Orthographic forms affect speech perception in a second language

Bassetti, Bene; Masterson, Jackie; Cerni, Tania; Mairano, Paolo

DOI:
10.1037/xhp0000949

Orthographic forms affect speech perception in a second language: Consonant and vowel length in L2 English

Bene Bassetti[a], Jackie Masterson[b], Tania Cerni[c], & Paolo Mairano[d]
[a]University of Birmingham, UK [b] UCL Institute of Education, UK
[c]University of Trento, Italy [d]University of Lille, France

Bene Bassetti, School of Education, University of Birmingham, Birmingham B15 2TT, United Kingdom. E-mail: b.bassetti@bham.ac.uk
Jackie Masterson, Department of Psychology and Human Development Institute of Education, 20 Bedford Way, London WC1H 0AL, United Kingdom. E-mail: j.masterson@ucl.ac.uk
Tania Cerni, Department of Psychology and Cognitive Science, University of Trento, Corso Bettini 84, 38068 Rovereto, Italy. E-mail: tania.cerni@unitn.it
Paolo Mairano, University of Lille, Domaine Universitaire de Pont de Bois, 3 Rue du Barreau, 59650 Villeneuve-d'Ascq, France. E-mail: paolo.mairano@univ-lille.fr

Author Note

Correspondence concerning this article should be addressed to Bene Bassetti, School of Education, University of Birmingham, Birmingham B15 2TT, UK.
Tel: (+44) (0)121 415 58221
E-mail: b.bassetti@bham.ac.uk

Acknowledgments

The authors are grateful to Alex Panicacci for contributing to data collection, to Becky Taylor for contributing to materials preparation, to Sheila Verrier for recording spoken materials, to Juhani Järvikivi for lending us a response box, to Paul Wakeling for evaluating our spoken materials, and to the editor and anonymous reviewers for insightful comments on a previous version of this article. This work was supported by a Leverhulme Trust Research Grant [grant number: RPG 2013 180] awarded to Bassetti and Masterson. The sponsor had no involvement in the conduct of the research or the preparation of the article. Preliminary findings were presented at the Sound to Word in Bilingual and Second Language Speech Perception conference, Iowa City, USA, in April 2016.

Abstract

Italian$_{L1}$ speakers of English$_{L2}$ produce the same English sound as longer if spelled with two than with one letter, following Italian grapheme-phoneme conversion rules. Do Italian listeners perceive short and long sounds in English homophonic word pairs that are spelled with a single letter or a digraph (*finish-Finnish*; *morning-mourning*)? In Experiment 1, 50 Italian$_{L1}$-English$_{L2}$ bilinguals and 50 English controls performed a Consonant Perception Task and a Vowel Perception Task. They heard English homophonic word pairs containing a target sound spelled with one or two letters and indicated whether the two words contained the same sounds or not. For half of the listeners a picture was used to activate target words (Auditory-Visual Input group). Bilinguals in this group perceived different sounds in homophonic pairs. Experiment 2 tested whether naturalistic exposure reduces orthographic effects on speech perception by comparing learners, sequential bilinguals and English controls (all *n* = 30) with Auditory-Visual Input. Orthographic form ('spelling') affected consonant perception in both of the second language listener groups. Learners were less affected than bilinguals. Analyses indicated that this was because of the learners' high proficiency. It appears that Italian$_{L1}$ speakers of English$_{L2}$ make a long-short contrast for consonants — unattested in English — and illusorily perceive it in spoken English homophonous words.

Keywords
Orthography; phonology; speech perception; second language; orthographic effects

Public significance

- The findings strongly suggest that spelling affects speech perception in a second language.
- Italian native speakers listened to pairs of English homophonic words (having the same sound) where a sound was spelled with one letter in one word and two letters in the other. The number of letters (one or two) in English spelling leads second language listeners to perceive an illusory contrast between short and long sounds in spoken English that does not exist in the English language.
- The findings highlight the importance of orthographic input in establishing mental representations of the phonological form (sound) of second language words.
- Findings also indicate that large amounts of naturalistic exposure do not reduce the effects of orthographic forms on second language speakers' speech perception.

Introduction

A rapidly growing body of research has established that the orthographic forms (spellings) of a second language (L2) can affect speech perception, speech production and sound categorisation in L2 speakers (e.g., Bassetti et al., 2015), such that L2 speakers can even produce a sound contrasts that is not attested in the target language (Bassetti, 2017; Bassetti et al., 2018, 2020). This paper addresses whether L2 orthographic forms can lead to L2 listeners perceiving a sound contrast that does not exist in the L2 spoken input.

Background

Effects of orthographic forms on non-native speech perception

Research on perception of novel languages (L0) and second languages (L2) shows that the orthographic forms ('spellings') of words and sounds can affect speech perception, both by facilitating it, and by interfering with it if there is incongruence between the spoken and orthographic input, or between native (L1) and L0/L2 orthography-phonology correspondences, as follows. Providing alphabetic orthographic input together with auditory input facilitates speech perception, both in naïve listeners first exposed to the sounds of a novel language (Erdener & Burnham, 2005) and in L2 listeners learning pseudowords

(Escudero et al., 2008), unless the target contrast is very difficult for the listener (Simon et al., 2010).

On the other hand, orthographic input can also negatively influence L0 and L2 perception. Looking at novel languages first, listeners perceive an extra sound in pseudowords in an artificial language that are spelled with a silent letter (Hayes-Harb et al., 2010). This orthographic effect is often a consequence of recoding the orthographic forms of a novel language according to the grapheme-phoneme correspondences of the L1, for instance L0 sound recognition is hindered if the L1 and the L0 have incongruent grapheme-phoneme conversion rules (Erdener & Burnham, 2005; Escudero & Wanrooij, 2010, Experiment 2; Mathieu, 2016; Rafat & Stevenson, 2019). While such effects on novel language perception are useful to explanation and theory-building, it is not at all clear that novel language perception can shed light on the perception of real words in L2 listeners.

Looking at second languages, word spelling can affect L2 listeners' perception resulting in the addition, omission and substitution of sounds in L2 speech perception, both of L2 pseudowords and real words. If a sound is represented in the orthographic form of an L2 word but is not present in its phonological form, L2 listeners may perceive a sound that is not present in the spoken input, for instance a syllable segmentation task showed that orthographic input leads Japanese$_{L1}$ listeners of French$_{L2}$ (and French$_{L0}$) to perceive an extra vowel and therefore segment one more syllable in consonant clusters inside spoken French pseudowords, as opposed to participants only exposed to auditory input (Detey & Nespoulous, 2008). If a sound is present in the phonological form of an L2 word but is not represented in its orthographic form, L2 listeners may fail to perceive a sound, for instance in a phoneme counting task Italian$_{L1}$ learners of Chinese$_{L2}$ counted one less vowel in the same triphthong if its romanisation spelling contains two letters than if it is spelled with three letters (Bassetti, 2006; Pytlyk, 2017). Crucially, if the L1 and L2 grapheme-phoneme correspondences are incongruent, and L2 listeners recode the L2 orthographic form according to L1 grapheme-phoneme correspondences, they may end up hearing a sound other than what is in the auditory input (Escudero et al., 2014; Hayes-Harb et al., 2018; Shea, 2017; Showalter, 2020). Orthographic effects were also found in loanword adaptation, for instance French$_{L1}$-English$_{L2}$ sequential bilinguals who hear English pseudowords adapt the same vowels in French loanwords differently if they only hear the pseudoword and if they also receive orthographic input, as the latter but not the former group is influenced by French grapheme-phoneme conversion rules (Vendelin & Peperkamp, 2006).

Findings of orthographic effects on L2 speech production can also indirectly support the evidence of studies on L2 speech perception. Although the relationship between perception and production in the L2 is debated (Colantoni et al., 2015), it is often believed that the two are linked, both in theory (for instance in Flege's Speech Learning Model (Flege, 1995), PAM-L2 (Best & Tyler, 2007)), and in empirical studies (Baker & Trofimovich, 2006; Flege et al., 1995; Flege et al., 1997). Hence, richer evidence from production studies can be used to corroborate relatively more limited evidence from perception studies. Such studies consistently show that L2 orthographic forms result in sound additions, omissions and substitutions in L2 speech (for a review, see Bassetti et al., 2015). Reflecting findings from the phoneme counting study reported above (Bassetti, 2006), Italian$_{L1}$ learners of Chinese$_{L2}$ also produce one less vowel in the same triphthong if its romanisation spelling contains two letters than if it is spelled with three letters (Bassetti, 2007). L2 speakers also produce epenthetic (i.e., additional) sounds corresponding to 'silent letters' (Bassetti & Atkinson, 2015), and there is much evidence of orthography-induced sound substitutions, at all levels of

L2 proficiency (Escudero et al., 2008; Piske et al., 2002; Vokic, 2011; Young-Scholten, 2002; Young-Scholten & Langer, 2015; Zampini, 1994).

There is then growing evidence that orthographic forms affect L2 speech perception, with evidence also coming from studies of L2 speech production. In contrast, to the best of our knowledge, no study has investigated whether orthography may affect the perception of sound duration in real words in L2 users. This is surprising, given that research on L2 speech production and metalinguistic awareness has repeatedly found that digraphs result in L2 speakers establishing a length contrast in English consonants and vowels. The following section presents these findings.

The perception of consonant and vowel duration in English and Italian

In the Italian language, consonant length is contrastive, meaning that geminate consonants are longer than singleton consonants (Bertinetto, 1981; Esposito & Di Benedetto, 1999). In Italian, a short or long consonant distinguishes otherwise identical word pairs, creating minimal pairs such as *note* (/ˈnɔte/, 'notes') and *notte* (/ˈnɔtːe/, 'night', Clark & Yallop, 1995). This contrast is not peculiar to Italian, but is also found in other languages such as Arabic (Khattab & Al-Tamimi, 2014), Bengali (Lahiri & Hankamer, 1988), Cypriot Greek (Arvaniti & Tserdanelis, 2000), Japanese (Idemaru & Guion-Anderson, 2010), and Turkish (Lahiri & Hankamer, 1988).

The main acoustic cue of gemination in Italian is consonant duration (Bertinetto, 1981; Esposito & Di Benedetto, 1999; Pickett et al., 1999; Rochet & Rochet, 1995; Tagliapietra & McQueen, 2010). In English, consonant length is not contrastive, although English has so-called 'fake geminates', that is to say long consonants that occur at morpheme boundaries to differentiate pairs such as *topic-top pick* (Oh & Redford, 2012).

Looking at orthography, consonant length in Italian is represented by the use of a single consonant letter for a short consonant and double consonant letters for a long consonant. In English, double consonant letters do not represent consonant length (Carney, 1994), for instance there is no difference in the length of the /n/ in *finish* and *Finnish*. Italian$_{L1}$-English$_{L2}$ speakers are affected by the presence of a double consonant letter in an L2 English word, and produce it as a long consonant (Bassetti, 2017; Bassetti et al., 2018) and categorise it as a long consonant in a metalinguistic awareness task (Bassetti et al., 2020; Cerni et al., 2019). However, it is unknown whether Italians also perceive a short-long contrast in English consonants.

The perception of L2 consonant length contrasts is difficult for L2 listeners, particularly but not exclusively if their L1 does not have such contrast (De Clercq et al., 2014; Hayes-Harb, 2005; Sorianello, 2014; Tsukada et al., 2018). For instance, looking at learners of Japanese$_{L2}$, English$_{L1}$ learners apply a fixed length criterion and do not adapt to changes in speech rate (Sonu et al., 2013), and German$_{L1}$ learners but not Japanese native speakers display decreased discrimination performance if acoustic complexity is added (a task-irrelevant pitch fall simultaneous with the length contrast; Asano, 2018), although, Japanese$_{L2}$ listeners' perception of Japanese consonant length contrasts can improve with training (Sadakata & McQueen, 2013) and higher proficiency (Hardison & Motohashi Saigo, 2010; Hayes, 2001). It remains an open question whether incongruent orthographic representations of length may interfere with the perception of L2 consonant length.

Looking at vowels, length is not contrastive in either English or Italian. However, English tense versus lax vowels (e.g., *tin* /tɪn/ and *teen* /tiːn/) are distinguished by both durational and spectral cues (Klatt 1976; Strange 1989). Italian stressed vowels can be phonetically longer in word-internal open syllables, but this phenomenon is conditioned prosodically (Bertinetto & Loporcaro, 2005) and is therefore not a phonological contrast. While English native speakers rely more heavily on spectral cues than duration cues for the discrimination of tense-lax vowels (Hillebrand, Clark & Houde, 2000; (Kondaurova & Francis, 2008), English$_{L2}$ listeners generally rely more on durational cues (Cebrian, 2006; Escudero & Boersma, 2004; Flege et al., 1997; Wang & Munro, 1999). L2 learners' reliance on duration is found whether their L1 uses duration as a phonological feature or not, for instance Japanese$_{L2}$ vowel perception did not differ among Italian$_{L1}$, American English$_{L1}$ and Thai$_{L1}$ learners, despite the different role of duration in their native languages (Tsukada et al., 2014). The desensitization hypothesis (Bohn, 1995) argues that listeners can detect temporal differences more easily than spectral differences in unfamiliar L2 vowels. Looking at orthography, Italian$_{L1}$-English$_{L2}$ speakers are affected by the presence of a vowel digraph in an L2 English word, and produce it as a long vowel (Bassetti et al., 2018) and categorise it as a long vowel in a metalinguistic awareness task (Bassetti et al., 2020). However, it is unknown whether Italians' perception of English vowel length is also affected by the vowel's spelling.

In conclusion, previous research investigated the perception of existing L2 contrasts and found that L2 listeners have difficulty perceiving a short-long contrast in a second language, but no study has investigated whether L2 listeners may illusorily perceive a short-long contrast that is unattested in their second language.

The Present Study

This study aimed to investigate whether second language (L2) orthographic forms (word spellings) affect English$_{L2}$ listeners' speech perception. We used a task where participants were asked to indicate whether spoken word pairs contained the same sounds or different sounds. The aim was to see whether orthographic forms lead Italian$_{L1}$ listeners of English$_{L2}$ to perceive the same English sound as two different sounds (a short and a long one) in a homophonic word pair if the sound is spelled with a single letter in one word and a digraph in the other. Experiment 2 investigated whether such orthographic effects on perception are modulated by naturalistic exposure, by comparing instructed learners and sequential bilinguals with lengthy naturalistic exposure.

With regards to consonant perception, the study tested whether Italian$_{L1}$ speakers of English$_{L2}$ perceive the same consonant as singleton ('short consonant') or geminate ('long consonant') in homophonic English words that are spelled with single or double consonant letters, for instance perceiving the same [n] as a singleton /n/ in *finish* and as a geminate /nː/ in *Finnish*. In the Italian language, consonant length is contrastive and represented orthographically by the use of double letters for long (geminate) consonants, whereas English has no contrastive length. If an English consonant is spelled with double letters, Italians produce it (Bassetti, 2017; Bassetti et al., 2018) and categorise it (Bassetti et al., 2020;Cerni et al., 2019) as a geminate. Therefore, they may perceive different consonants in homophonic English word pairs such as *finish-Finnish*.

With regards to vowel perception, the study tested whether Italian$_{L1}$ speakers of English$_{L2}$ perceive the same vowel as long or short in English homophonic words where it is spelled with a single vowel letter or a vowel digraph (double letters or a combination of two different letters), for instance perceiving the vowel [iː] as short in *scene* and long in *seen* (both /siːn/). Vowel length is not contrastive in either Italian or English, but Italians often use length to distinguish English tense and lax vowels in speech production. If an English vowel is spelled with a digraph, Italians produce it (Bassetti & Atkinson, 2015; Bassetti et al., 2020;

Bassetti et al., 2018) and categorise it (Bassetti et al., 2020) as a long vowel (e.g., *choose* vs. *lose*). Therefore, they may perceive different vowels in homophonic English word pairs such as *scene-seen*. The predicted outcome is the same as with consonants, but the reason is different. With consonants, Italians would perceive an illusory difference corresponding to two phonological categories in their native language (geminates and singletons), whereas with vowels they would perceive an illusory difference corresponding to two values (long and short) of a feature (length) of their second language.

These hypotheses were tested by means of a consonant perception task and a vowel perception task. In both tasks, participants heard homophonic word pairs where the same sound is spelled with one letter in one word and two letters in the other, for instance the C-CC pair *finish-Finnish* and the V-VV pair *soles-souls*. Both accuracy (number of correct responses) and response times (RTs) were measured. If the two homophones are perceived as containing the same sounds, this will mean that orthographic form does not affect perception. In contrast, we predicted that Italians would perceive C-CC pairs as containing a short and a long consonant, and V-VV pairs as containing a short and a long vowel, because of orthographic effects on L2 perception, as detailed below.

