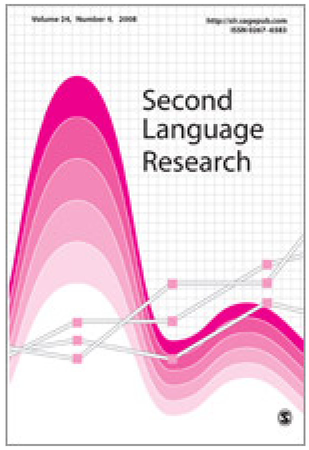


Illusory vowels in Spanish-English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production

Journal:	*Second Language Research*
Manuscript ID	SLR-18-0044.R3
Manuscript Type:	Original Article
Keywords:	phonotactic constraints, speech perception, speech production, Spanish, English, L2 use, grammatical proficiency, age of L2 acquisition, executive control
Abstract:	Spanish native speakers are known to pronounce onset /sC/ clusters in English with a prothetic vowel, as in esport for sport, due to their native language phonotactic constraints. We assessed whether accurate production of e.g. spi instead of espi, was related to accurate perceptual discrimination of this contrast in L2 speech of Spanish-English sequential bilinguals. A same-different discrimination task in stimulus pairs such as spi-espi assessed speech perception and a phonemic verbal fluency task elicited speech production. Logistic mixed model regressions revealed significant differences in accuracy between the bilinguals and the English monolinguals, although some bilinguals performed within the monolingual range. For the production task, but not for the perception task, bilinguals with more exposure to English and greater grammatical knowledge of English performed significantly more accurately than those with less exposure and lower grammatical knowledge. There was no significant correlation between production accuracy and perception accuracy. Through examining phonotactic constraints, these results expand a growing body of research into single sounds which suggests dissociations between L2 perception and production. In contrast to predictions made by L2 speech models, the findings indicate that accurate L2 perception is neither necessary nor sufficient for accurate L2 production, and instead are interpreted to indicate that the two capacities recruit different executive control mechanisms and are acquired – at least to a certain extent - independently in L2 acquisition.

Illusory vowels in Spanish-English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production

Esther de Leeuw[1], Linnaea Stockall[1],
Dimitra Lazaridou-Chatzigoga[2], Celia Gorba Masip[3]

[1]Queen Mary University of London

[2]Humboldt-Universität zu Berlin and University of Cambridge

[3]Universitat Autònoma de Barcelona

Acknowledgements

The authors are particularly grateful to Professor Juana Gil Fernández who welcomed us to Madrid, where we collected most of the data at her *Laboratorio de Fonética* of the University of Madrid, and to Professor Fernando Sánchez Miret at the University of Salamanca. Without their help, we would never have been able to conduct this study. We are also grateful to Annette Zhao for the data analysis, all of the participants who took part in this study, and to the British Academy for funding the project (BA grant SG112797: PI: EdeL; co-investigator: LS).

Abstract

Spanish native speakers are known to pronounce onset /sC/ clusters in English with a prothetic vowel, as in *esport* for *sport*, due to their native language phonotactic constraints. We assessed whether accurate production of e.g. *spi* instead of *espi*, was related to accurate perceptual discrimination of this contrast in L2 speech of Spanish-English sequential bilinguals. A same-different discrimination task in stimulus pairs such as *spi-espi* assessed speech perception and a phonemic verbal fluency task elicited speech production.

Logistic mixed model regressions revealed significant differences in accuracy between the bilinguals and the English monolinguals, although some bilinguals performed within the monolingual range. For the production task, but not for the perception task, bilinguals with more exposure to English and greater grammatical knowledge of English performed significantly more accurately than those with less exposure and lower grammatical knowledge. There was no significant correlation between production accuracy and perception accuracy.

Through examining phonotactic constraints, these results expand a growing body of research into single sounds which suggests dissociations between L2 perception and production. In contrast to predictions made by L2 speech models, the findings indicate that accurate L2 perception is neither necessary nor sufficient for accurate L2 production, and instead are interpreted to indicate that the two capacities recruit different executive control mechanisms and are acquired – at least to a certain extent - independently in L2 acquisition.

Introduction

While there is some research into bilinguals' perception and production of single sounds, there is little work investigating L2 acquisition of phonotactic constraints, and to our knowledge no work systematically investigating the relationship between perception and production of phonotactic constraints. Our study examined phonotactic constraints, i.e. language specific licit versus illicit sequences of sounds, in the speech perception and production of Spanish-English sequential bilinguals. As mentioned by Dupoux, Kakehi, Hirose, Pallier, & Mehler (1999), "in Spanish, /s/ + consonant clusters are always preceded by a vowel and we have informally heard reports by Spanish speakers who maintain that they hear the vowel [e] preceding English words that begin with an /sC/ cluster" (p. 1568). Accordingly, Spanish learners of English often mispronounce initial /sC/ clusters with a prothetic initial vowel (Hualde, 2005).

The main question of this study was whether accurately *perceiving* the /sC/ cluster in English (and therefore *not* perceiving an illusory preceding [e]) was associated with the accurate *production* of the /sC/ cluster (and therefore *not* pronouncing a preceding vowel in English speech where it is not required) within a group of Spanish–English sequential bilinguals. Furthermore, it was examined whether increased and earlier exposure to the L2, and improved L2 grammatical proficiency as measured through a C-Test (Klein-Braley, 1985; Raatz and Klein-Braley, 1981), would help to improve accurate perception and production of /sC/ clusters. Bilingual data were collected in Spain (Madrid and Salamanca, henceforth referred to as the Spain bilinguals), and in London, UK (henceforth referred to as the UK bilinguals). This study therefore aimed to shed light on how the "rules" of language specific sound sequences are implemented in L2 speech acquisition - when they violate rules in the L1 - and feeds into a growing body of research which examines the relationship between L2 perception and production.

L2 speech acquisition models

Often, when considering L2 speech acquisition, the premise is that problems accurately *producing* L2 speech arise from difficulties in accurately *perceiving* L2 speech. It seems logical that perception abilities would need to be in place *before* accurate production abilities are possible, and that difficulties producing L2 speech would have a perceptual basis, such that incorrect perception would lead to incorrect production (see e.g. Escudero, 2005; and Llisterri, 1995, for an overview of early studies). For example, according to the Speech Learning Model (SLM), it is claimed that those L2 sounds which have a similar (although not identical) counterpart in the L1 are the most difficult to learn to pronounce because a similar sound is more likely to be *perceived* within the same category as that of the L1 (Flege, 1995; Flege, Schirru, & MacKay, 2003). A new category, on the other hand, is more likely to be created for a dissimilar sound (originally termed 'new' (Flege, 1987)), which is not perceived to be the same sound. Long-term pronunciation problems are, according to the SLM, more likely in the case of similar sounds than in dissimilar sounds as perceptual equivalence classification prevents experienced L2 learners from producing similar, but not dissimilar sounds (Flege, 1995). This notion of equivalence classification has strong parallels with Kuhl's Native Language Magnet Theory which suggests that the nearer an L2 sound is to an L1 sound (specifically to its prototype), the more it will be perceptually assimilated to this L1 sound (Kuhl, 2004).

The notion of perceptual similarity is also intrinsic to Best's Perceptual Assimilation Model (PAM) (1995) (see also PAM-2 with regard to L2 acquisition of prosody, So and Best, 2010, 2014). According to this model, the similarity between L1 and L2 sounds is based on the perceived resemblance of articulatory gestures used to produce L2 sounds, in comparison to those used to produce the closest L1 sound. As such, perception entails the ability to detect articulatory properties of speech (i.e., tongue movement, vocal tract size, etc.), suggesting that speech perception and production are aligned. This model has strong parallels with the Motor Theory of speech perception, proposed by Liberman and Mattingly (1985), which claims that "the objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations" (p. 2) (see also Fowler (1996) and Ohala (1996)).

What unifies both the SLM and PAM is that perception is considered to be intrinsically linked to production, and that either perception precedes production (SLM), or that they develop in tandem (PAM). However, as will be discussed, there is a growing body of research examining the perception and production of individual L2 sounds which suggests dissociations between L2 perception and production. Moreover, languages differ not only because of differences in their respective phoneme inventories, but also in their realisation of phonotactic constraints. Indeed, very little is known regarding the acquisition of new permissible sound sequences in the L2 which violate L1 phonotactic constraints. The results from this study therefore inform both PAM and the SLM regarding the assertion that difficulties in producing L2 speech are linked to difficulties in perception.

Research into perception and production of L2 speech

Some previous research into the question of whether accurate perception leads to accurate production in L2 acquisition suggests that perception and production are intertwined. For example, it has been found that non-native English speakers who were considered to be experienced L2 speakers (who lived in the US for approximately 7 years) produced and perceived English vowels more accurately than relatively inexperienced non-native participants (who lived in the US for under 1 year) (Flege, Bohn, & Jang, 1997). Both production and perception accuracy appeared to depend on the perceived relation between English vowels and vowels in the participants' specific L1 inventory. Such findings provide support for the SLM in suggesting that perceptual similarity between the L1 and L2 is linked to speech production in the L2.

However, other research has not substantiated a clear dependency between perception and production of L2 speech, and it has been found that some learners are able to produce differences between L2 sounds that they cannot perceive (Baker & Trofimovich, 2006; Beach, Burnham, & Kitamura, 2001; Kartushina & Frauenfelder, 2014; Sheldon & Strange, 1982; Zampini, 1998). For example, early results from Sheldon and Strange (1985) revealed that native Japanese speakers who had learned to produce /l/ and /r/ appropriately still made perception errors in perceiving the contrast. The authors summarised that "perceptual mastery of a foreign contrast does not necessarily precede adult learners' ability to produce acceptable tokens of the contrasting phonemes, and may, in fact, sometimes lag behind production mastery" (p. 254). Similar results from Zampini (1998) into perception and production of voiced and voiceless plosives in Spanish by English native speakers confirmed that L2 production may in some cases precede perception and that learners do not begin to adjust perceptual boundaries until they have attained accurate production categories. Likewise, research examining the production of the three-way voicing difference in Thai bilabial stops by Greek-English bilinguals and English monolinguals has indicated that those bilinguals who "exaggerated" the voicing difference in their Thai speech production also best perceived those differences when listening (Beach et al., 2001). The study concluded that production profiles are an important adjunct to the assessment of bilingual speakers, and have important implications for the interface between perception and production.

The current investigation expands on such findings by examining whether accurately perceiving syllable onset /sC/ clusters in English (although such clusters are not allowed in Spanish) would likewise help to *avoid* production of a prothetic vowel before such clusters in Spanish-English sequential bilinguals, e.g. would those bilinguals who *perceive* a difference between stimuli such as *spi* and *espi* be less likely to *produce* a prothetic vowel in a phonemic verbal fluency task eliciting onset /sC/ clusters, whilst bearing in mind individual differences in English L2 exposure, age of acquisition, and grammatical proficiency in English?

This question is particularly interesting when viewed in relation to executive control, which has been linked to bilingualism and L2 acquisition (Bialystok, 2009, 2017; de Leeuw and Bogulski, 2016). Executive control processes are hypothesised to supervise “the selection, initiation, execution, and termination” of multiple task performance (Rubinstein et al., 2001: 763). Frequently postulated executive mechanisms underlying the overall system are the (1) *shifting* of mental sets, (2) *monitoring* and updating of working memory representations, and (3) *inhibiting* competing stimuli (Miyake et al., 2000). It may be that speech production recruits different executive control functions than speech perception, as the former is inherently more active than the latter, i.e. when producing speech, the articulators must be co-ordinated to match the intended output, whilst when perceiving speech, there is no output. Different underlying executive control processes may have influenced the findings from the aforementioned studies, and are therefore also considered to be potentially relevant in relation to the present research.

Research into perception of L2 phonotactic constraints

In seminal research on the perception of illusory vowels by Japanese native speakers, it was found in four experiments comparing French and Japanese hearers that phonotactic properties of Japanese induced native Japanese listeners to perceive "illusory" vowels inside consonant clusters in VCCV stimuli, where there were in fact no vowels, as Japanese does not allow such word medial consonant clusters (e.g. *ebzo* was perceived as *ebuzo*) (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; see also Harris, 1983; Itô and Mester, 1995).

In contrast, French native speakers had no difficulty perceiving this contrast (French allows such consonant clusters, e.g. 'observer' /obzɛrve/), However, the French native speakers had difficulties discriminating items that differed in vowel length (e.g. *ebuzo* vs. *ebuuzo*), while Japanese native speakers had no difficulty discriminating such stimuli (vowel length is contrastive in Japanese but not in French). The conclusion from this series of studies was that models of speech perception should be revised to account for language specific phonotactic constraints. However, although the research found significant differences between the Japanese and French native speakers, it was not further investigated whether, for example, *some* of the Japanese native speakers might have been able to acquire the French phonotactic constraints, nor whether accurate perception of the French phonotactic constraints (i.e. licit consonant clusters) may have been associated with accurate *production* of the consonant clusters.

In more recent research, several populations of Japanese-Brazilian bilinguals were tested in order to determine whether the perceived illusory vowel differed between bilingual groups (Parlato-Oliveira, Christophe, Hirose, & Dupoux, 2010). The results from an explicit metalinguistic task showed that, as expected based on the phonotactics of each language, monolingual Japanese participants mainly perceived the stimuli as containing the epenthetic vowel /u/ (e.g. in items like *ebna* and *ebuna*) while monolingual Brazilian participants mainly perceived the epenthetic vowel /i/ (e.g. in items like *ebna* and *ebina*). Interestingly, first-generation immigrants, who had Brazilian Portuguese as an L2, appeared to behave like Japanese monolinguals, suggesting that their immersion within a Brazilian-speaking country, which started in adulthood, "did not induce them to significantly modify their phonological settings" (p. 3742). In sharp contrast, second-generation immigrants behaved like Brazilian monolinguals and were more likely to perceive the epenthetic vowel /i/. Moreover, in an implicit perceptual task, it was found that all three groups of bilinguals behaved exactly like Brazilian monolinguals, "showing high confusability between *ebna* and *ebina*, while the distinction between *ebna* and *ebuna* proved to be easy to for them" (p. 3745). Parlato-Oliveira et al. (2010) summarised that an effect of L2 exposure was most clearly observed in the explicit task but not in the implicit task and suggested that L2 experience is more likely to affect explicit, or metalinguistic, perceptual tasks.

Another study, specifically examining Spanish–English sequential bilinguals, similarly found differences between bilinguals with regard to their capacity to perceive phonotactic constraints in the L2 (Carlson et al., 2016), noting that English not only permits #sC, the prohibited sequence in Spanish, but it favours this sequence over #VsC, the preferred Spanish repair. In both the identification and discrimination tasks, the Spanish–English bilinguals exhibited weaker perceptual repair effects relative to Spanish monolinguals, who came from Cuetos et al. (2011). This was true even for the bilinguals who were considered to be dominant in Spanish; however, the perceptual repair effects were least pronounced for English-dominant bilinguals, suggesting that the acquisition of English influenced the perception of phonotactic constraints in Spanish (Carlson et al., 2016). These results support related research which has similarly found that /sC/ onset clusters in English activate Spanish phonotactic constraints in Spanish–English bilinguals (Freeman et al., 2016) and most recent research suggesting that even the perception of L1 Brazilian Portuguese phonotactic constraints are modified by the acquisition of English as an L2 (Cabrelli et al., 2019).

