structure. For the approximants, segment duration was measured from the drop in intensity or, in cases where frication occurred, from the start of frication. The average RMS amplitude ratio for the bilabial stop variants was 3.04, for the velars it was 2.89, and for the dentals it was 2.98. For the approximant targets the average RMS amplitude ratio for the bilabial targets was 1.02, for the velars it was 1.09, and for the dentals it was 1.11. Thus, there was a difference between the target segments in terms of the intensity ratios.

In Experiment 1, participants made lexical decisions on two types of trials: cross-modal (written–auditory) and within-modal (auditory–auditory) (see Table 1). Participants were presented with the two conditions in separate blocks (counterbalanced across participants) with a 3–6 minute pause between the blocks. The experiment

Table 1. Experiment 1 trials.

Condition	Cross-modal: written–auditory		Within-modal: auditory–auditory	
	Prime	Target	Prime	Target
Approximant matched	CABELLO 'hair'	[kaβeɟo]	[naða] *nada* 'nothing'	[naða]
Stop matched	CABEZA 'head'	[kabesa]	[ɟega] *llega* 'arrives'	[ɟega]
Approximant control	ABEJA 'bee'	[toβiɟo] *tobillo* 'ankle'	[aβuela] *abuela* 'grandmother'	[aβuso] *abuso* 'abuse'
Stop control	SABOR 'flavour'	[aber] *haber* aux.verb	[estudio] *estudio* 'study'	[pasado] *pasado* 'past'
Fillers	FLOR 'flower'	[sol] *sol* 'sun'	[paɾke] *parque* 'park'	[kamino] *camino* 'road'
Filler nonwords	COCHE 'car'	[kami]	[papel] *papel* 'role'	[masu] *masu*
Nonwords	SOLU	[mesu]	[ombru]	[puelu]

included 7 different trial types for the cross-modal condition and 7 for the within-modal condition. We first describe the within-modal trials.

For the within-modal condition, we created a set of matched trials for which the prime and the target were the same word (different tokens), five for each place of articulation. The first set consisted of the approximant matched and had the approximant variant in intervocalic position (e.g. [kaðena]–[kaðena] *cadena* 'chain'). There were also 15 stop matched trials, with the stop variant in intervocalic position (e.g. [todo]–[todo] *todo* 'all' masculine singular). We did not include any trials where the prime and target differed across phonetic variant. The experiment focused on comparisons across the within-modal and cross-modal priming conditions, and there is no way of creating a condition for the cross-modal priming where the phonetic variants differ across prime and target. Thus, to maintain maximum comparability across the cross-modal and within-modal priming conditions, we only used matched trials for the latter.

The next set of trials included controls for which the lexical items differed across the prime and target, but the allophonic variant remained the same. For example, [lago]–[fuego] *lago–fuego* 'lake–fire' or [laðo]–[kaða] *lado–cada* 'side–each'. There were 15 approximant and 15 stop control trials across the three places of articulation. We also included 30 filler trials. The filler trials included different lexical items for the primes and targets, and the items did not have any of the target sounds or graphemes in any position in the word. There were also 30 filler–nonword trials and 30 nonword–nonword trials.

In total, there were 30 approximant and stop-matched trials, 30 approximant or stop control trials, 30 filler–filler trials, 30 filler–nonword trials and 30 nonword–nonword trials for the within-modal auditory form priming block of Experiment 1 (see Table 1).

For the cross-modal trials, the prime was always written and the target was auditory. For the 15 approximant matched trials (five for each place of articulation) the written prime and auditory target were the same word, with the approximant variant in intervocalic position (e.g. CABINA–[kaβina], 'cabin'). For the 15 stop matched trials, the written prime and

auditory target were the same word, with the stop variant in intervocalic position (e.g. CABELLO–[kabeɟo] 'hair'). There were also 15 approximant control items (e.g. AGUA–[paɣo] *pago* 'water-payment', for which the prime and target were different words, with the approximant in word-medial position. There were 15 stop control items for which the prime and target were different words, with the stop in intervocalic position. For the control trials, the intervocalic grapheme always matched the intervocalic allophone of the target.

There were 30 filler trials, for which the prime and target were different words that did not have the target graphemes or allophone in them, 30 filler–nonword trials and 30 nonword trials. Table 1 presents the trial types used in Experiment 1.

On a separate occasion following the experiment session, all L2 participants completed a rating/translation task for the lexical items used in the experiment. The familiarity ratings were on a scale of 1–5, where a score of 1 corresponded to the statement 'I am not familiar with this word' and 5 corresponded to 'I am familiar with this word and use it or hear it regularly.' For the approximant and stop matched and control items, the average familiarity rating was 4.3 (SD = 0.72) and the average for successful translation percentage was 94%. For the filler items, the average familiarity rating was 4.4 (SD = 0.52). The average for successful translation percentage was 91.3%.

To ensure that the stimuli were acoustically acceptable, two native Mexican Spanish-speaking judges rated all items on a scale of 1 (not natural) to 5 (natural) for naturalness and acceptability. For the tokens that followed the expected distributional patterns of Spanish (approximant-medial, stop-onset) the average score was 4.96 (SD = 0.011). For the other tokens, the native Mexican Spanish-speaking judges were asked to judge whether the item sounded like a plausible example of that word, even if they would not necessarily pronounce it that way themselves. For these tokens, the average score was 3.01 (SD = .02). The judges were from Mexico City, a highland dialect very similar in phonetic and phonological characteristics to the Spanish spoken in Bogotá, Colombia and also from the same dialect regions as the native Spanish-speaker participants.

c   Procedure. Participants saw a '+' on the computer screen and then saw the forward mask (#####) for 500 ms. After the forward mask, for the cross-modal trials, the written prime appeared in lower-case letters for 67 ms. Following this, participants heard an auditory target and made a lexical decision by pressing a key on a button box (RB834) connected to a MacBook Pro laptop computer. For the within-modal trials, the auditory prime was played after the forward mask, followed by a 500 ms pause, and then the auditory target was presented. Participants were tested in a quiet room with noise canceling headphones. Experimental stimuli were presented using Superlab 4.5 software and reaction times were recorded as soon as the button was pressed, following the target, whether auditory or written. Participants had 2 seconds to respond before the trial timed out. The start of the next trial was indicated by the '+' in the center of the screen.

2   Results

In total, there were 300 trials (150 for each mode) * 34 L2 participants, giving a total of 10,200 trials. We eliminated responses that were greater than 2 SD above or below the mean for the participant (3% of all responses) and also eliminated all incorrect responses

Table 2. Raw reaction times (ms) for L1 English / L2 Spanish and Spanish speaker listeners (SD and % error follow raw response time values).

