


Orthography affects L1 and L2 speech perception but not production in early bilinguals

Antje Stoehr^a  and Clara D. Martin^{a,b}^aBasque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain and ^bIkerbasque, Basque Foundation for Science, Bilbao, Spain.

Research Article

Supplementary material can be found online at <https://doi.org/10.1017/S1366728921000523>

Cite this article: Stoehr A, Martin CD (2022). Orthography affects L1 and L2 speech perception but not production in early bilinguals. *Bilingualism: Language and Cognition* 25, 108–120. <https://doi.org/10.1017/S1366728921000523>

Received: 13 August 2020
Revised: 14 July 2021
Accepted: 19 July 2021
First published online: 25 August 2021

Keywords:

orthographically-guided misperception; orthographically-guided mispronunciation; incongruent grapheme-to-phoneme correspondences; early bilinguals; Basque–Spanish bilinguals; auditory lexical decision task

Address for correspondence: Antje Stoehr, Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 20009, Donostia-San Sebastián, Spain a.stoehr@bcbl.eu

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

CAMBRIDGE
UNIVERSITY PRESS

Abstract

Orthography plays a crucial role in L2 learning, which generally relies on both oral and written input. We examine whether incongruencies between L1 and L2 grapheme-phoneme correspondences influence bilingual speech perception and production, even when both languages have been acquired in early childhood before reading acquisition. Spanish–Basque and Basque–Spanish early bilinguals performed an auditory lexical decision task including Basque pseudowords created by replacing Basque /s/ with Spanish /θ/. These distinct phonemes take the same orthographic form, <z>. Participants also completed reading-aloud tasks in Basque and Spanish to test whether speech sounds with the same orthographic form were produced similarly in the two languages. Results for both groups showed orthography had strong effects on speech perception but no effects on speech production. Taken together, these findings suggest that orthography plays a crucial role in the speech system of early bilinguals but does not automatically lead to non-native production.

Introduction

When bilinguals acquire the phonological and phonetic systems of their two languages, they are generally confronted with phoneme inventories that overlap to some degree. Sounds that differ in terms of phonetic realization or are only present in one of their languages are especially likely to cause difficulties in perception and production. Even bilinguals who have acquired their second language (L2) in early childhood may not consistently distinguish phonemes that are similar in their two languages (Pallier, Bosch & Sebastián-Gallés, 1997; Samuel & Larraza, 2015; Sebastián-Gallés, Echeverría & Bosch, 2005; Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer & Diaz, 2006; Sebastián-Gallés, Vera-Constan, Larsson, Costa & Deco, 2009). For instance, Samuel and Larraza (2015) found that Spanish–Basque early bilinguals did not always distinguish the unique Basque affricate /tʃ/ from the affricate /tʃ/, which exists in both Basque and Spanish. Among other tasks, they had Spanish–Basque early bilinguals perform an auditory lexical decision task (LDT). In this task, participants had to provide lexicality judgments on words where the critical affricate was mispronounced: for example, the Basque word /itʃal/ <itzal> “shadow” mispronounced as [itʃal] or the Basque word /kutʃa/ <kutxa> “box” mispronounced as [kutʃa]¹. Participants accepted mispronunciations as real words in about 30% of all cases. To investigate whether this was due to a perceptual deficit, participants performed an AXB discrimination task testing their ability to auditorily discriminate the critical sounds embedded in meaningless syllables. Performance was close to ceiling, suggesting acceptance of mispronunciations in the LDT was not merely the result of a perceptual deficit. Samuel and Larraza (2015) conducted their study in the Spanish Basque Country, where large parts of the population are native (L1) Spanish speakers with L2-Basque, who presumably mispronounce Basque affricates. The authors argued that frequent exposure to mispronounced variants had led listeners to treat the mispronounced form as an allophonic variant of the target form and acceptance of mispronunciations should be considered an efficient adaptation to the actual linguistic environment rather than an error. In a similar line of research, Spanish-Catalan early bilinguals were found to have difficulty distinguishing the Catalan vowel /ɛ/ from the adjacent Catalan and Spanish vowel /e/ (Pallier et al., 1997; Sebastián-Gallés et al., 2005, 2006, 2009). Sebastián-Gallés and colleagues (2005, 2006, 2009) found that early Spanish-Catalan bilinguals accepted Catalan words in which the vowel /ɛ/ was mispronounced as [e] (e.g., /gəʎɛðə/ <galleda> “bucket” mispronounced as [gəʎeðə]) and vice versa (e.g., /uʎɛrəs/ <ulleres> “glasses” mispronounced as [uʎeɾəs]) in approximately 75% of all cases. Even Catalan-dominant bilinguals accepted mispronounced words in about 40% of all cases, but only when /ɛ/ was

¹Hereafter, phonemes are represented between / /, phones between [] and graphemes between < >.

mispronounced as [e]. Spanish-Catalan bilinguals also struggled to discriminate these two sounds perceptually (Pallier et al., 1997). Sebastián-Gallés et al. (2009) suggested that bilinguals' acceptance of mispronounced word forms could either be due to their inability to perceive the sound contrast or because they maintained two lexical representations for each word: one containing the target vowel and one based on the mispronounced form to which they were presumably routinely exposed, since many inhabitants of Catalonia are L1-Spanish speakers who acquired Catalan as an L2. In fact, accepting mispronunciations in the L1 is an important prerequisite for understanding foreign-accented speech. Here, we speculate that Spanish-Catalan bilinguals may have a higher error rate than Spanish-Basque bilinguals in part because Catalan /*ε*/ and Catalan and Spanish /*e*/ share the grapheme <*e*>, whereas Basque /*tʃ*/ <*tx*> and Spanish /*tʃ*/ <*ch*> have unique spellings. In the following, we will support this speculation with evidence on the role of orthography in speech perception and production.

Orthography is known to play an important role in auditory language processing in monolingual adults. For example, L1-English speakers seem to rely on orthographic information in auditory rhyme judgements, detecting words that rhyme more quickly when their spellings match (e.g., *tie* and *pie*) than when they differ (e.g., *tie* and *rye*; Seidenberg & Tanenhaus, 1979). This finding, amongst others, suggests a close association between orthographic representations and auditory lexical representations in L1 listeners. More recently, researchers have started investigating the complex effects of orthography on L2 learning. A number of studies have not found orthographic effects on L2 speech processing (Dean & Valdés Kroff, 2017; Simon, Chambless & Alves, 2010). Others have provided evidence that exposure to orthography in addition to oral input enhances lexical learning, increasing phonemic accuracy in both perception and production (Bürki, Welby, Clément & Spinelli, 2019; Erdener & Burnham, 2005; Escudero, Hayes-Harb & Mitterer, 2008). Yet other research has offered evidence that orthography can have negative impacts. For instance, incongruent L1-L2 grapheme-to-phoneme correspondences (GPCs) appear to have detrimental effects on phonetic aspects of L2 speech perception and production (Bassetti, 2017; Bassetti & Atkinson, 2015; Bassetti, Sokolović-Perović, Mairano & Cerni, 2018; Bürki et al., 2019; Cerni, Bassetti & Masterson, 2019; Nimz & Khattab, 2020; Rafat, 2016; Stoehr & Martin, 2021; Young-Scholten & Langer, 2015). These mixed findings may be related to the use of different tasks, the presence or absence of orthographic information in these tasks, and the materials used. Yet, the general picture emerging from previous studies is facilitation when L1 and L2 share GPCs, and hinderance when GPCs differ. For instance, L1-English learners of Spanish are likely to mispronounce the Spanish word <*zumo*> /*θumo*/ "juice" as [zumo] because the grapheme <*z*> corresponds to the phoneme /*z*/ in English, not to /*θ*/ as in (Castilian) Spanish. This is intriguing because the phoneme /*θ*/ (<*th*>) also exists in English, indicating that production difficulty in L2 cannot account for this type of mispronunciation. Cross-linguistic incongruencies in GPCs are very common since the 26 letters of the Roman script are used to represent the phonemes of nearly all Western European languages, some Eastern European languages, and even non-European languages such as Vietnamese, Swahili and Tagalog, yet the phoneme inventories and inventory sizes of these languages differ greatly. Incongruent GPCs between languages appear to have particularly strong impacts on instructed L2 learning, as described below.

L2 learning in Western societies most commonly takes place in a classroom setting, where literate children, teenagers, or adults learn the L2 through simultaneous auditory and orthographic exposure. As these learners already have robust orthographic knowledge in their L1, the reported influence of L1 GPCs on the L2 is hardly surprising. Yet, even sequential bilinguals who have been immersed in an L2 environment with its wealth of native speaker input for many years appear to be affected by L1 orthography in L2 speech production. This highlights the robustness of orthographic effects on an L2 (Bassetti et al., 2018). In bilingual communities, children are typically exposed to a learning environment that features both languages from birth or early childhood. Sebastián-Gallés and colleagues (Pallier et al., 1997; Sebastián-Gallés et al., 2005, 2006, 2009) and Samuel and Larraza (2015) sampled their participants in such bilingual communities (Catalonia and the Basque Country, respectively). It remains unclear if incongruent cross-linguistic GPCs affect L2 and L1 speech perception and production in early sequential bilinguals who acquired both languages prior to reading acquisition.

The current study addresses orthographic effects in speech perception and production of Spanish-Basque and Basque-Spanish bilinguals in the Spanish Basque Country, who acquired Basque and Spanish in early childhood, before receiving formal reading instruction. In the Basque Country, both Spanish and Basque have official status. Both languages are used in the public educational system and large sectors of society. Individuals raised in the Basque Country are frequently exposed to spoken and written Basque and Spanish and are generally highly proficient in both languages. This provides a suitable test case to investigate the influence of incongruent cross-linguistic GPCs on speech perception and production and to test whether L2 orthography also affects L1 perception and production in early bilinguals who learned both of their languages before acquiring literacy.