However, in none of these studies was speech production investigated. Indeed, as it is quite often accepted that perception provides the foundation for production, it might be – potentially prematurely - interpreted that such results regarding perception provide insight into what those same participants might produce in their L2 speech. However, based on the previously discussed research which shows weak to not existent relationships between L2 speech perception and production (see also Kartushina & Frauenfelder, 2014), we cannot assume that such studies into the perception of L2 sound sequences provide direct insight into the capacity of L2 learners to produce those same L2 sound sequences.

In the present study, by focusing on a single phonotactic feature, we were able to systematically examine the degree to which perception and production of the prothetic vowel were associated with one another in Spanish native speakers with English as an L2. The research by Hallé et al. (2008) confirms that the phenomenon of an illusory prothetic vowel preceding sibilant consonant clusters in Spanish native speakers is real, but the study does not actually employ a straightforward, simple, auditory discrimination task along the lines of Dupoux et al. (1999), and, moreover, Hallé et al.'s (2008) research does not compare perception and production in the same experiment and same speakers, which is the aim of the current study.

Methodology

The study comprised two main tasks. In Task I, the perception of a prothetic vowel in minimal pairs was examined in the same two participant groups: English monolinguals, and Spanish-English bilinguals. In Task II, the production of prothetic vowels preceding /sC/ consonant clusters in onset position was assessed to determine whether there was an association between perceptual and production accuracy. Additionally, predictor variables were examined which might have played a role in determining production and perception in the sequential Spanish–English bilinguals, i.e. amount of daily exposure to L2, as self-assessed; age of L2 acquisition; and grammatical proficiency in the L2, as measured by a written C-Test (see Appendix I for questions and Appendix II for information). Grammatical proficiency was thought to be an interesting variable to examine because it may have been that bilinguals with higher grammatical knowledge were also more aware of the phonotactic rules of English, which allow /sC/ clusters in onset position (see Kivistö-de Souza, 2015 and Morales Pech and Izquierdo, 2011 for related research). This metalinguistic awareness may have improved their English speech production, and, potentially also their speech perception, if the two domains are related.

Participants

In total, 42 participants were examined. Eight participants were native English speakers, whose data were collected in London, United Kingdom. These participants listed no fluency in other languages, although some did have knowledge of other languages due to school education and vacations abroad. Seven participants were native Spanish speakers with English as an L2 who were recorded in London, United Kingdom. These participants were either working in London, or were students on a year abroad, and had lived in the UK for between 3 and 16 months. The remaining 27 participants were also all native Spanish speakers whose data were collected in Spain. These participants reported that they had learned English as an L2 in Spain, and used English to varying extents in their daily lives. Many of the Spain bilinguals had also spent time in an English speaking country with the average time being 7 months, and both the United Kingdom and Spain bilinguals were therefore collapsed into one group. All participants were between 40 and 18 years of age and most were either university students, researchers, administrators, or academic faculty. All of the bilinguals considered Spanish to be their native language, and English to be the language in which they had the greatest proficiency after Spanish. The mean amount of daily English use of the bilinguals, which was self-assessed, was 30% with a standard deviation of 18%; the mean grammatical proficiency score was 38% with a standard deviation of 16%; and the mean age of English acquisition was 8 years of age with a standard deviation of 5.4 years.

	English monolinguals	Spanish-English bilinguals
Number of participants	8	34
Average age at time of experiment, standard deviation in brackets	23.6 (6.3)	27.15 (11.9)
Females : males	4:4	24:10

Table 1: Participant background information, see Appendix II.

It is relevant to note that more data were collected initially than included in the analysis. This is because a number of English monolinguals turned out to have high proficiencies in additional languages; and a number of participants in Spain cited that they were bilinguals in either Catalan or Galician, and sometimes that they did not consider themselves to be native speakers of Spanish, which we had not expected when originally conducting the experiment. To ensure the reliability of our study, we therefore excluded 19 participants from the analysis.

Data collection procedure

The same procedure was followed in both Madrid, Salamanca and London, and in each setting the recordings were conducted in a sound attenuated room at respectively the *Laboratorio de Fonética* of the *Centro de Ciencias Humanas y Sociales*, the University of Salamanca, or the Phonetics Laboratory of Queen Mary University of London. Recordings were conducted entirely in English in all cases in order to ensure an English mode (Grosjean, 1998), i.e. this investigation tested L2 English perception and production; however, particularly with regard to the perception task, as the tokens were non-words, the participants' accuracy also provides insight into perceptual capacities more generally (Carlson et al., 2016) and therefore the results also have ramifications on the mechanisms of L1 perception (Cabrelli et al., 2019) and its plasticity in the context of bilingualism (de Leeuw and Celata, 2019).

Participants entered into the sound attenuated room where the outline of the procedure was described to them, and then filled in the participant consent form. Thereafter, the research investigator (either EdeL or DLC) filled in an adapted version of the MPI Language Background Questionnaire (Gullberg and Indefrey, 2003), by asking the participant the questions on the form (see Appendix I). In general, the completion of this form took between 15 and 20 minutes. Participants then completed the C-test, for which they were given a maximum of 10 minutes (see Appendix I), the onset elicitation task (i.e. the phonemic verbal fluency task), and then the perceptual discrimination task. Although the perceptual discrimination task was completed last to ensure /sC/ clusters were not made salient to participants before the production task, the perception task is presented initially in this manuscript, and the production task thereafter. In total, data elicitation took no more than 45 minutes.

Task 1: Perception Task Elicitation

The purpose of this task was to determine the extent to which Spanish native speakers with English as an L2 perceived a difference between stimuli such as *spi* and *espi* (see Table 3 for the full set of stimuli). To assess this question, participants were invited to take part in a same-different perceptual discrimination task (AX task). Participants wore Sennheiser over ear headphones, and were presented with two non-words directly after one another using DMDX software (Forster and Forster, 2003), which also recorded participant responses, e.g. (1) *spi* and *spu* or (2) *spi* and *espi* or (3) *spi* and *spi* (see Table 2). For ease of exposition, we will continue to use these tokens as the model stimuli. In line with what we know about Spanish (Hallé et al., 2008; Hualde, 2005), we expected the Spanish-English bilinguals to perceive a difference between the stimuli in (2) less accurately than English monolinguals, but we did not expect a difference between the Spanish-English bilinguals and English monolinguals in (1) and (3) (see Table 2). We did not expect to find a difference in Condition (3) because the items presented in the AX discrimination task were the same in these trials, and we did not expect a difference in Condition (1) because /i/ and /u/ are contrastive in both English (e.g. *beat* versus *boot*) and Spanish (e.g. *sí* versus *su*). However, as English favours /sC/ onset clusters over /VsC/ onsets (Carlson et al., 2016), we also expected that English monolinguals might have difficulties in Condition 2, albeit less than the Spanish native speakers. There were approximately twice as many stimuli pairs for Condition 3 as for the other two conditions in order to ensure an equal number of same stimuli pairs (Condition 3) versus different stimuli pairs (Conditions 1 & 2), see Table 2, and 197 trials in total.

	Contrast example	Nr. of trials	Prediction
Condition 1	*spi-spu*	49	No difference between Spanish–English bilinguals and English monolinguals
Condition 2	*spi-espi*	49	Spanish–English bilinguals will exhibit less accuracy than English monolinguals
Condition 3	*spi-spi*	99	No difference between Spanish–English bilinguals and English monolinguals

Table 2: Conditions in perception task and general predictions

Stimuli

Stimuli were produced by the first author, who is a native speaker of English. The author produced all of the *espi* tokens first, listed in column b in Table 3.

a. Edited token with no prothetic vowel	***b. Original token with prothetic vowel***
spi	*espi*
spu	*espu*
sli	*esli*
slu	*eslu*
ski	*eski*
sku	*esku*
smi	*esmi*
smu	*esmu*
sni	*esni*
snu	*esnu*
sti	*esti*
stu	*estu*

Table 3: Tokens used in same-different discrimination task.

Thereafter, the initial vowel was edited from the stimulus, such that a new stimulus was created: *spi*. These consonant clusters (i.e. /sp/, /st/, /sk/, /sm/, /sn/, /sl/) were chosen because they are the possible /sC/ clusters in English, and therefore also the same onsets which were elicited in the phonemic verbal fluency task.

Delimiting the initial vowel from the following sibilant was not difficult based on analysis of the spectrogram and waveform. The boundary was marked where the periodicity of the vowel ended and the aperiodicity of the fricative commenced. Therefore, these stimuli (a. *spi* and b. *espi*) were exactly the same except for the vowel in b (see Figure 1 and 2 of /spi/ and /espi/). Note that it may have been that formant remnants from the initial vowel remained in the sibilant due to coarticulation effects, although all actual periodicity from the vowel was cut out (see Fig. 1 & 2), which might potentially have led to more perceived instances of /e/ on the part of the bilinguals; however, all bilinguals heard the same items, and therefore this would have affected their discrimination of *spi* versus *espi* equally.

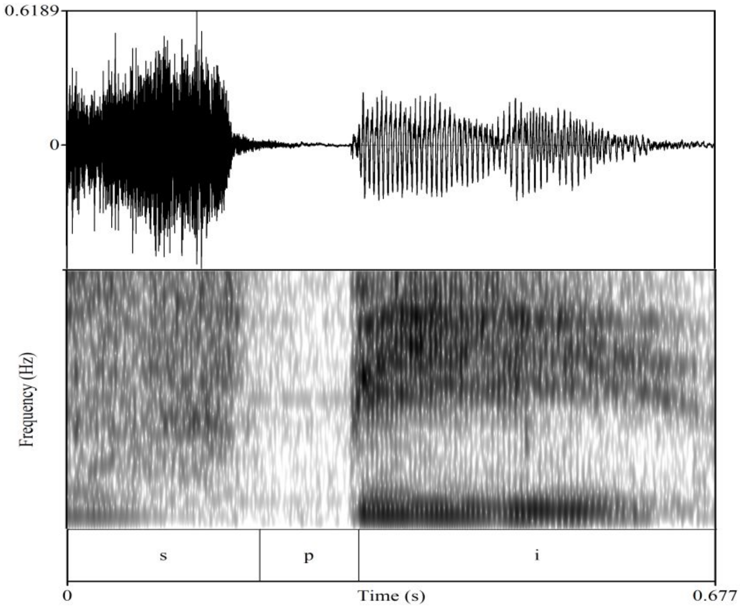


Figure 1: Token of *spi* used in perception task.

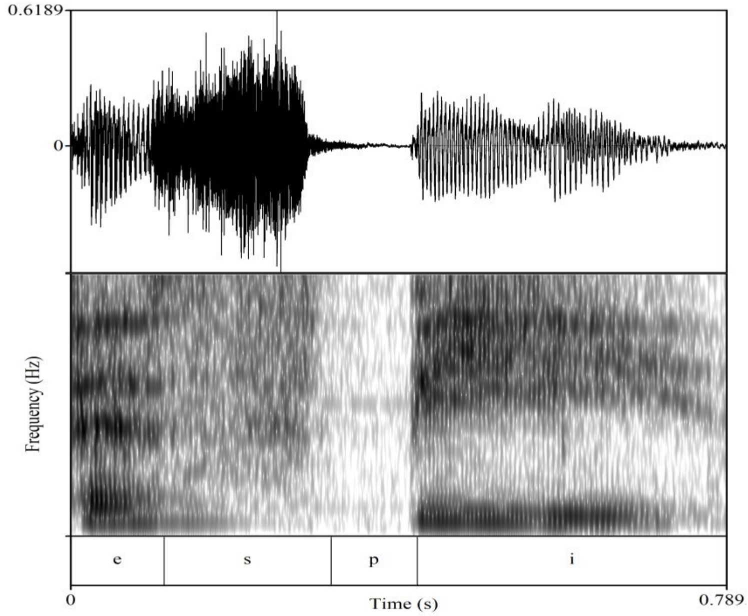


Figure 2: Token of *espi* used in perception task.

As this study examined the L2 acquisition of English, the initial vowel in *b* was the open-mid front unrounded vowel /ɛ/, which is present in the first syllable of English words like *escape* and *estate*. This vowel is also similar to Spanish /e/, and in some transcriptions has been considered to be an allophone of this phoneme (Tomás Navarro, 1918), although other transcriptions refute this (Martínez-Celdrán et al., 2003). This vowel was considered to be the most appropriate for the purpose of this study, i.e. in contrast to schwa, because (1) it is present in English words preceding /sC/ clusters, and hence the bilinguals would have been exposed to this sound sequence in their acquisition of English, and because (2) it is similar to Spanish /e/, and hence more likely to be identifiable on the part of the bilinguals than English schwa would have been. A final reason rationalising the choice of /ɛ/ was that (3) it was more naturalistic for the speaker to *produce*, as this vowel indeed occurs in English words, and therefore the recordings were likewise more likely to also *sound* naturalistic for the listeners. In essence, had we chosen schwa as the initial vowel, it may have been more likely that the *quality* of the vowel impeded discrimination, rather than the *presence* of the vowel itself. Sound files are online for interested readers.

The duration of the vowel approximated 80ms to 85ms, and was therefore considered to be controlled yet naturalistic. As in Table 2, only high vowels followed the /sC/ cluster as in stressed position these English vowels were thought to be more similar to the Spanish high vowels than English low vowels to Spanish low vowels. For example, in Spanish there is just one low vowel, /a/, whilst in English there are numerous, e.g. depending on dialect /a/, /ɑ/ and /ɒ/ (Hualde, 2005). As such, the task was considered to focus on the discrimination of the vowel prothesis in the original token, and no additional challenges were present which might have additionally encumbered the perception of the stimuli. Note that the speaker produced monophthongal realisations of the vowels in the *espi* items with low second formant frequencies (see Fig. 3 & 4), i.e. no GOOSE-fronting, which has been reported in new varieties of British English (Cheshire et al., 2011; Mills, 2014).

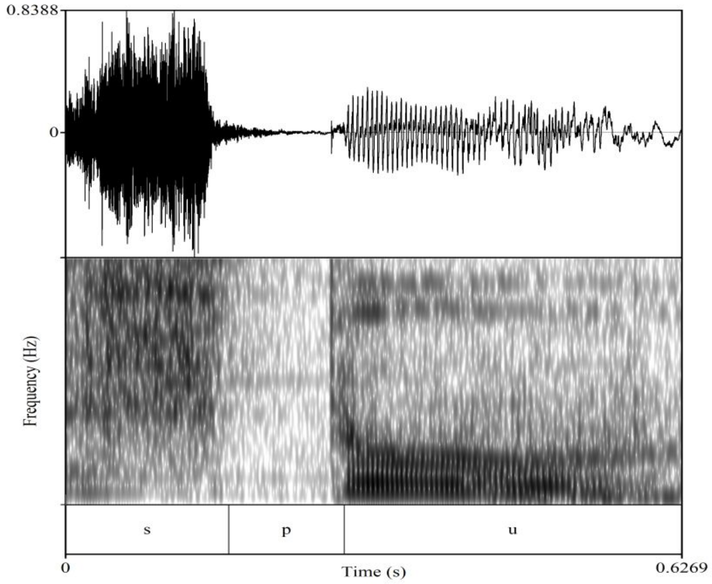


Figure 3: Token of *spu* used in perception task.