Condition	Cross-modal: written–auditory		Within-modal: auditory–auditory	
	L2	Spanish	L2	Spanish
Approximant matched	1,368 (368.6, 7.7%)	737.8 (78, 3.5%)	1,309.5 (345.1, 6.5%)	728.2 (208, 3%)
Stop matched	1,277.2 (136.1, 4.1%)	874.4 (164.6, 5%)	1,307.8 (151.6, 5.5%)	807.6 (87, 4.4%)
Approximant control	1,573 (415.8, 11%)	1,038.5 (147.2, 4%)	1,515.5 (351.8, 8.9%)	1,003.02 (299.3, 4.1%)
Stop control	1,517.5 (220, 9.1%)	1,199.6 (123.8, 4%)	1,571.1 (173.9, 8.1%)	1,160.6 (137.3, 5.1%)
Fillers	1,544 (360, 9.2%)	1,140.5 (195.4, 4.1%)	1,512 (337.7, 8.7%)	1,101.3 (240.1, 4%)
Filler nonwords	1,787 (544, 12%)	1,481.2 (434.1, 8%)	1,664 (408, 11%)	1,246.4 (352.2, 8%)
Nonwords	1,702 (521, 10%)	1,564.8 (267.7, 9%)	1,882 (388, 14%)	1,160.6 (137.4, 8.8%)

Notes. Significance codes: *** = .001; ** = .01; * = .05.

from the analysis (6%). The total number of trials analysed for the L2 listeners was 9435. For the Spanish listeners, there were 300 trials (150 for each mode) * 32 participants, giving a total of 9,600 trials. We eliminated responses that were greater than 2 SD above or below the mean for the participant (.8% of all responses) and also eliminated all incorrect responses from the analysis (1.2% of all responses). The total number of trials analysed for the Spanish listeners was 8,910. Reaction times were measured from the offset of the target word for the auditory trials and at the start of presentation for the cross-modal trials.

We created three models to analyse the data. The first included fixed effects for both groups (L2 and native Spanish), modes (cross-modal and within-modal) and all trial types. Following this, we created separate models for each language group with mode and trial type as fixed variables.

We first present the means for error rates and reaction times for the full data set. Following this we present the results from three linear mixed models. The first model includes both groups, all trial types and both cross-modal and within-modal trials. The second model includes only the L2 listeners and the third only the native Spanish listeners. Table 2 presents the mean reaction time and error rates for all conditions across both groups.

The lowest error rate was observed for the cross-modal stop matched trials (4.1%). The highest error rates occurred for the nonword trials in the within-modal condition (14%). Given this, we analysed the accuracy data by modeling response-type likelihood using a mixed effect logistic regression model (with logit link; see Jaeger, 2008) using the lme4 package (Bates et al., 2015b, v.5) in the R programming language and environment (R Core Team, 2016, v.3.3.2). We began with a maximal model that included the binomial dependent variable correct/incorrect, random intercepts for participants and items and interactions between the fixed factors of trial type, mode and group. We then carried out a backward reduction procedure, removing interactions before main effects, to locate the simplest model that did not differ significantly from the full model in terms of variance.

The models were assessed at each stage using chi-square tests on the log-likelihood values. Given that the highest error rate occurred for the nonword–nonword trials, these trials were used as the baseline, as was the L1 group and cross-modal condition. The fixed effects structure included group, mode and trial type with no interactions and the random effects structure included random intercepts for trial and participant, with no random slopes. Spanish speakers were significantly less likely to make errors than the L2 group on the within-modal condition for all trials ($\beta = 2.5$, SE = 1.68, z = 11.21, p < 0.0001). There were significant interactions between trial type and mode whereby all trial types were significantly more accurate than the nonword trials for within-mode (*ps* < 0.001). There were also significant interactions between group and trial type. The Spanish speakers were significantly more accurate than the L2 speakers on all within-mode trials.

An examination of the density and Q–Q plots revealed a non-normal distribution of reaction times that we addressed by applying a log10 transform of the data prior to the statistical analysis.[3] The model for reaction time was created using the function *lmer*, and *p*-values were obtained for the regression coefficients using Satterthwaite's approximation, available through the *lmerTest* package (Kuznetsova et al., 2013, v.2.0–29). We used the *anova()* function (part of the *lmerTest* package) to carry out model evaluation.

The full statistical model included log10 reaction time as the outcome variable,[3] fixed effects were group, mode, trial type and group, mode, trial type interactions. Random effects were participant and trial. Because we had a very high number of random effects parameters, it was important to avoid overparameterization when seeking model maximization. We followed the recommendations in Bates et al. (2015a)[4] and used the rePCA function in the package RePsychLing (Bates et al., 2015b) in R to determine the most parsimonious random effects structure. We began with the fully specified random effects structure, which included random slopes for the group*mode interaction with random intercept for trial and participant. After examining the results from the rePCA analysis, we removed the group*mode interaction for participant but retained the interaction for trial, given the principal components analysis results. Subsequently, we ran a zero-order correlation parameters random effects specialization to remove the correlations. The outcome of this process led to a maximal random effects structure that had group*mode slopes and random intercepts for trial and random intercepts for participant.

We verified that there were no violations of homogeneity of variance by using the *plot()* function in R, which allows an examination of the fitted vs. residual values. An examination of the plotted values did not reveal any patterning of the values. The predictors were analysed for collinearity by computing the condition index of the predictor matrix ('kappa'), using the package *mer.utls.R* (downloaded from github.com/aufrank/R-hacks/blob/master/mer-utils.R). The kappa value was 4.22, indicating mild collinearity (kappa < 10 is typically used as to determine predictor inclusion). Thus, collinearity was determined not to be an issue.

We then verified the model fit using the *anova* function in R, model fit starting with the maximum specified model (main and interaction effects for all fixed factors, random slopes for group and random intercepts for trial and no random slopes for participant. Thus, the final model included interactions among all fixed factors and random slopes and intercepts for trial and only intercepts for participant. Table 3 presents the results from Model 1.

Table 3. Results from Model 1: Reaction time for L2 and native Spanish listeners on all trials (reference category: approximant-matched trials, cross-modal mode, L2 group).