If Italian_{L1}-English_{L2} listeners perceive the same sound as two different sounds when spelled with single letter or digraph, the following predictions could be made. Accuracy should be lower in L-LL (single letter - digraph) homophonic word pairs than in control pairs for Italian_{L1}-English_{L2} speakers. There should be no differences in accuracy in the native speaking group, for whom single letters and digraphs are different spellings of the same sound. If these predictions are borne out, this would mean that orthographic forms result in Italian_{L1}-English_{L2} speakers not only producing (Bassetti, 2017; Bassetti et al., 2020; Bassetti et al., 2018) but also perceiving a phonological contrast that is not attested in the English phonological system.

In terms of RTs, when pictures activate two homophones, Italian_{L1}-English_{L2} speakers should be slower with L-LL than control pairs, because to them L-LL pairs contain different sounds. All listeners, including native speakers, may be slightly slowed down if orthography is co-activated with phonology, as they would need to resolve the incongruence between the similarity of the two phonological words and the differences between the two orthographic words. This would be similar to findings of metalinguistic awareness tasks, where native speakers' RTs for rhyme judgments are slower when rhymes have different rather than the same spellings, as in *pie-rye* versus *pie-tie* (all /aɪ/; Donnenwerth-Nolan et al., 1981; Pring, 1986; Seidenberg & Tanenhaus, 1979). This is a lexical effect, because it is not found with (arguably pre-lexical) shadowing (Ventura et al., 2004), and is not found with pseudowords (Ziegler & Ferrand, 1998). However, for native speakers the effect should be much smaller with C-CC pairs than with V-VV pairs, because vowel digraphs have many grapheme-phoneme correspondences (e.g., the *ie* in *pie* has seven readings, Carney, 1994), whereas double consonant letters can be decoded in only one way, which for native speakers is the same as the corresponding single consonant letter. In contrast, for Italians, double consonant letters and the corresponding single consonant letter could represent two different sounds (a geminate and a singleton consonant), therefore causing longer RTs.

In the first experiment, we tested whether orthographic input affects L2 listeners' L2 speech perception. We compared Italian_{L1}-English_{L2} sequential bilinguals and English native speakers performing a consonant and a vowel perception task either with only auditory input or with auditory plus visual input. Visual input consisted of two pictures, which were used to activate the two words of interest. If orthographic form affects perception, bilinguals would show lower levels of accuracy in the condition with auditory-visual input than in the condition with auditory input only. This would be due to a perceptual illusion, because the bilinguals' mental representations of the two homophonic words are not homophonic, but contain a short and a long sound respectively. Therefore, top-down processes would interfere with bottom-up processes, resulting in lower accuracy, due to perceiving a singleton and a geminate, that is to say two different phonological categories, and slower RTs due to the incongruence between the phonological and orthographic forms of the two words.

In the second experiment, we investigated whether naturalistic exposure reduces orthographic effects on perception by comparing instructed $Italian_{L1}$ learners of $English_{L2}$ and sequential $Italian_{L1}$-$English_{L2}$ bilinguals who had been living in England for years. We hypothesised a positive effect of lengthy exposure to speech by native speakers of English and of languages without gemination. Previous research found that naturalistic exposure does not eliminate orthographic effects on sound duration in the speech production of $Italian_{L1}$ speakers of $English_{L2}$ (Bassetti et al., 2018; 2020; Mairano et al., 2018). However, it is unclear whether naturalistic exposure reduces orthographic effects, as previous research found that double letter spelling affected consonant production and categorisation more in learners than in bilinguals in a word repetition task and a rhyme judgment task (Bassetti et al., 2020), but not in reading aloud tasks (Bassetti et al., 2018; Mairano et al., 2018) or in a qualitative study of rhyme judgment (Bassetti et al., 2020). As for vowels, no differences were found between learners and bilinguals in vowel production or categorisation (Bassetti et al., 2018; 2020), and in the case of vowel production one study even found a slightly stronger effects in bilinguals than learners (Bassetti et al., 2020). In the present study we compared learners, who were studying English as a foreign language in Italian schools, with sequential bilinguals who had studied English in school but then received lengthy naturalistic exposure while living in England (they are called 'sequential' here to clarify that they were not early bilinguals). There was also a control group of English native speakers. Since Experiment 1 had already ruled out orthographic effects on accuracy in the absence of activation of the two homophones of interest, Experiment 2 did not include an Auditory Input condition, but used Auditory plus Visual Input only. If bilinguals have higher accuracy rates than learners, this would indicate that naturalistic exposure reduces orthographic effects on L2 perception.

This study is innovative in both object of inquiry and methods, as follows. First, the study tested whether orthography results in L2 listeners perceiving a contrast that is not attested in the target language. This is innovative because L2 perception research has focussed on the perception of sounds and contrasts that are attested in the L2 phonological system, but there has been no research on the perception of a contrast that exists in the L1 phonological system and is not attested in the phonological system of the target language. In light of evidence, reported above, that L2 speakers produce sounds that do not exist in the L2 spoken input, it is important to investigate whether they also perceive such sounds. This has theoretical implications as follows. In the early days of L2 phonology research, the widely cited hierarchy of difficulty in L2 phonological acquisition posited that negative divergent transfer is very difficult, that is to say when two L2 phonological categories correspond to one L1 category (as the widely-studied $English_{L2}$ /l/-/r/ contrast for $Japanese_{L1}$ speakers, whose L1 only has one category, realised as [ɾ] or [l] depending on its position), but the opposite situation where the L1 has two categories and the L2 has one was considered easy for learners. Current models of L2 phonological development focus on the acquisition of L2 sounds and contrasts, particularly those that do not exist in the L1. For instance, Eckman et al. (2003) distinguish L2 phonological contrasts where the L1 has neither sound, the L1 has one sound, and the L1 has both sounds as allophones. In contrast, L2 speech production research shows that L2 learners produce sounds and contrasts that do not exist in the target language, whether producing a [v] sound in Spanish$_{L2}$ (Zampini, 1994), or a short-long contrast in English$_{L2}$ consonants. If L2 listeners turn out to illusorily perceive two different sounds where the L2 has no distinction, that would explain why they produce two different sounds in their L2 speech production. This would then show that the level of difficulty of this combination — two categories in the L1 corresponding to one category in the L2 — deserves to be investigated more than has been the case so far.

Second, the study introduces a task that can be used to test the perception of a contrast that is not attested in the language under analysis, using real words rather than pseudowords or an artificial language, and without orthographic input. Perception tasks usually investigate the perception of real contrasts, for instance the AXB task and its variations investigates the boundary between two phonological categories, but it cannot be used to investigate the geminate-singleton distinction in English unless consonant duration is artificially lengthened to create something that L2 listeners would perceive as a geminate. The present task eliminates the need for this kind of artificial manipulation, as it presents actual recordings of native speakers producing homophonic word pairs that L2 listeners should illusorily perceive as containing a different sound. The task also gets rid of the limitations of explicitly metalinguistic awareness tasks, such as phoneme segmentation or phoneme counting, which may invite the use of orthographic forms as a strategy to perform the test, and may therefore not be good measures of orthographic influence on actual perception. The task introduced in this study still involves a judgment, but by using images to activate the words of interest it manages to eliminate the need for participants to search their whole lexicon, and crucially it allows the target words to be activated without the need to provide orthographic input, which is important in light of some, albeit inconsistent, evidence that orthographic effects may be stronger when auditory input is accompanied by orthographic input, compared with auditory input alone (see Shea, 2017, for evidence for, and Bassetti, 2017, for evidence against). Such a task, if successful, could be used in other situations where L2 listeners' perception of real words in real speakers of a real language should be tested for a sound contrast that is not contrastive to native speakers.

Experiment 1

Method

Design

Participants took part in a consonant perception task and a vowel perception task with either auditory input or auditory plus visual input. There were two between-group factors: language background (Italian$_{L1}$-English$_{L2}$ bilinguals and English native speakers) and Input (auditory input, or auditory plus visual input). Type of word pair was a within-group factor with two levels: target pair (C-CC or V-VV) and control pair. The dependent variables were accuracy (correct/incorrect response) and RTs. Following the main experimental task, participants' knowledge of the correct spelling of the stimuli was tested with a spelling-to-dictation task.

Participants

Participants were 50 Italian$_{L1}$-English$_{L2}$ sequential bilinguals (mean age 34 years, *SD* = 8) and 50 British English native speakers (mean age 26, *SD* = 9), recruited in London. Within each language group (bilingual/English native speaker) participants were randomly assigned to one of two stimulus presentation groups (both *n*=25): one group took part in the perception tasks with auditory input and the other took part in the tasks with auditory plus visual input.

The bilinguals were native speakers of central and southern varieties of Standard Italian. Gemination is well documented in these varieties of Italian, where geminate consonants are phonetically about twice as long as corresponding singletons (Bertinetto & Loporcaro, 2005; Giordano & Savy, 2012; Mairano & De Iacovo, 2020, found similar durations in contemporary northern varieties). The bilingual participants had lived in Italy at least until high-school graduation, and had been living in England for an average of 8 years (*SD* = 6). Mean age of onset of acquisition of English was 9.67 years (*SD* = 2.82). The median self-rated English language proficiency for the group was the CEFR level C1 (Effective Operational Proficiency, i.e., advanced), including a quarter who rated themselves with the highest rating of C2 (Mastery, i.e., proficient). According to self-reported number of hours per week, the bilinguals spent on average 10% more time reading than listening to English (mean reading:listening ratio = 1.09, *SD* = 0.82), and spent more time using English than Italian (mean ratio of English reading time to total (Italian and English) reading time = 0.67, *SD* = 0.22; listening: 0.71(0.18); interacting: 0.63(0.23)).

The majority of English native speakers (86%) reported their native variety as Southern British English; seven reported another variety of British English (Yorkshire: *n* = 3; North Irish: *n* = 3; Manchester: *n* = 1).

The sample size was calculated using G*Power 3.1 (Faul et al., 2009). In order to detect a medium effect size (f = 0.25) in a mixed design with 2 between-subject factors measured in 2x2 within-subject factors, G*power suggested that 50 participants in each group would be sufficient to detect a power of 0.99 with a standard α = 0.05.. All participants reported normal or corrected-to-normal vision, and no reading or listening difficulties; 8% were left-handed. All participants were able to perceive the English lax-tense contrast as shown by their answers to 13 filler pairs opposing /ɪ/ and /iː/ (see below). Participation in the study was voluntary and rewarded with cash. The study received the approval of the University of Warwick Research Ethics Committee (Number: 118/14-15, Project Title: Effects of Orthography on Phonology in Second Language Speakers of English: Pronunciation, Phonological Awareness, Speech Perception and Spelling).

Materials

Materials consisted of 132 English word pairs: 18 C-CC pairs, 18 V-VV pairs, 36 control pairs (18 each for consonants and vowels) and 60 fillers (30 each for consonants and vowels). Materials are available at the Open Science Framework (https://osf.io/56ba4).

In *L-LL word pairs* (single letter – digraph pairs), the two words were homophonic: one sound (a consonant or a vowel) was written with a single letter in one word (*C-word* or *V-word*) and with a digraph in the other one (*CC-word* or *VV-word*). An example C-CC word pair is *finish-Finnish* and an example V-VV word pair is *soles-souls*.

L-LL pairs were the ones used to test orthographic effects on speech production by Bassetti et al. (2018), with the following changes. We removed three C-CC word pairs (*metal-mettle*, *medal-meddle*, *literal-littoral*) that were not imageable and had been included in our 2018 read-aloud experiment to measure VOTs. We also added two C-CC pairs (*bass-base*, *guest-guessed*) and four V-VV pairs (*B-bee*, *bore-boar*, *for-four*, *pore-poor*) in order to obtain 18 L-LL pairs for both consonants and vowels.

To obtain a representative sample of English words, in *C-CC word pairs* the target consonant was a plosive ([t] or [d] in three word pairs), a nasal ([n] in four word pairs), a fricative ([s] or [z] in six word pairs), or an approximant ([r] or [l] in five word pairs). The target consonant was in post-stressed position in 15 pairs and in pre-stressed position in the other three (post-stressed geminates are longer in Italian, Payne, 2005). Since geminate consonants in Italian occur only in certain positions (between vowels, or between vowel and liquid or glide), we included eight CC-words where gemination is legal in Italian (e.g., intervocalic *tenor-tenner*) and ten CC-words where gemination is not legal in Italian, that is to say it is not attested in native Italian words (either word-final, e.g., *add*, six pairs, or syllable coda, e.g., *missed*, four pairs, all ending in *-ed;* in the latter, legality of gemination depends on whether the participant produced *-ed* as [d] or as [Vd], as Italian$_{L1}$ speakers often do, see Bassetti & Atkinson, 2015). Bassetti et al. (2018) found that Italian$_{L1}$ speakers produce CC-consonants as longer than C-consonants in all positions, but the effect is stronger in intervocalic position than in positions that are illegal in Italian. Six C-CC pairs were orthographic minimal pairs — homophonic pairs that only differ in the spelling of the target sound (such as *ad-add*) — (including three pairs containing 'silent e') and 12 orthographic semi-minimal pairs — homophonic pairs that differed both in the spelling of the target consonant and in the spelling of another sound, such as the second vowel in *tenor-tenner*. This was due to the limited number of English orthographic minimal word pairs and the even smaller number whose words are both imageable and frequent enough to be familiar for L2 speakers. Also due to the difficulty of finding word pairs that satisfied our criteria, ten CC-words were cognates or loanwords in Italian. The proportion of cognates may then be higher in our materials than among CC-words in the English lexicon. In terms of frequency, Wilcoxon Signed-Ranks Tests revealed that C-CC word pairs had a higher written than spoken frequency (Mdn_{spoken} = 6.78, range: 0.10-7,264.17; $Mdn_{written}$ = 17.21, range: 0.63-1,873.26, V = 181, p = 0.016). There was no statistically significant difference in frequency between C-words and CC-words (Spoken frequency: $Mdn_{C\text{-}words}$ = 7.45, range: 1.15-7264.17; $Mdn_{CC\text{-}words}$ = 5.91, range: 0.10-106.82, W = 197.5, p = 0.268; Written frequency: $Mdn_{C\text{-}words}$ = 21.47, range: 0.80-1873.26; $Mdn_{CC\text{-}words}$ = 9.85, range: 0.63-128.19, W = 200, p = 0.239). We controlled for frequency in order to avoid confounds of two types: first, orthographic effects may be weaker with frequent than infrequent L2 words (Vokic, 2011); second, homophones may be pronounced at different rates depending on their frequency (more frequent items are pronounced faster, Gahl, 2008, although this effect seems to be smaller in experimental than naturalistic data, Guion, 1995).

In *V-VV word pairs* the target vowel was [iː] in five word pairs, [ɜː] in one word pair, [ɔː] in five word pairs, [uː] in one word pair, and [əʊ] in six word pairs. Except *morning-mourning*, all words were monosyllabic, therefore all but one target vowels were stressed (stress affects vowel durations, Klatt, 1976). There were eleven orthographic minimal pairs (e.g., *for-four*, including pairs spelled with <e_e> or 'silent e' such as *pore-poor*), and seven semi-minimal pairs (e.g., *scene-seen*). The digraph consisted of double vowel letters (VV, e.g., *seek*, *shoot*) in six pairs, and two different letters (V_1V_2, e.g., *board*, *soul*) in the remaining eight pairs (for the word *heard* in the pair *herd-heard* some learners may have a phonological representation with a diphthong [ea] or [eə] rather than the long vowel [ɜː], but this would still support our hypothesis). In 12 pairs the V-vowel was spelled with the grapheme <V_e>, which represents a long vowel in English, as in *nose*, however, Italian speakers are usually not aware of this correspondence (Bassetti & Atkinson, 2015), as confirmed by an informal survey of participants in this study, so that such cases count as single letter spellings from the participants' point of view. In one V-word, the target vowel was not represented in the orthographic form (<B>, / biː/). There was a loanword (*jeans*), which Italians produce with a shorter vowel than other vowels spelled with <ea> (Bassetti et al., 2018).

V-VV word pairs had a higher written than spoken frequency (Mdn_{spoken} = 13.60, range: 0.1-6,440.63; $Mdn_{written}$ = 31.00, range: 1.4-9,233.77, V = 204, p = 0.044). V- and VV-words did not differ in frequency (Spoken frequency: $Mdn_{V\text{-}words}$ = 8.606, range: 0.10-6,440.63; $Mdn_{VV\text{-}words}$ = 22.19, range: 1.06-1,188.10, W = 117.5, p = 0.164; Written frequency: $Mdn_{V\text{-}words}$ = 19.06, range: 1.40-9,233.77; $Mdn_{VV\text{-}words}$ = 38.76, range: 2.39-375.01, W = 125.5, p = 0.255).

