Language and Linguistics 79, 147-176 (2018)
DOI http://dx.doi.org/10.20865/20187906

The effect of 'illusory vowels' in Spanish-speaking second language learners of English

Maria Teresa Martinez Garcia
(Hankuk University of Foreign Studies)

Maria Teresa Martinez Garcia. 2018. The effect of 'illusory vowels' in Spanish-speaking second language learners of English. *Language and Linguistics* 79, 147-176. This paper shows that second-language (L2) spoken-word recognition is greatly influenced by differences between the native language (L1) and the second language (L2), possibly attributed to either L1-L2 syllable-structure or phonotactic differences. Spanish-speaking English learners (experimental group) and native English listeners (control group) completed an AXB task and a word-monitoring task in which they monitored /(ə)s+Consonant/-initial words in English. The results show a clear effect of L1 phonotactics, as the native speakers of English outperformed the Spanish group. These results indicate that L1-L2 syllable-structure differences or L1 phonotactics have pervasive consequences for spoken-word recognition, and effect that will be further explored in the discussion section of this paper.

Keywords : word recognition, syllable structure, phonotactics, consonant clusters, Spanish

1. Introduction

Learning a second language (L2) to the same level as a native speaker seems to be a hurdle difficult to overcome for adult L2 learners, particularly in the domain of phonology. Research in L2 acquisition has generally found that while children show great ability to perceive and produce foreign speech sounds, this ability decreases as age of acquisition increases, with adult L2 learners typically retaining a foreign accent. Among several factors (see Bohn & Munro, 2007), two have been held responsible for these learning difficulties in adulthood: (i) the so-called 'Critical Period'; and (ii) the entrenchment of the native language (L1) after years of exposure to and use of it.

On the one hand, the Critical Period Hypothesis (CPH) stipulates that changes in the neurological plasticity of the brain, which result in cerebral lateralization after the onset of puberty, lead to a reduction in the ability of adult L2 learners to learn a foreign language (e.g., Lenneberg, 1967; Patkowski, 1990; Scovel, 1969, 1988). This hypothesis is biological in nature and attributes foreign accent in adult L2 learners to the maturation of their brain: Unlike children, who have been exposed to the L1 since infancy and eventually reach native mastery of the L1 sound system (pathological cases notwithstanding), L2 learners whose exposure to the L2 began after the onset of puberty retain a foreign accent, irrespective of how long they have known the foreign language for and how proficient they are.

On the other hand, foreign accents have been attributed to the interaction between the phonetic system of the L2 and the one already established for the L1, suggesting that foreign accents are caused by the entrenchment of the L1 after years of exposure to and

use of it (e.g., Flege, 1995; Flege, Munro, & Mackay, 1995; Flege, Yeni-Komshian, & Liu, 1999). Then, the main difference between children acquiring their L1 and adults acquiring an L2 is that adults perceive and produce L2 sounds with reference to the already established L1 phonetic categories. Thus, learners' firmly established L1 phonetic system would be the main factor responsible for a foreign accent. Unlike the CPH, the L1 entrenchment account predicts a decline in human speech learning ability only for certain L2 sounds, those that are incorrectly perceived by learners due to their L1 phonetic categories. With enough exposure to and experience with the L2, adult learners can master the perception and production of novel L2 phones. However, even at advanced levels of proficiency, the perception and production of L2 sounds that are similar to L1 counterparts but realized in a phonetically different manner remains challenging for learners (e.g., Best, 1995; Best, McRoberts, & Sithole, 1988; Best & Tyler, 2007).

Importantly, in the acquisition process, L2 learners need to acquire not only the segmental contrasts of the new language, but also how segments fit together in permissible combinations, also known as phonotactics. For example, Spanish and English share the phonemes /s/ and /t/, but only English allows them to be combined in syllable- and word-initial position (e.g., Spanish: *estudiar* 'to study'; English: *study*). Research has found that the perception of an L2 is affected not only by the segmental properties of the L1 (e.g., Best, 1995; Best & Tyler, 2007; Best et al., 1988; Flege, 1995; Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999; Goto, 1971; Schmidt & Flege, 1995), but also by how segments can be combined in relation to word and syllable boundaries (e.g., Dehaene-Lambertz, Dupoux, & Gout, 2000;

Dupoux, Kaheki, Hirosi, Pallier, & Mehler, 1999; Dupoux, Pallier, Kaheki, & Mehler, 2001; Kabak & Idsardi, 2007; Matthews & Brown, 2004). Hence, language-specific rules governing how segments can be combined can have important effects on how listeners perceive auditory words.