	Variance	Standard deviation		
Random effects:				
Trial (Intercept)	.0134	.116		
Native Spanish speakers	.0192	.139		
Participant (intercept)	.0084	.0.0037		
Residual	.0.0014	.0.037		
Fixed effects:	β	S.E.	t	p
Intercept	3.19	.0097	358.7	< .001***
Native Spanish: within modal	−.022	.017	−1.34	.182
Native Spanish: approximant–matched	.037	.019	24.83	< .001**
Native Spanish: stop–control	.007	.0001	1.56	.076
Native Spanish: stop–matched	.009	.0002	1.48	.081
Native Spanish: filler–filler	.023	.0031	7.5	< 0.001***
Native Spanish: filler–nonword	−.08	.0043	−18.5	< .001***
Native Spanish: nonword–nonword	.019	.0042	4.5	< 0.001***
Native Spanish: within modal: approximant–matched	.097	.0054	−18.01	< .0001***
Native Spanish: within modal: filler–filler	.042	.0046	9.14	< .0001***
Native Spanish: within modal: filler–nonword	0.044	0.006	7.37	< .0001***
Native Spanish: within modal: nonword–nonword	.032	0.005	5.14	.0015**
Native Spanish: within modal: stop–control	.017	.0012	3.12	< .0018**
Native Spanish: within modal: stop–matched	.007	.0044	2.8	< .013*

Notes. Significance codes: *** = .001; ** = .01; * = .05. Due to space restrictions we do not present the mode:trial type interactions. Effect sizes were obtained using the MuMin package in R (Bartoń, 2016, v.1.15.6, based on Nakagawa and Schielzeth, 2013). The R^2m (marginal R^2, corresponding to fixed effects) was 0.58. The R^2c (conditional R^2, corresponding to the fixed effects and random effects) was 0.76.

The fixed factors were treatment coded and reference categories were L2 (group), cross-modal (mode) and approximant-control (trial type). We report only the significant interactions. For trial type, there was no significant difference between groups for the stop control trials in the cross-modal condition ($\beta = .007$, S.E. $= .0001$ $t = 1.56, p = .076$) nor for the stop matched trials ($\beta = .009$, S.E. $= . 0002$ $t = 1.48$ $p = .081$). Significant three-way interactions also emerged whereby the difference between the two groups on cross-modal trials was significantly greater for all trial types (p < .001) when compared to the approximant-control base level.

We created a second model that permitted a closer examination of the hypothesis driving this study, specifically, that L2 listeners will be faster on lexical decision for matched cross-modal trials where the auditory target is the stop variant. In Model 2 we again used log10 reaction times as our outcome variable and a subset of the trial types included in Model 1 above, namely, the matched and control trials only. The total number of trials was 4,165.

We started with a maximal random effects model (Barr et al., 2013) that included varying intercepts for participants, with random slopes for mode and trial type. We also included varying intercepts for trial, with random slops for mode (trial type was a

Table 4. Results from Experiment 1: L2 listeners.

	Variance	Standard deviation		
Random effects:				
Trial (Intercept)	.00986	.0993		
Residual	.001033	.03214		
Fixed effects:	β	S.E.	*t*	*p*
Intercept	3.17	.008838	358.7	< .001***
Within: stop matched	−0.0197	.0035	−4.8	.0052**
Within: approximant control	0.0181	.00351	5.155	< .001***
Within: stop control	0.0072	.00371	2.99	.0041**

Notes. Significance codes: *** = .001; ** = .01; * = .05. Effect sizes for the treatment-coded model (approximant-matched, cross-modal as reference levels for each factor) were obtained using the MuMin package in R (Bartoń, 2016, v.1.15.6, based on Nakagawa and Schielzeth, 2013). The R^2m (marginal R^2, corresponding to fixed effects) was 0.49. The R^2c (conditional R^2, corresponding to the fixed effects and random effects) was 0.72.

between-item variable and was thus excluded). The random structure was simplified using a backward model selection procedure. We removed first the random slopes for each predictor (but kept the random intercepts of participant and item) and at each step the simplified model was compared to the preceding model using a Chi-squared test. If the test results were nonsignificant, we stopped with the simplified random effects structure. The results revealed that the best model included random intercepts for trial only.

For the fixed effects structure, we followed the same backward model selection procedure. The final model included random slopes for trial only, and the fixed effects structure included an interaction term for trial type and mode. The kappa value was 7.22, indicating mild collinearity.

Our main hypotheses focused on the interaction between mode and trial type, and our main interest was to determine whether there was an interaction between the different levels of each factor. The reference levels for Model 2 were approximant-matched trial type and cross-modal mode. We determined that maintaining treatment coding would permit testing the interactions and comparisons corresponding to the hypotheses, namely, that the L2 listeners would show an interaction between mode and trial type. Table 4 shows these results, and Figure 2 portrays these results graphically.

Since all higher-level interactions were significant, we only report those results here. For the within-modal trials, the stop-matched trials were significantly faster than the approximant-matched trials ($\beta = -0.01968$, S.E. = 0.0035, t = −4.8, p = 0.0052). For the within-modal condition, the approximant-control ($\beta = -0.0181$, S.E. = 0.00351, t = −5.155, p < .001) and stop-control trials ($\beta = -.0072$ S.E. = 0.00371, t = 2.99, p = 0.0041) were significantly slower than the approximant matched trials. Overall, the results from the L2 listeners show an interaction between mode and phonetic variant. This is shown in Figure 3.

These results lend support to the hypothesis that when L2 listeners see the written grapheme that corresponds to a native language phonetic category, hearing a target word that is consistent with the L1 grapheme shortens lexical decision times, even when this phonetic variant is not consistent with the L2 input. In other words, native language
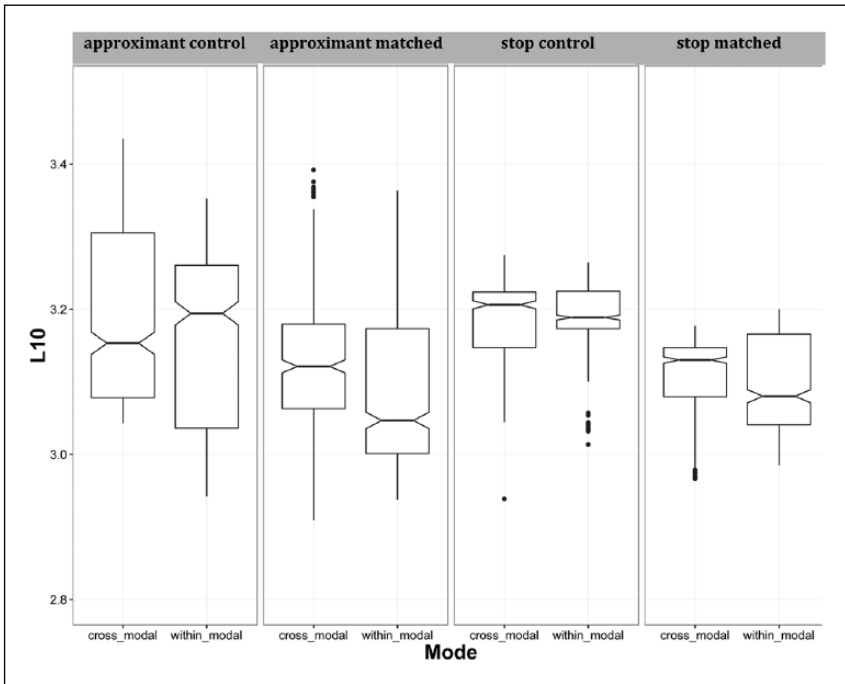
Figure 2. Results from Experiment 1: L2 listeners.

phonetic categories are activated when reading words in a second language and auditory targets that are consistent with the native language phonetic categories are also activated, to a greater extent (as shown by the significant difference between stop-matched trials and approximant matched trials for the within-modal condition) than the second language phonetic variant. This interaction is portrayed in Figure 3.