In *control word pairs*, the two words were homonyms (same phonological and orthographic form, different meaning), e.g., *fork-fork* and *leaves-leaves*. The *fillers* were minimal word pairs that differed in one sound. These items were included to vary the response pattern, and were not included in the analyses. Filler consonant pairs opposed an articulatory and perceptively similar consonant pair, such as *angle-ankle*. Filler vowel pairs opposed a lax and a tense vowel, or other vowels (e.g., /ɪ/ - /e/, /e/ - /æ/, /ɒ/ - /ʌ/). Thirteen filler vowel pairs opposing /ɪ/ and /iː/ — such as *ship-sheep* — were used to screen participants who could not perceive this contrast, as there could not be orthographic effects on vowel perception in these participants.

All words were recorded in a sound-proof booth by a female native speaker of British English, whose production had been identified as a good example of Received Pronunciation by an English_{L1} sociologist. In order to hide the purposes of the research, and to avoid potential exaggeration of phonetic detail, her reading list included C- and V- targets, controls and fillers in random order. Acoustic measurements confirmed that the target sounds within each pair had similar durations (M_C = 167 ms, SD = 86; M_{CC} = 168, SD = 89; $t_{(17)}$ = -0.34, p = 0.736; M_V = 312, SD = 105; M_{VV} = 324, SD = 103; $t_{(17)}$ = -1.69, p = 0.109).

Finally, a questionnaire was used to collect demographic and linguistic information. For bilinguals only, this included English proficiency self-ratings using the CEFR descriptors (Common European Framework of Reference for Languages, Verhelst et al., 2009). For each of the four skills (listening, reading, speaking, writing), participants chose one of six ‘can do’ that map to six levels, from A1 (Breakthrough) to C2 (Mastery). The four self-ratings were used to obtain a median proficiency score for each participant.

Tasks

Half of the participants in each language background group performed the perception task with auditory input only, and half with auditory plus visual input. For the visual input, an image representing the word of interest appeared on the computer screen simultaneously with the spoken word’s recording, and disappeared simultaneously with the end of the recording. The purpose was to activate the word of interest on each occasion. The images were tested for reliability in a pilot study with final debriefing involving 5 native speakers. Finally, a spelling task was used to check the participants’ knowledge of the words used in the perception tasks.

Perception tasks. To hide the purpose of the experiment, the consonant and vowel perception tasks were presented as a single perception task. In each trial, a fixation point appeared for 400 ms, then an audio recording of a word was played over headphones. Each recording lasted 900 ms (silence was added at the end of the word recording to adjust the duration). After that, another fixation point appeared for 400 ms, followed by another 900-ms audio recording. For the Auditory plus Visual Input participants, a colour image appeared in the centre of the screen simultaneously with the onset of the audio recording, and disappeared at the end of the recording. As an example, the audio recordings of the word /ˈfɪnɪʃ/ in the C-CC word pair *finish-Finnish* were accompanied by a picture of a finish-line flag and a Finnish flag. Figure 1 shows a schematic representation of the two conditions of the perception tasks. The participants’ task was to click on one of two designated buttons on the response box to indicate whether the two words contained the same sounds (right button) or different sounds (left button). RTs were calculated from the onset of the second word on each trial. The order of presentation of trials was randomised anew for each participant, and within each trial, the two audio recordings and their associated pictures were also presented in random order.

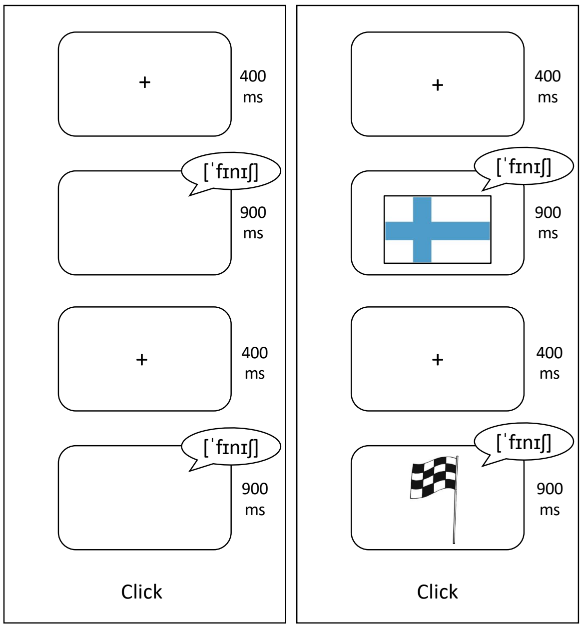

Figure 1. *Schematic representation of the two conditions of the Perception Tasks: Auditory Input (left panel) and Auditory plus Visual Input (right panel).*

Spelling-to-dictation task. In this task participants heard the same audio recordings as in the perception task, accompanied by the same images used for the auditory-input condition. Picture-audio file pairs were presented one by one in random order. The task was used to ensure that participants knew the spelling of the target sounds, as orthographic effects would not be expected if participants did not know the spellings (as shown by Sokolović-Perović et al., 2019, in a production task).

Instruments

Audio files were recorded in a sound-attenuated booth with a Røde NT2-A microphone and an Alesis Multimix 12 Firewire mixer at 44 kHz, and normalized to avoid intensity differences. Colour pictures were selected from the Art Explosion library (Nova Development, 2004). Psyscope X (Cohen et al., 1993) was used to present stimuli. Participants heard audio recordings through AKG HSD171 headphones. During the perception task, participants interacted with a MacIntosh laptop using an IoLab Response box, which recorded responses and response times. For the spelling tasks, responses were recorded using a keyboard.

Procedure

Participants were tested individually in a sound-attenuated or quiet room. All participants performed the perception task before the spelling task. Participants could rest at any point, as they controlled the start of the next trial with a button press.

For the perception task, participants decided whether the two words they heard had the same sounds or different sounds, and pressed the right button on the response box for 'same sounds' and the left button for 'different sounds'. Before the perception task, participants received written instructions and performed four practice trials, which included two homophonic and two different word pairs (e.g., *blue-blew* and *sing-song*), but no single letter-digraph contrasts. An $\text{Italian}_{\text{L1}}$-$\text{English}_{\text{L2}}$ researcher provided oral instructions and feedback on practice trials if needed.

For the spelling-to-dictation task, participants typed the word they heard using a keyboard. Simultaneously with the word's audio file, all participants saw a picture representing the word's meaning. The complete session (perception and spelling tasks) took about thirty minutes.

Data analysis

Before analysing accuracy and RTs for target and control word pairs in the perception task, we eliminated pairs where one or both words had been spelled incorrectly in the spelling-to-dictation task. As a result, for the consonant data 12% of target word pairs and 4% of control pairs were eliminated for the bilingual group and 7% of targets (and no controls) for the native speaker group. In the case of the vowel data, 15% of target word pairs and 6% of control pairs were eliminated for the bilingual group and 4% of target and 2% of control pairs for the native speaker group.[1]

Before analysing response times (RTs), we also removed outliers, considered as RTs that were beyond the 99th percentile and below the 1st percentile. For consonant data, this involved removing 3.27% of word pairs for the bilingual group and 2.02% for the native speaker group. For vowel data, it involved removing 2.04% of word pairs for the bilingual group and 2.01% for the native speaker group. The remaining RTs were log-transformed to approximate normal distribution.

Statistical analyses were performed using RStudio 1.1.456 (RStudio Team, 2016) and R-3.5.1 (R Core Team, 2018). Accuracy and RTs were analysed separately for the consonant and vowel perception tasks using the following procedure. In order to analyse accuracy, which was a binary variable (correct or incorrect), we fit two logit mixed-effect models (Jaeger, 2008) to consonant and vowel data using the function *glmer* in the *lmerTest* package (Kuznetsova et al., 2017). For RTs, we fit two linear mixed effects models using the *lmer* function in the same package. For both accuracy and RT analyses, the initial maximal models included the main effects and interaction between language background (native speaker, bilinguals), type of word pair (C-CC or V-VV, controls) and input (auditory, auditory plus visual). As random effects, we considered random intercepts for participants and for word pairs, a by-participant random slope for the effect of type of pair on participants, and a by-word pair random slope for the effect of input. We used likelihood ratio tests to simplify the

[1] For completeness, we report spelling error rates from the Spelling-to-dictation task for each group of participants. Consonant data – Native speakers: 4% (*SD* = 19%); Bilinguals: 8% (*SD* = 27%). Vowel data: Native speakers: 3% (*SD* = 18%); Bilinguals: 11% (*SD* = 30%)."

models when fixed effects were not significant, and the function *rePCA* in the *RePsychLing* package when the random structure was overfitted (Bates et al., 2015). We calculated conditional and marginal R^2 using function *r.squaredGLMM* in the *MuMIn* package (Bartoń, 2018); Wald chi-square tests for fixed effects using the *Anova* function in the *car* package (Fox & Weisberg, 2019); and post hoc contrasts on estimated marginal means with Tukey's adjustment using the function *emmeans* in *lsmeans* package (Lenth, 2016). Visual inspection of fitted and residual values did not show deviations from model assumptions.

Results

Effects of language background and consonant spelling on the perception of consonant length

Accuracy. Figure 2 shows the accuracy rate (mean percentage of correct responses) in the consonant perception task for control and C-CC word pairs under auditory only or auditory plus visual input and for native speakers and bilinguals.

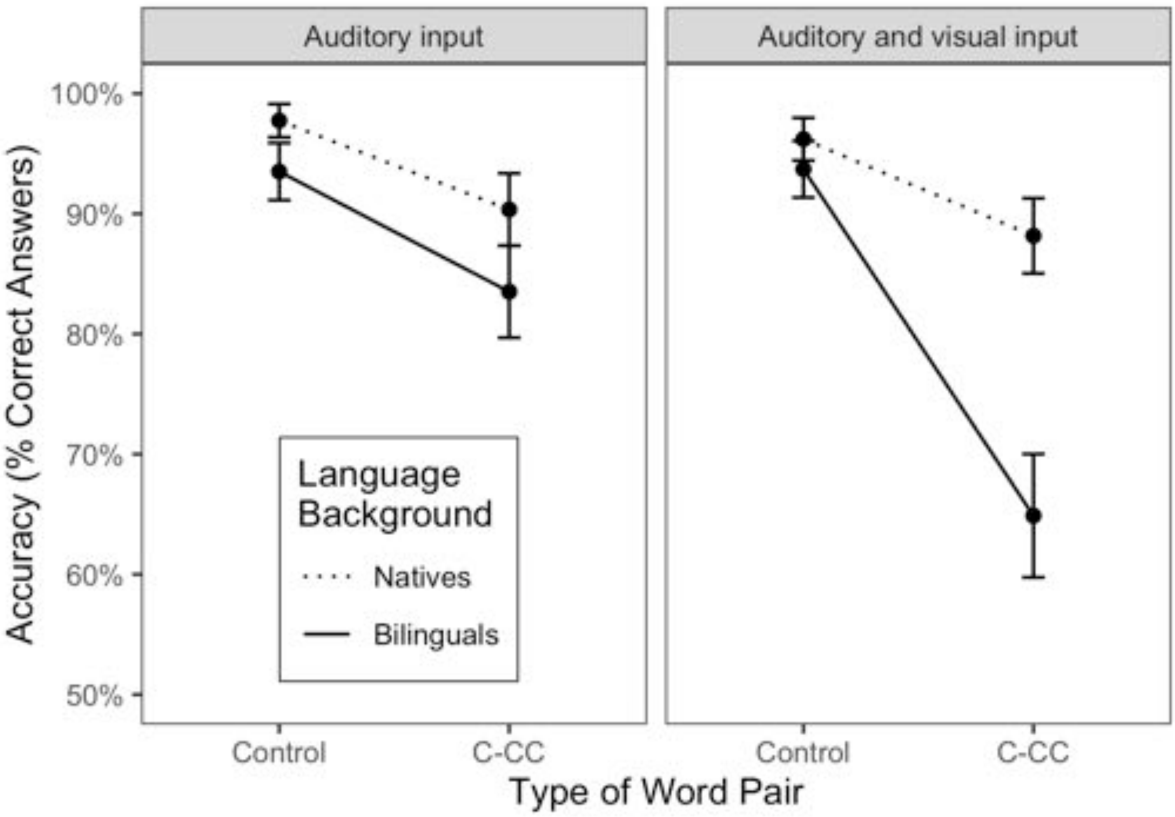

Figure 2. *Mean percentage of correct responses (bars represent 95%* CI*s) in the Consonant Perception Task in Experiment 1 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual) and type of word pair (control, C-CC).*

The model (see Online Supplementary Materials, Table 1) that best explained the likelihood of a correct response included as fixed effects language background, input, type of word pair and their interactions. As random effects, the model included the by-participant random slope for type of word pair and the by-word pair random slope for input, and random intercepts for participants and word pairs. The main effect of language background was significant ($\chi^2 = 22.72$, $p < 0.001$), with higher accuracy for native speakers than bilinguals. Type of word pair was also significant ($\chi^2 = 30.34$, $p < 0.001$), with responses more likely to be correct with control than C-CC pairs. Input approached but did not reach significance, $\chi^2 = 3.59$, $p = 0.058$). The interaction between input and type of word pair was significant ($\chi^2 = 3.97$, $p = 0.046$), as was crucially the three-way interaction of language background, input and type of word pair ($\chi^2 = 4.19$, $p = 0.041$). Post-hoc contrasts with Tukey adjustment revealed that bilinguals who received auditory plus visual input were more likely to give correct responses to control than C-CC word pairs ($z = -6.27$, $p < 0.001$). They were also more likely to respond to C-CC word pairs incorrectly, compared both with bilinguals who received auditory input ($z = -3.02$, $p = 0.030$) and with English native speakers in either group ($z_{visual} = -4.90$, $p < 0.001$; $z_{auditory\_plus\_visual} = -4.46$, $p = 0.001$). Other contrasts between C-CC word pairs and control pairs for native speakers in both input conditions and for bilinguals in auditory condition did not reveal significant differences (all $p > 0.05$).

Response times. Figure 3 shows the geometric mean of RTs in the consonant perception task with auditory and auditory-visual input for the two language background groups for control and C-CC pairs.

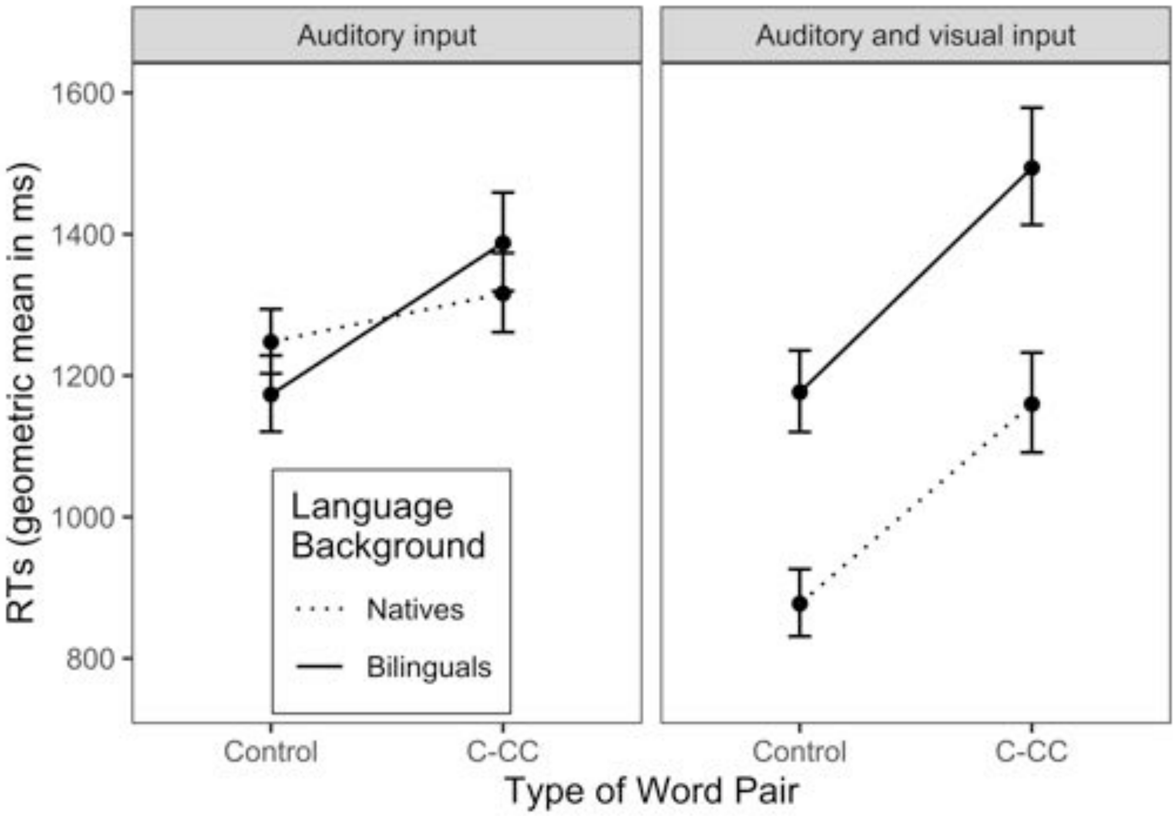

Figure 3. *Geometric mean of RTs (ms; bars represent 95% CIs) in the Consonant Perception Task in Experiment 1 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual) and type of word pair (control, C-CC).*

The final model (see Online Supplementary Materials, Table 2) included the fixed effects and interactions of language background, input, and type of word pair. The random structure included a by-participant random slope for type of word pair, a by-word pair random slope for input, and random intercepts for participants and word pairs. The effect of language background was significant ($\chi^2 = 5.63$, $p = 0.018$) with faster RTs for native speakers than bilinguals. The effect of input was significant ($\chi^2 = 4.25$, $p = 0.039$) with faster RTs for auditory-visual than auditory input. The effect of type of word pair was significant ($\chi^2 = 20.95$, $p < 0.001$), with faster RTs for control than C-CC word pairs. The interactions of language background and input and of input and type of pair were also significant ($\chi^2 = 6.56$, $p = 0.010$ and $\chi^2 = 17.36$, $p < 0.001$, respectively). Importantly, the three-way interaction of interest was significant ($\chi^2 = 6.59$, $p = 0.010$). Post-hoc contrasts with Tukey adjustments revealed faster RTs with controls than C-CC pairs in bilinguals, whether they saw visual input ($t = 4.86$, $p < 0.001$) or received only auditory input ($t = 3.98$, $p = 0.002$). For native speakers, the contrast between control and C-CC pairs was significant only when visual input was presented ($t = 5.92$, $p < 0.001$). This is probably because visual input reduced native speakers' processing time with control words, as native speakers who saw visual input responded to control word pairs faster than native speakers who only received auditory input ($t = -4.17$, $p < 0.001$), and faster than bilinguals who received visual input ($t = 3.61$, $p = 0.006$). They also responded to C-CC pairs faster than bilinguals who saw visual input, but the difference was smaller than with control pairs ($t = 2.97$, $p = 0.035$).