However, less clear is how L1 phonotactics impact the recognition of L2 words in continuous speech, as most existing studies on the topic have looked at speech perception in isolated words (e.g., Dupoux et al., 1999, 2001; Kabak & Idsardi, 2007). The aim of the current study is thus to investigate the effect of L1 phonotactics on late L2 learners' processing of the L2 continuous speech signal, focusing on /s/-initial consonant clusters in English as perceived by adult Spanish L2 learners of English.

2. Effects of L1 Phonotactics on L2 Speech Production and Perception

L1 phonotactics have been proposed as an explanation for some of the difficulties in the production of L2 sounds. L2 learners' foreign accent can be particularly noticeable when L2 learners try to produce a sequence of sounds not permitted in their L1. In L2 production, some possible repairs of L1 phonotactic violations include consonant deletion, prothesis, metathesis, and vowel epenthesis (e.g., Abrahamsson, 1999; Anderson, 1987; Broselow & Finer, 1991; Carlisle, 1997, 1998; Davidson, 2006; Davidson, Jusczyk, & Smolensky, 2004; Eckman, 1991; Eckman & Iverson, 1993; Hancin-Bhatt & Bhatt, 1998).

Vowel epenthesis is the most commonly used strategy for the production of L2 sounds (e.g., Carlisle, 1997) and is employed in

different situations, one of them being in loanword phonology. For example, while Spanish and English share phonemes such as /s/ and /l/, only English allows them to be combined in word-initial position. Importantly, English words introduced in the Spanish language undergo a vowel-epenthesis process so that they conform to the phonotactics of Spanish (e.g., the word for *slogan* in Spanish is *eslogan*, as described in the dictionary of the Real Academia de la Lengua[1]). Similarly, Japanese and Korean do not allow consonant clusters in onset position, so Japanese and Korean speakers also produce vowel epenthesis in borrowed words (e.g., the word for *basket* in Japanese is [basuketto]) (Rose & Demuth, 2006). Notice that the location of the vowel epenthesis is not the same across languages, with Spanish inserting a vowel *before* the two consonants (e.g., *eslogan*) and Korean and Japanese inserting a vowel *between* the two consonants (e.g., [sɨtɨraikɨ] in Korean; [sutoraiku] in Japanese). If we take the assumption that loanword phonology represents the initial stage of L2 acquisition (e.g., Broselow, 2004; Escudero & Boersma, 2004), this vowel-epenthesis process shows how non-native sound structure is adapted to L1 phonotactics.

Vowel epenthesis is also commonly used by L2 learners in speech production to repair sound sequences that otherwise would not be phonologically legal in the L1 (e.g., Abrahamsson, 1999; Anderson, 1987; Broselow & Finer 1991; Carlisle, 1997, 1998; Davidson et al., 2004; Eckman, 1991; Eckman & Iverson, 1993; Hancin-Bhatt & Bhatt 1998). Using the example described earlier, English but not Spanish allows /s/ and many consonants to be combined in syllable- and word-initial position. Thus, Spanish-speaking L2

1) The dictionary of the Royal Spanish Academy, the official royal institution responsible for overseeing the Spanish language.

learners of English commonly introduce an epenthetic vowel before /s/-initial consonant clusters in their oral productions, with a word like *school* being produced as [əskul] instead of [skul] (e.g., Abrahamsson, 1999; Carlisle, 1997, 1998).

L1 phonotactics constraints influence not only L2 speech production, but also L2 speech perception. For example, using a cross-linguistic speech perception experiment, Dupoux et al. (1999) examined the transfer of L1 phonotactic constraints in the perception of unfamiliar words. In this study, Japanese speakers were tested on the perception of consonant clusters, as their L1 has a rather restrictive syllable structure. Their perception results were compared with those of French speakers, whose L1 has a more permissible syllable structure. Dupoux et al. (1999) presented participants with a series of nonce words ranging from *ebzo* to *ebuzo*, in which the medial vowel was shortened in segments of five pitch periods, ranging from a full vowel to no vowel. Unlike French listeners, most of the Japanese heard an 'illusory vowel' in the illegal consonant sequence, even when there was no vocalic segment between the two consonants. These findings were taken as evidence that Japanese listeners were perceptually "repairing" an illegal sequence of sounds as a result of L1 phonotactics. These results with Japanese speakers have been replicated in several studies (e.g., Dehaene-Lambertz et al., 2000 using event-related potentials (ERPs); Dupoux, Pallier, Kakehi, & Mehler, 2001, testing the influence of "top-down" lexical effects; Matthews & Brown, 2004, using an AX discrimination task).