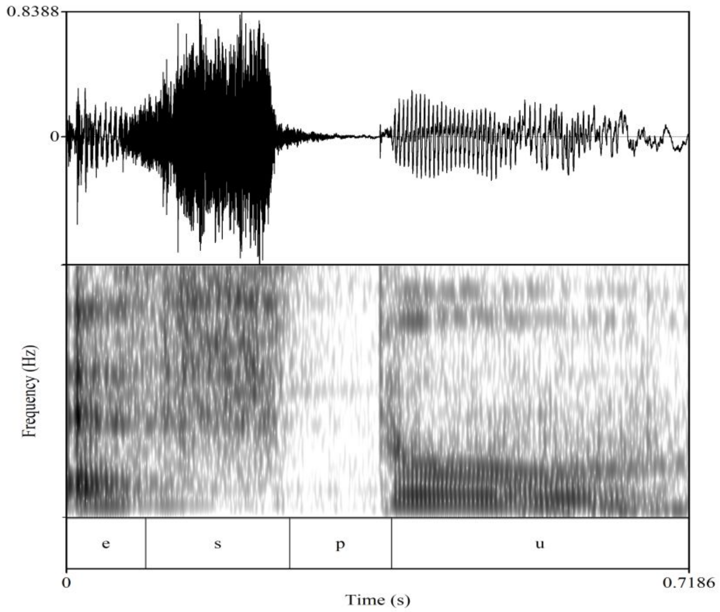


Figure 4: Token of *espu* used in perception task.

Procedure

The experiment comprised three conditions, which were presented randomly rather than in blocks. In each condition participants had to distinguish between the pairs of stimuli presented immediately after the other in conditions (1), (2) and (3). Participants were not specifically told what language the items would be in, but all of the instructions for this specific task were in English, and the entire experimental procedure was in English, as well as the questionnaire, so participants were considered to have been in an English mode (Grosjean, 1998). As such, it was thought that they would be more likely to discriminate *spi-espi* pairs than had the experiment been conducted entirely in Spanish.

Condition 1 contained stimulus pairs like *spi-spu*, which we expected to be relatively easy for both Spanish native speakers and English native speakers. Condition 2 was the target condition in which participants had to distinguish between *spi-espi* pairs. This condition was predicted to be more difficult for the Spanish native speakers with English as an L2 than for the English monolinguals. In Condition 3, participants had to determine whether *spi-spi* type pairs were the same or different, and this, like Condition 1, was expected to be relatively easy for both the Spanish-English bilinguals and the English monolinguals. However, because all three conditions were presented randomly to the participants, the task was on the whole expected to be more difficult for the bilinguals than for the monolinguals, such that the Spanish-English bilinguals may also have made more errors, or have had slower response times in Conditions 1 and 3.

Participants indicated their response using a handheld video game controller by pressing the left button for same and the right button for different. Participants were given 2500ms to complete their decision, and then the next stimulus pair was presented automatically. A 2500ms duration was allowed because it worked well in the pilot project as a duration that was long enough to allow responses from the full range of participants, but fast enough to keep the pace of the experiment proceeding without long lags, and to force a rapid, automatic response (rather than a more reflective slow response). Note that Dupoux et al. 1999 exp3 used a 4 sec cut-off, but we found that we lost less than 3% of trials with a 2500ms cut-off in the pilot (in the current experiment, we lost 4.5% of all trials due to no response within the time window).

There was a short practice block, consisting of 6 trials: 3 different, 3 same. Participants were given feedback after the practice block, and could ask questions if the instructions were not clear, although this never happened. The timing 'clock' started with the onset of the second wav file, so participants could press a button to indicate 'same' or 'different' as soon as the second file began. There were only 2 trials where a response was given before 300ms passed across the entire data set, so we included all trials in the analysis. There was no beep between stimuli; participants just heard the two sound files and then silence until they pressed the response button or 2500ms elapsed with no response (following Dupoux et al. 1999).

In total, the perception task lasted 12 minutes, which included two short breaks equally dispersed. Including the explanation of the perceptual discrimination task, the entire procedure took 15 minutes.

Task II: Prothetic Vowel Production Elicitation

For the prothetic vowel production task, participants took part in a phonemic verbal fluency task. In the standard version of this task, participants are asked to produce as many unique words as possible which start with a given letter within a given timeframe (Newcombe, 1969). The vast majority of studies that use this task use single letters as cues, but there is at least one study that used double-letter cues (e.g. "fa", "sm") when comparing Spanish-English bilinguals to English monolinguals (Sandoval et al., 2010). In the standard test, the participant's score is the number of unique words produced for each given letter, but for our purposes, the accuracy score reflected accurate /sC/ pronunciation (i.e. no prothetic vowel) for each unique word.

The phonemic verbal fluency task, also called the letter fluency task, is a standard task used in both normative and non-normative language assessments of native and non-native languages, often implemented together with the semantic verbal fluency task for which word categories are listed, e.g. words like *cat* and *dog* would be listed if "animal" were given as a category (Grogan et al., 2009; Newcombe, 1969). Interestingly, some research has shown that monolinguals and fluent bilinguals (in their L2) score similarly on the phonemic verbal fluency task, i.e. they are able to list approximately the same amount of words, but that monolinguals obtain higher scores on the semantic verbal fluency task (Gollan et al., 2002; Portocarrero et al., 2007, but see also Bialystok et al., 2008). For the purposes of our study, the bilinguals were considered to be capable of the phonemic verbal fluency task, but that the task would also be slightly more demanding than e.g. a word list reading task, for which we thought that the full orthographic representation of the word would enable them to focus on their pronunciation, and in this way not necessarily reflect what they would produce in normal conversation.

In the instructions for the phonemic verbal fluency task, it was stated that they would see one or two letters, like 'b' or 'gr' and should name as many English words as possible which begin with the sounds which those letters spell. An example of 'pr' was given, for which one should say 'print' but not 'pint' "because 'pint' does not begin with 'pr'. Likewise, if they saw 'sk', it was instructed that they could say 'scream' or 'sky' since "they both begin with the 'sk' sounds", but that 'scene' would be incorrect. Participants were given 10 seconds for each onset, and they were asked to only name unrelated words, e.g. they should avoid naming both "print" and "printed". A short amount of time of 10 seconds was given for each onset so that participants were not able to reflect too much on their pronunciation. In total, 22 onsets were elicited, but only the /sC/ clusters were of interest to the study at hand: /sp/, /st/, /sk/, /sl/, /sn/ and /sm/. As there were six clusters, participants were given in total 1 minute to produce the /sC/ clusters, which is the standard total time in a phonemic verbal fluency task (Newcombe, 1969).

The phonemic verbal fluency task is often used to assess executive control abilities, as participants need to retrieve words, which requires them to access their mental lexicon, whilst focusing on the task, selecting words meeting certain constraints and inhibiting repetition (Miyake et al., 2000). However, the amount of words produced is considered to be dependent on not only executive control abilities, but also on e.g.

vocabulary size (Sauzéon et al., 2011). As such, this task was considered to be a more demanding task than other potential production elicitation techniques, and more representative of normal everyday speech production, which would likewise require both monolinguals and bilinguals to access their mental lexicon whilst monitoring pronunciation.

Moreover, the task was also considered to be potentially more cognitively demanding and more reflective of "real world" language tasks than the previously described perception task. In this way, it could be argued that the tasks were not balanced, but we thought that it was interesting to investigate whether the "raw capacity" to discriminate *spi-espi* in the perception task, unimpeded by other cognitive tasks, would be shown to facilitate accurate production of /sC/ clusters in English. A crucial question was therefore whether bilinguals who scored higher in the production task would likewise also necessarily score higher in the perception task, if accurate perception is indeed a requirement for accurate production.

Acoustic analysis

In order to determine whether a prothetic vowel was present in the Spanish-English bilinguals' production of word-initial consonant clusters beginning with /sC/ in English, an acoustic analysis was conducted on recordings obtained from all participants taking part in the production task.

Using Praat (Boersma and Weenink, 2010), a textgrid file was created which included interval tiers (respectively, "word", "vduration") and one point tier ("notes"). In the top "word" tier, real words which were produced by participants were input directly while non-words were spelt out according to participants' actual pronunciations and were marked with an "X" at the end of the created word. The tier called "vduration" was used to annotate the prothetic vowel, if there was one at all, and the point tier was used to make comments on individual items, when considered necessary. To delimit the words, the cursor to indicate the start of the word was placed at the closest 0-crossing in the waveform where either phonation began from the prothetic vowel, or where aperiodic noise began from the sibilant. The cursor to indicate the end of the word was placed at the closest 0-crossing where there was considered to be a significant drop in waveform amplitude.

Both the spectrogram and the waveform were used to determine the onset and offset of the prothetic vowels preceding the onset of the sibilant. Where necessary, Praat's recognition of a pitch contour (i.e. lack of recognition for the voiceless sibilant) was also considered as an indicator of voicing in the case of the potential vowel. In most cases, the onset and offset of the potential prothetic vowel were located at the beginning and the end of periodicity where the waveform crossed the 0-axis (some prothetic vowels started with a glottal stop, or some other form of plosion – and this was also counted as part of the vowel). The presence of formants in the spectrogram was also used as an indicator of vowel prothesis. However, in some cases, the prothetic vowel was produced with creaky voice. In such cases, auditory perception of an initial vowel was the main cue which was used to determine whether there was a prothetic vowel. The offset of the vowel was determined by the onset of the following sibilant. Determination of the sibilant was relatively uncontroversial; marked by the onset of a high intensity band of aperiodic frication in the waveform and spectrogram. As was the case for the onset of the potential prothetic vowel, this transition from vowel to sibilant was marked at the nearest 0-crossing in the waveform. The duration of the prothetic vowel was then measured in milliseconds (ms), and data regarding duration can be found in the appendix; however, here due to space constraints, we only present binary data of both the perception and production results (findings from continuous durational analyses confirmed binary data).

Statistical analysis

For both the perception and production results, data were organised in CSV files using Excel software. Thereafter, R software (R Core Team, 2017) was used for the analyses and a series of binomial mixed-effects regression models were built for the accuracy results of the perception and production tasks and a series of linear regression models were built for the reaction time results of the perception task (using the lme4 package in R; Bates, Maechler, Bolker, & Walker, 2015) for the examination of the influence of fixed and random factors on the response. For the analysis of the perception data, timed out responses (i.e. trials which were not discriminated within 2500ms) were excluded resulting in the exclusion of 371 trials out of 8274 trials. For the analysis of the production data, 616 individual word tokens were analysed. We present the perception results initially, then the production results, then the results regarding perception as a predictor of speech production.

Results

Task I: Perception

The accuracy and reaction time results of the perception experiment are summarised per group and condition in Table 4 and the binary accuracy results are visualised in Figure 5, revealing that Spanish–English bilinguals performed least accurately in condition 2 compared to conditions 1 and 3, and less accurately than the English monolinguals.

	English monolinguals		***Spanish-English bilinguals***	
	Percent of correct responses	Response times for correct responses	Percent of correct responses	Response times for correct responses
Condition 1: spi-spu	96.05 (19.5)	1175.4 (217.12)	95.17 (21.45)	1193.48 (252.26)
Condition 2: spi-espi	87.4 (33.23)	1240.89 (302.94)	62.29 (48.48)	1311.41 (384.38)
Condition 3: spi-spi	97.36 (16.04)	1197.19 (308.67)	93.11 (25.33)	1228.09 (341.35)

Table 4: Percentage of correct responses and response times for each group (standard deviations in brackets).

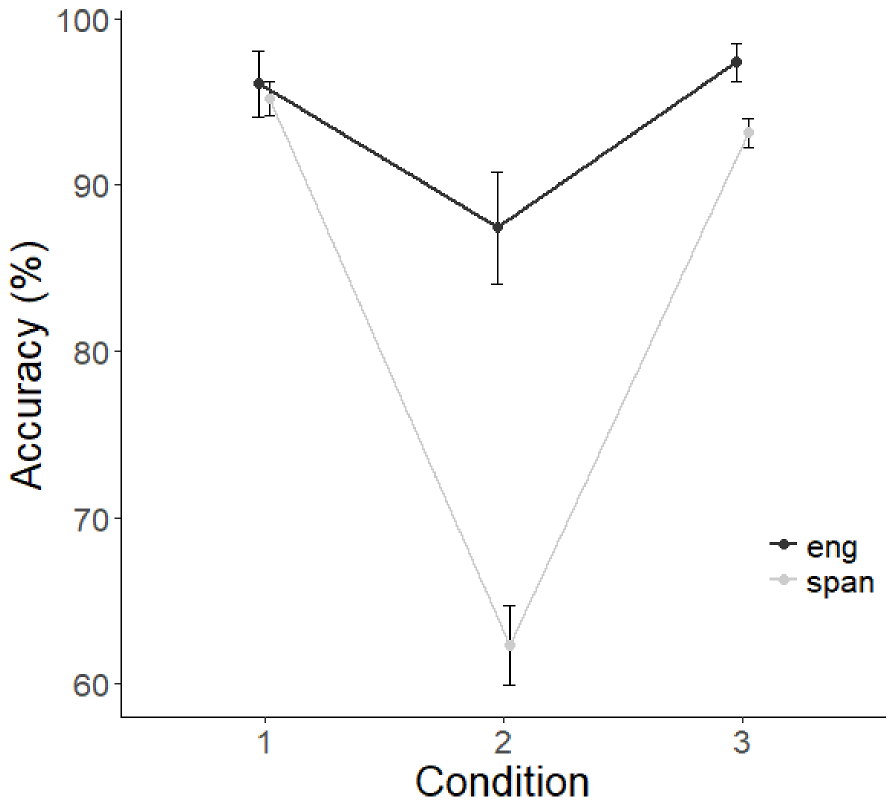


Figure 5: Perception accuracy in English monolinguals versus Spanish-English bilinguals.

For the data analysis of the accuracy results, the modelled response was accuracy (i.e. yes or no, whether the participant was accurate in the AX discrimination or not), and the fixed factors were language background (L1, English vs. Spanish) and condition

(C1: *spi-spu*, C2: *spi-espi*, C3: *spi-spi*). Random intercept was consonant, i.e. second consonant in the sC cluster (e.g. /p/ in /sp/) (participant was initially included but this model failed to converge). Models were manually stepped-down (using likelihood ratio tests) from maximal models containing all factors and possible interactions to the 'best' model that only contained significant predictors or predictors that participated in significant interactions (Barr et al., 2013). The best fitting model determined by step-down pairwise model comparison was acc ~ L1 * condition + (1 | consonant). The model parameters are in Table 5.

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept=L1=English, Cond=1)	3.1955	0.2689	11.884	< 2e-16 ***
L1=Spanish, Cond=1	-0.2099	0.2888	-0.727	0.46741
L1=English, Cond=2	-1.2577	0.3060	-4.111	3.95e-05 ***
L1=English, Cond=3	0.4174	0.3484	1.198	0.23084
L1=Span x Cond=2	-1.2300	0.3318	-3.707	0.00021 ***
L1=Span x Cond=3	0.7967	0.3742	-2.129	0.03325 *

Table 5: Estimates, standard errors, z values and p values of the best fitting model for the perception experiment.

As displayed in Table 5, the main finding was that, as predicted, the magnitude of the difference in the Spanish speakers between condition 1 and condition 2 was greater than the difference in the English speakers between condition 1 and condition 2 (β = -1.2300, z =-3.707, p < 0.001), although responses to condition 2 were also less accurate than for condition 1 for English speakers (β = -1.2577, z = -4.111, p < 0.0001). Furthermore, the magnitude of the difference in the Spanish speakers between condition 1 and condition 3 was also greater than the difference in the

English speakers between condition 1 and condition 3 ($\beta = 0.7967$, $z = -2.129$, $p < 0.05$). No other comparisons were significant.