Additional analysis of legality of position and cognate status. Finally, to check whether perception of English consonant length is affected by the legality of its position in Italian phonotactics, we compared bilinguals' accuracy in C-CC pairs where the target consonant was in a position that is legal (intervocalic) or illegal (word-final or syllable coda) in Italian native words, and found no effects of position.

Regarding cognate status, no differences were found between the word pair containing an orthographically-congruent cognate with the target consonant in legal position, and the other two pairs containing the same target consonant [n]. There were also no differences between the two pairs containing an incongruent cognate with intervocalic consonant and the other [r] and [n] pairs.

Effects of language background and vowel spelling on the perception of vowel length

Accuracy. Figure 4 shows the mean percentage of correct responses in the vowel perception task for control and V-VV word pairs under auditory only or auditory plus visual input and for native speakers and bilinguals.

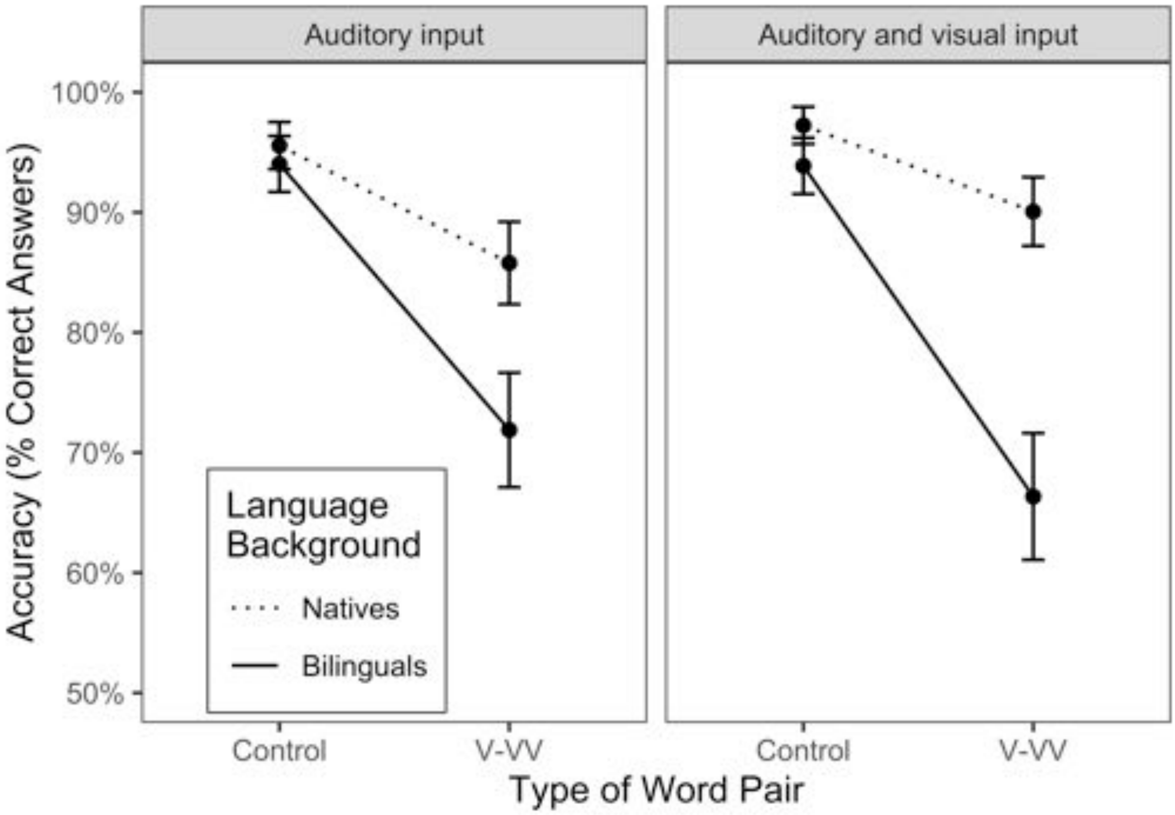

Figure 4. *Mean percentage of correct responses (bars represent 95%* CI*s) in the Vowel Perception Task in Experiment 1 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual) and type of word pair (control, V-VV).*

The final model (see Table 3 in the Online Supplementary Materials) included as fixed effects language background, type of word pair, and their interaction. The random structure included a by-participant random slope for type of word pair, a by-word pair random slope for input, and random intercepts for participants and word pairs. The effect of language background was significant ($\chi^2 = 29.94$, $p < 0.001$) with native speakers more likely to respond correctly than bilinguals. The effect of type of word pair was significant ($\chi^2 = 26.37$, $p < 0.001$) with more accurate responses for control than V-VV word pairs. The main effect of language background was qualified by the interaction ($\chi^2 = 8.64$, $p = 0.003$). Post-hoc contrasts with Tukey adjustment revealed that both native speakers and bilinguals were more accurate with control than V-VV pairs, but the effect was stronger for bilinguals ($z_{natives} = -2.56$, $p = 0.041$, $z_{bilinguals} = -5.76$, $p < 0.001$), and the native speakers were significantly more accurate than bilinguals in the case of V-VV pairs but not control pairs ($z_{v\text{-}vv\ pairs} = -6.21$, $p < 0.001$, $z_{controls} = -2.34$, $p = 0.074$).

Response times. Figure 5 shows the geometric mean of RTs for control and V-VV word pairs, for the two different input conditions and the two different language groups.

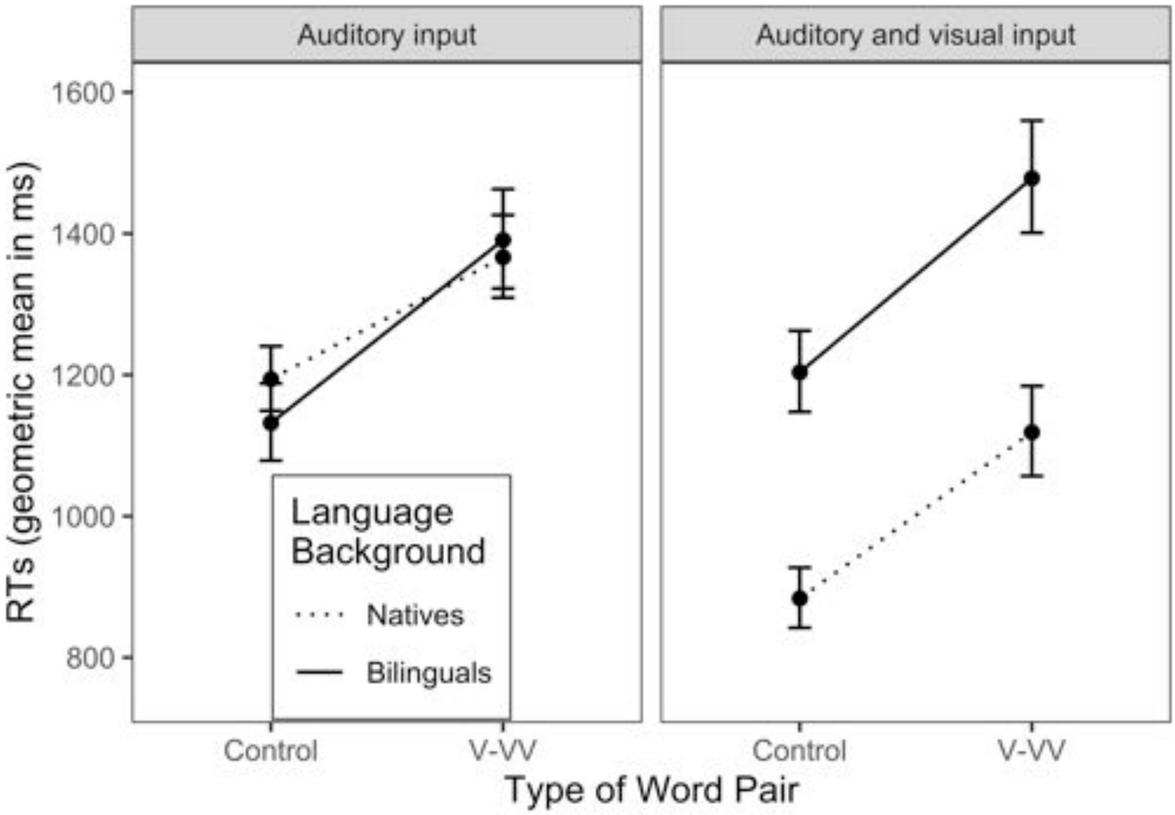

Figure 5.*Geometric mean of RTs (ms; bars represent 95% CIs) in the Vowel Perception Task in Experiment 1 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual) and type of word pair (control, V-VV).*

The final model (see Online Supplementary Materials, Table 4) included the interaction and fixed effects of input and language background, and the fixed effect of type of word pair. The random structure was the same as in the previous model for the accuracy data.

The effect of type of word pair (χ^2 = 31.44, p < 0.001) was significant with faster responses to control than to V-VV word pairs. The effect of language background was significant (χ^2 = 6.50, p = 0.011) with native speakers faster than bilinguals. The effect of input was also significant (χ^2 = 4.15, p = 0.042), with faster responses in the case of auditory plus visual input. The interaction between language background and type of pair did not improve the model fit, showing that both native speakers and bilinguals were faster with control than V-VV word pairs. The interaction between language background and input was significant (χ^2 = 8.58, p = 0.003) and post-hoc contrasts revealed that native speakers with auditory-visual input were faster than all other groups (native speakers with auditory input, t = -3.50, p = 0.002; bilinguals with auditory input, t = 3.23, p = 0.007; bilinguals with auditory-visual input, t = 3.87, p < 0.001).

Additional analysis of type of digraph and cognate status. Finally, we found no differences in bilinguals' accuracy between pairs spelled with double vowel letters and those spelled with two different vowel letters (Bilinguals: $M_{V\text{-}VV}$ = 68%, $M_{V\text{-}V1V2}$ = 70%), or between the pair containing a loanword (*genes-jeans*) and the other pair containing the same target vowel (*scene-seen*; $M_{scene\text{-}seen}$ = 53%, $M_{genes\text{-}jeans}$ = 62%).

Discussion

Experiment 1 aimed to test whether number of letters (single or digraph) in the spelling of English consonants and vowels affects speech perception in L2 listeners whose native language represents long consonants with double letters. Overall, results confirmed the predictions with regards to accuracy with consonant perception, and partially confirmed predictions regarding vowel perception, as detailed below.

Effects of orthographic forms on consonant perception

Accuracy results showed that Italian_{L1}-English_{L2} sequential bilinguals perceive the same English spoken consonant as two different sounds (a geminate and a singleton consonant) inside two homophonic words that represent the same consonant with a single consonant letter or with double letters. The three-way interaction between type of input, language background and type of word pair shows that bilinguals who heard recordings of English homophonic word pairs where the same sound is spelled with one or two letters were likely to report perceiving different sounds only if the two words had been pre-activated by two pictures. Those bilinguals who only heard audio recordings without visual input were more likely to report perceiving two similar spoken words, compared with bilinguals who 'knew' which two words they were hearing. Indeed, the mean percentage of correct responses for CC-C pairs was just 65% (*SD* = 48%) among the bilinguals who received visual input, compared with 84% (*SD* = 37%) among bilinguals who did not. Visual input made no difference in percentage of correct responses for control pairs, which was 94% in both bilingual groups (*SD*s = 24% for the group with visual input and 25% for the group without visual input). Although overall all groups descriptively showed lower percentages of correct responses to CC-C pairs than control pairs, the model showed that likelihood of correct response was not affected by type of word pair, as its main effect was justified by the interactions, and post-hoc tests confirmed that the difference in likelihood of correct response was lower for CC-C than control pairs only among bilinguals who received visual input. The main effect of language background was also justified by the interactions, so that native speakers were not more likely to respond correctly than bilinguals overall, but only with CC-C pairs presented with visual input.

Response time results revealed effects of spelling on both natives exposed to visual input and L2 listeners, but for different reasons. Bilinguals were overall faster with control than CC-C pairs, although the effect was stronger with visual input than without. It appears that bilinguals take longer to respond when two English homophonic words (e.g., /ˈfɪnɪʃ/) could be identified as two words with different phonological forms in their mental lexicons (/ˈfɪnɪʃ/ and /ˈfɪnːɪʃ/), compared with two homophonic words that have the same phonological form in their L2 mental lexicon (e.g., /palm/), and the effect is stronger when they 'know' that they are hearing two words with different phonological forms. An alternative explanation is that RTs were slower because the analysis included RTs for both 'same' and 'different' answers, and although from the perspective of the L2 listener both were correct responses, because pairs such as *finish-Finnish* could be legitimately considered different words, still processing time was longer for orthographically incongruent C-CC pairs. Among native speakers, visual input sped up responses to control pairs, probably because activating the two specific words speeds up lexicon searches as there is no need to search the whole lexicon for possible homophones, but in the case of CC-C pairs this positive effect is offset by the additional time required to resolve the incongruence between the two words having the same phonological form but different orthographic forms. It appears then that activating words with different spellings may slow down all listeners, in line with the negative effects of different spellings on RTs found with rhyme judgment tasks, which are faster when rhymes have the same spelling compared with different spelling (Donnenwerth-Nolan et al., 1981; Pring, 1986; Seidenberg & Tanenhaus, 1979). Yet, this was not expected for native speakers, because — unlike other graphemes that can be recoded in different ways — double consonant letters have only one recoding. It is only Italian$_{L1}$-English$_{L2}$ speakers who may recode single and double consonant letters as two different phonemes. It appears that incongruences between homophonic phonological forms and incongruent orthographic forms affect the time taken to respond for all listeners, whether native or non-native, regardless of whether the different graphemes can or cannot be decoded as different phonemes, but only L2 listeners are also affected in their likelihood of correct responses.

Effects of orthographic forms on vowel perception

Unlike consonants, accuracy results showed that English word pairs were more likely to be perceived as containing different vowels when the two words were homophones (with different spelling) than when they were homonyms (with the same spelling), both by native and L2 listeners, and regardless of whether they received visual input or not. The two-way interaction showed that native listeners were more likely than L2 listeners to respond correctly to V-VV pairs, whereas there were no differences between native and L2 listeners with control pairs. The three-way interaction was not significant, indicating that the negative effect of spelling on likelihood of correct response was the same for bilinguals who did and did not receive visual input that activated the two words of interest.