The present study uses a similar experimental design to Samuel and Larraza (2015) and Sebastián-Gallés et al. (2005) to test whether incongruent GPCs in Spanish and Basque affect Spanish-Basque and Basque-Spanish bilinguals' speech perception and speech production. Participants were first tested in an auditory LDT in which the Basque lamino-alveolar fricative /*ʃ*/ was mispronounced as the Spanish interdental fricative /*θ*/; importantly, both phonemes are represented by the same grapheme, <*z*>. The same participants then completed an AXB speech sound discrimination task to test whether they were capable of perceptually distinguishing these two sounds. Finally, the same participants were tested on their speech production in Basque and Spanish to ascertain whether they commonly mispronounced Basque /*ʃ*/ as Spanish /*θ*/. Below, we briefly discuss the Basque and Spanish phonological and writing systems before moving on to describe the study.

The phonemic inventories of Basque and Spanish largely overlap and both use the Roman script. (Castilian) Spanish has four fricative phonemes (/f/ <*f*>, /θ/ <*c*>, <*z*>, /s/ <*s*>; /x/ <*g*>, <*j*>), while (standard) Basque has five (/f/ <*f*>; /*ʃ*/ <*z*>; /*ʃ*/ <*s*>; /*ʃ*/ <*j*>²). These fricatives are all voiceless but differ in their place of articulation. The center of gravity, measured in Hertz (Hz), is a reliable cue to determine differences in the place of articulation across voiceless fricatives (Gordon, Barthmaier & Sands, 2002). It is measured as the average frequency on a

²The grapheme <*j*> in Basque corresponds to either /*x*/ or /*j*/ depending on the regional dialect. /*x*/ is the most common pronunciation in Gipuzkoa, where the present study was conducted.

spectrum, weighted by the amplitude. The apico-alveolar fricative /s/, produced with the tip of the tongue placed against the alveolar ridge, corresponds to <s> in both Spanish and Basque. The lamino-alveolar fricative /ʃ/, produced with the blade of the tongue placed against the alveolar ridge, corresponds to <z> in Basque, but is absent from the Spanish phoneme inventory and is notoriously difficult for L1-Spanish learners of Basque to acquire. L1-Spanish learners of Basque are less accurate in discriminating the lamino-alveolar fricative /ʃ/ from the apico-alveolar fricative /s/ compared to control sound contrasts, especially when Basque has been learned at a later age (Larraza, Samuel & Oñederra, 2016). Although not empirically tested in their study, Larraza et al. (2016) also note that L2-Basque speakers often produce the acoustically similar apico-alveolar fricative /s/ instead of the lamino-alveolar fricative /ʃ/.

Given these reported perception and production patterns, it appears that Basque/Spanish /s/ <s> most closely acoustically resembles (thus, is most likely to replace) Basque /ʃ/ <z>, a particularly difficult sound for L1-Spanish learners of Basque (Larraza et al., 2016). Crucially, Basque /ʃ/ is connected to Spanish /θ/ by the grapheme <z>. If orthographic effects override sound similarity effects to impact speech perception and/or production in early bilinguals, they might accept mispronunciations of /ʃ/ as [θ] in Basque and/or show the same mispronunciation pattern in speech production. The present study tests this hypothesis by investigating whether early bilinguals whose L2 or L1 is Basque and who acquired both Basque and Spanish before becoming literate nevertheless accept orthographically-guided mispronunciations of Basque words and produce such orthographically-guided mispronunciations themselves. Such an effect would reveal that the striking impact of orthography on phonology is not limited to late bilinguals who already have strong L1 GPCs that interfere with L2 learning. It would also demonstrate, for the first time, that GPCs established during reading acquisition at about six years of age can still modify a phonological system acquired previously – in early childhood or even from birth. In particular, the present study tests the following hypotheses:

- (1) If L2 perception and production are impacted by incongruent L1-L2 GPCs, we expect L1-Spanish-L2-Basque bilinguals to accept Basque words in which the target phoneme /ʃ/ is mispronounced as [θ] and to use this mispronunciation in speech production.
- (2) If the L1 is similarly influenced by incongruent L1-L2 GPCs, the same pattern should be found in L1-Basque-L2-Spanish bilinguals. It has previously been shown that bilinguals accept some degree of L1 mispronunciation, demonstrating the flexibility of the sound perception system even in a native language (Sebastián-Gallés et al., 2005, 2006). However, previous studies have argued flexibility in the L1 perceptual system is based on habituation to mispronunciations present in the environment. Here, we will investigate whether flexibility in L1 perception can also be triggered by the orthographic influence of a highly proficient L2.
- (3) Given the acoustic difference between Basque /ʃ/ and Spanish /θ/, both groups are expected to perceptually distinguish these two sounds.

General methods

The present study consists of a lexical decision task (Experiment I), an AXB speech sound discrimination task (Experiment II), and

a speech production task (Experiment III). The same participants completed all three experiments in the same fixed order in which they appear in this article.

Participants

Thirty L1-Spanish-L2-Basque and thirty L1-Basque-L2-Spanish bilinguals participated in the three experiments ($M_{\text{age}} = 22.6$ years, range = 18–34 years). The L1-Spanish-L2-Basque bilinguals had acquired Basque in early childhood (henceforth, L2-Basque speakers), and the L1-Basque-L2-Spanish bilinguals had acquired Basque from birth (henceforth, L1-Basque speakers). Only bilinguals who reported speaking either no dialect or the Gipuzkoan or Upper Navarese dialects of Basque were recruited for this study. Participants from other dialectal regions would likely be affected by the Basque sibilant merger (Hualde, 2010; Muxika-Loitzate, 2017). All participants also spoke English but reported no knowledge of any other foreign language.

As displayed in Table 1, the L1-Basque and L2-Basque speakers differed significantly on age of acquisition and self-reported exposure to Basque and Spanish. They further differed on Basque language skills but not on Spanish language skills, as measured through interviews³, the Basque and Spanish version of the LexTALE (de Bruin, Carreiras & Duñabeitia, 2017; see Lemhöfer & Broersma, 2012 for the original version), and the BEST⁴ (de Bruin et al., 2017). Participant groups were matched on age, gender, age of L2 acquisition (i.e., Spanish for L1-Basque speakers and Basque for L2-Basque speakers), verbal and non-verbal IQ as evaluated by the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 2004), age of acquisition, and self-reported exposure and proficiency in English.

An additional four participants were tested but not included in data analyses due to technical problems ($N = 3$) and experimenter error ($N = 1$). Participants were recruited from the Basque Center on Cognition, Brain and Language (BCBL) subject pool. They received 8€ compensation and a stamp on their fidelity card (ten stamps merit an additional gift). Informed consent was obtained from all participants prior to starting the experiments. The study had previously been approved by the BCBL's Ethics Committee.

General apparatus and procedure

Participants were tested individually in sound-attenuating chambers at the BCBL satellite laboratory at the University of the Basque Country in Donostia-San Sebastián. All experiments were run on a desktop computer using Open Sesame software (version 3.2.8; Mathôt, Schreij & Theeuwes, 2012). Stimuli in Experiments I and II were presented binaurally over Sennheiser GSP 350 headphones. These auditory stimuli were recorded multiple times by a female native speaker of Basque from Gipuzkoa, while a different female native speaker of Basque from Gipuzkoa selected the best

³Structured interviews were conducted by professional research assistants who had received standardized training for these interviews. The scores can be interpreted as follows: 5: native speaker competence. 4: speakers are highly fluent, able to talk about a wide range of topics, but make occasional errors in long and difficult sentences. 3: speakers are fluent, able to speak at length using a wide range of vocabulary, and generally easy to understand, although they make some mistakes. 2: speakers have limited fluency, able to convey basic meaning using limited vocabulary, with frequent errors that may lead to misunderstandings.

⁴The Basque, English, and Spanish test (BEST) comprises picture naming tasks in the respective languages. The test consists of 65 pictures per language corresponding to non-cognate words.

Table 1. Participant characteristics (in parentheses: SD; range).

	L1-Basque	L2-Basque	<i>p</i> -value	
Age	23 years (4.06; 18–33)	22 years (4.06; 18–34)	>.250	
Gender	21 F & 9 M	21 F & 9 M	>.250	
L2 age of acquisition	2.77 years (2.33; 0–6)	3.37 years (0.85; 3–6)	.190	
Verbal IQ	102 (10; 79–122)	100 (10; 82–120)	>.250	
Non-verbal IQ	110 (10; 86–126)	107 (9; 79–121)	.176	
Basque	Age of acquisition	0 years [birth]	3.37 years (0.85; 3–6)	<.001
	Self-reported exposure	58% (15; 30–90)	25% (11; 10–50)	<.001
	Interview scores	4.87/5 (0.35; 4–5)	3.60/5 (0.72; 3–5)	<.001
	LexTALE scores	91% (8; 66–99)	83% (13; 49–97)	.006
	BEST scores	62.70/65 (3.32; 52–65)	47.10/65 (9; 28–63)	<.001
Spanish	Age of acquisition	2.77 years (2.33; 0–6)	0 years [birth]	<.001
	Self-reported exposure	32% (15; 10–60)	67% (12; 40–90)	<.001
	Interview scores	4.97/5 (0.18; 4–5)	5/5	>.250
	LexTALE scores	92% (5; 79–100)	93% (5; 79–100)	>.250
	BEST scores	64.07/65 (1.57; 58–65)	64.47/65 (0.90; 62–65)	.232
English	Age of acquisition	5.37 years (1.90; 2–9)	5.73 years (1.39; 3–10)	>.250
	Self-reported exposure	9% (8; 0–30)	8% (5; 0–20)	>.250
	Interview scores	3.23/5 (0.73; 2–4)	3.03/5 (0.72; 2–4)	>.250
	LexTALE scores	63% (7; 50–81)	66% (8; 49–91)	.175
	BEST scores	44.13/65 (11.82; 23–62)	39.47/65 (12.54; 11–56)	.143

exemplar for each stimulus (see sections on Experiments I and II for stimulus details). The best exemplar was defined as a recording that clearly matched the pronunciation conditions and did not contain noise or list intonation. Fifty milliseconds of silence were added to the beginning of each audio file to allow for sufficient loading time in the experimental software. This allowed us to avoid any loss of auditory information. All oral and written instructions were given in Basque unless stated otherwise.