Further extending this line of research, Kabak and Idsardi (2007) looked at the perception of illusory vowels in word-medial clusters by Korean L2 learners of English. The authors wanted to determine whether the perception of illusory vowels is best explained by: (i)

the non-occurrence of certain consonants in coda position (e.g., *[c], *[g]); or (ii) consonantal contact restrictions that prohibit the combination of certain heterosyllabic consonants (e.g., *[k.m]; *[l.n]). The results of an AX discrimination paradigm suggest that L1 syllable structure constraints (i.e., (i)) rather than consonantal contact restrictions (i.e., (ii)) is responsible for these L2 learners' perception of epenthetic vowels. Further research on this topic indicated that Korean-speaking L2 learners' perception of initial consonant clusters is further influenced by the acoustic properties of the stops used in these complex consonant clusters (e.g., Lee, 2016).

However, there is still an important question, which remains unanswered. This question regards the potential effect of L1 phonotactics on the recognition of L2 words in continuous speech, not just in isolated words. If epenthetic vowels are perceived in isolated words, they should also be perceived in continuous speech. This misperception in speech could thus potentially activate competitor words, making lexical access less efficient and L2 speech processing more difficult (see Broersma & Cutler, 2011; Escudero, 2007; Weber, & Cutler, 2004).

With the purpose of better understanding how perceptual difficulties may affect word recognition in continuous speech and where these perceptual difficulties come from, the current study investigates the perception of /s/-initial consonant clusters by Spanish L2 learners of English, using an AXB task and a word-monitoring task.

3. Experiment 1: AXB Task

3.1. Participants

Thirty-two native speakers of English (16 females; mean age=23 years) and 32 Spanish-speaking L2 learners of English (12 females; mean age=24 years) participated in this study. The native speakers were students at a midwestern university in the USA. Spanish speaking L2 learners were tested in the same institution (20 participants) and in Spain, with a very large range of proficiencies (from really low proficient scoring 5 to 48 out of 50 possible points in a cloze test, Brown (1980)).

3.2. Materials

Participants first completed an AXB discrimination task. The AXB task was chosen over other discrimination tasks (e.g., AX or ABX task), because this paradigm is less likely to be subject to response biases, and because it creates a more balanced memory load between A and X and B and X than an ABX task.

Sixteen nonce word stimuli that began with an /s/-initial consonant cluster were created for this task. Nonce words were chosen so that participants could not make use of lexical information when judging the presence or absence of an epenthetic vowel. Each stimulus pair either contained a schwa (e.g., /əslɛn/) or did not contain a schwa (e.g., /slɛn/) before the consonant cluster. The nonce words followed the phonotactics of the English language without resembling any real English word.

The study also included thirty-two fillers. These nonce word fillers were divided in two conditions: The first filler condition

included stimulus pairs that contained either /b/ or /v/ (e.g., /mibɹɛz/ vs. /mivɹɛz/), a contrast that is difficult for Spanish L2 learners of English as these two sounds are not phonemic in their L1; the second filler condition included stimuli that differed in the presence of one phoneme (e.g., /snun/ vs. /snu_/). To make sure participants were paying close attention to the complete word, the sound contrasts of these two filler conditions appeared in different positions in the word (initial, middle, or final).

The resulting words—stimuli and fillers—were checked by a native English speaker to confirm that they followed the phonotactic rules of English. To ensure that participants would not rely on the physical (i.e., acoustic) properties of the stimuli to do the task, three different speakers were recorded uttering the stimuli. The stimuli were always presented in the same order, with A being produced by Speaker 1 (Midwest dialect), X being produced by Speaker 2 (East coast dialect), and B being produced by Speaker 3 (Midwest dialect). The test items were presented in a Latin square design, such that participants would hear either A or B as X but not both.

3.3. Procedure

The stimuli were presented using Paradigm by Perception Research Systems, Inc. (Tagliaferri, 2005). Participants were comfortably seated in a quiet room and they were instructed to listen carefully to a series of three nonce words and to choose whether the second word (X) was more similar to the first or to the third word (A or B). An experiment trial would look like: /əslɛn/ (A) /əslɛn/ (B) /slɛn/ (C), with the correct answer being A in this case. The inter-stimulus interval was 1,000 ms. Participants made

their decision by pressing one of the two buttons of the mouse: If they thought the middle word was more similar to the first word (AX), they pressed the left button; if they thought that the second and third words were more similar (XB), they pressed the right button. The next trial started immediately after the participants entered their response. A practice session of six stimuli with feedback preceded the main session of the experiment (which did not have any feedback). All trials were randomized across participants.