In a second analysis, we investigated possible differences among the Spanish–English bilinguals in their responses to the critical condition 2 (e.g. *spi-espi*) using the predictor variables of percentage of English daily use, age of English acquisition, and C-Test proficiency score, i.e. grammatical score, which were scaled using the scale function in R. However, none of these predictors proved to be significant (*p*s > 0.05).

For the reaction time results, in our first analysis, the modelled response was log-transformed RT and the fixed factors were language background (L1, English vs. Spanish) and condition (C1: *spi-spu*, C2: *spi-espi*, C3: *spi-spi*). Random intercepts were participant and again consonant. None of the predictors or their interaction proved to be significant (*p*s > 0.05). Nevertheless, the numerical RT pattern was consistent with the accuracy results pattern with the Spanish–English bilinguals having longer RTs in condition 2 (e.g. *spi-espi*).

In a second analysis, we investigated possible differences among the Spanish–English bilinguals in their responses to the critical condition 2 (e.g. *spi-espi*) using the predictor variables of percentage of English daily use, age of English acquisition, and C-Test proficiency score, which were scaled using the scale function in R; however, none of the predictors proved to be significant (*p*s > 0.05).

Task II: Production

Table 6 summarises the rates of prothesis, for each of the six cues of interest, for the two participant groups and Figure 6 visualises the results.

Cluster	*English monolinguals (%)*	*Spanish–English bilinguals (%)*
sk	8.70 (28.8)	55.56 (50.0)
sl	11.1 (32.3)	55.13 (50.1)
sm	0.00 (0.0)	57.53 (49.8)
sn	6.67 (25.8)	52.86 (50.3)
sp	4.35 (20.9)	62.63 (48.63)
st	8.33 (28.2)	67.37 (47.1)

Table 6: Percentage of words produced with prothesis for each sC cue, by participant group (standard deviations in brackets).

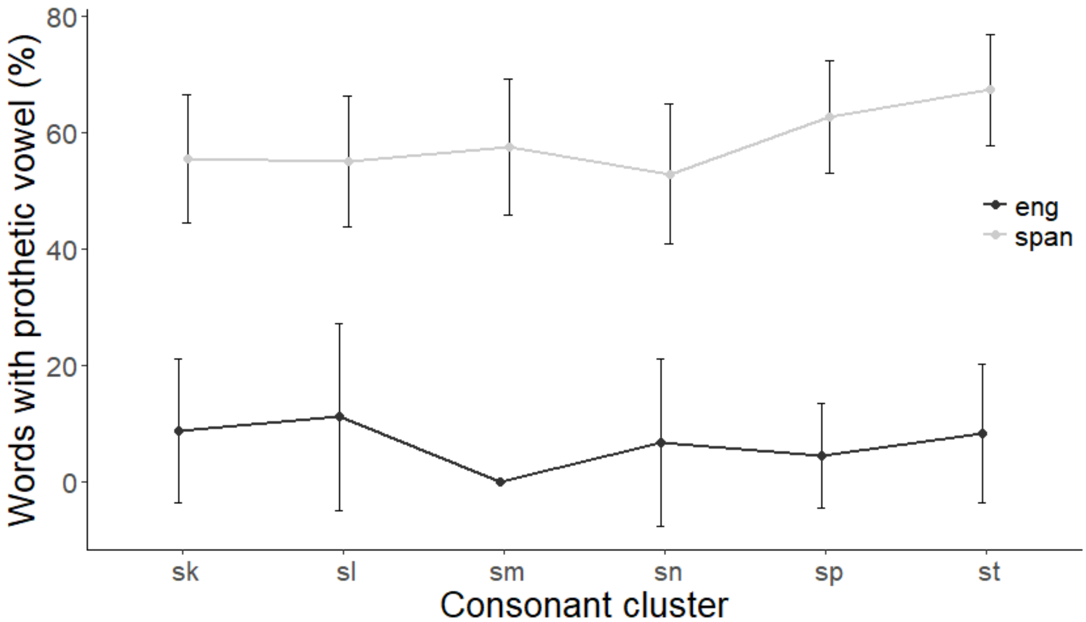


Figure 6: Production accuracy in English monolinguals versus Spanish-English bilinguals.

For the production experiment, the modelled response was whether a prothetic vowel was produced or not (i.e. prothesis), for each word produced, and the fixed factor was language background (L1 English vs. L1 Spanish). Random intercepts included participant and consonant cluster (i.e. /st/, /sp/, /sk/, /sm/, /sn/ and /sl/). The maximal model motivated by our hypotheses was prothesis ~ L1 + (1 | participant) + (1 | consonant.cluster).

	Estimate	***Std. Error***	***z value***	***Pr(>|z|)***
(Intercept) (L1 = English)	-3.7248	0.8353	-4.459	8.23E-06 ***
L1 = Spanish	4.2509	0.886	4.798	1.61E-06 ***

Table 7: Estimates, standard errors, z values and p values of the best fitting model for the production experiment.

As displayed, for the production analysis, the main finding was that as predicted, there was a significant difference between the English monolinguals and the Spanish-English bilinguals ($\beta = 4.2509$, $z = 4.798$, $p < 0.00001$) with the bilinguals producing significantly more prothetic vowels than the monolinguals. To exemplify this difference, individual words (i.e. *speak*, *slain*, *smile*) produced by the Spanish–English bilinguals are displayed in figures 7-9.

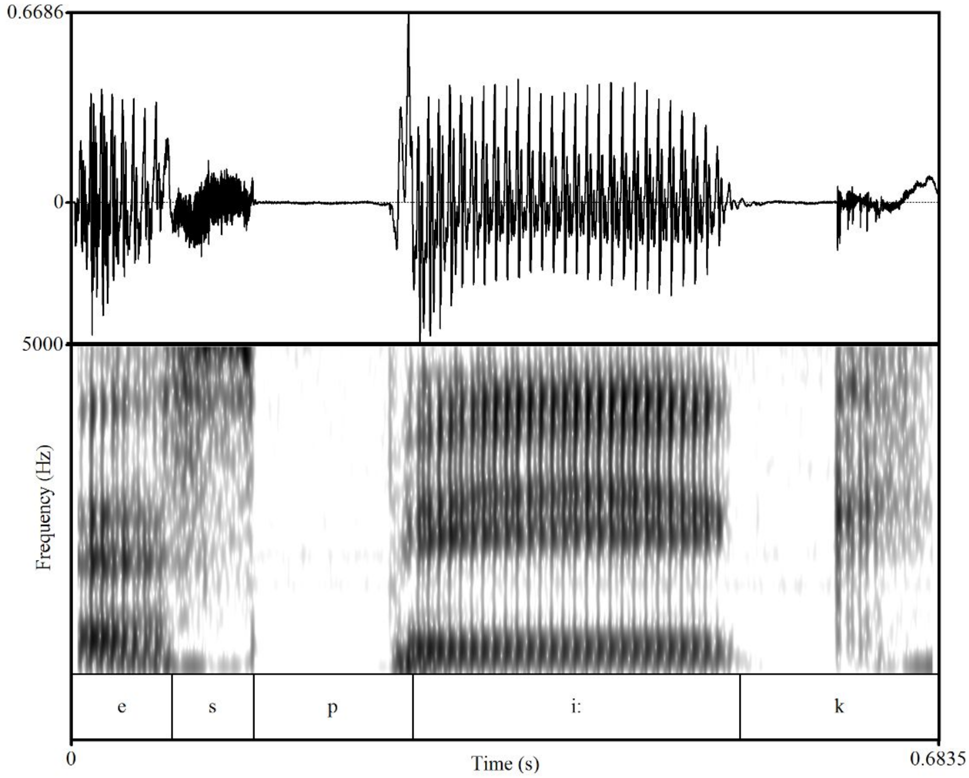


Figure 7: The word "speak" produced by a Spanish-English bilingual with prothesis.

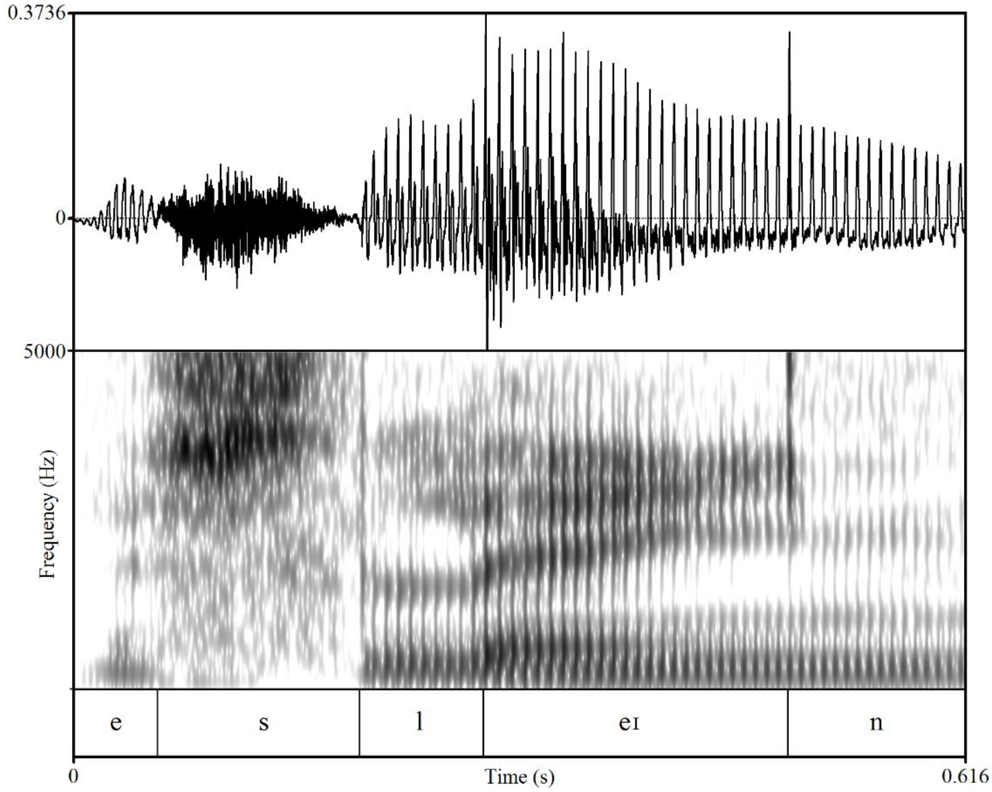


Figure 8: The word "slain" produced by a Spanish-English bilingual with prothesis.

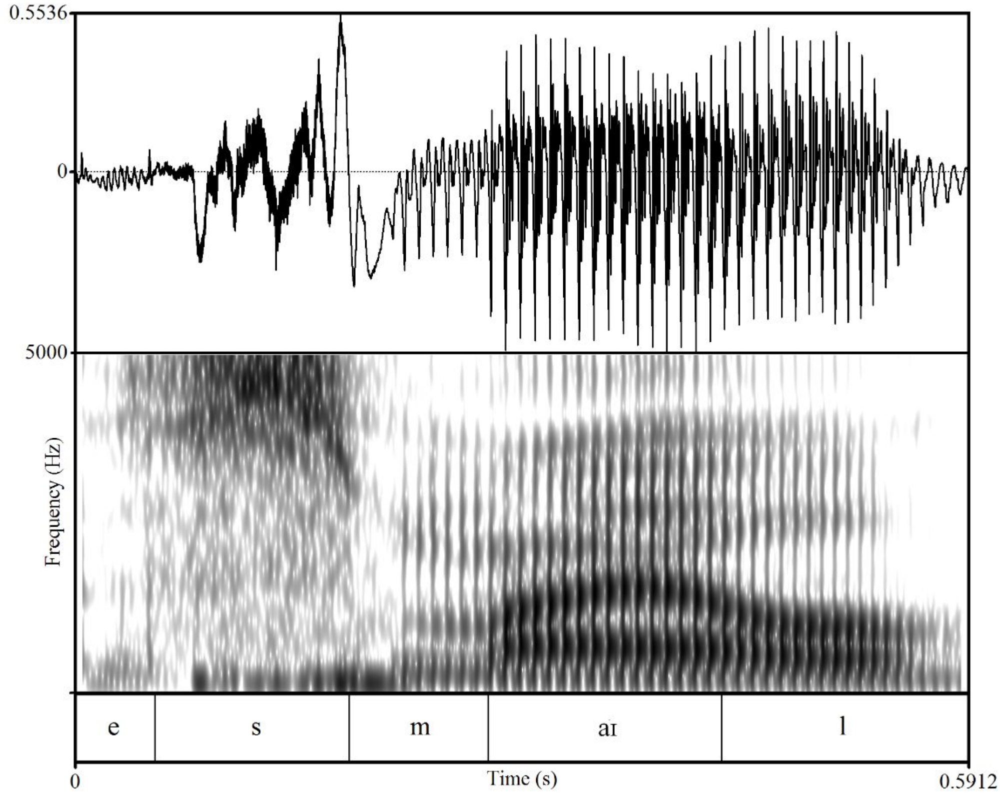


Figure 9: The word "smile" produced by a Spanish-English bilingual with prothesis.

Perception as a predictor for production

For the next analysis stage, we calculated a mean perception accuracy score on the critical Condition 2 contrast (e.g. *spi*-*espi*) in the perception experiment for each bilingual speaker. This score was included as a possible predictor in this second analysis of the production results, along with percentage of English daily use, age of English acquisition, and grammatical proficiency score (C-test result) in order to examine whether these factors would be influential in determining accuracy of pronunciation (i.e. either producing the prothetic vowel in the phonemic verbal fluency task or not). The maximal model tested was: prothesis ~ Percent of English in Daily Use + Age of Acquisition + English Proficiency +

Condition2PerceptionAccuracy + (1 | participant) + (1 | consonant.cluster)[1]. As before, all continuous predictor variables were scaled using the scale function in R. Models were manually stepped-down (using likelihood ratio tests) from maximal models containing all factors to the best fitting model that only contained significant predictors (Barr et al., 2013). The best fitting model determined by step-down pairwise model comparison was prothesis ~ Percent of English in Daily Use + English Proficiency + (1 | participant) + (1 | consonant.cluster). Neither age of English acquisition nor accuracy rate on condition 2 of the perception task was a significant predictor in the maximal model, and removing those terms from the maximal model did not decrease the model fit (as assessed by likelihood ratio tests) (all *p*s > 0.05). The parameters of the winning model are displayed in Table 8. As shown, both percentage of daily English use (β = -0.5454, z =-2.523, $p < 0.05$) and English proficiency (β = -0.6648, z =-3.057, $p < 0.01$) significantly influenced the amount of prothesis produced by the Spanish-English bilinguals. N. b. an analysis of duration of prothesis revealed similar overall findings.

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept) Prothesis	0.03271	0.41492	0.079	0.93716
Daily Eng Use	-0.5454	0.2162	-2.523	0.01165*
Eng Proficiency	-0.6648	0.2174	-3.057	0.00223**

Table 8: Estimates, standard errors, z values and p values of the best fitting model for prothesis produced by the Spanish–English bilinguals.