Analyses

Data analyses for all experiments were conducted in R software (R Core Team, 2013) using the lme4 package (Bates, Mächler, Bolker & Walker, 2015). Data were analyzed using logistic mixed-effects models for accuracy (Experiments I & II), linear mixed-effects models for reaction time (RT; Experiments I & II), and center of gravity (Experiment III). In linear mixed-effects models, *p*-values for *t*-statistics were obtained using Satterthwaite's method for denominator degrees of freedom through the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017). In Experiments I and II, data points with standardized residuals more than 2.5 standard deviations from 0 were removed using the LMERConvenienceFunctions package (Tremblay & Ransijn, 2020). The complete model outputs of all analyses are provided in the Supplementary Materials (Supplementary Materials).

Experiment I: Lexical Decision Task

The aim of this experiment was to test whether speech perception in L2-Basque and L1-Basque speakers who had acquired both Spanish

and Basque before the onset of reading acquisition were influenced by orthographic incongruencies between their L1 and L2. If L2 perception is impacted by incongruent L1-L2 GPCs, we expected L2-Basque speakers would accept Basque words in which the target phoneme /s/ was mispronounced as [θ]. If the L1 is similarly influenced by incongruent L1-L2 GPCs, we expected L1-Basque speakers would also accept these mispronunciations. Together these results would indicate that incongruent Spanish-Basque GPCs influence both L2-Basque and L1-Basque early bilinguals.

Stimuli

The LDT consisted of 232 Basque stimuli: half were existing Basque words, while the other half were pseudowords created by replacing a single sound of an existing Basque word. Experimental items contained the lamino-alveolar fricative /s/ <z> in the syllable-initial position. In total, 84 experimental items were selected from the BaSp database (Duñabeitia, Casaponsa, Dimitropoulou, Martí, Larraza & Carreiras, 2017): 1/3 were correct pronunciations (henceforth, correctly-pronounced items), 1/3 were orthographic mispronunciations, in which /s/ had been replaced by /θ/ (henceforth, critical items), and 1/3 were control mispronunciations, in which /s/ had been replaced by /x/ (henceforth, control items). In terms of phonological features, both critical /θ/ and control /x/ differed from /s/ in place of articulation but shared manner of articulation and voicing. Each of the 84 items appeared once in each of the three conditions, resulting in six different lists that were counterbalanced across participants. The lists were carefully matched on the following variables derived from the BaSp database (Duñabeitia et al., 2017): position of the critical sound, vocalic context of the critical sound in the

word, number of syllables, frequency, orthographic length, number of neighbors, age of acquisition, concreteness, orthographic Levenshtein distance, Basque–Spanish cognate rate, number of senses⁵, number of translations, RT of Basque bilinguals, and error rate of Basque bilinguals⁶. In addition, the lists were matched on the duration of the sound file and the position of the critical sound (stressed or unstressed position; see Tables S1 & S2 in Supplementary Materials for an overview of the matched variables). Each list thus contained 28 correctly-pronounced items, 28 critical items, and 28 control items. In each list, the remaining 148 stimuli were fillers that were either correctly-pronounced (88 stimuli) or mispronounced (60 stimuli) for a total of 116 words and 116 pseudowords. As in the critical trials, mispronounced fillers contained a single sound substitution (/b/ as [p], /k/ as [g], and /m/ as [n]). These sounds and substitutions were the same as those used in Samuel and Larraza (2015). They were adopted for the current study because they belonged to different sound classes from the critical items (i.e., plosives and nasals instead of fricatives) and, like the critical trials, included only a single-feature deviation from the target sound (voicing for /b/→[p] and /k/→[g]; place of articulation for /m/→[n]). None of the filler items contained /ʃ/ in any position. Stimuli were recorded either pronounced correctly or mispronounced.

Apparatus and procedure

In the LDT, participants had to indicate whether each auditorily presented item was an existing Basque word by pressing one of two keys labeled on the computer keyboard. Half of the participants pressed the left and the other half pressed the right key as soon as they heard a real Basque word. The instructions provided to participants, based on Samuel and Larraza (2015), established a very high threshold for accepting items as real words. Participants were informed that pseudowords would sound very similar to real words, and that they should only accept items as words if they were convinced they were completely correct (see complete instructions in Appendix S1, Supplementary Materials). Each trial started with a fixation dot displayed at the center of the screen for 300ms, then the auditory stimulus was played. The next trial began 700ms after a response was made. Items were presented in randomized order. Participants were instructed to respond as quickly and as accurately as possible, but there was no time limit. The experiment began with a practice block of 12 trials, using the same manipulation as the main task. Feedback was provided during practice but not during the main task. The entire LDT took approximately 12 minutes.

After finishing the LDT, participants' orthographic knowledge of the experimental items was verified in a spelling task. Participants listened again to the correctly-pronounced form of all 84 experimental items and 42 filler items. They were asked to write each word down as accurately as possible. For each participant, only the correctly-spelled experimental items were included in the analysis of the LDT.

Results

In total, 92.46% of the 5,040 critical, control, and correctly-pronounced trials were included in the final analyses. Critical

⁵Number of senses refers to the number of different concepts or entries in the Basque dictionary for the Spanish word.

⁶RT and error rates of Basque bilinguals correspond to the mean RT and error rates for Basque words obtained from 28 completely balanced Basque-Spanish bilinguals tested on the BaSp dataset.

trials contained the orthographic mispronunciation (<z> as [θ]), while control trials contained the control mispronunciation (<z> as [x]). Correctly-pronounced trials (CPs) contained the target pronunciation (<z> as [ʃ]). First, items that were spelled incorrectly during the spelling task were excluded. This was done to ensure that only those items for which participants had a correct orthographic representation were included in the final analyses. This led to the exclusion of 245 trials (4.86% of the data). Data were screened for unreasonably long (>5,000ms) or short (<100ms) reaction times (RTs), but none were found (Baayen & Milin, 2010). Afterwards, 135 trials (2.68% of the data) were removed as outliers (see Analyses section in the General methods). Accuracy on critical trials (<z> as [θ]) was 64% for L1-Basque speakers and 54% for L2-Basque speakers⁷. Both groups were highly accurate on control trials (<z> as [x]; 98% for L1-Basque speakers; 95% for L2-Basque speakers) and correctly-pronounced trials (<z> as [ʃ]; 98% for L1-Basque speakers; 94% for L2-Basque speakers).

Accuracy

The logistic mixed-effects model had Accuracy (1,0) as the dependent variable with fixed effects for Condition (using polynomial coding to compare critical [coded as -1] to control [coded as 1] and critical to CP [coded as 1]) and Group (L1-Basque, coded as 1; L2-Basque, coded as -1) and an interaction term. The model also included random intercepts for Subjects and Items, as well as by-subject and by-item random slopes for Condition, and by-item random slopes for Group. The model detected significant main effects of Condition (critical vs. CP: $\beta = 1.949$, $SE = 0.447$, $z = 4.357$, $p < .001$; critical vs. control: $\beta = 1.022$, $SE = 0.340$, $z = 3.011$, $p = .003$), showing that participants were less accurate in detecting critical mispronunciations than correct and control mispronunciations. A significant main effect of Group ($\beta = 0.531$, $SE = 0.149$, $z = 3.571$, $p < .001$) shows that, overall, L1-Basque speakers were more accurate than L2-Basque speakers. No significant interaction between Condition and Group was detected, which suggests that the effect of Condition was present for both L2-Basque and L1-Basque speakers. The complete model output is provided in Table S3, Supplementary Materials; results are visualized in Figure 1.

Reaction times

RT data was positively skewed (skewness score = 2.562) and therefore log-transformed. This procedure resulted in a moderate skewness score of 0.720. The linear mixed-effects model used log-transformed RTs in ms as a continuous dependent variable, and the remaining structure was identical to the logistic mixed-effects model on Accuracy reported above. The model⁸ detected a significant main effect of Condition between critical and CP trials ($\beta = -0.030$, $SE = 0.008$, $t = -3.694$, $p < .001$), showing that critical

⁷Given the variability in accuracy in the critical condition, individual data were explored to verify that this variability was not due to differential responses in two sub-groups of participants. Overall, most participants showed large variability in their responses in the critical condition: In the L2-Basque group, 19/30 participants scored between 25% and 75% correct; 5/30 scored below 25% (between 4–21% correct), and 6/30 scored above 75% (between 81–93% correct). In the L1-Basque group, 14/30 participants scored between 25–75% correct; 2/30 scored below 25% correct (only 4% correct); and 14/30 scored above 75% correct (between 77–100% correct – only 1 participant reached 100% correct).