3.4. Data Analysis

The participants' accuracy was analyzed with a logistic regression model (cf. Baayen, 2008), using the *glm* package (Everitt & Hothorn, 2006) in *R* (R Development Core Team, 2009). L1 was considered as a categorical predictor with two levels (English vs. Spanish), with the English group representing the baseline. The effect of the predictor was assessed using log-likelihood tests comparing models with and without the predictor. Two sets of models were run—one on all the accuracy rates with L1 as predictor, and one on the L2 learners' accuracy rates with proficiency as a predictor. The effect of L2 proficiency was assessed by comparing models that included proficiency with models that did not include it; in each case, the model with the best fit was kept. Since proficiency did not improve the L2 model, only the analysis of all the accuracy data is reported. Participants, item and list were included as random variables.

3.5. Results

Figure 1 presents the mean accuracy results for the two groups, and Table 1 presents the results of the logit mixed-effects model for all participants' accuracy.

Figure 1. Mean accuracy (standard errors) of the two groups in the AXB task

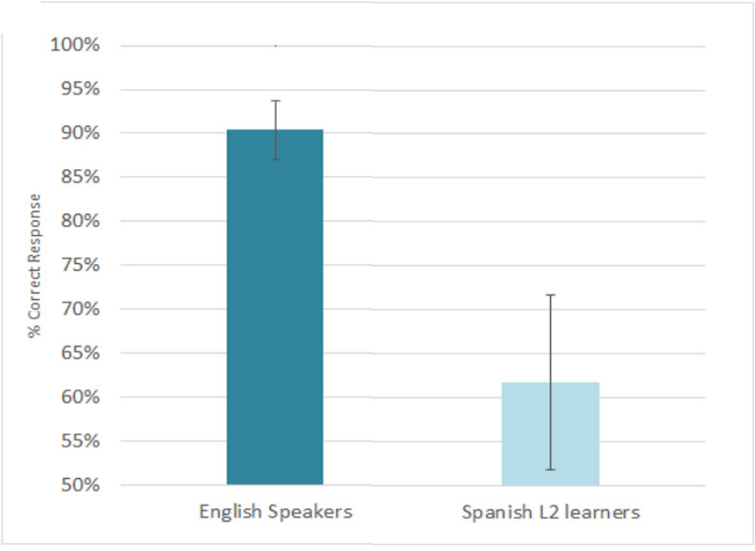

Table 1. Logit Regression Model on All Participants' Accuracy Results

Variable	Estimate (SE)	*z*	*p*
(Intercept)	2.24 (.15)	14.95	〈.001
L1: Spanish L2 learners	-1.77 (.2)	-10.07	〈.001

Note: *df = 1536; a = .05*

The model summarized in Table 1 revealed a main effect of language group, indicating that Spanish L2 learners of English were

statistically less accurate than native English speakers in discriminating stimuli with /s/-initial clusters that contained or did not contain a word-initial schwa.

3.6. Discussion

In this first experiment, we used an AXB task to examine whether Spanish L2 learners of English would show some difficulty in the perception of /s/-initial clusters in English. We found that native speakers of Spanish showed less accurate perception of the distinction between these two forms. The main conclusion we can draw from this set of results is the importance of the more abstract phonotactic constraints of each L1. The Spanish speakers' perception of the stimuli was shaped to conform to the phonotactic constraints of their L1, in which /s/-initial (and all sibilant-initial) consonant clusters are illegal. This perception problem mirrors the production errors that are commonly associated with Spanish L2 learners of English (e.g., Carlisle, 1997).

The results of this discrimination task are in line with results that emphasize the influence of syllable structure on L2 learners' speech perception (e.g., Matthews & Brown, 2004; Kabak & Idsardi, 2007). However, without a comparison without another group of L2 learners whose L1 allows this syllable structure, it is impossible to determine whether L1 syllable structure or L1 phonotactics constraint L2 speech perception (being this a limitation of this study). Moreover, it remains to be determined how these perception errors affect L2 word recognition. From the literature, we know that difficulty in correctly perceiving L2 sounds can increase lexical competition and make word recognition less efficient (e.g., Broersma, & Cutler, 2011; Escudero, 2007; Weber,

& Cutler, 2004). Experiment 2 was thus created to examine how the perception of illusory vowels might impact L2 word recognition.