Looking at native listeners, there was a small decrease in likelihood of correct response, with both types of input, and the percentage of correct responses was lower with V-VV than control pairs (auditory input group: $M_{controls}$ = 96%, *SD* = 21%; $M_{V\text{-}VV}$ = 86%, *SD* = 35%; audio and visual input group: $M_{controls}$ = 97%, *SD* = 16%; $M_{V\text{-}VV}$ = 90%, *SD* = 30%). Caution should be exerted when making inferences in the absence of stimulus randomisation, as there may have been effects of unfathomable variables making V-VV pairs more difficult than control pairs. However, this small reduction in accuracy among native speakers cannot explain the huge differences in likelihood of a correct response among bilinguals (auditory input group: $M_{controls}$ = 94%, *SD* 24%; $M_{V\text{-}VV\,=}$ 72%, *SD* = 45%; audio and visual input group: $M_{controls}$ = 94%, *SD* 24%; $M_{V\text{-}VV\,=}$66%, *SD* = 47%). The negative effect of different spellings among bilinguals is best explained as an orthographic effect, in line with findings from production and metalinguistic awareness studies, that show that Italians produce and categorise a short-long contrast in English$_{L2}$ when the same vowel is spelled with a single vowel letter or with a digraph such as <ou> and <oo> (Bassetti et al., 2018, 2020).

Analysis of the response times confirmed that — as with consonant stimuli — native speakers respond faster if the two words of interest have been activated with pictures, compared with listening to spoken words only, presumably because activating the two specific words speeds up lexicon searches for potential homophones. Bilinguals were not facilitated by visual input but — unlike the consonant task — were faster with control than V-VV pairs regardless of whether the words of interest had been activated by images or not. This unexpected finding may be linked to L2 speakers' greater difficulty dealing with the spelling of vowels, as shown by smaller number of correct spellings and slower RTs than with consonants.

Summary

Experiment 1 showed orthographic effects in the perception of consonants and — less clearly — of vowels. In line with predictions, Italian$_{L1}$-English$_{L2}$ speakers perceived homophonic words as containing different consonants in words that are spelled with one or two consonant letters, but only after the activation of the two specific homophonic words. Orthographic effects on perception of vowels were less clear, due to other factors probably also affecting perception in native speakers and in bilinguals not exposed to visual input.

In light of findings that orthographic effects are resistant to naturalistic exposure, as they are reduced but do not disappear (Bassetti et al., 2018; Mairano et al., 2018) and may even occasionally be stronger in those with naturalistic exposure than without (Bassetti et al., 2020), the next question was whether orthographic effects on perception could also be modulated by naturalistic exposure. Experiment 2 tested this hypothesis by comparing instructed English$_{L2}$ language learners (high-school students) with Italian$_{L1}$- English$_{L2}$ sequential bilinguals who had learnt English$_{L2}$ in school and had been living in England for years. Since Experiment 1 established that orthographic effects on consonant perception were only found after lexical activation, to test whether orthographic effects are modulated by naturalistic exposure, all Experiment 2 participants performed the perception task with visual input. If naturalistic exposure reduces orthographic effects on perception, L2 learners should show lower levels of accuracy compared with sequential bilinguals.

Experiment 2

Method

Design

Participants took part in a consonant perception task and a vowel perception task with auditory plus visual input. The between-group factor was language background with three groups: Italian$_{L1}$ learners of English$_{L2}$, Italian$_{L1}$-English$_{L2}$ bilinguals, and English native speakers. Type of word pair was a within-group factor with two levels: target pair (C-CC or V-VV) and control pairs. The dependent variables were accuracy (correct/incorrect responses) and RTs. As in Experiment 1, following the main experimental tasks participants spelled all words used in the perception task to dictation.

Participants

There were 90 participants, 30 in each group. The Italian$_{L1}$ learners of English$_{L2}$ were Italian high-school learners of English, aged on average 16.83 years (*SD* = 0.38, range = 16-17). They were attending the fourth year of *liceo scientifico* ('scientific high-school') in one of eight classes in three different schools in Rome. In these schools English language and literature is a compulsory subject, taught for three hours a week using textbooks and related audio and video files in standard British English. On average they had been studying English as a school subject for 9.9 years (*SD* = 2.31), with a mean age of onset of acquisition of 5.87 years (*SD* = 2.15). Median proficiency in the group was B2 (upper intermediate), ranging from the lowest to the highest level (A1, beginner, to C2, proficient). The learners reported spending similar amounts of time listening to and reading English, with a mean self-reported reading:listening ratio of 1.02 (*SD* = 2.09). They considered a native-like pronunciation very important (median rating = 6 on a seven-point scale).

The bilinguals were native speakers of southern and central varieties of Italian, aged on average 36.15 years (*SD* = 8.34), who had studied English in Italy, had moved to England after the age of 18, and had been living there for an average of 7.1 years (*SD* = 7.22). Their mean age of onset of acquisition was higher than the learner group (*M* = 9.32, *SD* = 1.97, $t(53) = 6.16$, $p < 0.001$). They had studied English for an average of 8.56 years (*SD* = 2.90), and their median self-rated English language proficiency was C1 (advanced; five missing answers); over a third rated themselves as C2 (proficient).

According to their self-reported number of hours per week, they spent on average 25% more time reading than listening to English (mean reading:listening ratio = 2.31, *SD* = 4.30), and spent more time using English than Italian (mean English reading time out of total (Italian and English) reading time = 0.67, *SD* = 0.19; listening: 0.62(0.24); interacting: 0.52(0.19)).

The English native speakers had a mean age of 30.07 years (*SD* = 13.07). The majority (77%) reported being native speakers of Southern British English, five spoke other varieties of British English (Midlands: *n* = 4; Yorkshire = 2; Manchester = 1).

Sample size was calculated as in Experiment 1. All participants reported normal or corrected-to-normal vision, and no reading or listening difficulties; 7% were right-handed. Bilinguals and native speakers were recruited and tested mostly in London, with some in the Midlands; learners were recruited and tested in Rome, Italy. Participation was voluntary and rewarded (with book tokens for learners and cash for the others). The study received the approval of the University of Warwick Research Ethics Committee (Number: 118/14-15, Project Title: Effects of Orthography on Phonology in Second Language Speakers of English: Pronunciation, Phonological Awareness, Speech Perception and Spelling).

Materials, tasks and procedure

Materials and procedure were the same as in Experiment 1, except that all participants performed the perception task with auditory plus visual input, and learners took an English proficiency test (Oxford Placement Test, Allan, 1992). Additionally, as L2 learners may have been less familiar with some of the words of interest than bilinguals, a month before the beginning of data collection all classes involved in the project received a one-hour vocabulary learning session, administered to the whole class by their English language teachers during a normal English language class. Three short stories containing all target and control words were presented in print and as audio recordings, complete with colour photographs of some words and vocabulary-building exercises. This was presented as a vocabulary building session, and not linked explicitly to the forthcoming experimental session. Each student received a copy of the texts and exercises, then listened to a native speaker's recording of the story while reading the text, then performed tasks in pairs that involved pronouncing and spelling the words (gap-filling and picture naming). We aimed to mimic the way students normally come across words in Italian school settings as far as possible. This was achieved by the procedure already outlined, that is, by presenting words in the context of texts with both written text and audio recordings, and requiring both silent reading and active oral production.

Data analysis

Data preparation followed the same procedure as in Experiment 1. Thus, before analysing accuracy and RTs for target and control pairs we eliminated pairs where one or both words had been spelled incorrectly in the spelling to dictation task. For Experiment 2 consonant data word pairs removed due to spelling errors were 20.93% for the learner group, 15.35% for the bilingual group and 8.52% for the native speaker group. For the vowel data, 25.83% of word pairs were removed for the learner group, 17.31% for the bilingual group and 6.20% for the native speaker group[2]. These data were used to analyse accuracy.

As in Experiment 1, we further removed outliers before analysing RTs. For consonant data, we removed 2.11% of word pairs for the learner group, 2.19% for the bilingual group and 2.02% for the native speaker group. For vowel data, we removed 2.00% of word pairs for the learner group, 2.02% for the bilingual group and 2.17% for the native speaker group. The remaining RTs were log-transformed to approximate normal distribution.

Accuracy and RTs were analysed using two mixed models for consonant and vowel data and following the same model reduction procedures used for Experiment 1. For both accuracy and RTs analyses, the initial maximal models included the main effects and interactions between language background (learners, bilinguals, native speakers), type of word pair (C-CC or V-VV targets, C or V controls) and their interactions. As random effects, we considered random intercepts for participants and word pairs, and a random slope for the effect of type of pair on participants.

Results

Effects of language background and consonant spelling on the perception of consonant length

Accuracy. Figure 6 shows mean percentage of correct responses in the consonant perception task for control and C-CC word pairs in learners, bilinguals and native speakers.

[2] For completeness, we report spelling error rates from the Spelling-to-dictation task for each group of participants. Consonant data – Native speakers: 4% (*SD* = 19%); Bilinguals: 8% (*SD* = 27%); Learners: 11% (*SD* = 31%). Vowel data: Native speakers: 3% (*SD* = 18%); Bilinguals: 11% (*SD* = 30%); Learners: 15% (*SD* = 36%).

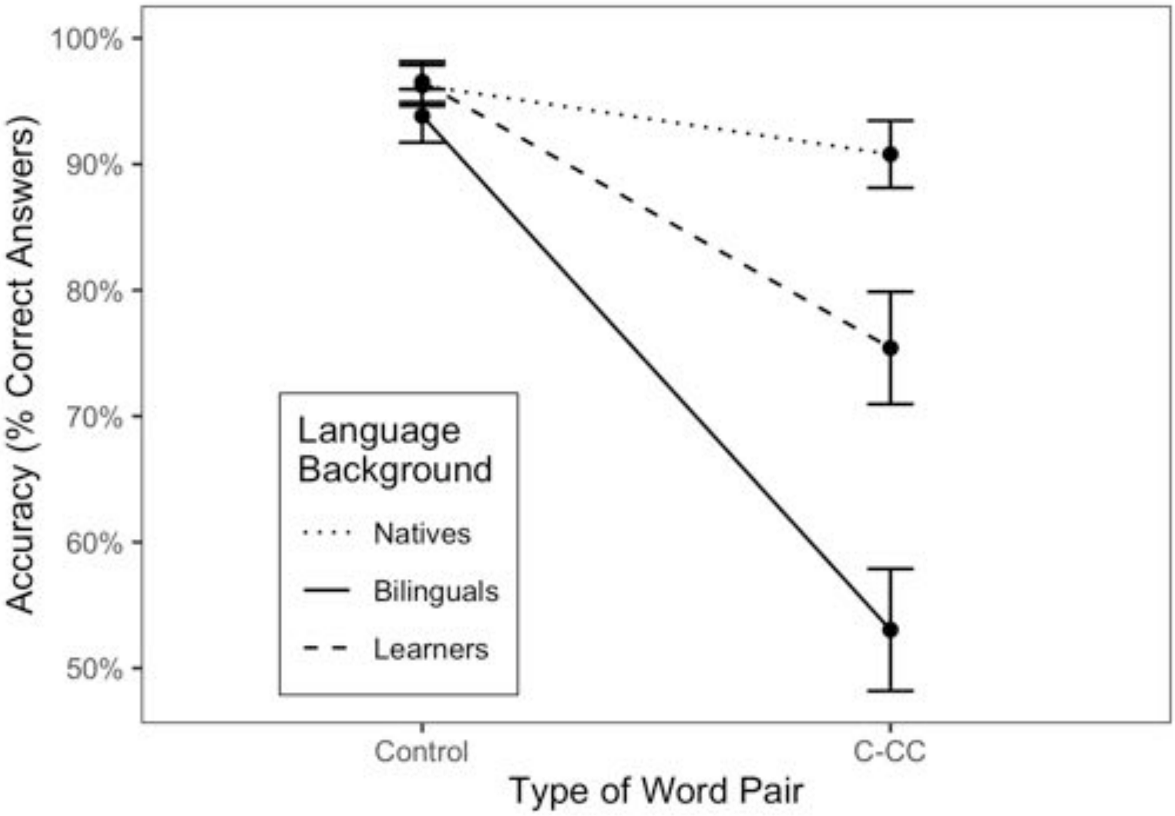

Figure 6. *Mean percentage of correct responses (bars represent 95%* CI*s) in the Consonant Perception Task in Experiment 2 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, C-CC).*

The final model (see Online Supplementary Materials, Table 5) included as fixed effects language background, type of pair and their interaction. The random structure included the by-participant random slope for type of pair and the intercepts for participants and word pairs. The main effects of language background ($\chi^2 = 33.31$, $p < 0.001$) and type of word pair ($\chi^2 = 62.02$, $p < 0.001$) were both significant, and qualified by the interaction ($\chi^2 = 16.33$, $p < 0.001$). Post-hoc contrasts revealed that L2 speakers were more likely to respond to control pairs correctly, compared with C-CC pairs ($z_{bilingual} = 8.23$, $z_{learners} = 5.84$, both $p < 0.001$; there were no differences in native speakers, $z = 2.53$, $p = 0.117$). Contrary to predictions, learners were more likely to respond to C-CC pairs correctly than bilinguals ($z = -3.35$, $p = 0.011$).

Response times. Figure 7 shows the geometric mean of RTs for the three language background groups for control and C-CC pairs.

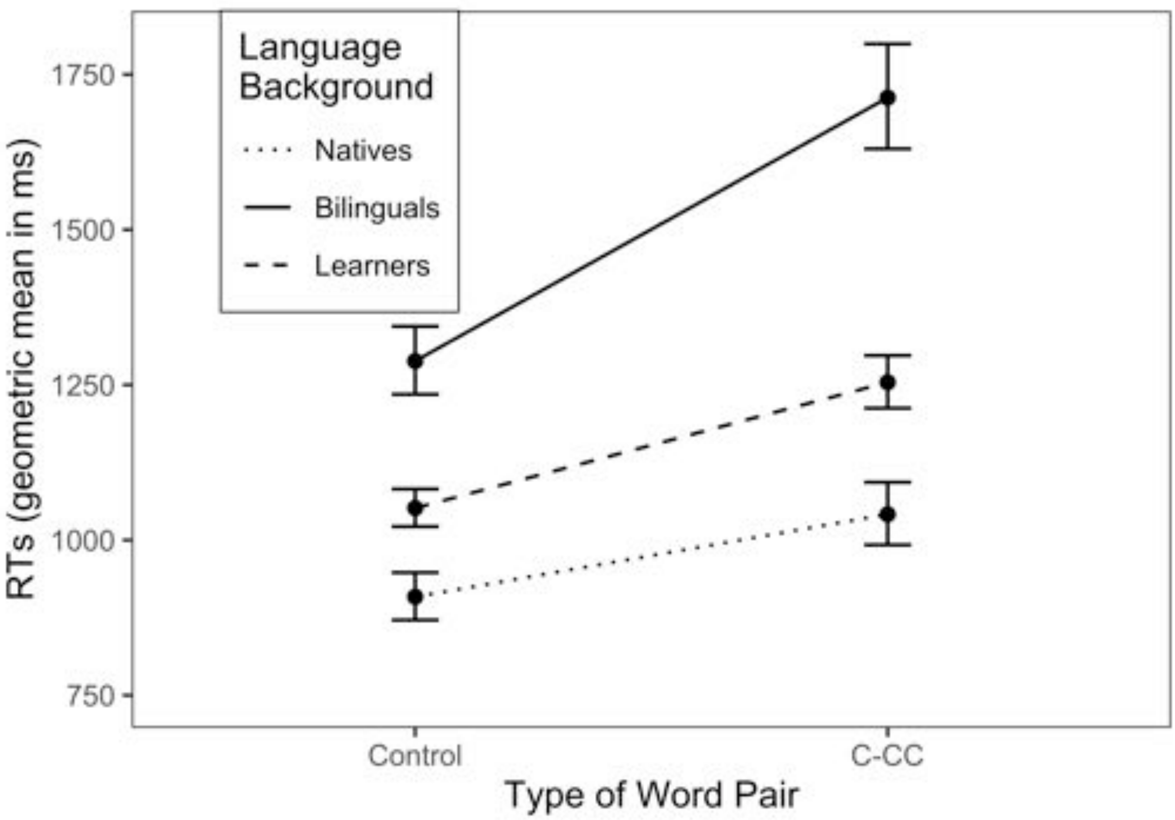

Figure 7. *Geometric mean of RTs (ms; bars represent 95%* CI*s) in the Consonant Perception Task in Experiment 2 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, C-CCs).*

The final model (see Online Supplementary Materials, Table 6) included the same fixed and random effects as the accuracy model. The effects of language background (χ^2 = 35.15, p < 0.001and type of pair (χ^2 = 19.30, p < 0.001) were qualified by the interaction (χ^2 = 9.43, p = 0.009). Post-hoc contrasts with Tukey adjustments revealed that all groups responded to control pairs faster than C-CC pairs, and bilinguals were the most affected ($t_{natives}$ = -2.796, p = 0.007; $t_{bilinguals}$ = -5.329, p < 0.001, $t_{learners}$ = -3.167, p = 0.002). Bilinguals were slower than native speakers and learners with both controls ($t_{natives\text{-}bilinguals}$ = -4.972, p < 0.001; $z_{bilinguals\text{-}learners}$ = 2.849, p = 0.015) and C-CC pairs ($t_{natives\text{-}bilinguals}$ = -6.545, p < 0.001; $t_{bilinguals\text{-}learners}$ = 4.195, p < 0.001), whereas RTs of learners did not differ from those of native speakers for either control pairs (t = -2.116, p = 0.093) or C-CC pairs (t = -2.308, p = 0.060).