We created a third model, using reaction time data from the Spanish listeners only, using the function *lmer* within the statistics software package R. The total number of trials was 3,767. We followed the same model testing procedure outlined above, starting with a maximal random effects model (Barr et al., 2013) that included varying intercepts for participants, with random slopes for mode and trial type. We also included varying intercepts for trial, with random slops for mode (trial type was a between-item variable and was thus excluded). When the maximal model failed to converge, the random structure was simplified using a backward model selection procedure. We removed first the random slopes for each predictor (but kept the random intercepts of participant and item), and at each step the simplified model was compared to the preceding model using a Chi-squared test. If the test results were nonsignificant, we stopped with the simplified random effects structure. The results revealed that the best model included random intercepts for trial only.

For the fixed effects structure, we followed the same backward model selection procedure described above for Model 2. The final model included random slopes for trial
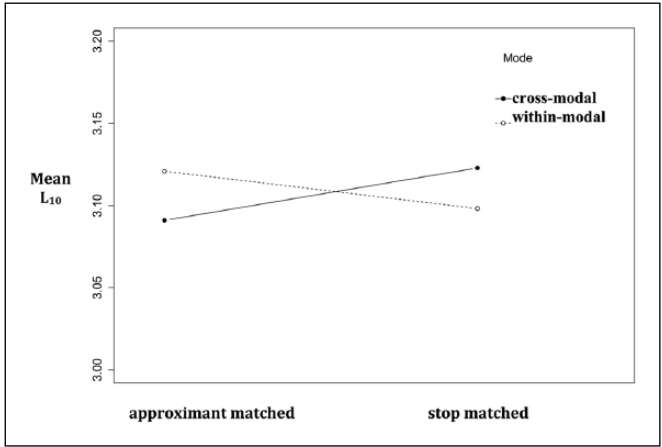
Figure 3. Interaction between mode and phonetic variant for L2 listeners.

only and the fixed effects structure included an interaction term for trial type and mode. The kappa value was 11.04, indicating mild collinearity Table 5 portrays the results from this model.

The results from Model 4 support the hypotheses laid out at the beginning of the study. For the within-modal trials, the native Spanish speakers were significantly slower on the stop-matched trials than on the approximant-matched trials ($\beta = 0.1025$, S.E. $= 0.00333$, $t = 30.74$, $p < 0.0001$). They were also slower on both control trial types, compared to the baseline cross-modal approximant-matched trials (approximant control: $\beta = 0.1325$, S.E. $= 0.0051$, $t = 39.6.$, $p < 0.0001$; $\beta = 0.0988$, S.E. $= 0.00436$, $t = 22.64$, $p < 0.0001$). To test for a main effect of mode, corresponding to the hypothesis that the native Spanish speakers would only show an effect for phonetic variant, we converted trial type to a sum-coded numeric representation (based on procedure described in Levy, 2014) and conducting a likelihood-ratio test between mixed-effects models differing only in the presence or absence of a fixed main effect of mode. Both models included an intercept, a main effect of trial type, and an interaction between mode and trial type. The likelihood-ratio test showed no evidence for a main effect of mode ($p = 0.533$). Figure 4 portrays these results.

Figure 5 is an interaction plot showing the lack of interaction between mode and trial type for the native Spanish listeners. The results from Model 3 show a lack of interaction between mode and trial type for the native Spanish speakers, confirming the hypothesis that native listeners are slower on the phonetic variant that does not occur in their language and the presence of an orthographic symbol does not affect the processing of auditory input in favor of the non-native variant. This is in contrast to the results reported above for the L2 listeners.

3   Discussion: Experiment 1

The results from Experiment 1 revealed an interaction between mode and phonetic variant for the L2 listeners (faster on the stop variants in the cross-modal condition) and no

Table 5. Results from Experiment 1: Native Spanish listeners.

	Variance	Standard deviation		
Random effects:				
Trial (Intercept)	.00868	.0932		
Fixed effects:	β	S.E.	*t*	*p*
Intercept	2.9131	.009537	305.44	< 0.001***
Within: stop matched	0.1025	.003334	30.744	< .001***
Within: approximant control	0.1325	.00335	39.56	< .001***
Within: stop control	0.0988	.004363	22.64	< .001***

Notes. Significance codes: *** = .001; ** = .01; * = .05. Effect sizes were obtained for the treatment-coded model (approximant-matched, cross-modal as baseline) using the MuMin package in R (Bartoń, 2016, v.1.15.6, based on Nakagawa and Schielzeth, 2013). The R^2m (marginal R^2, corresponding to fixed effects) was 0.51. The R^2c (conditional R^2, corresponding to the fixed effects and random effects) was 0.81.
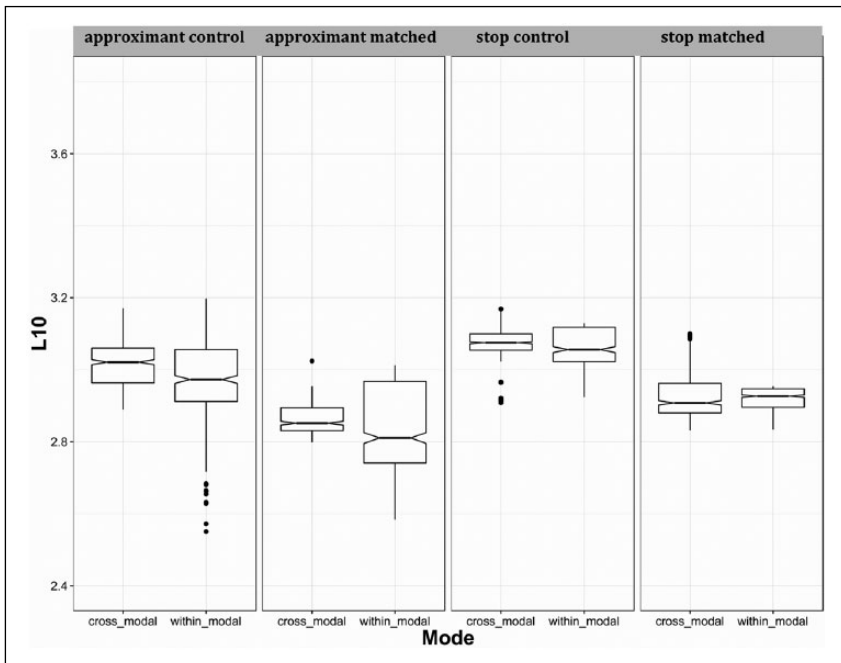