4. Experiment 2: Word Monitoring Task

4.1. Participants

The participants who completed Experiment 1 also took part in Experiment 2.

4.2. Materials

All participants completed a word-monitoring task. The word-monitoring paradigm is used to study aspects of speech processing such as how orthographic information is identified in visual words, how (semantic, syntactic, and pragmatic) context contributes to word identification, and how acoustic and phonetic information influences spoken word recognition and lexical access. In such paradigms, participants are required to monitor ongoing language input for a pre-designated target word (for a review, see Kilborn, & Moss, 1996).

In Experiment 2, two independent variables were manipulated: whether or not the word to be monitored contained a vowel (e.g., *specially* vs. *especially*), and whether or not the word heard in the speech signal matched the word to be monitored (i.e., match, mismatch). Table 2 illustrates these four conditions with an example.

Table 2. Example of Stimuli Used in the Word Monitoring Task

Condition	Written Target	Auditory Stimulus
Match Stimulus with vowel	*ESPECIALLY*	*I prepared that especially for you.*
Match Stimulus without vowel	*SPECIALLY*	*I prepared that specially for you.*
Mismatch Stimulus with vowel	*SPECIALLY*	*I prepared that especially for you.*
Mismatch Stimulus without vowel	*ESPECIALLY*	*I prepared that specially for you.*

Forty-eight experimental pairs (36 minimal pairs, 12 near-minimal pairs) that began with (e)/s/-initial cluster consonants were used. Each stimulus pair either contained or did not contain a schwa before the consonant cluster (e.g., *specially* vs. *especially*). In order to reach the number of forty-eight experiment pairs, inflected and derived forms of the same words were included (e.g., *estate* vs. *state* and *estates* vs. *states* were included). The critical words were inserted into semantically ambiguous sentences. Semantically ambiguous sentences were chosen to avoid participants making use of lexical information when judging the presence or absence of the word in the auditory stimuli. The sentences were manipulated so that the critical word appeared in different regions of the sentence (initial, medial, or final) and never appeared after a word-final schwa.

Experiment 2 also included ninety-six fillers. These fillers had the same characteristics as those discussed for Experiment 1. The first filler condition included stimulus pairs that differed in whether they contained /b/ or /v/ (e.g., *boat* vs. *vote*). The second filler condition included stimuli that differed in the presence of one phoneme (e.g., *bead* vs. *bee*), to make sure participants were

paying attention to all stimuli presented to them. The sound contrasts of these two filler conditions appeared in different positions in the word (word-initial, word-medial, or word-final).

All sentences were checked by two native English speakers to ensure that they were plausible and completely ambiguous for either word in the stimulus pairs. The sentences were then recorded by a female native speaker of American English with a Midwestern accent (Speaker 2 of Experiment 1). A Latin square design was used so that participants would not see or hear the same test item in more than one condition. All trials were randomized across participants.

4.3. Procedure

The stimuli were presented using Paradigm by Perception Research Systems, Inc. (Tagliaferri, 2005). Each trial was as follows. First, participants would see the target word appear in the middle of the screen in capital letters; the word was present on the screen for 1,000 ms and disappeared as the audio started playing (e.g., ESPECIALLY). Second, participants would listen to a sentence that would be played immediately after the presentation of the word. The sentence may or may not have included the word they just saw (e.g., *I prepared that especially for you* or *I prepared that specially for you*). Participants were asked to decide whether the sentence contained the word they saw on the screen by pressing the button "SI" ("yes" labelled on the left button of the mouse) as soon as they recognized the target word in the sentence, or otherwise press "NO" (labelled on the right button of the mouse) after listening to the whole sentence. The next trial began immediately after participants entered their response. A practice

session of six stimuli with feedback preceded the main session of the experiment (which did not have any feedback).