⁸The described model combined correctly-answered and incorrectly-answered trials. The same results were obtained in a model based only on the 3,926 (84.25% of the cleaned data) correctly-answered trials (Table S5).

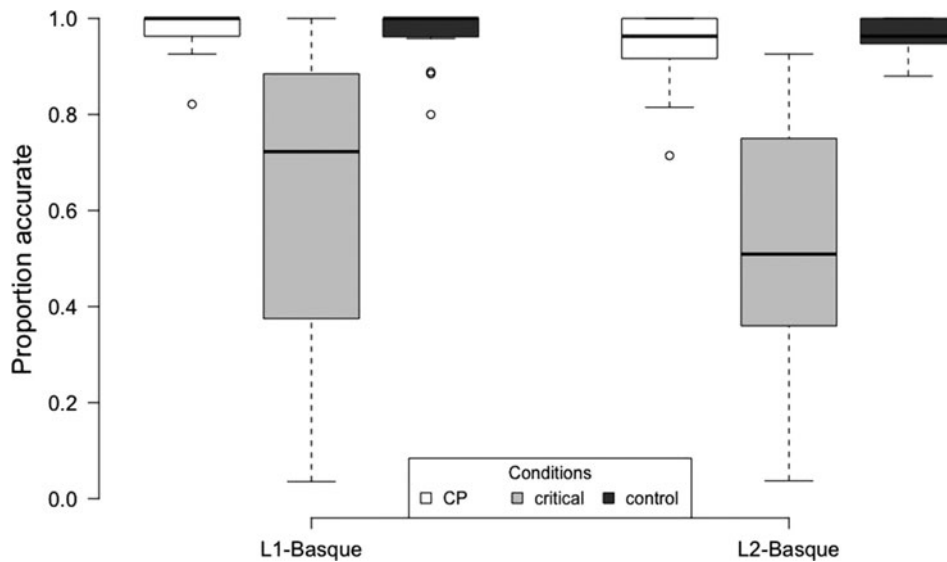


Figure 1. Accuracy in the LDT by group and condition (aggregated over participants).

mispronunciations elicited longer RTs than CPs. No significant differences in RT were observed between critical and control trials. A significant main effect of Group ($\beta = -0.062$, $SE = 0.014$, $t = -4.397$, $p < .001$) shows that L2-Basque speakers overall responded more slowly than L1-Basque speakers. No significant interaction between Condition and Group was detected, suggesting that the effect of Condition (critical vs. CP) was present for L2-Basque and L1-Basque speakers alike. The complete model output is available in Table S4, Supplementary Materials; results are visualized in Figure 2.

Discussion

Both L2-Basque and L1-Basque speakers were less accurate in rejecting words with orthographic than control mispronunciations. L2-Basque speakers performed at chance when responding to orthographic mispronunciations (54% accuracy). While L1-Basque speakers performed slightly better (64% accuracy), this difference was not statistically significant. No significant differences in RT were observed between critical and control conditions in either group, although both groups responded more slowly on critical than control trials (mean difference: L2-Basque: 122ms; L1-Basque: 57ms).

Overall, accuracy for the L1-Basque speakers was similar to that of the L1-Catalan speakers investigated by Sebastián-Gallés et al. (2005, 2006), but the L2-Basque speakers in the present study performed better than the L2-Catalan speakers, who attained a mean accuracy of only approximately 25% on mispronounced trials. This may be because the Catalan vowels / ϵ / and / e / are adjacent in vowel space, and Spanish / e / has [ϵ] as an allophonic variant, making them more similar than the sounds in the present study: Basque / ζ / and Spanish / θ / are distinct phonemes with no allophonic relationship. The L2-Basque speakers in the present study, however, performed less accurately than the L2-Basque speakers in Samuel and Larraza (2015), who detected mispronounced words with a mean accuracy of 67%. This overall poorer performance could reflect the fact that the sound contrast between / ζ / and / θ / tested here is simply more difficult to distinguish than the sound contrast between / t_s / and / t_f / tested in Samuel and Larraza (2015). In Experiment II, we rule out the possibility that our results stem from difficulty in discriminating / ζ /

and / θ /. Instead, we argue that the orthographic link provided by <z> increased task difficulty and led to higher error rates. This would make the present results for early bilinguals particularly striking: orthographic representations affected speech perception in both groups. These participants had acquired phonological representations in Basque either from birth or during early childhood before acquiring Basque or Spanish orthographic representations during reading acquisition. Nevertheless, these presumably stable phonological representations were strongly influenced by the incongruent Basque–Spanish GPCs. The effect of orthography on early bilinguals' speech perception is addressed further in the General discussion.

Experiment II: Discrimination

An AXB speech sound discrimination task was conducted to ensure that participants were able to perceive the phonetic difference between the critical sounds / ζ / and / θ / without lexical context.

Stimuli

The critical and control sounds were presented in disyllabic e_u syllables, which have no lexical meaning in either Basque or Spanish. Each trial consisted of a sequence of three stimuli, in which the first (A) and third (B) stimulus were phonologically different. Participants had to decide whether the second stimulus (X) matched the A or B stimulus. In critical trials, the X-stimulus always corresponded to / $e\theta u$ /, while the A-stimulus either corresponded to / $e\theta u$ / and the B-stimulus to / $e\zeta u$ / (AAB trials) or vice versa (ABB trials). Recall that / $e\theta u$ / and / $e\zeta u$ / have the same orthographic form, as both would be spelled <ezu> in Spanish and Basque, respectively. In control trials, the X-stimulus was always / exu / and had to be discriminated from / $e\zeta u$ /. In both Spanish and Basque, the orthographic form for / exu / would be <exu>. In total, there were 8 critical and 8 control trials. In addition, trials with the same sound contrasts previously used in the LDT (/p/-/b/, /k/-/g/, and /m/-/n/) were included as fillers. None of the tokens were repeated in the experiment. For this reason, the X tokens were never acoustically identical to either the A or B tokens. This procedure elicits a categorical

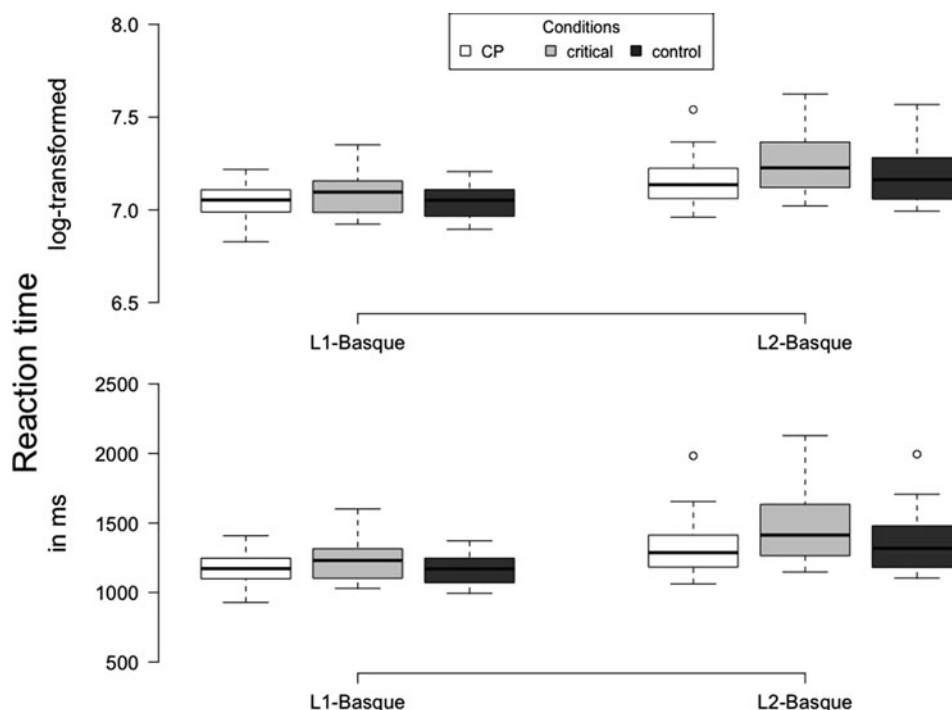


Figure 2. Log-transformed (top) and non-transformed (bottom) RTs in the LDT by group and condition (aggregated over participants).

judgement from the participants instead of relying on acoustic discrimination.

Apparatus and procedure

Each AXB-trial consisted of three tokens that were presented with a 300ms inter-stimulus-interval. Participants had to indicate via keyboard response whether the second item was the same as the first or third. The next trial began 1,000ms after participants provided a response. Participants were instructed to respond as quickly and as accurately as possible with no time limit. The main task was preceded by a practice block of 6 trials with feedback. The entire task lasted less than 5 minutes.

Results

In total, 93.33% of the 960 critical and control trials entered the final analyses. Trials with unreasonably long (>4,000ms) or short (<100ms) RTs were removed (40 trials; 4.17% of the data). Subsequently, 24 trials (2.50% of the data) were identified as outliers and removed from the data (see Analyses section in the General methods). Participants were highly accurate on both critical trials (L1-Basque: 94%; L2-Basque: 96%) and control trials (L1-Basque: 97%; L2-Basque: 99%).

Accuracy

The logistic mixed-effects model had Accuracy (1,0) as the dependent variable with fixed effects for Group (L1-Basque, coded as 1; L2-Basque, coded as -1), X-Sound (critical, coded as 1; control, coded as -1), and Position (ABB, coded as 1; AAB coded as -1), with a three-way interaction term including lower-level interactions. A random intercept for Subjects was included, as were by-subject random slopes for X-Sound and Position. The model did not detect any significant effects, suggesting that participants in both groups were as accurate in discriminating

Table 2. Proportion of accurate discrimination by group and sound (SDs in parentheses).