Additional analysis of legality of position and cognate status. As in Experiment 1, we examined whether perception of English consonant length was affected by the legality of its position (in Italian words), as well as cognate status. The analyses, employing the same stimuli, revealed no significant effect of either variable.

Effects of language background and vowel spelling on the perception of vowel length

Accuracy. Figure 8 shows the mean percentage of correct responses in the vowel perception task for the language background groups for control and V-VV pairs.

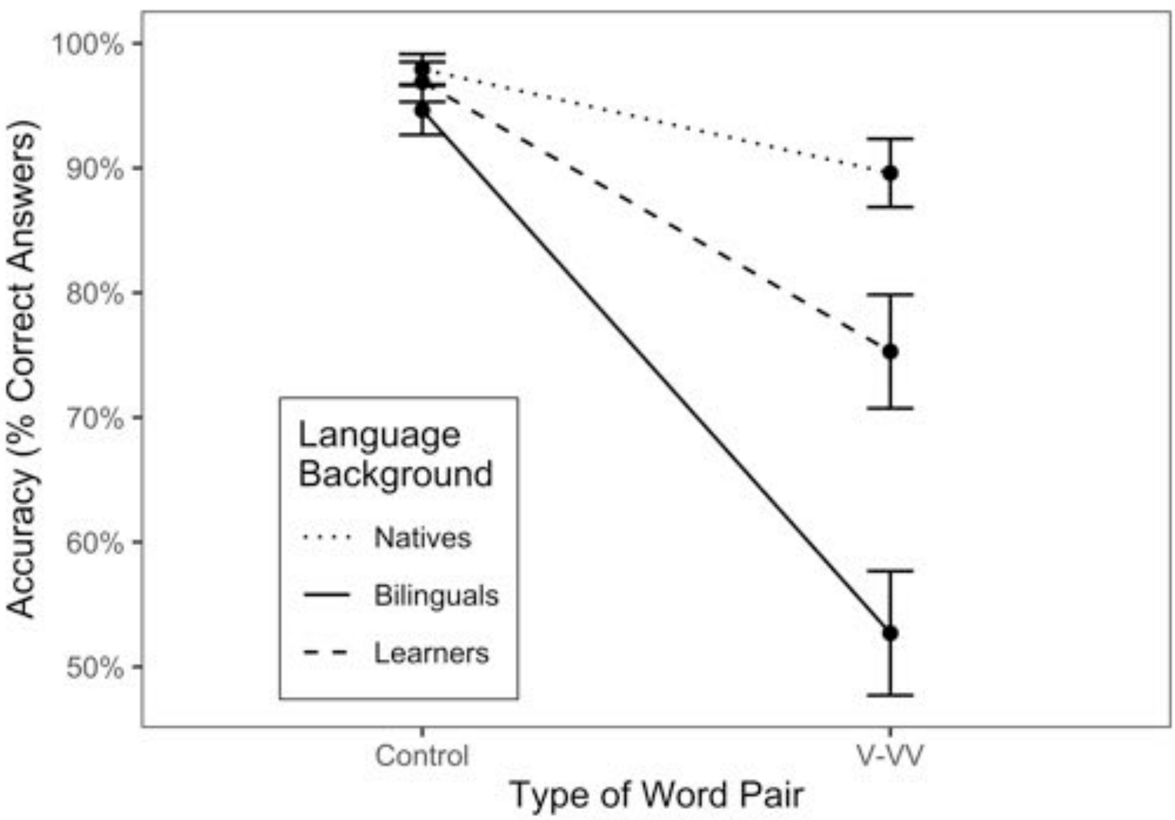

Figure 8. *Mean percentage of correct responses (bars represent 95%* CIs*) in the Vowel Perception Task in Experiment 2 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, V-VV).*

The final model (see Online Supplementary Materials, Table 7) included as fixed effects language background, type of pair and their interaction. The intercepts for participants and word pairs and the by-participant random slope for type of pair were included as random structure. The effect of language background was significant ($\chi^2 = 33.95$, $p < 0.001$) with native speakers most likely to respond correctly and bilinguals least likely. The effect of type of word pair was significant ($\chi^2 = 46.89$, $p < 0.001$) with responses to control pairs more likely to be accurate than those to V-VV pairs. The interaction was also significant ($\chi^2 = 15.58$, $p < 0.001$). All groups were more likely to give correct responses to control than V-VV word pairs, although the effect was strongest in bilinguals and weakest in native speakers ($z_{natives} = 2.95$, $p = 0.003$; $z_{bilinguals} = 7.61$, $p < 0.001$; $z_{\text{learners}} = 5.23$, $p < 0.001$). With control word pairs there were no group differences. As predicted, with V-VV pairs both L2 groups were less likely to respond correctly than native speakers ($z_{\text{native-bilingual}} = 6.97$, $p < 0.001$, $z_{native\text{-}learner} = 3.76$, $p = 0.005$), although contrary to predictions the effect was stronger in bilinguals than in learners ($z = -3.52$, $p = 0.001$).

Response time. Figure 9 summarizes the geometric mean of RTs for the language background groups for control and V-VV pairs.

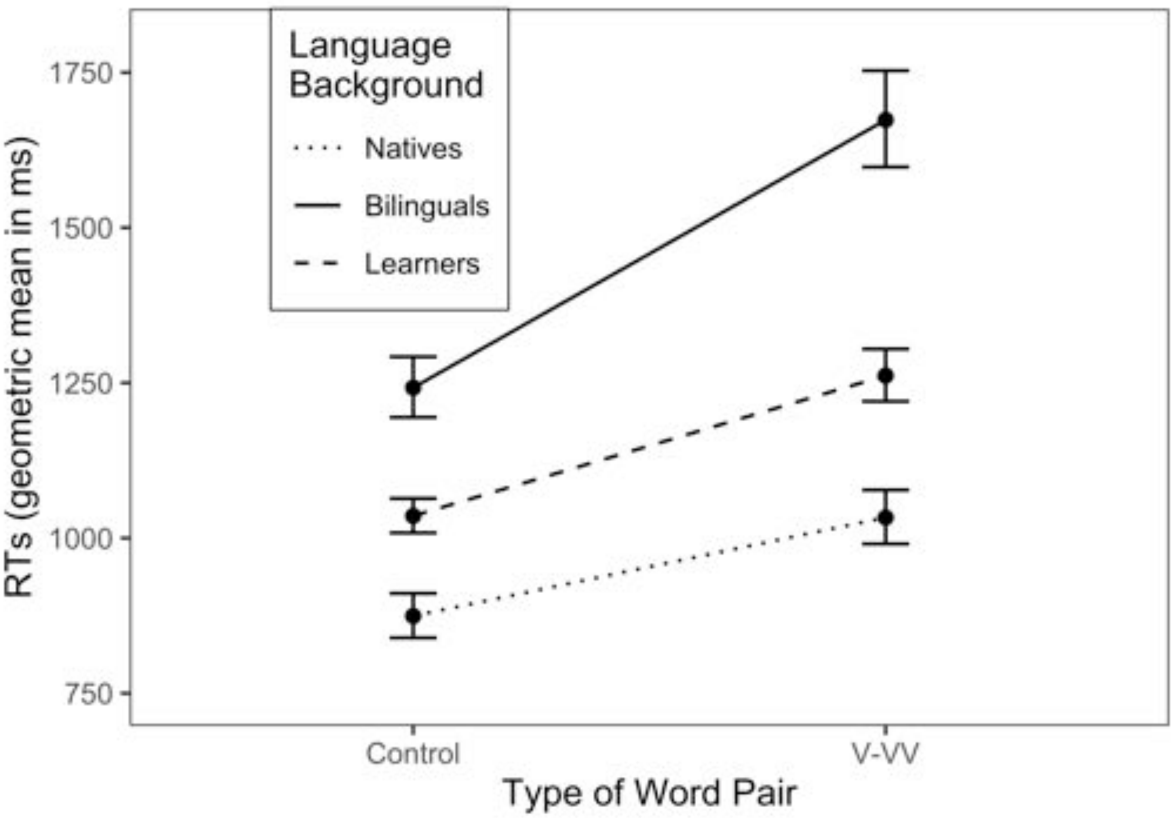

Figure 9. *Geometric mean of RTs (ms; bars represent 95% CIs) in the Vowel Perception Task in Experiment 2 by language background (native speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, V-VV).*

The final model (see Online Supplementary Materials, Table 8) included the same fixed and random structure as the accuracy model. The effect of language background was significant ($\chi^2 = 44.46$, $p < 0.001$) with responses of native speakers fastest and those of bilinguals slowest. The effect of type of word pair was significant ($\chi^2 = 53.76$, $p < 0.001$) with faster RTs for controls than V-VV pairs. The interaction was also significant ($\chi^2 = 9.04$, $p = 0.011$). Post-hoc contrasts revealed that all groups responded to control pairs faster than V-VV pairs, but bilinguals were most affected ($t_{natives} = 4.61$, $p < 0.001$; $t_{bilinguals} = 7.53$, $p < 0.001$; $t_{learners} = 4.72$, $p < 0.001$). Bilinguals were slower than both native speakers and learners, particularly with V-VV pairs (bilinguals vs. native speakers: $t_{controls} = -5.09$, $p < 0.001$; $t_{V\text{-}VV\ pairs} = -7.128$, $p < 0.001$; bilinguals vs. learners: $t_{controls} = 2.56$, $p = 0.032$; $t_{V\text{-}VV} = 4.25$, $p < 0.001$). Learners were overall slower than native speakers ($t_{controls} = -2.51$, $p = 0.037$; $t_{V\text{-}VV\ pairs} = -2.82$, $p = 0.016$).

Additional analysis of type of digraph and cognate status. Finally, as in Experiment 1, we found no differences in L2 speakers' accuracy between pairs spelled with double vowel letters and those spelled with two different vowel letters (Bilinguals: $M_{V\text{-}VV} = 54\%$, $M_{V\text{-}V1V2} = 52\%$; Learners: $M_{V\text{-}VV} = 76\%$, $M_{V\text{-}V1V2} = 75\%$). There were also no differences between *genes-jeans* and *scene-seen* (Bilinguals: $M_{scene\text{-}seen} = 32\%$, $M_{genes\text{-}jeans} = 37\%$; Learners: $M_{scene\text{-}seen} = 85\%$, $M_{genes\text{-}jeans} = 71\%$).

Discussion

Experiment 2 aimed to test whether naturalistic exposure may modulate orthographic effects on the perception of sounds in Italian$_{L1}$ listeners of English$_{L2}$ by comparing the accuracy likelihoods of sequential bilinguals and language learners. Results revealed no advantage of naturalistic exposure, and in fact the orthographic effect was stronger among sequential bilinguals than instructed learners, as discussed below.

Looking at consonants, in line with Experiment 1, orthography affected consonant perception in L2 listeners, whose likelihood of correctly considering homophones as containing the same sounds was lower compared with their likelihood of correct response in control pairs with no consonant spelling differences. However, contrary to predictions, bilinguals were more affected than learners, with lower likelihood of correct response, and a mean percentage of correct responses just above chance level. RTs were also in line with Experiment 1, as all listeners responded faster to controls than C-CC pairs, with a small decrease among native listeners probably caused by interference between phonological similarity and orthographic differences. Once more, against our predictions, bilinguals were more negatively affected by spelling than learners.

Looking at vowels, in line with Experiment 1, orthography affected vowel perception in all participants, as the likelihood of correct responses was lower with V-VV than control pairs, and L2 listeners were affected more than native listeners. Contrary to predictions, but in line with the results for the consonant perception task, bilinguals were more affected than learners, with lower likelihood of correct responses, and a mean percentage of correct responses just above chance level. RTs were also in line with Experiment 1, as all listeners responded faster to control than V-VV pairs, however — as was the case with consonants — bilinguals were more affected by spelling than learners.

Results showed stronger orthographic effects on sequential bilinguals with lengthy naturalistic exposure than on instructed learners with no or minimal naturalistic exposure. This runs contrary to the prediction that naturalistic exposure may reduce orthographic effects on L2 phonology, which was based on Bassetti et al.'s (2020) findings of stronger effects of double letters on consonant production and categorisation in learners than bilinguals. Various explanations are possible, for instance learners may pay more attention to phonetic detail, compared with more experienced bilinguals. It should be noticed that Experiment 2 participants in this study are a subgroup of participants who took part in Bassetti et al. (2020). The most likely explanation is then that learners in the present study were not representative of the general population of Italian high-school learners of English, as they had higher than average levels of proficiency, phonological awareness, and exposure to spoken input, as follows. Looking at proficiency, the criterion for inclusion in the present study was the ability to spell a large number of infrequent words, and only about a third of learners in Bassetti et al. (2020) qualified for inclusion in the present study. Indeed, further analyses revealed that the subgroup of 30 learners who met the criteria for inclusion in the present study had higher proficiency levels than the 61 peers who were not included, scoring on average 32.28 versus 18.97 on the Oxford Placement Test ($p < 0.001$), including stronger performance in grammar ($p < 0.001$), listening ($p = 0.004$), and phonological awareness ($p = 0.016$), but not better awareness that English digraphs do not represent long sounds ($p = 0.258$). Looking at amounts of written and spoken input, this subgroup reported spending a higher proportion of time listening to rather than reading English (reading:listening ratio: $p = 0.016$). Looking at orthographic effects on speech production, the subgroup did not differ from their peers (CC:C ratio: $p = 0.351$; VV:V ratio $p = 0.993$). It appears that learners with sufficient proficiency to know some of the relatively infrequent words used in the present experiment are above average in terms of proficiency and amounts of spoken input, yet do not differ from average in terms of orthographic effects on production and awareness.

On the other hand, bilinguals would instead be representative of the general population of Italian-English sequential bilinguals with lengthy naturalistic exposure, because knowing these words would not be unusual for people living in an English-speaking country as it would be for instructed learners. Indeed, bilinguals in this study's sample did not differ from the larger sample in Bassetti et al. (2020; phonological awareness: $p = 0.898$; reading:listening ratio: $p = 0.130$; CC:C ratio: $p = 0.268$; VV:V ratio: $p = 0.805$). Although learners had started learning English earlier than bilinguals, this may not be a crucial factor, because they had been studying English with Italian teachers, and indeed Bassetti and colleagues (2020) found stronger orthographic effects in learners than in bilinguals in spite of the former group's earlier age of onset of acquisition. With regards to vowels, the evidence more consistently shows that naturalistic exposure does not reduce orthographic effects on vowel perception and, on the contrary, effects may be stronger in bilinguals than in learners. This is in line with previous studies that found no differences (Bassetti et al., 2018), or slightly stronger effects in bilinguals than learners (Bassetti et al., 2020) in vowel production, and no differences in categorisation (Bassetti et al., 2020). While it is possible that the learners were unrepresentative of the general population of Italian instructed L2 learners, it is also possible that the effect is genuine, and learners are indeed less affected by orthography than sequential bilinguals, in spite of being in some cases more affected in production and awareness. This is because there are intrinsic differences between learners and sequential bilinguals that may explain these findings: learners may rely on phonetic detail in the auditory input – such as duration – more than the more advanced sequential bilinguals, as proposed by PAM-L2 (Best & Tyler, 2007) with regards to naïve compared with L2 listeners. Also, learners may rely less on mental representations of words compared with bilinguals, because learners' L2 speech processing is less automatised, as proposed by Strange's (2011) Automatic Selective Perception (ASP) model, and/or because they have weaker mental representations of words (including both phonological and orthographic forms). In conclusion, more research should ascertain whether L2 orthographic input affects perception in sequential bilinguals less than in learners, and if so why.

Finally, analyses of word-level variables showed that orthographic effects on consonant perception are not stronger in cognates than in non-cognates, or indeed in consonants in positions that are legal or illegal in the Italian phonology, confirming that these are orthographic effects rather than effects of the L1 phonological system or the phonological form of L1 cognates.