Figures 10 and 11 plot the relationships between the two significant predictor variables (respectively English daily use and English grammatical proficiency) and

[1] Interaction terms were not included in the maximal model as there were not enough observations to allow for so many parameters.

the proportion of words produced with prothesis for each speaker, with the grey ribbon representing 95% confidence interval. As displayed in Figure 10, a higher amount of English use was correlated with less prothetic vowel productions, i.e. a more English-like pronunciation (t = -2.1289, df = 32, $p < 0.05$, r=-0.35). As displayed in Figure 11, a higher grammatical proficiency in English was correlated with less prothetic vowel productions, i.e. a more English-like pronunciation (t = -2.5228, df = 32, p = 0.05, r=-0.41).

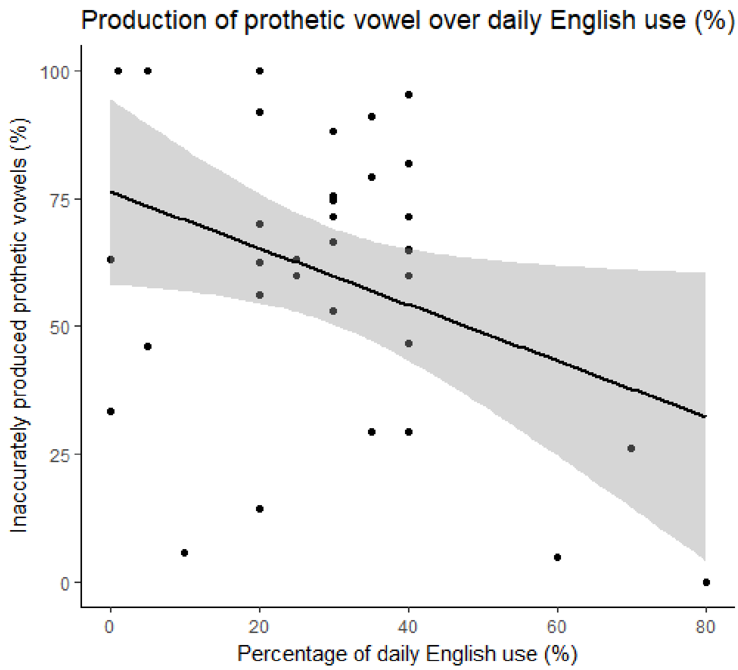


Figure 10: Production of prothetic vowel over daily English use (%).

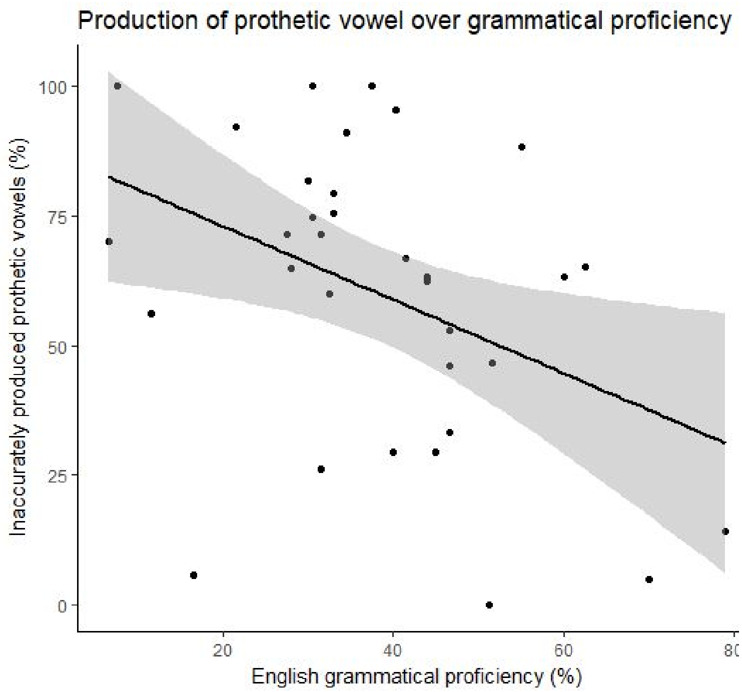


Figure 11: Production of prothetic vowel over grammatical proficiency C-test result (%).

Figure 12 plots the non-significant relationship between perception accuracy and prothesis proportion (t=1.4906, df = 32, p=0.1458). If the four quarters of the scatterplot are inspected, it is possible to see that the bottom right-hand quarter is almost empty, indicating that hardly anyone displayed accurate production and perception. Alternatively, in the top left-hand corner, many bilinguals displayed perception accuracy rates below chance level (i.e. beneath 50%), as well as high prothetic vowel rates. In the top right-hand quarter, bilinguals are visible who showed perception accuracy rates above chance, combined with high prothetic vowel rates, indicating that accurate perception was not sufficient for accurate production. In the bottom left-hand corner, bilinguals are visible who showed perception accuracy rates below chance, combined with low prothetic vowel rates, indicating that perception was not necessary for accurate production.

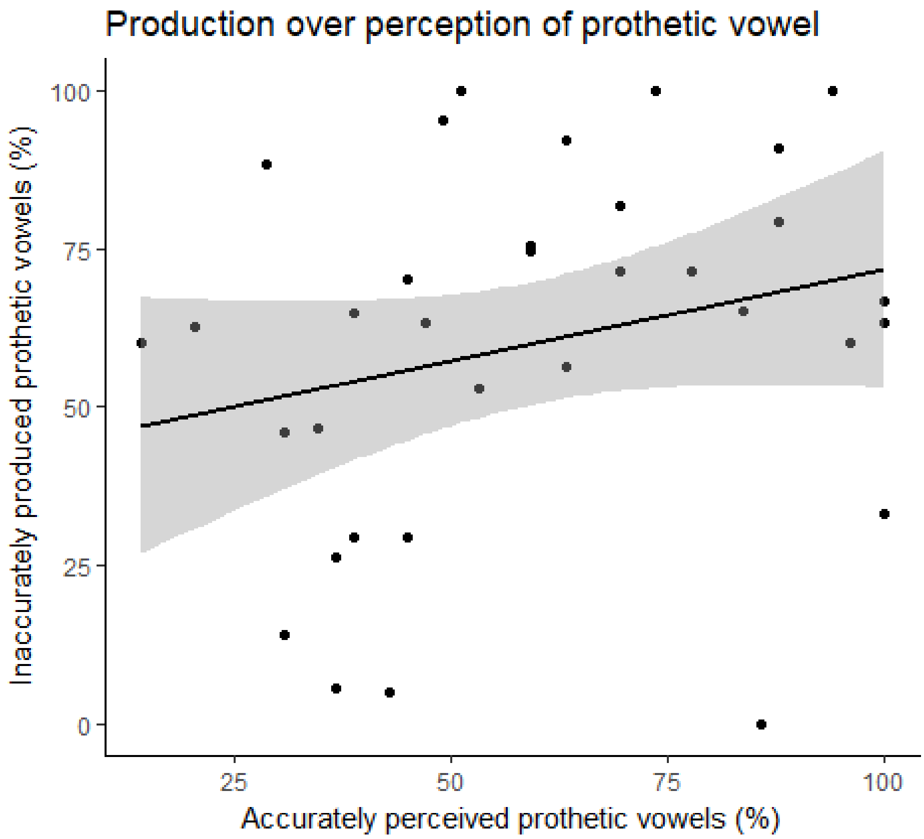


Figure 12: Production over perception of prothetic vowel (non-significant relationship).

Discussion of findings

In this study, speech perception was assessed through a same-different discrimination task in stimulus pairs such as *spi-espi* and speech production of vowel prothesis was assessed through a phonemic verbal fluency task in Spanish-English sequential bilinguals. The primary objective was to examine whether accurately perceiving the prothetic vowel preceding /sC/ clusters would likewise help to *not* produce this prothetic vowel in the L2. In general, the findings revealed that there was a significant difference between the monolingual English speakers and the Spanish-English bilinguals, and that there was no clear relationship between the perception and production of /sC/ clusters in onset position (which are restricted by Spanish phonotactic rules) by the Spanish-English bilinguals. Results indicated that the bilinguals were less successful at discriminating the difference between tokens like *spi-espi* than the English monolinguals, although, interestingly, the monolinguals were, like the bilinguals, also more accurate in the *spi-spu* and *spi-spi* trials than in the *spi-espi* trials. The reason the monolingual English speakers may have performed less accurately on the *spi-espi* trials than the *spi-spu* and *spi-spi* trials is potentially because, as observed by Carlson (2016), English favours #sC over #VsC. Nevertheless, the English monolinguals performed significantly more accurately than the Spanish bilinguals in discriminating *spi-espi* in comparison to the other two conditions. Furthermore, in this population of sequential Spanish-English bilinguals accurately *not* producing the prothetic vowel in the phonemic fluency task was *not* associated with accurate perception of the prothetic vowel.

Moreover, although a greater amount of English daily use and improved grammatical proficiency were associated with better production (i.e. *not* producing the prothetic vowel), these same predictor variables did not lead to better discriminatory abilities in the perception task. Therefore, the results suggest that with regard to this particular phonotactic constraint, production appears to be modified by increased exposure to the L2 and grammatical proficiency, whereas perception does not.

Moreover, the analysis of the individual bilinguals did not fully verify the assertion that accurate speech perception would be a prerequisite for accurate speech production. For example, some participants scored poorly on the production task, but above chance level on the perception task; alternatively, many participants scored highly on the production task, but below chance level on the perception task. Although many participants differed with regard to their accuracy on the perception versus production tasks, some sequential Spanish-English bilinguals achieved high accuracy results on both the perception and production tasks. Alternatively, other participants scored poorly on both tasks. These different profile patterns suggest that accurate L2 perception was neither necessary nor sufficient for accurate L2 production in the case of these sequential Spanish-English bilinguals, but that, as previously noted, increased exposure to English and improved grammatical proficiency could improve production results.

Therefore, as in previous research into the relationship between speech perception and production, no clear overall pattern was evident in terms of whether perception or production necessarily “comes first” (Baker & Trofimovich, 2006; Beach, Burnham, & Kitamura, 2001; Llisterri, 1995; Kartushina & Frauenfelder, 2014; Sheldon & Strange, 1982; Zampini, 1998). With regard to increased L2 exposure, it seems plausible that bilinguals exposed to more English would have been more likely to develop English-like pronunciation, and, likewise, it seems plausible that bilinguals with higher grammatical proficiency would likewise be more metalinguistically *aware* that in English /sC/ clusters in onset position are licit, and that this knowledge and exposure would enhance their L2 speech production, but not necessarily their performance in the speeded perceptual discrimination task.

Why did the bilinguals seem to be able to produce English L2 speech without the vowel in front of /sC/ clusters, but increased English use and improved English grammatical proficiency did not seem to improve their perception of this contrast? Firstly, in answering this question, it is important to remember that some participants *did* perceive a difference but nevertheless produced /sC/ clusters in the Spanish manner – so it would not be entirely correct to say that perceptual processes are always more robust and pervasive as far as L1 influence in L2 processing than is speech production. If that were the case then we would only have found evidence for enhanced speech production, but not for speech perception. What is more likely is that either the tasks assessed were actually quite different - and assessed speech perception and speech production differently - or, alternatively, other predictor variables, not assessed within this study, impact speech perception, rather than L2 use and L2 grammatical proficiency, which proved to enhance speech production. For example, it may be that some individuals simply *hear* differences between sounds better than others. We know as well that some people suffer in musical terms from amusia, i.e. tone deafness (Powell, 2016). Similarly, someone who is less "apt" at hearing differences between sounds might nonetheless be able to *not* produce the prothetic vowel in front of English /sC/ clusters with the metalinguistic knowledge that this sound combination is *possible* in English, and, with practice, become quite good at consistently *not* producing the vowel. Note that what is different about this speech production task is that accuracy was reached when *nothing* was pronounced, rather than when a sound was produced more "native-like", e.g. such as a more similar vowel quality. Alternatively, someone who is very "apt" at hearing differences between sounds (Powell, 2016) might essentially have a talent for hearing differences, which would be revealed in the perception task, but, when taking part in the production task, that same person might not perform particularly highly if not focussed, or if he or she didn't know the grammatical rule in English, or had not internalised that rule through practice. The point here is that different factors might lead to improved speech perception than lead to improved speech production, and the current study might only have examined factors which lead to improved speech production, i.e. no production.

It may also be that the striking interpersonal differences between bilingual participants arose due to different individual approaches to the tasks. Parlato-Oliveira et al. (2010) summarised in their work that an effect of L2 exposure was most clearly observed in their explicit task, but not in the implicit task, and suggested that L2 experience is more likely to affect explicit, or metalinguistic, perceptual tasks. With regard to our AX discrimination task, participants may have been metalinguistically aware of the phonotactic constraint it assessed (note that grammatical knowledge was correlated with production of the prothetic vowel), and the interpersonal differences may have been a result of some participants simply being more aware of this metalinguistic difference than others.

Furthermore, AX discrimination tasks require low-level perceptual judgments on whether auditory input is the same or different. This is arguably easier to discriminate than, for example, *whether* a vowel is present at the beginning of /sC/ or *whether* the auditory input forms a word or non-word, and hence the perception task may have assessed a rather low-level ability to acoustically discriminate, rather than the prothetic vowel - as a grammatical constraint - as such. Therefore, the interpersonal differences in the perception results may have been the result of some participants simply acoustically discriminating better than others, with no direct insight into their grammatical knowledge; however, the bilinguals were just as good as the English monolinguals at perceiving the *spi-spu* contrast, whereas there was a significant difference in their ability on the *spi-espi* task, so it cannot simply be the case that some participants simply acoustically discriminated better than others.

It may also be that some of the prothetic vowels were actually hesitation markers, and not reflective of grammatical constraints. It has been found that in spontaneous speech hesitation markers are often used and there are discrepant interpretations regarding their function, e.g. whether "um" and "uh" are used to "hold the floor" or whether they are symptomatic of cognitive processes on the part of the speaker, or both (Clark and Fox Tree, 2002; de Leeuw, 2007; Fox Tree, 2001; Maclay and Osgood, 1959; Shriberg and Lickley, 1993). The present methodology included all vocalic utterances preceding the sibilant as prothetic, and thus, a future analysis may consider including a read task as well, where hesitations are less likely.

It may also be, moreover, that the perception and production tasks recruited different cognitive functions. As previously discussed, the phonemic verbal fluency task is often used to assess executive control, as participants need to retrieve words, which requires them to access their mental lexicon, whilst focusing on the task, selecting words meeting certain constraints and inhibiting repetition (Miyake et al., 2000). This particular production task of eliciting words which begin with /sC/ clusters would require the Spanish L1 participants to shift, as they moved from one cluster to the next, monitor, to ensure that they didn't repeat words, and to remember new words; as well as *inhibit* their native Spanish language phonotactic constraint, as well as previously named words. In contrast, it could be argued that the perception task might have recruited a shifting mechanism, but that working memory and inhibition would have been less involved, i.e. the participants only had two stimuli to remember. Likewise, it may be more generally that the two capacities – L2 speech production and L2 perception - recruit different executive control mechanisms to varying extents, which could be an interesting research avenue to pursue in the future.

Essentially, the perception task in the current study was a 'can you do it' type task, whilst for the production task the participants were required to think quickly within a time constraint, which takes their focus off of their pronunciation. As such, in the current experiment, the production task was cognitively more costly, and it therefore seems possible that the relationship between production and perception might have been more clear cut if the tasks had required similar cognitive demands. That said, if it is accepted that the production task was more cognitively demanding than the perception task, and accurate perception is indeed a prerequisite for accurate production, we should nevertheless *not* have observed bilinguals who had high accuracy rates on the production task, but low accuracy rates on the perception task (i.e. bottom left-hand quarter in Figure 12). The fact that this quarter was indeed quite full of participants, indicates that accurate L2 speech perception is not necessary for accurate L2 speech production.