Figure 4. Results from Experiment 1: Native Spanish listeners.

such interaction for the Spanish listeners (slower on the stop variants across both conditions). These results suggest that the L2 listeners are primed by the presence of the grapheme to activate the stop allophone, which corresponds to the grapheme–orthography convention of their native language. The Spanish listeners, on the other hand, did not show an effect for mode, merely for variant. They took longer to respond to trials with the stop variants, which are less consistent with the input in their native language.
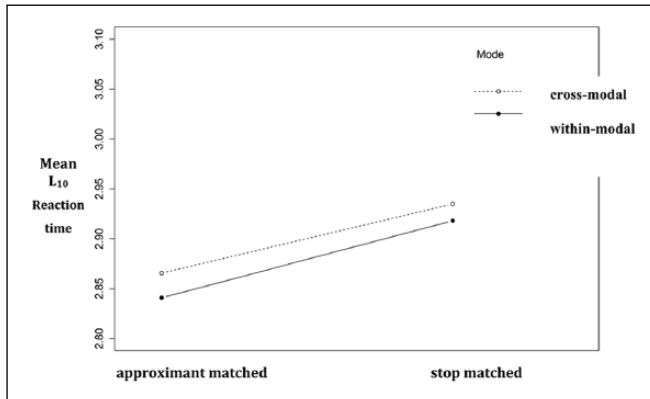
Figure 5. Interaction plot for Experiment 1: Native Spanish listeners.

Experiment 1 used repetition priming, where the same word was presented twice, with or without modifications. The time lapse between presentations was brief (500 ms) and the prime-target activation occurred quickly and directly. In Experiment 2 we ask what happens with a much longer lapse between the presentation of the prime and target, more specifically, we ask which of the two priming patterns will hold? To answer this question, we used long-term repetition priming, a method that allows us to tap longer-term representations and does not strictly require orthography for its completion.

III Experiment 2: Long-term repetition priming

In long-term priming, there are typically a high number of intervening items between the occurrences of the prime and target items, accompanied by a brief break between the two blocks. By examining whether one auditory form can prime another across a high number of intervening trials, we gain insight onto whether L2 listeners perceive the two forms as equivalent (Pallier et al., 2001). This task complements the cross- and within-modal priming methodology used in Experiment 1 by testing the role of L1 orthography in L2 lexical processing on a task that taps into long-term representations and for which orthography is not strictly needed.

1 Method

a Participants. Participants were the same as those in Experiment 1. Experiments 1 and 2 were always counterbalanced and the vocabulary tests for the L2 listeners were applied between the two computer-mediated experiments.

b Stimuli and design. Stimuli were created in the same fashion as those for Experiment 1. Two different speakers were used: one for Block 1 and one for Block 2. Both speakers were female and from Bogota, Colombia. The items used for Experiment 2 were repeated from Experiment 1. This was necessary for two reasons. First, there are a limited number of target words that fulfill the characteristics required for the experiment and second, of that list, there are even fewer that can be considered familiar to L2 Spanish learners. If indeed familiarity

effects played a role, they were uniform across all participants and counterbalanced across experiments. In all models a random effect for item was also included, as were stimuli lists that guaranteed balanced presentation of all stimuli across variants and blocks.

Block 1 consisted of primes and fillers while Block 2 consisted of targets and fillers. For the targets and the corresponding primes, we included five items for each place of articulation, giving 30 items matched across word and variant (15 approximant-matched items and 15 stop-matched items). We included 60 nonword primes and targets (not matched across blocks) and 30 control pairs that had different lexical items across prime-target blocks, but the same phonetic variant (half with stop primes and half with approximant primes). Finally, we included 200 filler items that differed across the two blocks. The entire experiment consisted of 320 words per block. The stimuli were organized across four different lists in which each word appeared once as a prime, once as a target and once with the approximant variant and once with the stop variant. These lists were randomly assigned to participants.

c   Procedure. Participants were run individually in a quiet room using the Superlab 4.5 experimental software program. In both the prime and target blocks, the stimuli were presented binaurally through noise-cancelling headphones. Participants were told to listen to the words and decide if they were real words in Spanish. Inter-stimulus interval (ISI) was 1,500 ms and the program continued to the next item after 2,000 ms if no response was registered. The matched and control items were always separated by 29 items across the two blocks in a pseudo-randomized order. This guaranteed a consistent number of intervening items between the primes and targets and permitted a more controlled comparison across the different conditions. Each list had different combinations of fillers and nonwords intervening. There was a five-minute break between blocks, during which participants were encouraged to get up and walk around. The entire experiment lasted approximately 17 minutes in total.

2   Results

There were 10,880 trials for the L2 Spanish group (34*320) and 10,240 for the L1 Spanish group (32*320). Only correct trials were included in the statistical analysis. To determine the priming effect across the two blocks, we subtracted the reaction time for Block 2 from the reaction time for Block 1. This gave us a difference score that reflected the priming that occurred between the two blocks. Reaction times were calculated from the offset of the target to the moment participants pressed the response button. Reaction times that were more than two standard deviations above the participant's mean for each condition were treated as an error and not included in the analysis. Table 6 presents the mean raw reaction time for each condition, the standard deviation for reaction time and the error rate:

We are most interested in the prime-target pairs that correspond to the matched and unmatched items and therefore only included those critical trials in the linear mixed effects model. The analysis was carried out with R (v.3.01; R Core Team, 2016) and lme4 (Bates et al., 2015b). We created two separate models, one for each group.

For both groups of listeners, model selection and fit were carried out as described above for Experiment 1, using the maximum model and backward verification until

Table 6. Experiment 2 results (SD and % error follow raw response time values).

Group	Approximant matched	Stop matched	Approximant control	Stop control	Nonwords	Fillers
L2	68.9	76.8	8.04	9.05	4.32	8.42
	(73.7, 4.2%)	(44.9, 4.9%)	(16.1, 5.8%)	(28.8, 5.2%)	(112.7, 11.9%)	(64.9, 8.6%)
Native	86	35.5	12.4	16.48	9.17	10.19
Spanish	(23.4, 2.8%)	(18.8, 3.1%)	(81, 5.2%)	(84.2, 4.9%)	(81, 8.1%)	(85, 6.6%)

arriving at the maximum converging model. For the L2 group the final model included trial type as fixed factor and trial as random factor (kappa = 11.2).