X-Sound	Group	
	L1-Basque	L2-Basque
/eθu/ (critical)	0.94 (0.24)	0.96 (0.20)
/exu/ (control)	0.97 (0.18)	0.99 (0.09)

critical /eθu/-tokens from /eʂu/-tokens as control /exu/-tokens from /eʂu/-tokens (see Table 2; Table S6, Supplementary Materials shows the complete model output).

Reaction times

RT data was positively skewed (skewness score = 3.538) and therefore log-transformed. This resulted in a 0.403 skewness score, indicating an approximately symmetric distribution of RTs. The linear mixed-effects model had log-transformed RTs (in ms) as the continuous dependent variable; the remaining structure was identical to the logistic mixed-effects model on Accuracy reported above. The model detected a significant main effect of X-Sound ($\beta = 0.103$, $SE = 0.020$, $t = 5.159$, $p < .001$), showing that participants in both groups were slower at discriminating critical /eθu/-tokens from /eʂu/-tokens than control /exu/-tokens from /eʂu/-tokens (see Figure 3). A significant main effect of Position ($\beta = -0.042$, $SE = 0.020$, $t = -2.111$, $p = .039$) shows that participants responded faster to AAB than to ABB trials. No other effects or interactions were significant (see Table S7, Supplementary Materials for complete model output).

Discussion

The results show that both groups were highly accurate in discriminating the critical and control sound contrasts presented

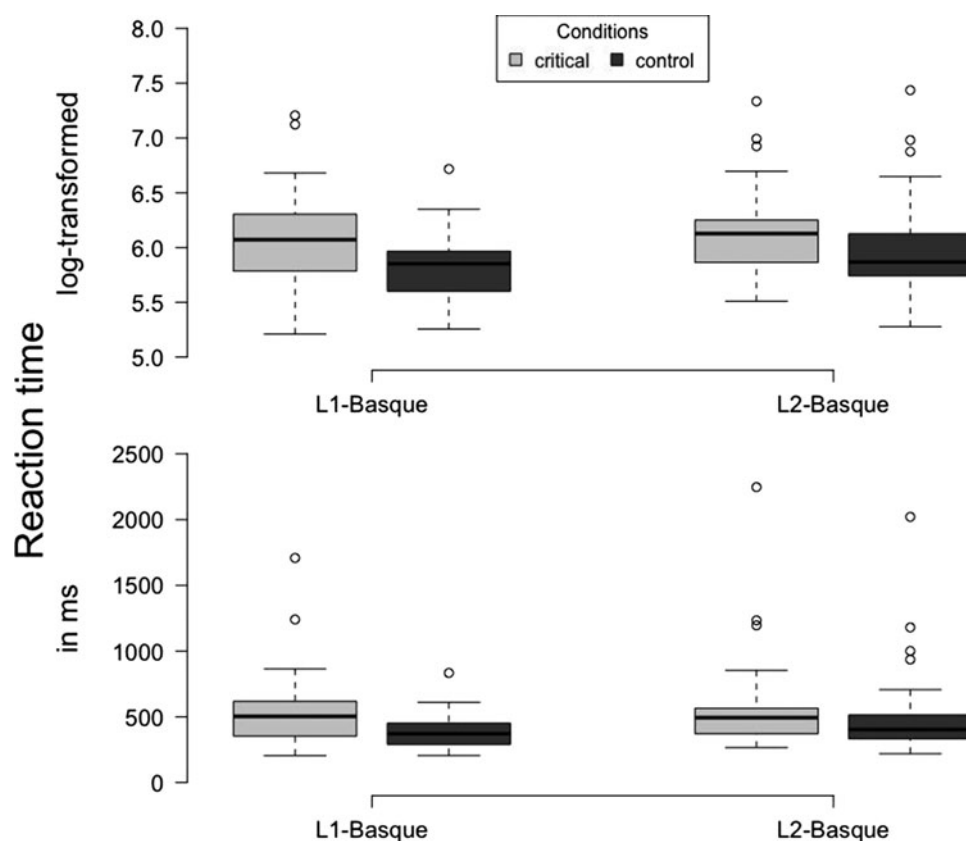


Figure 3. Log-transformed (top) and non-transformed (bottom) RTs in the discrimination task by group and condition (aggregated over participants).

without lexical context. This finding suggests that poorer performance on critical versus control trials in the LDT cannot simply be explained by difficulty perceiving mispronunciations. Interestingly, both groups were slower in discrimination judgements on critical than control AXB trials, possibly due to the orthographic link between the sounds /s/ and /θ/. As critical and control sounds were presented in meaningless syllables, this would imply that orthographic representations may be encoded not only at the lexical but also at the phonological level. This possibility is elaborated in the General discussion.

Experiment III: Speech production

Basque and Spanish reading-aloud tasks were administered with two aims: first, to ascertain whether the orthographic effects observed on speech perception were likewise present in the speech production of early bilinguals who had acquired their two languages prior to reading acquisition; second, to empirically test Larraza et al.'s (2016) claim that L2-Basque speakers do not produce Basque /s/ <z> and /s/ <s> distinctly.

Stimuli

The experiment was divided into a Basque reading-aloud task and a Spanish reading-aloud task. The Basque task consisted of 60 stimuli. Twenty contained <z>, 20 contained <s>, and another 20 served as filler items. The graphemes <z> and <s> always occurred in the stressed syllable-initial position. The <z> and <s> words were selected from the BaSp database (Duñabeitia et al., 2017) and closely matched on the vocalic context of the critical graphemes. Moreover, items were matched on the following variables accessed

through the BaSp database: frequency, orthographic length, phonological length, number of neighbors, age of acquisition, concreteness, orthographic Levenshtein distance, Spanish–Basque cognate rate, number of senses, number of translations, RT of Basque bilinguals, and error rate of Basque bilinguals (see Table S8, Supplementary Materials for an overview of the matched variables). Due to restrictions in the availability of suitable stimuli, 18 <z> words and 14 <s> words had already been used in the LDT.

The Spanish task consisted of 36 stimuli. Twelve of them contained <z>, 12 contained <s>, and an additional 12 served as fillers. Since Spanish /θ/ is only spelled as <z> when followed by either /a/, /o/ or /u/, the Spanish list contained fewer stimuli than the Basque list, which contained 20 instead of 12 items per condition. As in the Basque reading-aloud task, the graphemes <z> and <s> always occurred in the stressed syllable-initial position. The 12 Spanish <z> and <s> words were matched with a subset of 12 Basque <z> and <s> words for the cross-language comparison. The 12 Spanish and 12 Basque words in the reading-aloud tasks were matched on vocalic context, orthographic length, phonological length, and Spanish–Basque cognate rate (variables accessed through the BaSp database; see Table S9, Supplementary Materials for an overview of the matched variables). In addition, the Spanish <z> and <s> words were matched on Spanish–Basque cognate rate, log frequency, orthographic length, number of higher frequency neighbors, orthographic Levenshtein distance, phonological length, number of syllables, position of the accented syllable, and number of higher frequency phonological neighbors (see Table S10, Supplementary Materials for an overview of the matched variables). The values of these variables were accessed through the EsPal database (Duchon, Perea, Sebastián-Gallés, Martí & Carreiras, 2013).

Apparatus and procedure

All participants started with the Basque reading-aloud task, with all oral and written instructions provided in Basque. The experimenter then engaged each participant in a five-minute conversation in Spanish before administering the Spanish reading-aloud task, with all oral and written instructions provided in Spanish. The remaining procedure for both tasks was identical.

Stimuli were orthographically presented on the computer screen. Participants were asked to read each word at the volume they would use if addressing another person. Recordings were made using the integrated microphone of the Sennheiser GSP 350 headset, and digitized at 44,100Hz. Each trial started with a 500ms blank screen, followed by a fixation dot displayed for 500ms, after which the written word appeared on screen. The word remained on screen for three seconds, and the microphone was activated during this period. The next trial then started automatically. The main task was preceded by five practice trials to familiarize participants with the procedure. The Basque reading-aloud task lasted approximately five minutes and the Spanish reading-aloud task took approximately three minutes.

Data processing and acoustic measurements

Recordings were high-pass filtered at 300Hz to minimize interference from voicing and other low-frequency noise at the center of gravity (File-Muriel & Brown, 2010; Maniwa, Jongman & Wade, 2009; Muxika-Loitzate, 2017). Fricatives were segmented manually using Praat software (Boersma & Weenink, 2017). The onset of the fricative was defined as the moment when the high intensity frication noise began, and the offset of the fricative was defined as the point when the frication noise started to decrease (Figure S1, Supplementary Materials).

Results

In total, 97.73% of the 3,840 <z> and <s> trials were included in the final analyses. Seventy-two trials (1.88% of the data) were excluded because of noise during the recording, such as coughing or yawning. Moreover, 15 trials in which the fricative's center of gravity (CoG) was below 1,000Hz were excluded from the analyses (0.39% of the data). Values below this threshold are likely to be faulty and may represent glottal pulses rather than turbulent noise (e.g., Jongman, Wayland & Wong, 2000; Silbert & de Jong, 2008), providing no information on a fricative's place of articulation.

Two separate linear mixed-effects models were conducted. The first compared the CoG for Basque /s̺/ <z> and /s̺/ <s> versus Spanish /θ/ <z> and /s̺/ <s> (henceforth, the Basque–Spanish model), while the second compared the CoG for Basque /s̺/ <z> and /s̺/ <s> (henceforth, the Basque model). Both models had CoG in Hz as the continuous dependent variable.