Perhaps it was not only the task which varied in terms of different executive control mechanisms, or whether it was explicit or implicit (although this is of course the case too), but that the individuals approached the same tasks differently. For example, in the case of the perception task, participants may have differed in the amount of attention they devoted to the task. Those participants who were very focused may have performed more accurately than those who were less focused, regardless of how proficient they were in their L2. Likewise, those participants who were more focussed on their pronunciation in the production task might have produced no prothetic vowel, whilst those participants who focussed more on the task of naming as many words as possible could have more frequently produced a prothetic vowel. Again, this would have little to do with their proficiency in English as such, but rather with how they approached the task, and may help to explain the differences observed between participants. Nonetheless, at face value, the results do not lend direct support to the notion that perception precedes production in L2 acquisition.

As such, the findings suggest that L2 speech models such as SLM and PAM, which postulate to different extents that perception is linked to production, may need revisiting. Indeed, some of the participants in the present study were able to *not* produce the prothetic vowel (i.e. accurately, as in English), but performed below chance with regard to the perception of the prothetic vowel in the same-different discrimination task. The findings from the correlational analysis, which indicated no significant relationship between perception and production, as well as in the mixed model analyses, which likewise did not indicate that perception abilities predicted production, both suggest that perception and production were not clearly related to one another with regard to the tasks at hand.

In brief, the results confirm a growing body of research examining the perception and production of individual L2 sounds which suggests dissociations between L2 perception and production. All in all, the results reveal that it is possible for sequential bilinguals to acquire knowledge of new permissible phoneme sequences, which violate their L1 phonotactic constraints, but that there is a great amount of interpersonal variability within regard to how the L2 acquisitional process unfolds.

References

Baker W and Trofimovich P (2006) Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL – International Review of Applied Linguistics in Language Teaching* 44(3): 231–250.

Barr D J, Levy R, Scheepers, C and Tily, HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3): 255-278.

Beach, E. F., Burnham, D. & Kitamura. C. (2001). Bilingualism and the relationship between perception and production: Greek/English bilinguals and Thai bilabial stops. *International Journal of Bilingualism*, *5*(2), 221–235.

Best CT (1995) A direct realist view of cross-language speech perception. In: Strange W (ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Baltimore: York Press, pp. 171–204.

Bialystok E, Craik F and Luk G (2008) Cognitive control and lexical access in younger and older bilinguals. *Journal of experimental psychology. Learning, memory, and cognition* 34(4): 859–873.

Boersma P and Weenink D (2010) *PRAAT*. University of Amsterdam. Available at: http://www.praat.org.

Cabrelli J, Luque A and Finestrat-Martínez I (2019) Influence of L2 English phonotactics in L1 Brazilian Portuguese illusory vowel perception. *Journal of Phonetics* 73: 55–69.

Carlson MT, Goldrick M, Blasingame M, et al. (2016) Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism: Language and Cognition* 19(5): 939–954.

Cheshire J, Kerswill P, Fox S, et al. (2011) Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2): 151–196.

Clark HH and Fox Tree JE (2002) Using uh and um in spontaneous speaking. *Cognition* 84(1): 73–111.

Cuetos F, Hallé P, Dominguez A, et al. (2011) Perception of prothetic /e/ in #sC utterances: Gating data. In: *ICPhS XVII Conference Proceedings*, Hong Kong, 17 August 2011.

de Leeuw E and Bogulski CA (2016) Frequent L2 language use enhances executive control in bilinguals. *Bilingualism: Language and Cognition* 19(5): 907–913.

de Leeuw E (2007) Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics* 19(2): 85–114.

de Leeuw E and Celata C (2019) Plasticity of native phonetic and phonological domains in the context of bilingualism. *Journal of Phonetics* 75: 88–93.

Dupoux E, Kakehi K, Hirose Y, et al. (1999) Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25(6). American Psychological Association: 1568.

Escudero P (2005) *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. University of Utrecht, Utrecht.

Flege JE (1987) The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.

Flege JE (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Maryland: York Press, pp. 233–277.

Flege JE, Bohn O-S and Jang S (1997) Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25(4): 437–470.

Flege JE, Schirru C and MacKay IRA (2003) Interaction between the native and second language phonetic subsystems. *Speech Communication* 40(4): 467–491.

Forster KI and Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35(1): 116–124.

Fowler CA (1996) Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America* 99(3): 1730–1741.

Fox Tree JE (2001) Listeners' uses ofum anduh in speech comprehension. *Memory & Cognition* 29(2): 320–326.

Freeman MR, Blumenfeld HK and Marian V (2016) Phonotactic constraints are activated across languages in bilinguals. *Frontiers in Psychology* 7.

Gollan TH, Montoya RI and Werner GA (2002) Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology* 16(4): 562–576.

Grogan A, Green DW, Ali N, et al. (2009) Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral Cortex* 19(11): 2690–2698.

Grosjean F (1998) Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1(02): 131–149.

Gullberg M and Indefrey P (2003) Language History Questionnaire. Developed in The Dynamics of Multilingual Processing. Max Planck Institute for

Psycholinguistics, Nijmegen. Available at: http://www.mpi.nl/departments/other-research/research-projects/the-dynamics-of-multilingual-processing/tools/Lang-Hist-Quest-Engl.pdf.

Hallé PA, Dominguez A, Cuetos F, et al. (2008) Phonological mediation in visual masked priming: Evidence from phonotactic repair. *Journal of Experimental Psychology: Human Perception and Performance* 34(1): 177–192.

Hualde JI (2005) *The Sounds of Spanish*. Cambridge, UK. Available at: http://www.cambridge.org/gb/academic/subjects/languages-linguistics/semantics-and-pragmatics/sounds-spanish (accessed 3 January 2018).

Kartushina N and Frauenfelder UH (2014) On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology* 5.

Kivistö-de Souza H (2015) *Phonological awareness and pronunciation in a second language*. PhD. Universitat de Barcelona, Barcelona. Available at: https://core.ac.uk/download/pdf/78632484.pdf.

Klein-Braley C (1985) A cloze-up on the C-Test: a study in the construct validation of authentic tests. *Language Testing* 2(1): 76–104.

Kuhl PK (2004) Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience* 5(11): 831–843.

Liberman AM and Mattingly IG (1985) The motor theory of speech perception revised. *Cognition* 21(1): 1–36.

Llisterri J (1995) Relationships between speech production and speech perception in a second language. In: *Proceedings of the 13th International Congress of Phonetic Sciences*, Stockholm, 1995.

Maclay H and Osgood CE (1959) Hesitation Phenomena in Spontaneous English Speech. *WORD* 15(1): 19–44.

Martínez-Celdrán E, Fernández-Planas AM and Carrera-Sabaté J (2003) Castilian Spanish. *Journal of the International Phonetic Association* 33(2): 255–259.

Mills L (2014) *An articulatory study of GOOSE-fronting in varieties of English*. PhD. Queen Margaret University, Edinburgh. Available at: https://eresearch.qmu.ac.uk/handle/20.500.12289/7678 (accessed 4 September 2018).

Miyake A, Friedman NP, Emerson MJ, et al. (2000) The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology* 41(1): 49–100.

Morales Pech D and Izquierdo J (2011) L2 phonology learning among young adult learners of English: Effects of regular classroom-based instruction and L2 proficiency. Revista Electrónica de Investigación Educativa 13(1). Available at: http://redie.uabc.mx/vol12no2/contentsmoralesizquierdo.html.

Newcombe F (1969) *Missle Wounds of the Brain. A Study of Psychological Deficits*. Oxford, UK: Oxford University Press.

Ohala JJ (1996) Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America* 99(3): 1718–1725.

Parlato-Oliveira E, Christophe A, Hirose Y, et al. (2010) Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. *The Journal of the Acoustical Society of America* 127(6): 3738–3748.

Portocarrero JS, Burright RG and Donovick PJ (2007) Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of clinical neuropsychology : The official journal of the National Academy of Neuropsychologists* 22(3): 415–422.

Powell J (2016) *Why We Love Music: From Mozart to Metallica - The Emotional Power of Beautiful Sounds*. London: John Murray. Available at: http://www.hodder.co.uk/HodderStoughton/books/detail.page?isbn=9781473613751 (accessed 10 September 2018).

Raatz U and Klein-Braley C (1981) *The C-Test--A Modification of the Cloze Procedure*. Not available separately; see FL 013 014.

Rubinstein JS, Meyer DE and Evans JE (2001) Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance* 27(4): 763–797.

Sandoval TC, Gollan TH, Ferreira, VS and Salmon, DP (2010). What causes the bilingual disadvantage in verbal fluency: The dual-task analogy. *Bilingualism: Language and Cognition*, 13: 231–252.

Sauzéon H, Raboutet C, Rodrigues J, et al. (2011) Verbal knowledge as a compensation determinant of adult Age differences in verbal fluency tasks over time. *Journal of Adult Development* 18(3): 144–154.

Sheldon A and Strange W (1982) The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3(3): 243–261.

Shriberg EE and Lickley RJ (1993) Intonation of Clause-Internal Filled Pauses. *Phonetica* 50(3): 172–179.

So CK and Best CT (2010) Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* 53(2): 273–293.

So CK and Best CT (2014) Phonetic influences on English and French listeners' assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition* 36(2): 195–221.

Tomás Navarro T (1918) *Manual de Pronunciación Española*. Reprinted in 1982. Madrid: CSIC.

Zampini ML (1998) The relationship between the production and perception of L2 Spanish stops. *Texas Papers in Foreign Language Education*: 1–18.

Appendix I

Language use question regarding daily use

Consider your present amount of English and Spanish language use. Divide 100% into English and Spanish use, e.g. if you only speak English presently, write "English=100%; Spanish=0%". If you also use another language, or other languages, include them in your division of 100%.
Alternatively, if you speak both English and Spanish equally, write "English=50%; Spanish=50%". Divide according to your own estimates, considering all aspects of language use, e.g. both passive (watching television) and active (speaking with parents).

____

C-tests

On the next pages you will find 2 small texts in total. Each text contains gaps where parts of some words haves been left out (no whole words are missing, though). Please try and fill in the gaps. In many cases there are several possibilities, so there are no right or wrong answers. Thank you very much for your help.

Text 1:
Two former US navy ships contaminated with chemicals were expected to arrive in the English Channel last night. The Maritime and Coastguard Agency sa____________________ the ves____________________ , at the cen____________________ of an enviro____________________ row, we____________________ being to____________________ through the cha____________________ , before hea____________________ up the east co____________________ to Hartlepool. Pl____________________ to dism____________________ them in north-east England have been she____________________ after being dee____________________ to fl____________________ international ru____________________. Last we____________________ , the gover____________________ said the ships could be sto____________________ in Hartlepool before go____________________ back acr____________________ the Atlantic.
Two former US navy ships contaminated with chemicals were expected to arrive in the English Channel last night. The Maritime and Coastguard Agency says the vessels, at the centre of an environmental row, were being towed through the channel, before heading up the east coast to Hartlepool. Plans to dismantle them in north-east England have been shelved after being deemed to flaunt international rules. Last week, the government said the ships could be stowed/stored in Hartlepool before going back across the Atlantic.

Text 2:
Don't get me wrong. I love magazines. I've been addicted to them since my teenage years. There's some____________________ about wom____________________ magazine superfi____________________ that I of____________________ enjoy. But

oh b____________________ , they are ju____________________ so, so frustr____________________ predictable. I rec____________________ you co____________________ cobble o____________________ together very eas____________________ in five min____________________. Take the co____________________ for example: the cover im____________________ , get a he____________________ and shou____________________ shot of a smi____________________ , heavily make-uped and airbr____________________ model (or optio____________________ a fam____________________ person).

Don't get me wrong. I love magazines. I've been addicted to them since my teenage years. There's something about women's magazine superficiality that I often enjoy. But oh boy, they are just so, so frustratingly predictable. I reckon you could cobble one together very easily in five minutes. Take the cover for example: the cover image, get a head and shoulder shot of a smiling, heavily make-uped and airbrushed model (or optionally a famous person).

Appendix II

Participant	Recorded in Spain (2) vs UK (1)	Amount of Daily English Use (%)	C-Test Result	Age of English Acquisition
1	2	25	32.5	4
2	2	20	44	6
3	2	30	55	6
4	2	20	79	7
5	2	5	46.5	6
6	1	40	51.5	5
7	2	70	31.5	6
8	1	10	16.5	30
9	2	40	28	11
10	2	35	45	8
11	1	60	70	7
12	2	40	40	1
13	1	20	6.5	10
14	2	25	60	3
15	1	40	40.25	5
16	2	20	37.5	8
17	2	30	46.5	8
18	2	30	30.5	10
19	1	30	33	7
20	2	20	21.5	6
21	2	20	11.5	7
22	2	40	27.5	7
23	2	40	30	3
24	2	1	7.5	5
25	2	30	31.5	5
26	2	40	62.5	7
27	1	80	51.25	7
28	2	35	34.5	8
29	2	35	33	6
30	2	5	30.5	6
31	2	40	32.5	10
32	2	0	46.5	10
33	2	0	44	11
34	2	30	41.5	4