In conclusion, results clearly show that naturalistic exposure does not eliminate orthographic effects on L2 sound perception, confirming and extending findings from studies of production and awareness. It is also unlikely that naturalistic exposure reduces orthographic effects, or at least that it reduces them in a way that is not confounded by other, stronger factors. Given the inconsistent results, and in particular the fact that the learner sample in the present experiment was not representative of the population of instructed learners, the modulating effect of naturalistic exposure on orthographic effects on L2 phonology requires further investigation.

General Discussion

This study aimed to test the effects of L2 sound spelling (single letter or digraph) on L2 speech perception in L2 listeners whose L1 uses single and double letters to distinguish short and long consonants. Results can be summarised as follows. Looking at consonants, $Italian_{L1}$-$English_{L2}$ listeners perceived different consonants in homophonic words spelled with single or double consonant letters, as shown by their likelihood of correctly identifying as homophones the two words in an orthographic minimal pair, that is to say two homophones that spell one sound differently. The effects were found in those bilinguals who had the two words of interest activated by means of two pictures, but not in those who only heard the two words. This shows that the effect was due to top-down interference from the mental representation of the two words — one containing a short sound and one containing a long sound — on the perception of the two words' spoken forms. Similar effects were found both in L2 speakers with lengthy experience of naturalistic immersion and in instructed L2 learners with very limited exposure to native input, showing that naturalistic experience does not eliminate orthographic effects on L2 perception. The RT results were less clear, as both native and L2 listeners were slower with C-CC pairs than control pairs. It is possible that the incongruence between the similar phonological forms and the differences in the orthographic forms slows down the decision process in all listeners, although it is only L2 listeners who actually perceive different sounds. Looking at vowels, V-VV pairs were less likely to be considered homophones than control pairs across both L1 and L2 listeners, but the effect was stronger among L2 than L1 listeners, and especially among bilinguals. It is possible that vowels are more susceptible to misperception than consonants, or perhaps vowel spelling may affect vowel perception in native speakers, due to the inconsistent phoneme-grapheme correspondences of vowels. Finally, the effects found in this study cannot be ruled out as effects of the L1 phonological system or the phonological form of L1 cognates, because there were no stronger effects in cognate words that contain a geminate in the L1, or in words where gemination is legal in the L1.

In conclusion, with consonants and to a lesser extent with vowels, Italians perceive an illusory difference in duration because of orthography, but in the case of consonants the difference corresponds to two L1 phonological categories (geminates and singletons), whereas with vowels it corresponds to two values (long and short) of an L2 feature (length, which is one of the correlates of the lax-tense distinction in English). This is discussed in more detail below.

Overall, it appears that $Italian_{L1}$ listeners of $English_{L2}$ perceive different consonants in words with different consonant spellings, namely a singleton or geminate consonant in correspondence with single letter or double letters spellings. These results extend to perception previous results from studies of speech production, awareness, and word learning, and confirm that the interaction between second language orthographic forms and first language grapheme-phoneme correspondences leads L2 speakers to establish a long-short phonological contrast that is not attested in the English language (Bassetti, 2017; Bassetti et al., 2018, 2020; Cerni et al., 2019; Mairano et al., 2018; Sokolović-Perović et al., 2020). Italians can perceive differences in closure duration as small as 17ms (Esposito & Di Benedetto, 1999), but they have difficulty perceiving the lack of difference in duration.

This orthographic effect appears to be a top-down interference caused by the incongruence between the (short) consonant in the spoken input and the geminate in the mental representation of words spelled with double consonant letters. The effect could be compounded by the processing of an orthographic form co-activated in the process of spoken word recognition.

Orthography then adds an additional source of perceptual illusion in L2 speech perception. There is much evidence that L1 phonology introduces perceptual illusions resulting in sound omission, addition, and replacement. Indeed, if a L2 consonant cluster is illegal in L1, L2 listeners perceive a different sound, for instance perceiving /tl/ as /kl/, and if a sound sequence is not allowed in their L1 without a vowel, L2 learners perceive an illusory vowel that is not present in the input, such that for instance Japanese$_{L1}$ but not French$_{L1}$ listeners of English$_{L2}$ report perceiving the vowel [ɯ] (Matthew & Brown, 2004; see also Darcy & Thomas, 2019, for epenthesis in Korean$_{L1}$ listeners of English$_{L2}$). As Matthews and Brown (2004, p. 5) stated with reference to L1 phonology effects, in orthographic effects on perception as well "a learner's percept of the L2 input, or L2 intake, actually includes more segmental material than is physically present".

Results support Bassetti's (2017) proposal that English native speakers' consonant duration appears to Italian$_{L1}$-English$_{L2}$ speakers as in-between the durations of singleton and geminate consonants. Therefore, when they hear for instance /ˈfɪnɪʃ/ in isolation, they can perceive it as either *finish* or *Finnish*. However, if they know from context which word they heard, they activate a phonological representation that contains a singleton or a geminate consonant. If one believes that perception is the basis for production, then these results in conjunction with previous results from metalinguistic awareness tasks could explain why Italians produce short and long consonants in English$_{L2}$, and learn new English words as containing geminates. Bassetti (2007, 2008) argued that L2 phonological representations are formed from both L2 auditory input, partly reinterpreted according to L1 phonology, and L2 orthographic input, partly reinterpreted according to L1 orthography-phonology correspondences. Such orthography-influenced phonological representations in turn result in non-targetlike L2 pronunciation and sound categorisation. Various articles by Bassetti and colleagues provided evidence that supports this interpretation. Furthermore, such results could explain why many English loanwords that are spelled with double consonant letters have entered Italian as containing a geminate (Hamann & Colombo, 2017), as illusory perception could reinforce written input.

In the study of L2 phonological development, it may be necessary to reconsider the difficulty of unlearning an L1 contrast that is not attested in the target language. While it is generally recognised that geminates are hard to perceive for L2 listeners whose L1 does not have consonant length contrasts (De Clercq et al., 2014; Hayes-Harb, 2005; Sorianello, 2014; Tsukada et al., 2018), it appears that non-geminates are also hard to perceive for L2 listeners whose L1 has length contrasts in the presence of clues pointing to gemination (spelling). Therefore, models of L2 phonological development should address not only situations where two phonological categories in the L2 correspond to one category in the L1 (for instance, see the PAM-2 model, Best & Tyler, 2007), but also the opposite situation where the L1 has two categories corresponding to one category in the L2. Finally, it is worth investigating whether a consonant length contrast may be a feature of the Italian variety of English, an original creation at the interface between the orthographies and phonologies of their two languages that is shared by Italian speakers of English.

In conclusion, results confirmed the prediction that L2 orthographic forms affect L2 speech perception, resulting in the perception of the same spoken sound as two different sounds, in the case of consonants corresponding to different phonological categories. This confirms previous evidence of orthographic effects on L2 speech production and sound categorisation, and show that the same effects are found in perception.

Limitations and future research

The study suffers from a number of limitations. Looking at participants, our study investigated perception in highly proficient L2 learners with an unusual vocabulary knowledge, as only about a third of learners who were originally tested could pass the stringent vocabulary knowledge criterion of being able to spell more than half of the target words, hence the study's L2 learners may not be representative of the general population of Italian high-school students of English. It is possible that testing those with lower levels of English may show stronger orthographic effects on perception, in line with results from production studies (Bassetti et al., 2020) which found that proficiency is negatively correlated with orthographic effects on production. Hence the actual orthographic effects on L2 perception may be stronger than those reported above. Looking at the task, it involved an explicit decision component, such that the use of orthography as a task-solving strategy cannot be ruled out. Looking at the materials, some of the words were unusual and it is likely that learners would not have heard them often or not at all, although they needed to know the word's spelling for the word to be entered in analysis, they may only have encountered the word in writing. It may even have been the case that, for some words, the learners' sole exposure was in the pre-experimental vocabulary learning session in which the words were introduced, thus constituting a tenuous basis for a lexical representation. It will be informative in future work to investigate word familiarity in relation to orthographic effects on L2 learners' word perception. Finally, the study could not disentangle the effects of naturalistic exposure from proficiency and length of study. It remains to be seen whether explicit instruction, rather than naturalistic exposure, can eliminate or reduce this orthographic effect.

Conclusion

The study provides evidence that second language written forms, recoded according to L1 orthography-phonology correspondence rules, affect the perception of second language speech. This results in perceptual illusions whereby listeners report hearing different sounds in recordings of homophonic word pairs. This study then confirms previous findings of orthographic influences on production, metalinguistic awareness, and word learning in a second language.

The results have implications for debates about the relationship between L2 perception and production. Views differ substantially across models of L2 phonology development, with some arguing for a close relationship between perception and production (Best & Tyler's PAM-L2, 2007; Flege's Speech Learning Model, 1995) and others denying this relationship. This study supports the view that perception and production are closely linked, by demonstrating the same orthographic effects in perception as were already demonstrated in production and awareness and word learning.

Finally, the study introduces a novel method that has the potential to be used widely in the field. We provided a task that uses real words in order to assess the perception of sound differences not attested in the language under analysis. The task eliminates the need for artificial manipulation of experimental stimuli, as it involves the presentation of word pairs that L2 listeners should perceive as different, and it avoids the limitation of techniques such as phoneme segmentation or phoneme counting, which may encourage the use of orthographic representations for task performance. By using pictures to activate target words, the need to provide orthographic input is overcome. The task can be used in other situations where L2 listeners' perception of words in speakers of a real language should be tested for a sound contrast that is not contrastive to native speakers.

References

Allan, D. (1992). *Oxford Placement Test*. Oxford University Press.

Altmann, H., Berger, I., & Braun, B. (2012). Asymmetries in the perception of non-native consonantal and vocalic length contrasts. *Second Language Research*, *28*(4), 387-413. https://doi.org/10.1177/0267658312456544

Arvaniti, A., & Tserdanelis, G. (2000). *On the phonetics of geminates: evidence from Cypriot Greek*. Sixth International Conference on Spoken Language Processing,

Asano, Y. (2018). Discriminating non-native segmental length contrasts under increased task demands. *Language and Speech*, *61*(3), 409-429. https://doi.org/10.1177/0023830917731907

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics*, 231-250. https://doi.org/10.1515/IRAL.2006.010

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a Foreign Language. *Written Language and Literacy*, *9*(1), 95-114. https://doi.org/10.1075/wll.9.1.07bas

Bassetti, B. (2007). Effects of hanyu pinyin on pronunciation in learners of Chinese as a Foreign Language. In A. Guder, X. Jiang, & Y. Wan (Eds.), *The cognition, learning and teaching of Chinese characters* (pp. 155-179). Beijing Language and Culture University Press.

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191-206). Multilingual Matters. https://doi.org/10.21832/9781847691118-013

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1835-1842. https://doi.org/10.1037/xlm0000417

Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners *Applied Psycholinguistics*, *36(2)*, 67-91. https://doi.org/10.1017/s0142716414000435

Bassetti, B., Hayes-Harb, R., & Escudero, P. (2015). Second language phonology at the interface between acoustic and orthographic input. *Applied Psycholinguistics*, *36*(1), 1-6. https://doi.org/10.1017/s0142716414000393

Bassetti, B., Masterson, J., Cerni, T., & Mairano, P. (2021). Materials for "Orthographic forms affect speech perception in a second language: Consonant and vowel length in L2 English". Retrieved from osf.io/56ba4

Bassetti, B., Mairano, P., Masterson, J., & Cerni, T. (2020). Effects of orthographic forms on L2 speech production and phonological awareness, with consideration of speaker-level predictors. *Language Learning 70(4), 1218-1256.* DOI: 10.1111/lang.12423.

Bassetti, B., Sokolović-Perović, M., Mairano, P., & 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. https://doi.org/10.1177/0023830918780141

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. In arXiv preprint arXiv:1506.04967.

Bertinetto, P. M. (1981). *Strutture prosodiche dell'Italiano*. Accademia della Crusca.

Bertinetto, P. M., & Loporcaro, M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association*, *35*(2), 131-151. https://doi.org/10.1017/s0025100305002148

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second-language speech learning: In honor of James Emil Flege* (pp. 13-34). John Benjamins.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. earch. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language res* (pp. 279-304). York Press.

Carney, E. (1994). *A survey of English spelling*. Routledge.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372-387. https://doi.org/10.1016/j.wocn.2005.08.003

Cerni, T., Bassetti, B., & Masterson, J. (2019). Effects of orthographic forms on the acquisition of novel spoken words in a second language [Original Research]. *Frontiers in Communication*, *4*(31). https://doi.org/10.3389/fcomm.2019.00031

Clark, J., & Yallop, C. (1995). *An introduction to phonetics and phonology*. Blackwell Publishing.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257-271. https://doi.org/10.3758/bf03204507

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech*. Cambridge University Press. https://doi.org/10.1017/cbo9781139087636

Darcy, I., & Thomas, T. (2019). When *blue* is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism: Language and Cognition, 22*, 1141–1159. doi: 10.1017/S1366728918001050

De Clercq, B., Simon, E., & Crocco, C. (2014). Rosa versus rossa: The acquisition of Italian geminates by native speakers of Dutch. *Phrasis: Studies in Language and Literature*, *2*, 3-29.

Detey, S., & Nespoulous, J.-L. (2008). Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua*, *118*(1), 66-81. https://doi.org/10.1016/j.lingua.2007.04.003

Dias, J. W., Cook, T. C., & Rosenblum, L. D. (2017). The McGurk effect and the primacy of multisensory perception. In A. G. Shapiro & D. Todorović (Eds.), *The Oxford compendium of visual illusions* (p. 787–792). Oxford University Press.

Donnenwerth-Nolan, S., Tanenhaus, M. K., & Seidenberg, M. S. (1981). Multiple code activation in word recognition: Evidence from rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 170-180. https://doi.org/10.1037/0278-7393.7.3.170

Eckman, F. R., Elreyes, A., & Iverson, G. K. (2003). Some principles of second language phonology. *Second Language Research*, *19*(3), 169-208. https://doi.org/10.1191/0267658303sr2190a

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*(2), 191-228. https://doi.org/10.1111/j.0023-8333.2005.00303.x

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(04), 551-585. https://doi.org/10.1017/S0272263104040021

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

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, *17*, 384-395. https://doi.org/10.1017/s1366728913000436

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, *53*(3), 343-365. https://doi.org/10.1177/0023830910371447

Esposito, A., & Di Benedetto, M. G. (1999). Acoustical and perceptual study of gemination in Italian stops. *Journal of the Acoustical Society of America*, *106*(4), 2051-2062. https://doi.org/10.1121/1.428056

Faul, F., et al. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160.

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-277). York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437-470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125-3134. https://doi.org/10.1121/1.413041

Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (Third ed.). Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84, 474-496.

Giordano, R., & Savy, R. (2012). Sulla standardizzazione del consonantismo dell'italiano: Consonanti geminate, rafforzate e fricative alveolari in contesto intervocalico. In P. Bianchi, N. De Blasi, C. De Caprio, & F. Montuori (Eds.), *La variazione nell'italiano e nella sua storia. Varietà e varianti linguistiche e testuali [Variation in Italian and its history: Linguistic and textual varieties and variations]* (pp. 431-445). Franco Cesati Editore.

Guion, S. G. (1995). Word frequency effects among homonyms. *Texas Linguistic Forum*, *35*, 103-116.

Hamann, S., & Colombo, I. E. (2017). A formal account of the interaction of orthography and perception. *Natural Language & Linguistic Theory*, *35*(3), 683-714. https://doi.org/10.1007/s11049-017-9362-3

Hardison, D. M., & Motohashi Saigo, M. (2010). Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics*, *31*(1), 81-99. https://doi.org/10.1017/S0142716409990178

Hayes, R. L. (2001). The perception of novel phoneme contrasts in a second language: A developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. *Coyote Papers: Working Papers in Linguistics*, *12*, 28-41.

Hayes-Harb, R. (2005). Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language and Linguistics*, *4*(1), 1-29.