The Basque–Spanish model included fixed effects for Sound (binary coding using the graphemes <z> coded as 1; <s> coded as -1), Group (L1-Basque coded as 1; L2-Basque coded as -1), Language (Spanish coded as 1; Basque coded as -1), and Gender (male coded as 1; female coded as -1) with a three-way interaction term between Sound, Group and Language including lower-level interactions. Random intercepts for Subjects and Items were included, as were by-subject random slopes for Sound and Language and by-item random slopes for Group and Gender. The model detected a significant main effect of

Gender ($\beta = -137.960$, $SE = 46.707$, $t = -2.954$, $p = .005$), showing that females produced all fricatives with higher CoGs than males. In addition, the model detected significant main effects of Sound ($\beta = -266.471$, $SE = 3.090$, $t = -5.019$, $p < .001$), Language ($\beta = -359.320$, $SE = 48.405$, $t = -7.423$, $p < .001$), and a significant interaction between Sound and Language ($\beta = -314.314$, $SE = 39.847$, $t = -7.888$, $p < .001$). No other significant main effects or interactions were observed (see Table S11, Supplementary Materials for the complete model output). Analyses on the data split by sound using a Bonferroni-adjusted α -level of .025 revealed that productions of Basque /s̺/ <z> and Spanish /θ/ <z> were statistically different ($\beta = -703.050$, $SE = 69.700$, $t = -10.086$, $p < .001$, see Table S12, Supplementary Materials for complete model output), but productions of Basque and Spanish /s̺/ <s> were not (see Table S13 for complete model output and Table S14 for group means, Supplementary Materials). Figure 4 visualizes the CoGs by group and sound in Basque and Spanish.

The Basque model included fixed effects for Sound (/s̺/ <z> coded as 1; /s̺/ <s> coded as -1), Group (L1-Basque coded as 1; L2-Basque coded as -1), and Gender (male coded as 1; female coded as -1) with an interaction term between Sound and Group. Random intercepts for Subjects and Items were included, as were by-subject random slopes for Sound and by-item random slopes for Group and Gender. The model detected a significant main effect of Gender ($\beta = -181.680$, $SE = 52.970$, $t = -3.430$, $p = .001$), showing that females produced all fricatives with higher CoGs than males. No significant effects of Sound, Group, and no significant interaction between Sound and Group were observed, suggesting that participants in both groups produced Basque /s̺/ <s> and /s̺/ <z> with similar CoGs, as visualized in Figure 4 (see Table S15 for complete model output and Table S16 for group means, Supplementary Materials).

Discussion

The speech production results show that neither L2-Basque nor L1-Basque speakers frequently mispronounced the Basque lamino-alveolar fricative /s̺/ as the Spanish interdental fricative /θ/, although both fricatives are represented by the grapheme <z>. Instead, both groups seemed to merge their production categories for the Basque lamino-alveolar fricative /s̺/ and the Basque (and Spanish) apico-alveolar fricative /s̺/, as previously assumed by Larraza et al. (2016). These results suggest that orthography does not necessarily influence speech production in either the L2 or L1 in early bilinguals who have acquired both languages prior to reading acquisition. Instead, both groups of bilinguals merged their production of Basque /s̺/ with the phonologically and perceptually close /s̺/, present in both Basque and Spanish. These results indicate that the Basque lamino-alveolar fricative /s̺/ is not routinely mispronounced as the Spanish interdental fricative /θ/ and suggest that participants' acceptance of this type of mispronunciation in Experiment I is unlikely to be driven by assimilation of the two sounds in production, or frequent exposure to this mispronunciation. Further implications of these findings are put forward in the General discussion.

General discussion

The present study was inspired by two earlier lines of research. One group of studies found that early bilinguals acquiring Spanish and either Catalan or Basque erroneously accepted mispronounced words as correct (Samuel & Larraza, 2015;

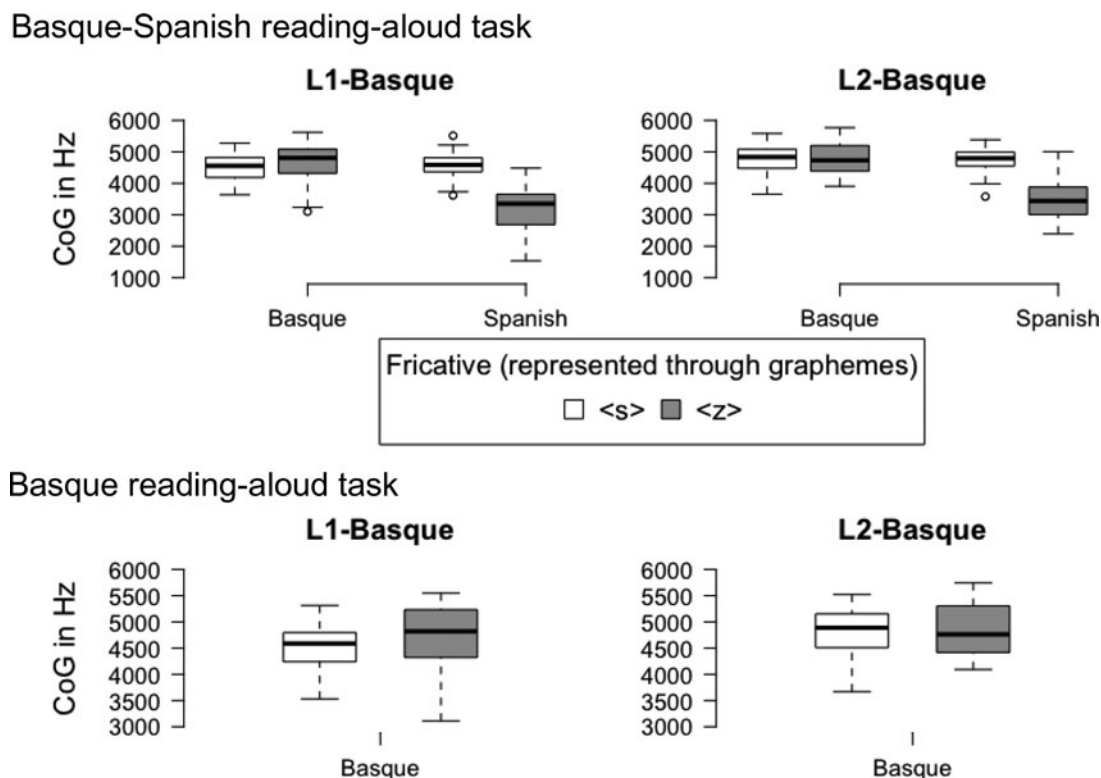


Figure 4. Center of gravity (CoG) in Hz by group, language, and fricative in the Basque-Spanish reading-aloud task (top) and by group and fricative in the Basque reading-aloud task (bottom; aggregated over participants).

Sebastián-Gallés et al., 2005, 2006, 2009). In these studies, error rates were much higher in Spanish-Catalan than Spanish-Basque bilinguals, presumably because the critical sound contrast between / ϵ / and / e / is represented by a single grapheme <e> in Catalan, while the critical sound contrast between / t_s / and / t_f / is represented by the distinct graphemes <tz> and <tx> in Basque. A second group of studies demonstrated that incongruent L1-L2 grapheme-to-phoneme correspondences (GPCs) predominantly affected L2 speech production (Bassetti, 2017; Bassetti & Atkinson, 2015; Bassetti et al., 2018; Bürki et al., 2019; Cerni et al., 2019; Nimz & Khattab, 2020; Rafat, 2016; Stoehr & Martin, 2021; Young-Scholten & Langer, 2015) but also influenced L2 speech perception (Stoehr & Martin, 2021). These studies were conducted with bilinguals who had already formed strong L1 GPCs and only learned their L2 later in life. It remained unclear whether orthography would similarly affect speech production and perception in bilinguals who had acquired both of their languages in early childhood long before reading acquisition.

To investigate this learning scenario, L1-Spanish-L2-Basque (L2-Basque) and L1-Basque-L2-Spanish (L1-Basque) early bilinguals who had acquired both languages before the onset of formal reading instruction were tested on their perception and production of the Basque fricative / s / and the Spanish fricative / θ /, both represented by the grapheme <z>. Importantly, the Basque fricative / s / appears to be mapped onto / s / <s> by L2-Basque learners, suggesting that there is no interference from Spanish / θ / (Larraza et al., 2016). This implies that Basque / s / and Spanish / θ / are not as difficult to distinguish as the notorious Catalan vowels / ϵ / and / e /. Previously reported high acceptance rates for mispronunciations of / ϵ / as / e / in Spanish-Catalan and Catalan-Spanish bilinguals (Sebastián-Gallés et al., 2005, 2006,

2009) might be driven by: (1) the perceptual similarity of these two sounds and the associated presence of this type of mispronunciation in everyday input; (2) the fact that they share the grapheme <e>; (3) a combination of these two factors. The sound contrast employed in the present study does not appear to be particularly difficult for L2 learners of Basque, suggesting that acceptance of mispronunciations of this type are most likely driven by orthographic influence rather than perceptual difficulty. For this reason, the contrast between Basque / s / and Spanish / θ / constitutes an appropriate case for testing whether the impact of orthography on speech perception and production goes beyond perceptual difficulty. The results of an auditory LDT, in which pseudowords were created by replacing the Basque fricative / s / by the Spanish fricative / θ / (critical trials) or by the Spanish and Basque fricative / x / (control trials) showed that L2-Basque and L1-Basque speakers were indeed more likely to accept critical mispronunciations as real words than control mispronunciations. The same participants completed an AXB speech sound discrimination task, in which they had to discriminate critical / θ / or control / x / sounds from / s / in the absence of lexical context. Both groups were highly accurate on critical and control trials alike. This ceiling performance in phonological discrimination demonstrates that the LDT results cannot be explained by a perceptual deficit. Interestingly, both groups responded more slowly to critical than control AXB trials, raising the possibility that orthographic representations may be part of phonological representations, as discussed in more detail below. Finally, the same participants completed reading-aloud tasks in Basque and Spanish to test whether incongruent L1-L2 GPCs affected speech production. Acoustic measures of participants' speech production showed that Basque / s / and Spanish / θ / were produced distinctly,

despite sharing the grapheme <z>. Participants in both groups produced Basque /s/ in line with their Basque/Spanish category for /s/ <s>, which shows that they had neutralized this sound contrast in favor of the articulatorily easier member present in both of their languages. Although these findings cannot rule out the possibility that bilinguals with L2-Basque and L1-Basque are occasionally exposed to mispronunciations of /s/ as [θ], they do indicate that this type of mispronunciation is not common.