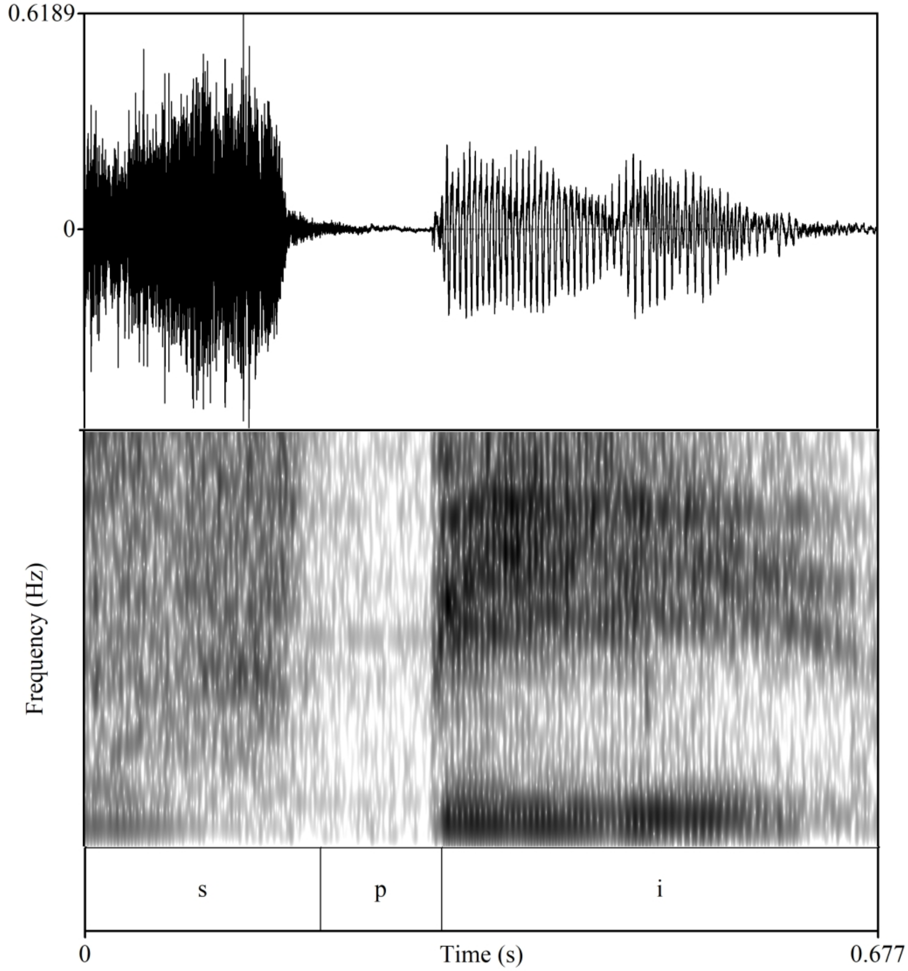


Figure 1

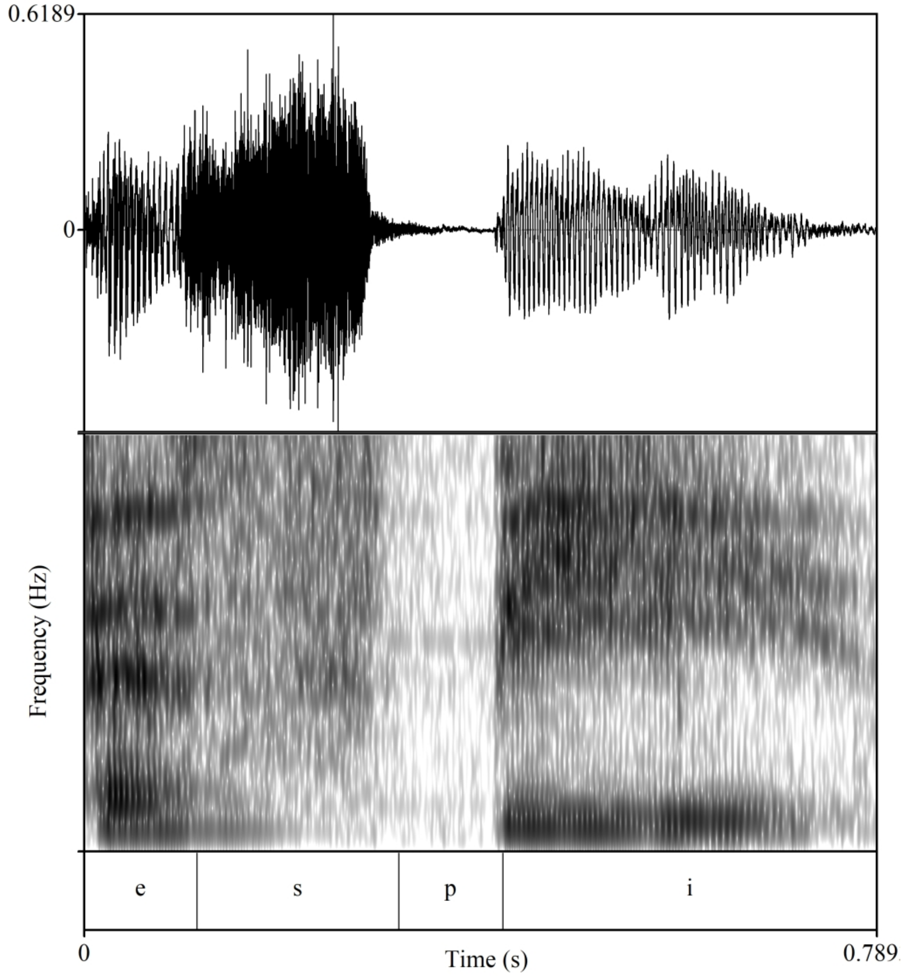


Figure 2

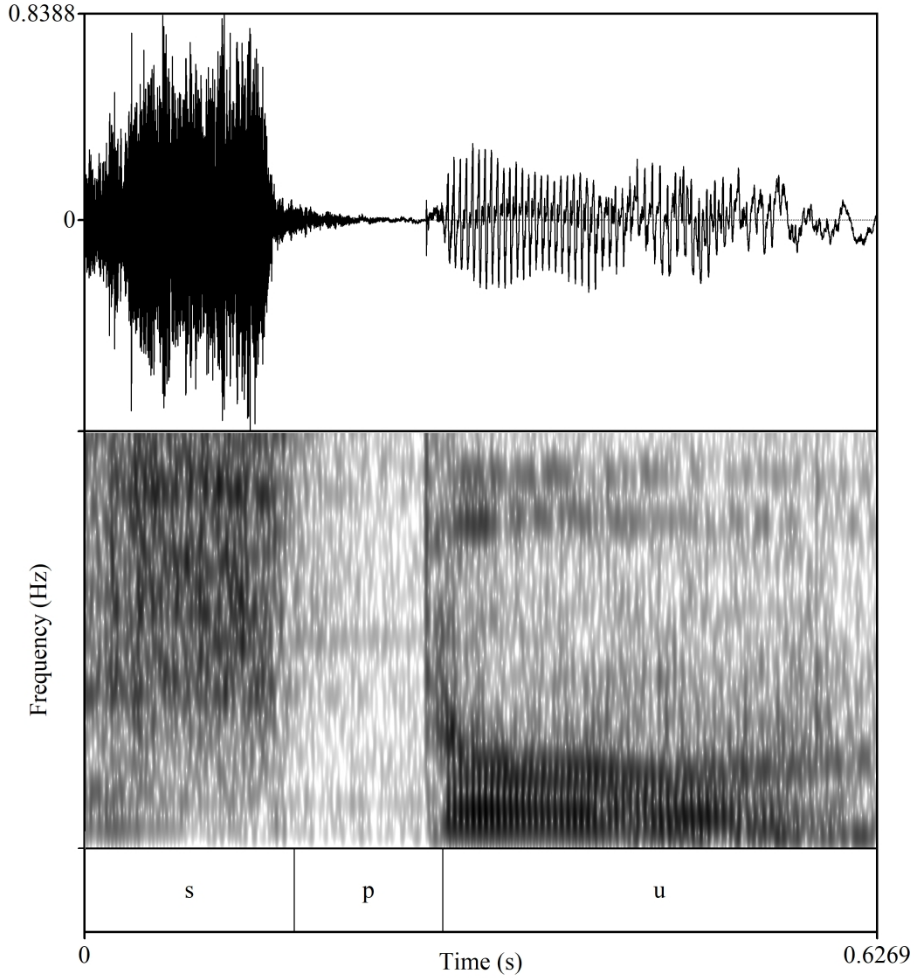


Figure 3

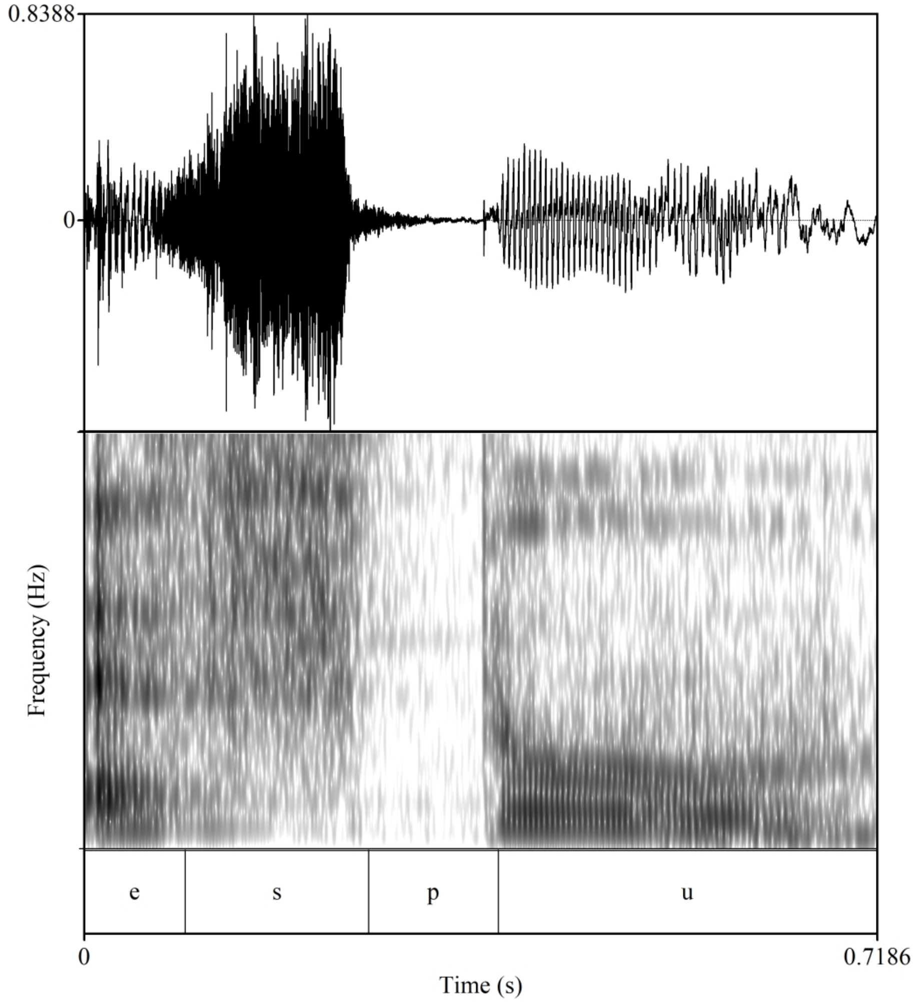


Figure 4

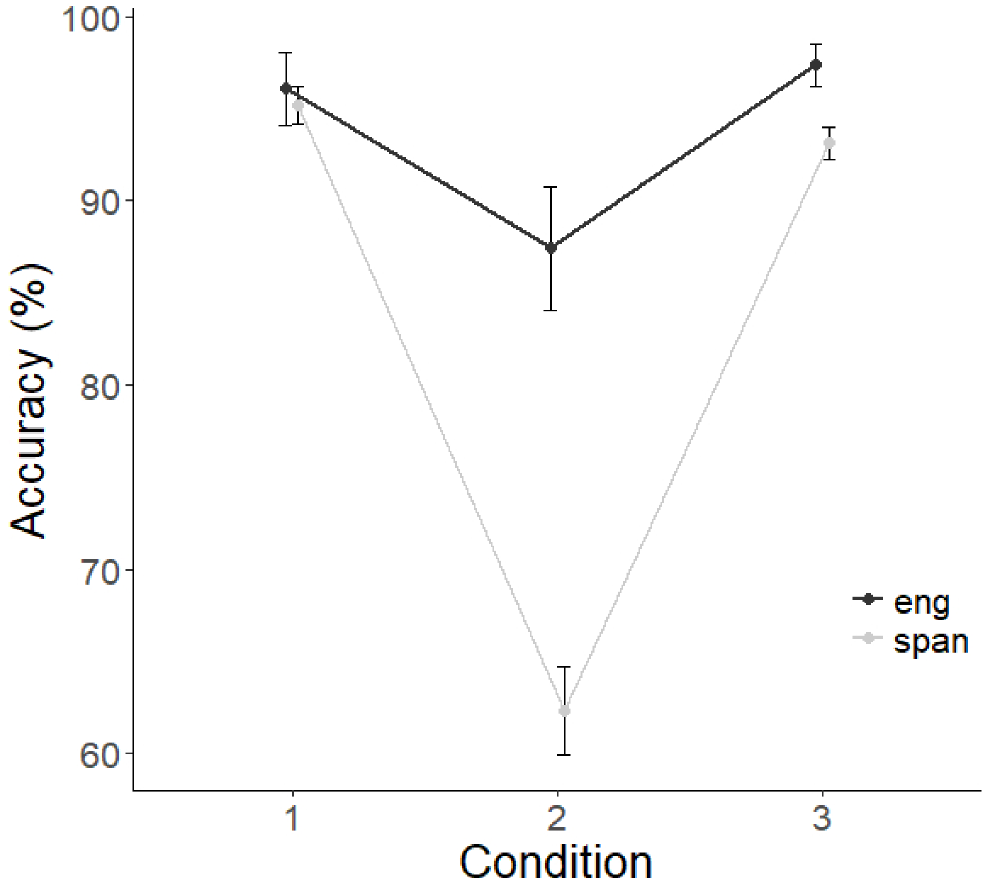
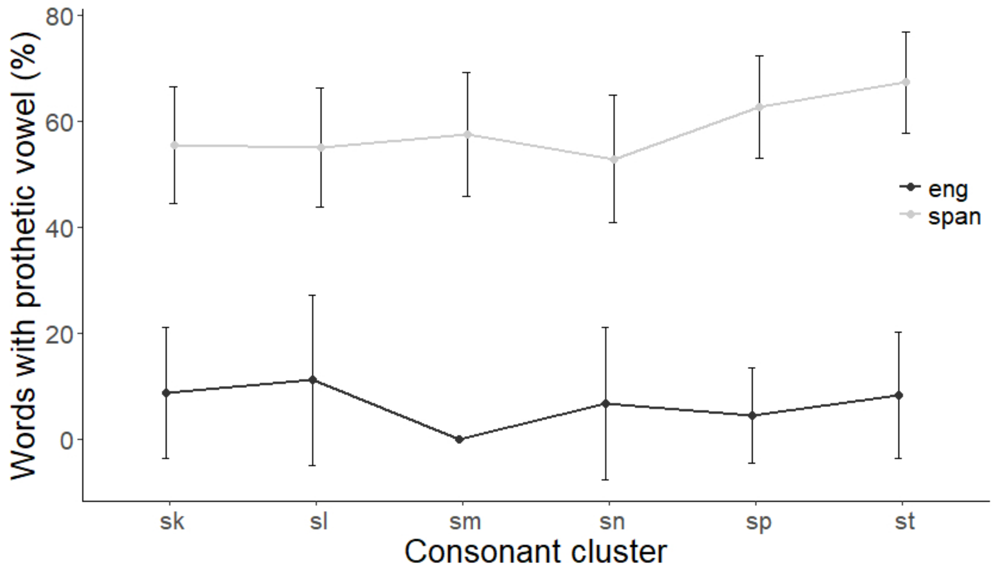


Figure 5



Words with prothetic vowel (%)
80
60
40
20
0
sk
sl
sm
sn
sp
st
Consonant cluster
eng
span

Figure 6

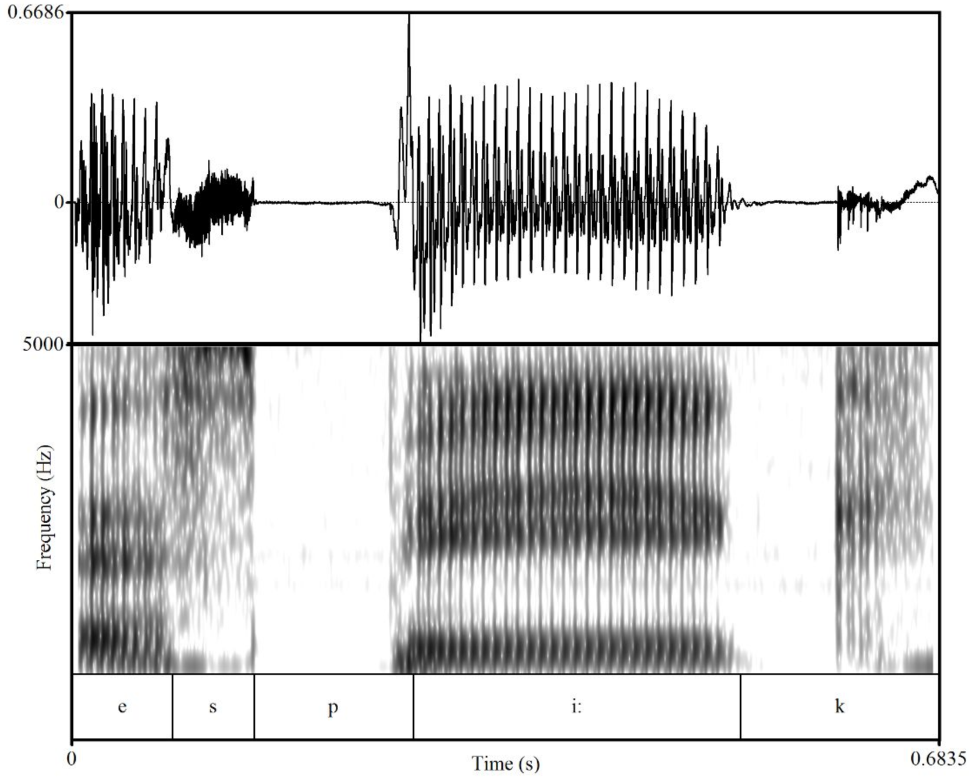


Figure 4: Example of a prothetic vowel, 'speak', produced by Participant 18.