The data were treatment coded, with the approximant-matched trials as baseline. For the L2 group, the difference scores between Block 1 and Block 2 for the stop-matched trials were not significantly different from the difference scores for the approximant-marched trials across both blocks ($\beta = 8.1$, S.E. = 7, p = 0.25). However, the approximant control and stop control difference scores were significantly lower ($\beta = -60.03$, S.E. = 6.2, p < 0.0001; $\beta = -41.2$, S.E. = 9.1, p < 0.0001), indicating an attenuated priming effect.

For the native Spanish listeners, the final model included the fixed factor trial type and random intercepts for trial. The kappa value was very low (5.66) so there was no need to be concerned about collinearity. There were significantly lower priming effects for all trials across the two blocks. All trial types showed significantly lower priming effects compared to the approximant matched trials (stop-matched: $\beta = -22.8$, S.E. = 5, p = 0.0044; approximant-control: $\beta = -74.1$, S.E. = 10, p < 0.0001; stop-control: $\beta = -60$, S.E. = 3.2, p < 0.0001). Tables 7 and 8 show the results from the long-term repetition priming experiment.

Figure 6 shows the difference scores across all conditions. These results show that the native Spanish listeners had a significantly greater priming effect for the approximant-matched trials than for any other trial type. This supports the hypothesis that these listeners encode the approximant variant in their lexicon.

3 Discussion: Experiment 2

The results from the long-term repetition priming with lexical decision revealed different patterns across the two language groups. The native Spanish speakers have differentiated these two phonetic variants in their representations and do not perceive lexical items with stops and approximants in word-medial position as exactly the same items. These results are consistent with other studies and show that not only do native speakers encode the relevant phonological contrasts present in their native language but they also encode phonetic contrasts that do not serve to distinguish amongst lexical items (McLennan et al., 2003; Sumner and Samuels, 2009).

The L2 listeners, on the other hand, did not show any significant difference between the matched trials and the related trials, suggesting that in contrast to the native speakers, the L2 group does not encode the stop–approximant distinction in their lexicon. In the following section we discuss Experiments 1 and 2 and how models of L2 lexical activation might account for them.

Table 7. Linear mixed model results from Experiment 2: L2.

	Variance	Standard deviation		
Random effects:				
Trial (Intercept)	21.4	14.6		
Residual	169.2	109.5		
Fixed effects:	β	S.E.	t	p
Intercept	68.2	5.8	11.8	< .001***
stop-matched	8.1	7	1.6	.25
approximant-control	−60.03	6.2	−8.61	< .001***
stop-control	−41.2	9.1	−9.2	< .001***

Notes. Effect sizes were obtained for the treatment-coded model (difference scores for the approximant-matched trials across Blocks 1 and 2) using the MuMin package in R (Bartoń, 2016, v.1.15.6, based on Nakagawa and Schielzeth, 2013). For the L2 listener model, the R^2m (marginal R^2, corresponding to fixed effects) was 0.48. The R^2c (conditional R^2, corresponding to the fixed effects and random effects) was 0.65. For the native Spanish listeners, the values were 0.39 and 0.71 respectively.

Table 8. Linear mixed model results from Experiment 2: Native Spanish listeners.

	Variance	Standard deviation		
Random effects:				
Trial (Intercept)	107.6	28		
Residual	89	22		
Fixed effects:	β	S.E.	t	p
Intercept	87.42	7.987	11	< .001***
stop-matched	−22.798	4.959	−10.5	< .001***
approximant-control	−74.112	9.958	−15	< .001***
stop-control	−60.338	3.52	−12.2	< .001***

Notes. Effect sizes were obtained for the treatment-coded model (difference scores for the approximant-matched trials across Blocks 1 and 2) using the MuMin package in R (Bartoń, 2016, v.1.15.6, based on Nakagawa and Schielzeth, 2013). For the L2 listener model, the R^2m (marginal R^2, corresponding to fixed effects) was 0.48. The R^2c (conditional R^2, corresponding to the fixed effects and random effects) was 0.65. For the native Spanish listeners, the values were 0.39 and 0.71 respectively.

IV General discussion

In the current study we examined how orthography primes allophones in L1 English / L2 Spanish learners. We predicted that L1 English / L2 Spanish listeners would exhibit greater priming for their native allophonic variant than for the L2 variant on a cross-modal priming task (written–auditory) than on a within-modal (auditory–auditory) priming task. The results supported this hypothesis, suggesting that an orthographic symbol consistent with an L1 allophone activates that allophone when it occurs in the input more strongly than the corresponding L2 allophone. Experiment 2 used long-term repetition priming and showed that these effects are attenuated when no orthographic symbol is present and priming occurs over longer periods of time.
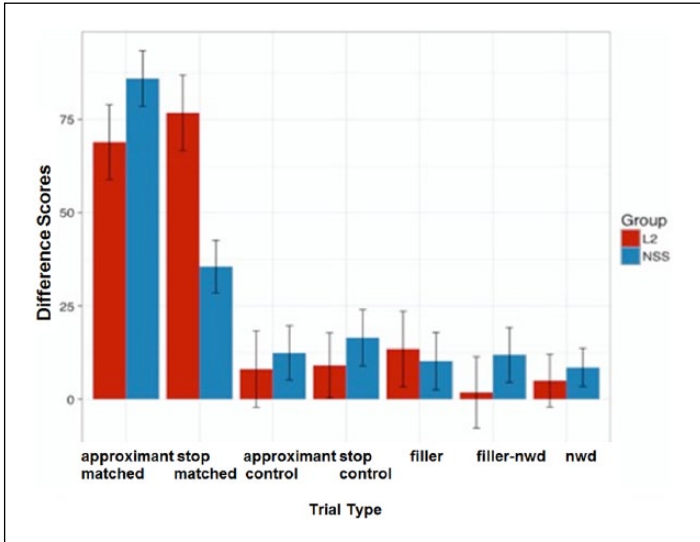
Figure 6. Difference Scores with ±.95CI for Experiment 2.

The results from Experiment 1 showed a contrast between the two listener groups. The native Spanish listeners were slower responding to targets with stops in intervocalic position across both the cross-modal and within-modal trials. There was no interaction between mode and phonetic variant for this group. The L2 listeners showed an interaction between mode and phonetic variant. However, for Experiment 2, the L2 listeners did not show any differentiation in priming effects across the two blocks for either variant. For Experiment 2, the native Spanish speakers showed the greatest priming effect for approximant-matched trials, as predicted.

The question we are left with now is how best to account for these results. We require a model that can take into account cross-linguistic influences between the sub-lexical (subphonemic) level and restrict these influences to the cross-modal condition, where the presence of an orthographic symbol activates the corresponding L1 phonetic variant.