Participants were able to discriminate the sound contrast at ceiling and produced /s/ and /θ/ distinctly. Thus, the high acceptance rates found for orthographic mispronunciations in the LDT demonstrate that speech perception is affected by incongruent cross-linguistic GPCs – even in early bilinguals who acquired both their languages before learning to read. This is especially remarkable considering that no orthographic information was provided in the task. It is likely that a similar task additionally featuring visually available orthographic forms would lead to even higher acceptance rates of orthographically-guided mispronunciations because of the direct influence of orthography through the visual modality. The observed accuracy rates of 54% for L2-Basque speakers and 64% for L1-Basque speakers fall between the relatively low accuracy rates of approximately 25% for Spanish-Catalan bilinguals and approximately 60% for Catalan-Spanish bilinguals observed in Sebastián-Gallés et al. (2005, 2006) and the overall higher accuracy rate of 67% for Spanish-Basque bilinguals observed in Samuel and Larraza (2015). The Catalan-Spanish participants in Sebastián-Gallés et al. (2005, 2006) were confronted with the perceptually difficult sound contrast between the two adjacent Catalan vowels /ɛ/ and /e/. This sound contrast may be particularly difficult for L1-Spanish speakers, since [ɛ] is an allophonic variant of the Spanish vowel /e/. On top of this perceptual difficulty, the Catalan vowels /ɛ/ and /e/ and the Spanish vowel /e/ share the grapheme <e>. The participants in the present study performed with higher accuracy on a sound contrast that, despite sharing a grapheme, was easier to perceive – as confirmed by the ceiling performance on AXB discrimination. The accuracy of L2-Basque speakers in our LDT was overall lower than the accuracy rate Samuel and Larraza (2015) observed for the Basque affricates /tʃ/ and /tʃ/. However, these affricates are not only perceptually easy to discriminate but also orthographically represented by distinct graphemes. Considered together, the results of the present study, Sebastián-Gallés et al. (2005, 2006), and Samuel and Larraza (2015) provide important insight into the additive effects of phonological and orthographic factors on speech perception: mispronunciations are accepted if the two critical sounds are allophones in the other language or frequently occur as mispronounced variants in the input, or if the two critical sounds share the same grapheme in the bilinguals' two languages. Acceptance of mispronunciations is further enhanced when these scenarios are found in combination.

Overall, the high acceptance rate of orthographic mispronunciations found in the present study is remarkable, especially considering that mispronunciations of /s/ as [θ] do not seem to occur frequently in real life. This is in stark contrast to previous studies conducted by Sebastián-Gallés et al. (2005, 2006) and Samuel and Larraza (2015), who found high acceptance rates for mispronunciations commonly produced by L2 speakers of Catalan and Basque, respectively. As these L2 populations represent a substantial part of local society in Catalonia and the Basque Country, it can be assumed that the mispronunciations investigated by Sebastián-Gallés et al. (2005, 2006) and Samuel and Larraza

(2015) are frequently encountered. The present results show acceptance of uncommon mispronunciations, using a non-similar critical sound contrast, linked only visually by a shared grapheme. This finding highlights the robustness of orthographic effects on both L2 and L1 speech perception. The speech perception of early bilinguals can be affected by GPC incongruencies such that an uncommon mispronunciation involving auditorily distinct sounds becomes acceptable.

Discrimination by both groups of participants was close to ceiling for the critical and control sound contrasts in the non-lexical context. An incidental finding showed that both groups were slower to discriminate critical /θ/ from /s/ than control /x/ from /s/. This indicates that although the critical contrast was easy to discriminate, it was nevertheless cognitively costlier than the control sound contrast. A possible explanation is that orthographic information is not only encoded on the lexical level as previously assumed (e.g., Dijkstra, Roelofs & Fieus, 1995; Rastle, McCormick, Bayliss & Davis, 2011; Ziegler & Ferrand, 1998) but also on the phonological level as GPCs are learned. This might result in improved access to unequivocal phonological representations. Indeed, phonological awareness tasks show an advantage in phoneme deletion and reversal tasks for phonemes with unequivocal compared to phonemes with multiple spellings (Castles, Holmes, Neath & Kinoshita, 2003). It is therefore possible that discriminating /θ/ from /s/ is cognitively costlier because their phonological representations have been retuned and linked through the shared grapheme <z> during literacy acquisition. Alternatively, orthographic representations may be co-activated whenever lexical information is perceived or rapidly created for unknown or non-existent words. In these scenarios, orthographic representations would not be part of phonological representations but rather closely associated and systematically co-activated during speech sound perception. As the present study was not designed to test these two competing hypotheses, it does not allow us to determine which alternative is most likely. Importantly, the outcome is the same whether orthographic representations form part of phonological representations or are co-activated (or rapidly created) when speech is processed. Listening to the meaningless syllables /eθu/ and /eʃu/ may have led to the activation of the orthographic representation <z> directly through the phonological representations /θ/ and /s/ or to the creation of the orthographic representation <ezu>. In both cases, the matching orthographic form of these two meaningless syllables may have slowed down participants' decision-making compared to /exu/ - /eʃu/ syllables, which would be represented by the distinct orthographic forms <ēju> and <ezu>.

Despite the strong impact of orthography on speech perception, no effect of orthography was observed in early bilinguals' speech production. Participants in both groups produced Basque /s/ and Spanish /θ/ distinctly, although both phonemes are represented by the grapheme <z>. This was true even though participants' speech was elicited in reading-aloud tasks, where they were directly confronted with the word's spelling. In sum, it appears that orthographic effects are sufficiently robust to influence early bilinguals' speech perception, but do not automatically lead to less native-like speech production. This finding is particularly striking since no orthographic information was present in the auditory LDT, while orthographic information was used to elicit participants' speech production. There are two possible explanations for this result. The first is based on the finding that perceptual categories are highly malleable and, for example, can be recalibrated after only a few minutes of exposure to

ambiguous or foreign-accented speech (for a review on perceptual learning, see Samuel & Kraljic, 2009). While production categories are also prone to phonetic drift, they may be less malleable than perceptual categories (for a review on phonetic drift in bilingual speech production, see Kartushina, Frauenfelder & Golestani, 2016). The second explanation is related to the finding that /s/ appeared to be difficult for both L2-Basque and L1-Basque speakers: neither group produced Basque /s/ <z> and Basque /s/ <s> distinctly. This finding complements the previous observation that Spanish–Basque bilinguals have difficulty perceptually distinguishing these two sounds (Larraza et al., 2016). It is possible that the assimilation of Basque /s/ <z> and Basque /s/ <s> is stronger than any potential orthographic influence that might establish a link between Basque /s/ <z> and Spanish /θ/ <z>. If this is the case, orthographic influence in production would be expected in a constellation in which two sounds that are not yet assimilated to any other sounds are linked through a shared GPC.

Taken together, the present speech production results have three important implications. First, speech production in early bilinguals who acquired their languages before becoming literate does not seem to be affected by orthography, as consistently reported for L2 speakers who learned their L2 when they were already literate (Bassetti, 2017; Bassetti & Atkinson, 2015; Bassetti et al., 2018; Bürki et al., 2019; Cerni et al., 2019; Nimz & Khattab, 2020; Rafat, 2016; Stoehr & Martin, 2021; Young-Scholten & Langer, 2015). Second, early bilinguals may resort to the phonologically closest alternative to compensate for production difficulty, while orthographic influence appears to play a minor role. Third, the results of the LDT are not likely to be due to frequent input of incorrect forms in daily life since mispronunciations rooted in orthographic similarity were not detected in participants' speech production.

In conclusion, the combined results of the present study suggest that the speech of L2-Basque and L1-Basque early bilinguals who acquired Basque and Spanish prior to becoming literate is affected by orthography. Yet, this orthographic influence seems to differ from the orthographic influence reported for bilinguals who learned their L2 after L1 reading acquisition. In the present study, while participants' speech perception was affected by orthography, the robust effect of orthography on speech production – previously reported for L2 speakers – was not observed. This may be due to greater malleability in perceptual compared to production categories in both the L1 and an early acquired L2. Interestingly, L2-Basque and L1-Basque speakers behaved similarly showing effects of orthography on perception but not production. To our knowledge, the present study is the first to show that incongruent L1-L2 GPCs can also affect L1 speech perception. These findings enrich previous research reporting strong orthographic effects on the speech production of late bilinguals who learn an L2 through simultaneous auditory and visual input after strong L1 GPCs have been established. The current study demonstrates the strength of orthographic impact on speech perception through two major findings. First, orthographic impact extends even to scenarios where L1 and L2 phonological representations have been acquired long before orthographic representations are learned. Second, incongruent cross-linguistic GPCs are strong enough to affect L1 perception.