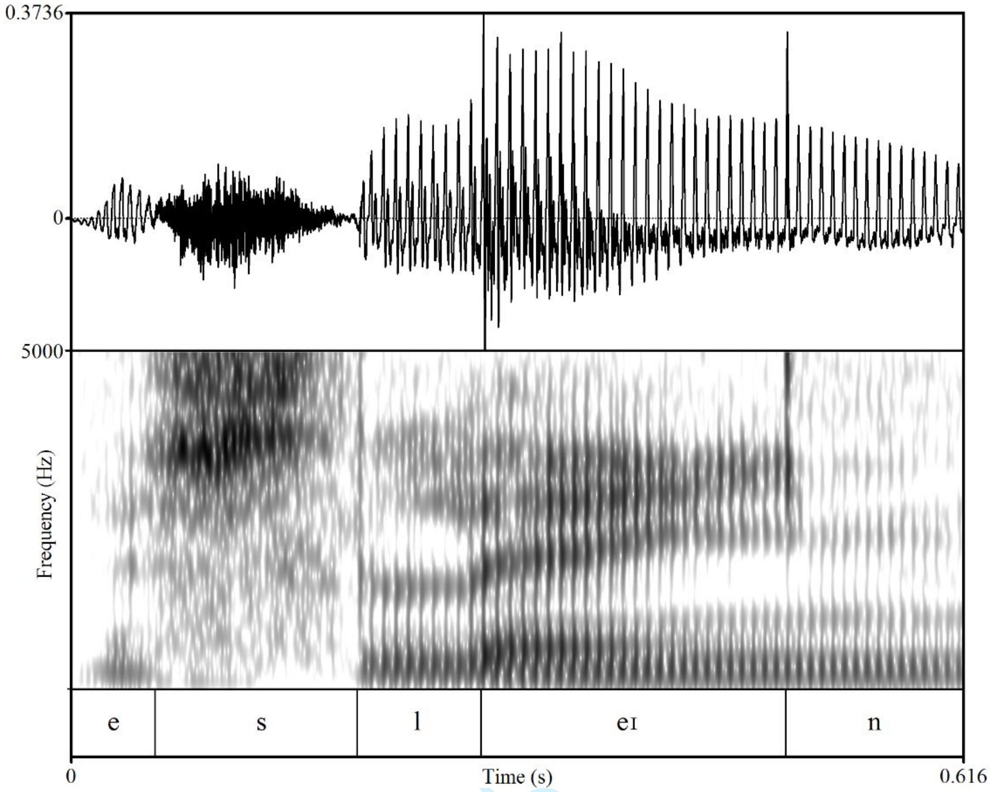


Figure 5: Example of a prothetic vowel, ‘slain’ produced by Participant 16.

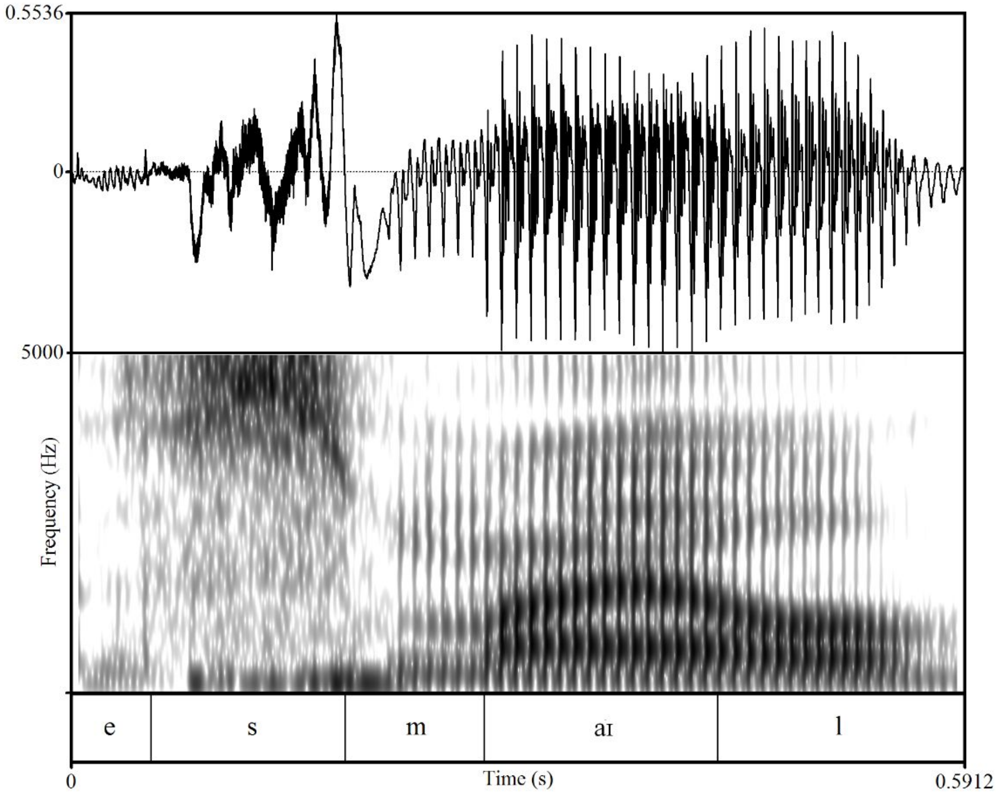


Figure 6: Example of a prothetic vowel, 'smile' produced by Participant 1.

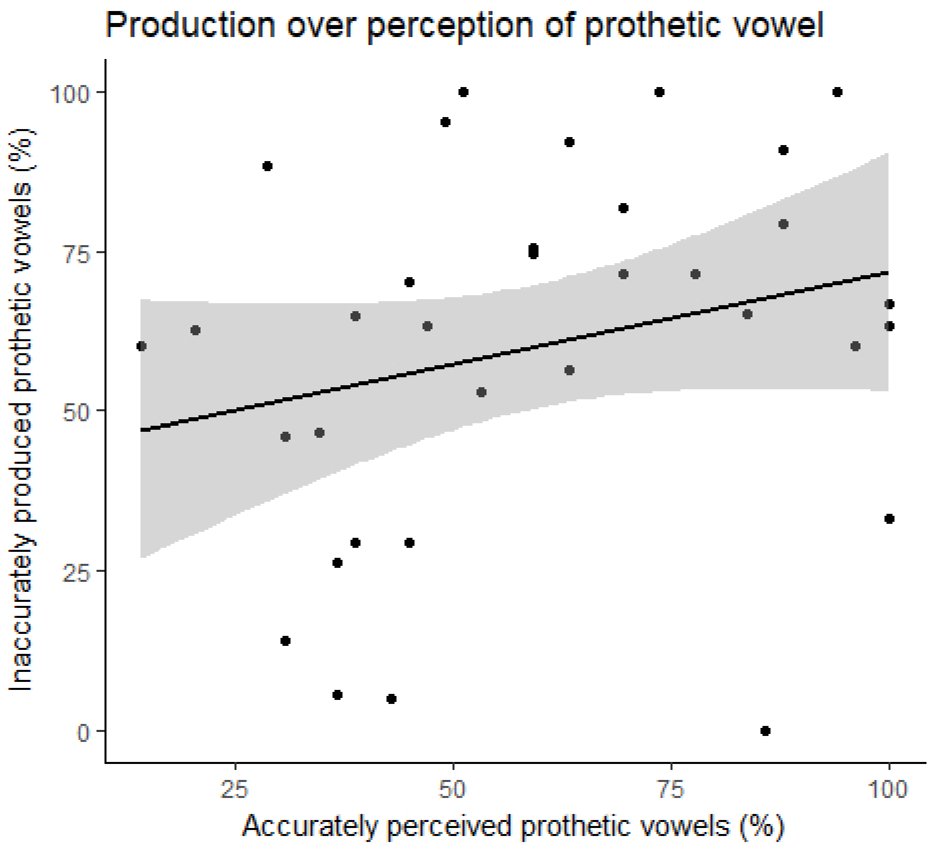

Production over perception of prothetic vowel
Inaccurately produced prothetic vowels (%)
100
75
50
25
0
25
50
75
100
Accurately perceived prothetic vowels (%)

Figure 10

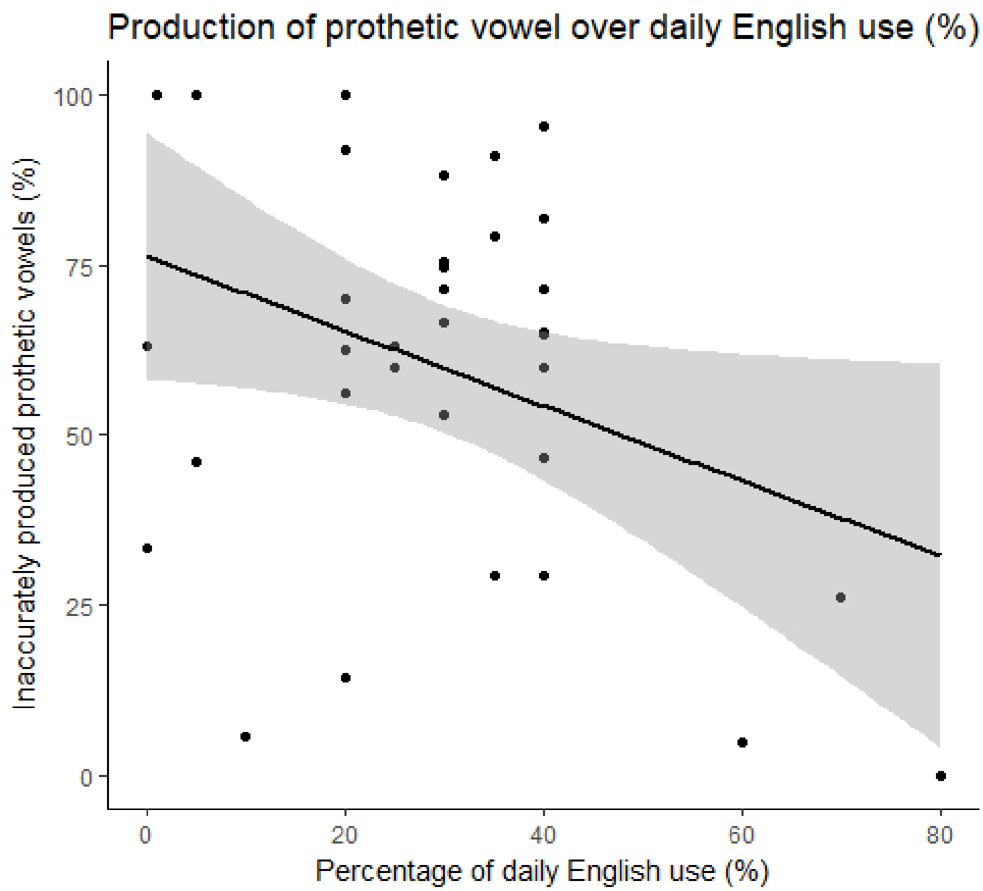

Production of prothetic vowel over daily English use (%)
Inaccurately produced prothetic vowels (%)
100
75
50
25
0
0
20
40
60
80
Percentage of daily English use (%)

Figure 11

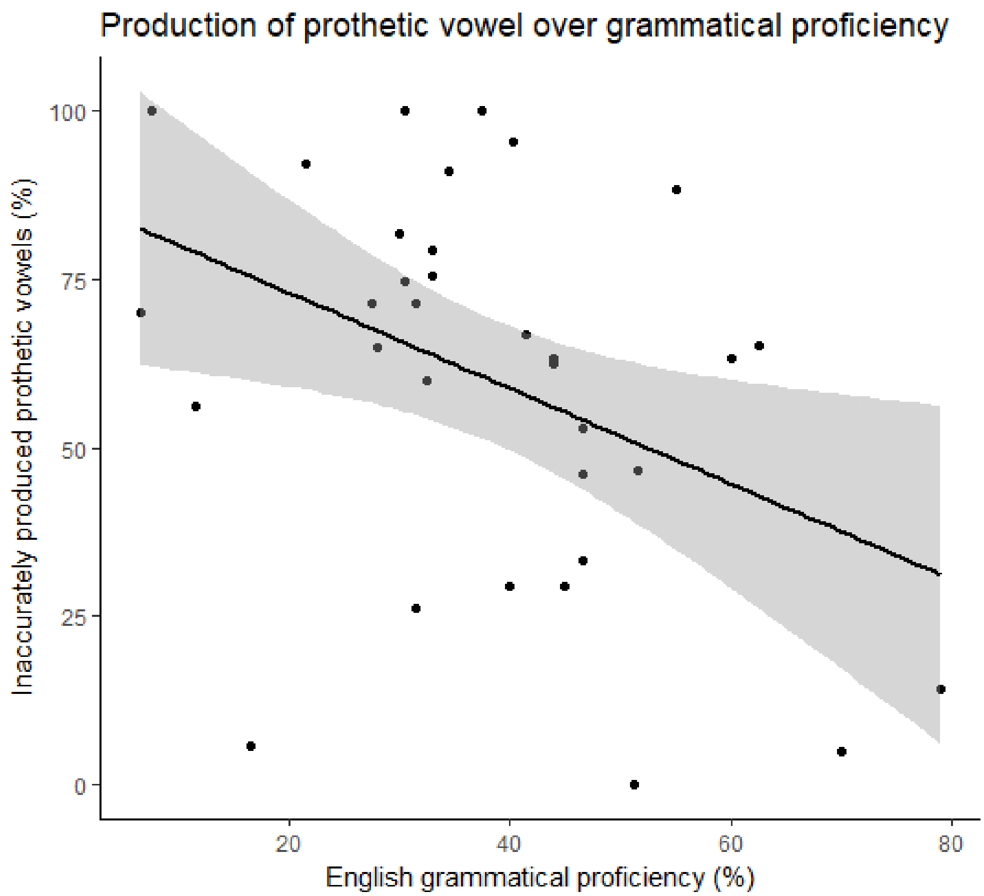

Production of prothetic vowel over grammatical proficiency
Inaccurately produced prothetic vowels (%)
100
75
50
25
0
20
40
60
80
English grammatical proficiency (%)

Figure 12