One model that does incorporate cross-modal activation and also incorporates subphonemic information is the BLINCS model (Bilingual Language Interaction Network for the Comprehension of Speech; Shook and Marian, 2013). BLINCS allows for the encoding of phonemes as three-element vectors, consisting of phonological features that are encoded in a phonolexical level that respects syllable positions of all segments. BLINCS incorporates self-organizing maps (SOMs) that are similar to unsupervised learning algorithms. Self-organizing maps work on the Hebbian learning principle that 'what fires together wires together' (Hebb, 1949) and, as the SOM receives input, it is mapped to the node that is best matched to it, altering the value of the node itself to become more similar to the input (Shook and Marian, 2013: 305). The BLINCS model is highly interactive and includes phonological, phonolexical,

ortholexical and semantic levels. At the phonolexical level, SOMs in BLINCS allow for the organization of words from the bilingual's two languages according to their phonological similarity.

There are two key advantages to the BLINCS model in terms of the accounting for the results in this study. First, as stated, it captures the dynamic nature of bilingual lexical processing by using self-organizing maps and, second, it allows for shared phonological representations for segments that occur in both languages and language-specific activation of sounds that exist in only one language. In the present context, we propose that the approximant and stop are close in the phonological space of the listener (both L1 listeners and, crucially, L1 English / L2 Spanish listeners) and may even be in the same category in the case of the L1 listener, undifferentiated in the lexical representations but still differentiable in a task such as auditory repetition priming.

The level of activation for shared sounds will be higher, given the higher frequency and greater co-activation across both languages. This is particularly relevant in the present context since the Spanish input received by native English speakers does include stops; stops are infrequent in intervocalic position but do occur in post-pause and other phonetic contexts. Importantly, L1 English / L2 Spanish learners receive reinforcement from the Spanish input regarding the orthography–phonology connection between the graphemes 'b d g' and the stop variant, given that they are consistent with at least part of the input received in the second language.

For the auditory form priming condition, on the other hand, the lack of activation of the stop variant for the L2 listeners can be understood as a function of the fewer number of mappings involved, or fewer number of intervening representations and activations, and therefore less competition. Repetition priming fosters a direct comparison between prime and target. This direct comparison means a lower threshold of activation for the approximant variant, given that there is no orthography mediating access and therefore no grapheme intervening that could potentially activate the stop variant. This explanation could also account for why orthography is not always activated, i.e. why it only plays an influential role in the cross-modal masked priming context and not in the auditory prime context. In the auditory prime condition, the approximant requires a lower threshold of activation that may not permit the intervention of the orthographic connection to the L1. Instead, either orthography is not activated at all or it is activated weakly and does not influence lexical activation to the same extent as in the cross-modal condition. Given that BLINCS assumes shared phonological space for the bilingual's two languages, this model can account for the activation of the stop variant by the grapheme, given its higher activation across English and its activation in Spanish as well.

In the current study, the L2 participants were not balanced bilinguals and had never spent longer than 6 weeks in a country where Spanish was spoken. Moreover, they were all late L2 learners who had acquired the language in classroom contexts, which has important repercussions for orthography–phonology activation. Many L2 learners acquire their target language in classroom contexts and are encouraged to read in their second language (mostly in the form of vocabulary learning activities) right from the first day of class. Thus, the development of target language reading and writing skills occurs before phonological categories are in place, the opposite of what occurs with children becoming literate in their first language. Veivo and Järvikivi (2013) propose that for

literate L2 learners the 'restructuring' of phonological representations by orthography (as occurs in L1 literacy acquisition by children) would be better conceptualized as 'co-structuring', whereby L2 orthographic and phonological representations are created together with previously established L1 representations. As stated in Bassetti (2008: 204), L2 acoustic input is affected by the existence of another phonological system and the 'L2 orthographic input … is also modulated by the presence of another orthography.' L2 learners interpret the L2 orthography–phonology relationship through their first language, which can help or hinder L2 phonological development. We can liken this to a double filter effect: orthography encourages the formation of certain categories (and discourages others) and existing L1 categories underscore this, making the establishment of new orthography–phonology connections especially challenging, particularly when there is no explicit functional need. Orthographic input, sometimes reinterpreted according to L1 orthography–phonology correspondences, interacts with acoustic input in shaping learners' L2 phonological representation and leading to nontarget like lexical representations (Showalter and Hayes-Harb, 2013).

In closing, the results from this study add to others that show a need for models of bilingual lexical activation that take into account how orthographic symbols map onto sub-phonemic phonetic and phonemic categories in their second, or nondominant language. As well, given the differences in priming effects across the different conditions, we need to keep in mind that the relationship between a bilingual's two languages is rarely – if ever – stable, and any model that attempts to capture it must be dynamic and flexible to reflect changes in proficiency, language dominance and context of interaction. When a bilingual uses language to carry out different tasks, different language resources will be called upon and different outcomes will result.

Declaration of conflicting interests

Funding

Notes

1. Cross-modal priming has been elusive at prime durations shorter than 50 ms. Kouider and Dupoux (2001) examined this issue in a study with native French speakers and determined that cross-modal priming occurred only at 67 ms prime duration. The authors concluded that conscious perception of the prime was required for it to activate the target, which did not occur at 33 ms but did occur at 67 ms. Grainger et al. (2003) also used cross-modal and within-modal priming to test the effects of different prime durations in a lexical decision task. These authors found that repetition priming effects (the prime and target were the same items) at 50 ms and 67 ms across modalities but only found within-modality effects for pseudohomophones at 67 ms.
2. The vocabulary test was the Bilingual Expressive One-word Picture Test and its Receptive counterpart (EOWPT/ROWPT; Brownell, 2012). The grammar test was an in-house

evaluation of grammatical items covered in the pre-major-level Spanish courses at the same university.
3. Log frequencies reflect the base-10 logarithm of the overall frequency in the database, plus one. Log transformation is a normalization procedure that allows statistical tests to be run on the frequency data.
4. This article was in response to Bates et al.'s (2015b) call for maximal fixed effects structure under all circumstances, which could potentially lead to overparamaterization of models with a high number of parameters, as is the case here.

References

Alcock KJ, Ngorosho D, Deus C, and Jukes MCH (2010) We don't have language at our house: disentangling the relationship between phonological awareness, schooling, and literacy. *British Journal of Educational Psychology* 80: 55–76.

Barr DJ, Levy R, Scheepers C, and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–78.

Bartoń K (2016) MuMIn: multi-model inference. R package, version 1.15.6. Available at: http://r-forge.r-project.org/projects/mumin (accessed December).

Bassetti B (2006) Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy* 9: 95–114.

Bassetti B (2008) Orthographic input and second language phonology. In: Piske T and Young-Scholten M (eds) *Input matters in SLA*. Clevedon: Multilingual Matters, pp. 191–206.

Bates D, Kliegl R, Vsishth S, and Baayen H (2015a) Parsimonious Mixed Models. *arXiv:1506.04967* [stat.ME]. Available at: https://arxiv.org/pdf/1506.04967 (accessed December 2016).