Hayes-Harb, R., Brown, K., & Smith, B. L. (2018). Orthographic input and the acquisition of German final devoicing by native speakers of English. *Language and Speech*, *61*(4), 547-564. https://doi.org/10.1177/0023830917710048

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*(3), 367-381. https://doi.org/10.1177/0023830910371460

Idemaru, K., & Guion-Anderson, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica*, *67*(1-2), 25-46. https://doi.org/10.1159/000319377

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformations or not) and towards Logit Mixed Models. *Journal of Memory and Language*, *59*, 434-446. https://doi.org/10.1016/j.jml.2007.11.007

Khattab, G., & Al-Tamimi, J. (2014). Geminate timing in Lebanese Arabic: The relationship between phonetic timing and phonological structure. *Laboratory Phonology*, *5*(2), 231-269. https://doi.org/10.1515/lp-2014-0009

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, *124*(6), 3959-3971. https://doi.org/10.1121/1.2999341

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest: Tests in linear mixed effects models*. In https://CRAN.R-project.org/package=lmerTest

Lahiri, A., & Hankamer, J. (1988). The timing of geminate consonants. *Journal of Phonetics*, *16*, 327-338. https://doi.org/10.1016/S0095-4470(19)30506-6

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1-33. https://doi.org/10.18637/jss.v069.i01

Mairano, P., Bassetti, B., Soković-Perović, M., & Cerni, T. (2018). Effects of L1 orthography and L1 phonology on L2 English pronunciation. *Revue Française de Linguistique Appliquée*, *23*(1), 45-57. https://doi.org/10.3917/rfla.231.0045

Mairano, P., & De Iacovo, V. (2020). Gemination in northern versus central and southern varieties of Italian: A corpus-based investigation. *Language and Speech*, 63, 608-634.

Massaro, D. W., Cohen, M. M., & Thompson, L. A. (1988). Visible language in speech perception: Lipreading and reading. *Visible Language*, 22(1), 8.

Mathieu, L. (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, *32*(2), 145-170. https://doi.org/10.1177/0267658315601882

Matthews, J., & Brown, C. A. (2004). When intake exceeds input: Language specific perceptual illusions induced by L1 prosodic constraints. *International Journal of Bilingualism*, *8*(1), 5-27. https://doi.org/10.1177/13670069040080010201

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748. https://doi.org/10.1038/264746a0

Nova Development. (2004). *Art Explosion library*. In Nova Development Co.

Oh, G. E., & Redford, M. A. (2012). The production and phonetic representation of fake geminates in English. *Journal of Phonetics*, *40*(1), 82-91. https://doi.org/10.1016/j.wocn.2011.08.003

Pickett, E. R., Blumstein, S. E., & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica*, *56*(3-4), 135-157. https://doi.org/10.1159/000028448

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, *59*, 49-71.

Pring, L. (1986). Orthographic effects of braille and print in the auditory modality. *Journal of Visual Impairment and Blindness*, *80*(10), 993-998. https://doi.org/10.1159/000056205

Pytlyk, C. (2017). Are orthographic effects language specific? The influence of second language orthography on second language phoneme awareness. *Applied Psycholinguistics*, *38*(2), 233-262. https://doi.org/10.1017/S0142716416000175

R Core Team. (2018). *R: A language and environment for statistical computing*. In R Foundation for Statistical Computing. https://www.R-project.org

Rafat, Y., & Stevenson, R. A. (2019). Auditory-orthographic integration at the onset of L2 speech acquisition. *Language and Speech*, *62*(3), 427-451. https://doi.org/10.1177/0023830918777537

Rochet, L. B., & Rochet, A. P. (1995). The perception of the single-geminate consonant contrast by native speakers of Italian and Anglophones. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIII International Congress of Phonetic Sciences (14-19 August 2019)* (Vol. 3, pp. 616-619).

Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. Pisoni & R. Remez (Eds.), *The handbook of speech perception*. John Wiley.

RStudio Team. (2016). *RStudio: Integrated development for R*. In RStudio, Inc. http://www.rstudio.com

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *Journal of the Acoustical Society of America*, *134*(2), 1324-1335. https://doi.org/10.1121/1.4812767

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 546-554. https://doi.org/10.1037/0278-7393.5.6.546

Shea, C. (2017). L1 English/L2 Spanish: Orthography–phonology activation without contrasts. *Second Language Research*, *33*(2), 207-232. https://doi.org/10.1177/0267658316684905

Showalter, C. E. (2020). Russian phono-lexical acquisition and orthographic input: Naive learners, experienced learners, and interventions. *Studies in Second Language Acquisition*, 42, 255-277. https://doi.org/10.1017/S0272263119000585

Simon, E., Chambless, D., & Alves, U. K. (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences*, *32*(3), 380-394. https://doi.org/10.1016/j.langsci.2009.07.001

Sokolović-Perović, M., Bassetti, B., & Dillon, S. (2020). English orthographic forms affect L2 English speech production in native users of a non-alphabetic writing system. *Bilingualism: Language and Cognition*, *23*(3), 591-601. https://doi.org/10.1017/s136672891900035x

Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R., & Sagisaka, Y. (2013). Non-native perception and learning of the phonemic length contrast in spoken Japanese: Training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics*, *22*(4), 373-398. https://doi.org/10.1007/s10831-013-9107-1

Sorianello, P. (2014). Italian geminate consonants in L2 acquisition. In L. Costamagna & C. Celata (Eds.), *Consonant gemination in first and second language acquisition* (pp. 25-46). Pacini Editore.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *The Journal of the Acoustical Society of America*, *85*(5), 2135-2153.

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39*, 456–466. doi: 10.1016/j.wocn.2010.09.001

Tagliapietra, L., & McQueen, J. M. (2010). What and where in speech recognition: Geminates and singletons in spoken Italian. *Journal of Memory and Language*, *63*(3), 306-323. https://doi.org/10.1016/j.jml.2010.05.001

Tsukada, K., Cox, F., Hajek, J., & Hirata, Y. (2018). Non-native Japanese learners' perception of consonant length in Japanese and Italian. *Second Language Research*, *34*(2), 179-200. https://doi.org/10.1177/0267658317719494

Tsukada, K., Hirata, Y., & Roengpitya, R. (2014). Cross-language perception of Japanese vowel length contrasts: Comparison of listeners from different first language backgrounds. *Journal of Speech, Language, and Hearing Research*, *57*(3), 805-814. https://doi.org/10.1044/2014_JSLHR-S-12-0416

Vendelin, I., & Peperkamp, S. (2006). The influence of orthography on loanword adaptations. *Lingua*, *116*(7), 996-1007. https://doi.org/10.1016/j.lingua.2005.07.005

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, *19*(1), 57-95. https://doi.org/10.1080/01690960344000134

Verhelst, N., Van Avermaet, P., Takala, S., Figueras, N., & North, B. (2009). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* CUP.

Vokic, G. (2011). When alphabets collide: Alphabetic first-language speakers' approach to speech production in an alphabetic second language. *Second Language Research*, *27*(3), 391-417. https://doi.org/10.1177/0267658310396627

Wang, X., & Munro, M. J. (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 3, pp. 125-128). University of California.

Young-Scholten, M. (2002). Orthographic input in L2 phonological development. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honour of Henning Wode* (pp. 263-279). Wissenschaftlicher Verlag Trier.

Young-Scholten, M., & Langer, M. (2015). The role of orthographic input in second language German: Evidence from naturalistic adult learners' production. *Applied Psycholinguistics*, *36*(01), 93-114. https://doi.org/10.1017/s0142716414000447

Zampini, M. L. (1994). The role of native language transfer and task formality in the acquisition of Spanish spirantization. *Hispania*, *77*(3), 470-481. https://doi.org/10.2307/344974

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review*, *5*, 683-689. https://doi.org/10.3758/bf03208845

Bassetti, B., Masterson, J., Cerni, T., & Mairano, P. (2021, online first) Orthographic forms affect speech perception in a second language: Consonant and vowel length in L2 English. *Journal of Experimental Psychology: Human Perception and Performance*.

ONLINE SUPPLEMENTARY MATERIALS
Orthographic forms affect speech perception in a second language: Consonant and vowel length in L2 English

Bene Bassetti[a], Jackie Masterson[b], Tania Cerni[c], & Paolo Mairano[d]
[a]University of Birmingham, UK [b] UCL Institute of Education, UK
[c]University of Trento, Italy [d]University of Lille, France

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	0.75	0.86			
	Type of pair (C-CC)	0.48	0.69	0.22		
Word pair	Intercept	0.67	0.82			
	Input (Auditory & Visual)	0.15	0.39	-0.79		
Fixed Effects		**Estimate**	***SE***	**95%CI**	***z*-value**	***p***
Intercept		4.39	0.46	[3.49, 5.28]	9.59	< 0.001
Group (Bilinguals)		-1.30	0.47	[-2.23, -0.38]	-2.76	0.006
Type of pair (C-CC)		-1.33	0.53	[-2.36, -0.29]	-2.51	0.012
Input (Auditory & Visual)		-0.49	0.53	[-1.53, 0.54]	-0.93	0.350
Group (Bilinguals) * Type of pair (C-CC)		0.34	0.52	[-0.68, 1.35]	0.65	0.514
Group (Bilinguals) * Input (Auditory & Visual)		0.45	0.64	[-0.81, 1.71]	0.71	0.480
Type of pair (C-CC) * Input (Auditory & Visual)		0.18	0.57	[-0.94, 1.31]	0.32	0.750
Group (Bilinguals) * Type of pair (C-CC) * Input (Auditory & Visual)		-1.42	0.70	[-2.79, -0.06]	-2.05	0.041
	Marginal R^2:	0.07		**Conditional R^2:**	0.19	

Table 1. *Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual), and type of word pair (control, C-CC) on accuracy (0 = error, 1 = correct) in the Consonant Perception Task.*

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	0.08	0.29			
	Type of pair (C-CC)	0.01	0.07	0.02		
Word pair	Intercept	0.01	0.08			
	Input (Auditory & Visual)	0.01	0.05	0.21		
Residual		0.16	0.40			
Fixed Effects		**Estimate**	***SE***	**95%CI**	***t*-value**	***p***
Intercept		7.13	0.06	[7.00, 7.26]	111.02	< 0.001
Group (Bilinguals)		-0.05	0.09	[-0.22, 0.11]	-0.61	0.544
Type of pair (C-CC)		0.05	0.04	[-0.03, 0.13]	1.16	0.252
Input (Auditory & Visual)		-0.36	0.09	[-0.54, -0.19]	-4.17	< 0.001
Group (Bilinguals)*Type of pair (C-CC)		0.12	0.05	[0.04, 0.20]	2.66	0.009
Group (Bilinguals)*Input (Auditory & Visual)		0.37	0.12	[0.12, 0.61]	2.98	0.004
Type of pair (C-CC) *Input (Auditory & Visual)		0.23	0.05	[0.15, 0.32]	4.84	< 0.001
Group (Bilinguals)* Type of pair (C-CC) * Input (Auditory & Visual)		-0.17	0.06	[-0.28, -0.5]	-2.57	0.012
	Marginal R^2:	0.09		**Conditional R^2:**	0.43	

Table 2 *Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual), and type of word pair (control, C-CC) on RTs in the Consonant Perception Task.*

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	0.84	0.91			
	Type of pair (V-VV)	0.45	0.67	-0.07		
Word pair	Intercept	0.67	0.82			
	Input (Auditory & Visual)	0.02	0.15	0.44		
Fixed Effects		**Estimate**	***SE***	**95%CI**	***z*-value**	***p***
Intercept		3.95	0.36	[3.24, 4.65]	10.89	< 0.001
Group (Bilinguals)		-0.77	0.33	[-1.41, -0.13]	-2.34	0.019
Type of pair (V-VV)		-1.11	0.43	[-1.96, -0.26]	-2.56	0.010
Group (Bilinguals)* Type of pair (V-VV)		-1.03	0.35	[-1.72, -0.34]	-2.94	0.003
	Marginal R^2:	0.09		**Conditional R^2:** 0.23		

Table 3.*Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual), and type of word pair (control, V-VV) on accuracy (0 = error, 1 = correct) in the Vowel Perception Task.*

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	0.07	0.27			
	Type of pair (V-VV)	0.01	0.11	-0.09		
Word pair	Intercept	0.01	0.09			
	Input (Auditory & Visual)	0.01	0.06	-0.23		
Residual		0.15	0.39			
Fixed Effects		**Estimate**	***SE***	**95%CI**	***t*-value**	***p***
Intercept		7.07	0.06	[6.95, 7.19]	117.36	< 0.001
Group (Bilinguals)		-0.02	0.08	[-0.18, 0.13]	-0.27	0.79
Type of pair (V-VV)		0.19	0.03	[0.13, 0.26]	5.61	< 0.001
Input (Auditory & Visual)		-0.28	0.08	[-0.44, -0.12]	-3.50	< 0.001
Group (Bilinguals)* Input (Auditory & Visual)		0.33	0.11	[0.11, 0.54]	2.93	0.004
	Marginal R^2:	0.09		**Conditional R^2:** 0.42		

Table 4

Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals), input (auditory, auditory plus visual), and type of word pair (control, V-VV) on RTs in the Vowel Perception Task.

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	1.67	1.29			
	Type of pair (C-CC)	0.23	0.48	-0.26		
Word pair	Intercept	0.50	0.71			
Fixed Effects		**Estimate**	***SE***	**95%CI**	***z*-value**	***p***
Intercept		4.44	0.49	[3.48, 5.41]	9.01	< 0.001
Group (Bilinguals)						
	Bilinguals	-1.06	0.51	[-2.06, -0.05]	-2.06	0.040
	Learners	-0.28	0.54	[-1.33, 0.77]	-0.53	0.597
Type of pair (C-CC)		-1.29	0.51	[-2.28, -0.29]	-2.53	0.011
Group*Input						
	Bilinguals*C-CC	-1.84	0.46	[-2.75, -0.94]	-4.01	< 0.001
	Learners*C-CC	-1.34	0.48	[-2.28, -0.40]	-2.80	0.005
	Marginal R^2:	0.16		**Conditional R^2:** 0.33		

Table 5. *Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, C-CC) on accuracy (0 = error, 1 = correct) in the Consonant Perception Task.*

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	0.07	0.26			
	Type of pair (C-CC)	0.02	0.13	-0.16		
Word pair	Intercept	0.01	0.12			
Residual		0.12	0.35			
Fixed Effects		**Estimate**	***SE***	**95%CI**	***t*-value**	***p***
Intercept		6.81	0.06	[6.69, 6.92]	118.23	< 0.001
Group (Bilinguals)						
	Bilinguals	0.35	0.07	[0.21, 0.49]	4.97	< 0.001
	Learners	0.15	0.07	[0.01, 0.29]	2.12	0.037
Type of pair (C-CC)		0.15	0.05	[0.04, 0.25]	2.80	0.007
Group*Input						
	Bilinguals*C-CC	0.13	0.05	[0.04, 0.23]	2.89	0.005
	Learners*C-CC	0.02	0.05	[-0.07, 0.12]	0.48	0.634
	Marginal R^2:	0.16		**Conditional R^2:** 0.51		

Table 6. *Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, C-CC) on RTs in the Consonant Perception Task.*

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	0.91	0.95			
	Type of pair (V-VV)	0.21	0.46	0.46		
Word pair	Intercept	0.86	0.93			
Fixed Effects		**Estimate**	***SE***	**95%CI**	***z*-value**	***p***
Intercept		4.67	0.48	[3.73, 5.61]	9.72	< 0.001
Group (Bilinguals)						
	Bilinguals	-1.04	0.46	[-1.96, -0.13]	-2.23	0.026
	Learners	-0.62	0.51	[-1.61, 0.37]	-1.23	0.219
Type of pair (C-CC)		-1.60	0.54	[-2.66, -0.53]	-2.95	0.003
Group*Input						
	Bilinguals* V-VV	-1.85	0.48	[-2.78, -0.91]	-3.88	< 0.001
	Learners* V-VV	-0.93	0.51	[-1.93, 0.07]	-1.83	0.067
	Marginal R^2:	0.17		**Conditional R^2:** 0.33		

Table 7. *Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, V-VV) on accuracy (0 = error, 1 = correct) in the Vowel Perception Task.*

Random Effects		**Variance**	***SD***	**Corr**		
Participants	Intercept	0.07	0.26			
	Type of pair (V-VV)	0.01	0.12	-0.36		
Word pair	Intercept	0.01	0.07			
Residual		0.11	0.33			
Fixed Effects		**Estimate**	***SE***	**95%CI**	***t*-value**	***p***
Intercept		6.77	0.05	[6.67, 6.87]	129.06	< 0.001
Group						
	Bilinguals	0.36	0.07	[0.22, 0.50]	5.09	< 0.001
	Learners	0.18	0.07	[0.04, 0.31]	2.51	0.014
Type of pair (C-CC)		0.18	0.04	[0.10, 0.25]	4.61	< 0.001
Group*Input						
	Bilinguals* V-VV	0.12	0.04	[0.04, 0.20]	2.78	0.007
	Learners* V-VV	0.01	0.04	[-0.07, 0.10]	0.30	0.769
	Marginal R^2:	0.19		**Conditional R^2:** 0.50		

Table 8. *Results of mixed-model analysis of the effects of language background (Native Speakers, Italian$_{L1}$-English$_{L2}$ bilinguals, Italian$_{L1}$ learners of English$_{L2}$) and type of word pair (control, V-VV) on RTs in the Vowel Perception Task.*