Acknowledgments. We would like to thank Amets Esnal and Amaia Cano for their help in selecting the stimulus materials; Oihane Iturbe for recording the stimulus materials; Mina Jevtović for thoroughly testing the experimental scripts; Amets Esnal and Itziar Basterra for administering the present study;

and Magda Altman for proofreading this article. We are grateful to the members of the Speech and Bilingualism group and the Spoken Language group at the Basque Center on Cognition, Brain and Language for valuable discussions and feedback on this study.

This work was supported by the Basque Government [BERC 2018–2021 program]; the Spanish State Research Agency [BCBL Severo Ochoa excellence accreditation SEV-2015-0490]; the H2020 European Research Council [Marie Skłodowska-Curie grant 843533; ERC Consolidator Grant ERC-2018-COG-819093]; the Spanish Ministry of Economy and Competitiveness [PSI2017-82941-P; Europa-Excelencia ERC2018-092833; RED2018-102615-T]; and the Basque Government [PIBA18-29].

Supplementary Material. For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S1366728921000523>

The supplementary material comprises one file containing:
 Appendix S1: Instructions of the LDT (Experiment I)
 Tables S1-S2: Variable matching for the LDT (Experiment I)
 Tables S3-S5: Model output of the LDT (Experiment I)
 Tables S6-S7: Model output of the Discrimination task (Experiment II)
 Figure S1: Waveform, spectrogram and measurement interval of a Basque fricative (Experiment III)
 Tables S8-S10: Variable matching for the Reading tasks (Experiment III)
 Tables S11-S16: Model output and averaged results of the Reading tasks (Experiment III)

Competing interests. The authors declare none.

References

- Baayen RH and Milin P (2010) Analyzing reaction times. *International Journal of Psychological Research* 3, 12–28. doi:10.21500/20112084.807
- Bassetti B (2017) Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43, 1835–1842. doi:10.1037/xlm0000417
- Bassetti B and Atkinson N (2015) Effects of orthographic forms on pronunciation in experienced instructed second language learners. *Applied Psycholinguistics* 36, 67–91. doi:10.1017/S0142716414000435
- Bassetti B, Sokolović-Perović M, Mairano P and Cerni T (2018) Orthography-induced length contrasts in the second language phonological systems of L2 speakers of English: Evidence from minimal pairs. *Language and Speech* 61, 577–597. doi:10.1177/0023830918780141
- Bates D, Mächler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48. doi:10.18637/jss.v067.i01
- Boersma P and Weenink D (2017) *Praat: Doing phonetics by computer*. Version 6.0.24. Retrieved from <http://www.praat.org/>.
- Bürki A, Welby P, Clément M and Spinelli E (2019) Orthography and second language word learning: Moving beyond “friend or foe”? *Journal of the Acoustical Society of America* 145(4), EL265–EL271. doi:10.1121/1.5094923
- Castles A, Holmes VM, Neath J and Kinoshita S (2003) How does orthographic knowledge influence performance on phonological awareness tasks? *The Quarterly Journal of Experimental Psychology* 56A, 445–467. doi:10.1080/02724980244000486
- Cerni T, Bassetti B and Masterson J (2019) Effects of orthographic forms on the acquisition of novel spoken words in a second language. *Frontiers in Communication* 4, 31. doi:10.3389/fcomm.2019.00031
- de Bruin A, Carreiras M and Duñabeitia JA (2017) The BEST dataset of language proficiency. *Frontiers in Psychology* 8, 522. doi:10.3389/fpsyg.2017.00522
- Dean CA and Valdés Kroff JR (2017) Cross-linguistic orthographic effects in late Spanish/English bilinguals. *Languages* 2, 24. doi:10.3390/languages2040024
- Dijkstra T, Roelofs A and Fieuws S (1995) Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology* 49, 264–271. doi:10.1037/1196-1961.49.2.264

- Duchon A, Perea M, Sebastián-Gallés N, Martí A and Carreiras M** (2013) EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods* **45**, 1246–1258. doi:10.3758/s13428-013-0326-1
- Duñabeitia JA, Casaponsa A, Dimitropoulou M, Martí A, Larraza S and Carreiras M** (2017) BaSp: A Basque–Spanish database of translation equivalents. Ms., Basque Center on Cognition, Brain and Language.
- Erdener VD and Burnham DK** (2005) The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning* **55**, 191–228. doi:10.1111/j.0023-8333.2005.00303.x
- Escudero P, Hayes-Harb R and Mitterer H** (2008) Novel second-language words and asymmetric lexical access. *Journal of Phonetics* **36**, 345–360. doi:10.1016/j.wocn.2007.11.002
- File-Muriel RJ and Brown EK** (2010) The gradient nature of s-lenition in Caleño Spanish. *University of Pennsylvania Working Papers in Linguistics* **16**, 44–55. doi:10.1017/S0954394511000056
- Gordon M, Barthmaier P and Sands K** (2002) A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* **32**, 141–174. doi:10.1017/S0025100302001020
- Hualde JI** (2010) Neutralización de sibilantes vascas y seseo en castellano. *Oihenart* **25**, 89–116.
- Jongman A, Wayland R and Wong S** (2000) Acoustics of English fricatives. *Journal of the Acoustical Society of America* **108**, 1252–1263. doi:10.1121/1.1288413
- Kartushina N, Frauenfelder UH and Golestani N** (2016) How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning* **66**, 155–186. doi:10.1111/lang.12187
- Kaufman AS and Kaufman NL** (2004) *Kaufman brief intelligence test KBIT-2*. Bloomington, MN, US: Pearson.
- Kuznetsova A, Brockhoff PB and Christensen RHB** (2017) lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software* **82**, 1–26. doi:10.18637/jss.v082.i13
- Larraza S, Samuel AG and Oñederra ML** (2016) Listening to accented speech in a second language: First language and age of acquisition effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **42**, 1774–1797. doi:10.1037/xlm0000252
- Lemhöfer K and Broersma M** (2012) Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* **44**, 325–343. doi:10.3758/s13428-011-0146-0
- Maniwa K, Jongman A and Wade T** (2009) Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America* **125**, 3926–3973. doi:10.1121/1.2990715
- Mathôt S, Schreij D and Theeuwes J** (2012) OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* **44**, 314–324. doi:10.3758/s13428-011-0168-7
- Muxika-Loitzate O** (2017) Sibilant merger in the variety of Basque spoken in Amorebieta-Etxano. *Languages* **2**, 25. doi:10.3390/languages2040025
- Nimz K and Khattab G** (2020) On the role of orthography in L2 vowel production: The case of Polish learners of German. *Second Language Research* **36**, 623–652. doi:10.1177/0267658319828424
- Pallier C, Bosch L and Sebastián-Gallés N** (1997) A limit on behavioral plasticity in speech perception. *Cognition* **64**, B9–B17. doi:10.1016/S0010-0277(97)00030-9
- R Core Team** (2013) *R: A language and environment for statistical computing*. Retrieved from <http://www.R-project.org/>.
- Rafat Y** (2016) Orthography-induced transfer in the production of English-speaking learners of Spanish. *The Language Learning Journal* **44**, 197–213. doi:10.1080/09571736.2013.784346
- 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**, 1588–1594. doi:10.1037/a0024833
- Samuel AG and Kraljic T** (2009) Perceptual learning for speech. *Attention, Perception and Psychophysics* **71**, 1207–1218. doi:10.3758/APP.71.6.1207
- Samuel AG and Larraza S** (2015) Does listening to non-native speech impair speech perception? *Journal of Memory and Language* **81**, 51–71. doi:10.1016/j.jml.2015.01.003
- Sebastián-Gallés N, Echeverría S and Bosch L** (2005) The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language* **52**, 240–255. doi:10.1016/j.jml.2004.11.001
- Sebastián-Gallés N, Rodríguez-Fornells A, de Diego-Balaguer R and Diaz B** (2006) First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience* **18**, 1277–1291. doi:10.1162/jocn.2006.18.8.1277
- Sebastián-Gallés N, Vera-Constán F, Larsson JP, Costa A and Deco G** (2009) Lexical plasticity in early bilinguals does not alter phoneme categories: II. Experimental evidence. *Journal of Cognitive Neuroscience* **21**, 2343–2357. doi:10.1162/jocn.2008.21152
- Seidenberg MS and Tanenhaus MK** (1979) Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory* **5**, 546–554. doi:10.1037/0278-7393.5.6.546
- Silbert N and de Jong K** (2008) Focus, prosodic context, and phonological feature specification: Patterns of variation in fricative production. *Journal of the Acoustical Society of America* **123**, 2769–2779. doi:10.1121/1.2890736
- Simon E, Chambless D and Alves UK** (2010) Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences* **32**, 380–394. doi:10.1016/j.langsci.2009.07.001
- Stoehr A and Martin CD** (2021) The impact of orthographic input on speech production and perception: An artificial vowel learning study. Ms., Basque Center on Cognition, Brain and Language.
- Tremblay A and Ransijn J** (2020) Package ‘LMERConvenienceFunctions’. *Model selection and post-hoc analysis for (G)LMER models. Version 3.0*. Retrieved from <https://CRAN.R-project.org/package=LMERConvenienceFunctions>.
- Young-Scholten M and Langer M** (2015) The role of orthographic input in second language German: Evidence from naturalistic adult learners’ production. *Applied Psycholinguistics* **36**, 93–114. doi:10.1017/S0142716414000447
- Ziegler JC and Ferrand L** (1998) Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review* **5**, 683–689. doi:10.3758/BF03208845