Bates D, Maechler M, Bolker B, and Walker S (2015b) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014. Institute for Statistics and Mathematics of WU website. Available at: http://CRAN.R-project.org/package=lme4 (accessed December 2016).

Brownell R (ed.) (2012) *Expressive One-Word Picture Vocabulary Test manual*. 3rd edition. Novato, CA: Academic Therapy Publications.

Brysbaert M, Van Dyck G and Van De Poel M (1999) Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance* 25: 137–48.

Carrasco P, Hualde JI, and Simonet M (2012) Dialectal differences in Spanish voiced obstruent allophony: Costa Rican versus Iberian Spanish. *Phonetica* 69: 149–79.

Castles A, Wilson K, and Coltheart M (2011) Early orthographic influences on phonemic awareness tasks: Evidence from a preschool training study. *Journal of Experimental Child Psychology* 108: 203–10.

Church BA and Schacter DL (1994) Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 521.

Cutler A and Davis C (2012) An orthographic effect in phoneme processing, and its limitations. *Frontiers in Psychology* 3.

Cutler A and Davis C (2012) An orthographic effect in phoneme processing, and its limitations. *Frontiers in psychology* 3. Electronic publication. Published online: 7 February 2012 (doi: 10.3389/fpsyg.2012.00018).

Dehaene S, Pegado F, Braga LW et al. (2010) How learning to read changes the cortical networks for vision and language. *Science* 330(6009): 1359–64.

Dich N (2011) Individual differences in the size of orthographic effects in spoken word recognition: The role of listeners' orthographic skills. *Applied Psycholinguistics* 32: 169–86.

Escudero P, Hayes-Harb R, and Mitterer H (2008) Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60.

Escudero P and Wanrooij K (2010) The effect of L1 orthography on non-native vowel perception. *Language and Speech* 53: 343–65.

Forster KI and Davis C (1984) Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10: 680–98.

Goldinger SD (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105: 251–79.

Grainger J, Diependaele K, Spinelli E, Ferrand L, and Farioli F (2003) Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29: 1256–69.

Guasch M, Boada R, Ferré P, and Sánchez-Casas R (2013) NIM: A web-based Swiss army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods* 45: 765–71.

Hallé PA, Chéreau C, and Seguí J (2000) Where is the /b/ in 'absurde' [apsyrd]? It is in French listeners' minds. *Journal of Memory and Language* 43: 618–39.

Hayes-Harb R, Nicol J, and Barker J (2010) Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech* 53: 367–81.

Hebb DO (1949) *The organization of behavior: A neuropsychological approach*. New York: John Wiley.

Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59: 434–46.

Kouider S and Dupoux E (2001) A functional disconnection between spoken and visual word recognition: Evidence from unconscious priming. *Cognition* 82: B35–B49.

Kuznetsova A, Brockhoff P, and Christensen R (2013) lmerTest: Tests in Linear Mixed Effects models. Version 2.2–20. Available at: http://cran.r-project.org/web/packages/lmerTest/index (accessed December 2016).

Lavoie L (2001) *Consonant strength, phonological patterns and phonetic manifestations*. New York: Garland.

Levy R (2014) Using R formulae to test for main effects in the presence of higher-order inter-actions. *arXiv:1405.2094* [stat.ME]. Available at: https://arxiv.org/pdf/1405.2094 (accessed December 2016).

Luce PA, McLennan CT, and Charles-Luce J (2003) Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In: Bowers J and Marsolek C (eds) *Rethinking implicit memory*. Oxford: Oxford University Press, pp. 197–214).

McLennan CT, Luce PA, and Charles-Luce J (2003) Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29: 539–53.

Nakagawa S and Schielzeth H (2013) A general and simple method for obtaining R2 from general-ized linear mixed-effects models. *Methods in Ecology and Evolution* 4: 133–42.

Pallier C, Colomé A, and Sebastián-Gallés N (2001) The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science* 12: 445–49.

Pattamadilok C, Knierim IN, Kawabata Duncan KJ, and Devlin JT (2010) How does learning to read affect speech perception? *Journal of Neuroscience* 30: 8435–44.

Pattamadilok C, De Morais JJ, and Kolinsky R (2011) Naming in noise: The contribution of ortho-graphic knowledge to speech repetition. *Frontiers in psychology* 2. Electronic publication. Published online 5 December 2011 (doi: 10.3389/fpsyg.2011.00361).

Perre L, Bertrand D, and Ziegler JC (2011) Literacy affects spoken language in a non-linguistic task: an ERP study. *Frontiers in psychology* 2. Electronic publication. Published online 19 October 2011 (doi: 10.3389/fpsyg.2011.00274).

Perre L, Pattamadilok C, Montant M, and Ziegler JC (2009) Orthographic effects in spoken language: on-line activation or phonological restructuring? *Brain Research* 1275: 73–80.

Rafat Y (2015) The interaction of acoustic and orthographic input in the acquisition of Spanish assibilated/fricative rhotics. *Applied Psycholinguistics* 36: 43–66.

Rastle K, McCormick SF, Bayliss L, and Davis CJ (2011) Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 1588–94.

R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org (accessed December 2016).

Sebastián-Gallés N, Martí MA, Carreiras M, and Cuetos F (2000) *LEXESP: Una base de datos informatizada del español*. Barcelona: Universitat de Barcelona.

Serniclaes W, Ventura P, Morais J, and Kolinsky R (2005) Categorical perception of speech sounds in illiterate adults. *Cognition* 98: B35–B44.

Shea CE and Curtin S (2010) Discovering the relationship between context and allophones in a second language. *Studies in Second Language Acquisition* 32: 581–606.

Shea CE and Curtin S (2011) Experience, representations and the production of second language allophones. *Second Language Research* 27: 229–50.

Shook A and Marian V (2013) The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism: Language and Cognition* 16: 304–24.

Showalter CE and Hayes-Harb R (2013) Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research* 29: 185–200.

Sumner M and Samuel AG (2009) The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60: 487–501.

Taft M, Castles A, Davis C, Lazendic G, and Nguyen-Hoan M (2008) Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language* 58: 366–79.

Tanenhaus MK, Leiman JM, and Seidenberg MS (1979) Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior* 18: 427–40.

Van Wijnendaele I and Brysbaert M (2002) Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance* 28: 616–27.

Veivo O and Järvikivi J (2013) Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism: Language and Cognition* 16: 864–83.

Ventura P, Kolinsky R, Fernandes S, Querido L, and Morais J (2007) Lexical restructuring in the absence of literacy. *Cognition* 105: 334–61.

Ventura P, Kolinsky R, Pattamadilok C, and Morais J (2008) The developmental turnpoint of orthographic consistency effects in speech recognition. *Journal of Experimental Child Psychology* 100: 135–45.