4.4. Data Analysis

Participants' accuracy in the word-monitoring task was analyzed with a logistic regression model (cf. Baayen, 2008), using the *glm* package (Everitt & Hothorn, 2006) in *R* (R Development Core Team, 2009). L1 was a categorical predictor with two levels (English vs. Spanish), with English serving as baseline. The models also examined the effects of the presence or absence of vowel (e.g., *especially* vs. *specially*) in the word to be monitored and the match between the word to be monitored and the word in the auditory stimulus (match vs. mismatch). The presence vs. absence of a vowel was coded as -0.5 if participants heard a vowel in the auditory stimulus and 0.5 if there was no vowel in the auditory stimulus. The vowel conditions were contrast coded rather than dummy coded because it was not clear which condition should be the baseline. The "match" condition was coded as 0 and the "mismatch" condition as 1. The "match" condition was treated as baseline, because it was predicted to be the condition in which the two groups would perform more similarly. Two sets of models were run—one on all the accuracy rates with L1 as predictor, and one on the L2 learners' accuracy rates with proficiency as a predictor. The effect of L2 proficiency was assessed by comparing models that included this variable with models that did not include it; in each case, the model with the best fit was kept. Since proficiency did not improve the L2 model, only the analysis of all the accuracy data is reported. Participant, item, and list were included as random variables.

4.5. Results

Figure 2 presents the mean accuracy of the two L1 groups, and Table 3 presents the results of the fit linear mixed-effects model using generalized least squares for all participants' accuracy.

Figure 2. Mean accuracy of the two groups in the word monitoring task

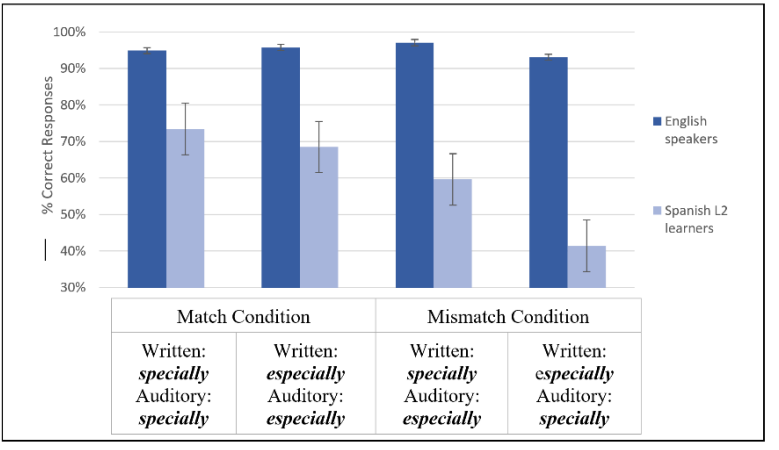

Table 3. Logit Regression Model on All Participants' Accuracy Results

Variable	Estimate (SE)	*z*	*p*
(Intercept)	3.02 (.2)	17.58	〈.001
Match	.03 (.3)	0.14	〉.1
Vowel	-.19 (.3)	〈|1|	〉.1
L1: Spanish L2 learners	-2.62 (.2)	-14.1	〈.001
Match x Vowel	1.1 (.5)	2.19	〈.05
Match x L1: Spanish	.06 (.3)	〈|1|	〉.1
Vowel x L1: Spanish	-.86 (.4)	2.29	〈.05
Match x Vowel x L1: Spanish	-1.46 (.5)	-2.68	〈.01

Note: *df = 4608; a = .05*

The results of the model summarized in Table 4 revealed the following effects: A main effect of L1, indicating that Spanish-speaking L2 learners of English behaved differently from the native English speakers; an interaction between vowel and L1, indicating that unlike native English speakers, Spanish speakers showed different accuracies when the words contained vs. did not contain a vowel; and a three-way interaction between match, vowel, and L1, indicating that Spanish speakers' difficulty in identifying whether or not a vowel was present differed depending on whether the auditory word matched or mismatched the written word.

Hence, two follow-up fit linear mixed-effects models were run to test for the effect of vowel separately for the match and mismatch conditions only for the Spanish speakers' data. These models are shown in Table 4 and Table 5 for the match and mismatch conditions, respectively.

Table 4. Logit Regression Model on Spanish Speakers' Accuracy Results, Match Condition.

Variable	Estimate (SE)	*z*	*p*
(Intercept)	.62 (.06)	11.15	〈.001
Vowel	.08 (.03)	2.41	〈.02

Note: *df = 768; a = .05*

Table 5. Logit Regression Model on Spanish Speakers' Accuracy Results, Mismatch Condition.

Variable	Estimate (SE)	*z*	*p*
(Intercept)	.6 (.05)	12.85	〈.001
Vowel	-.15 (.04)	4.12	〈.001

Note: *df = 768; a = .05*

The results of the models summarized in Tables 4 and 5 indicate that the effect of vowel is significant in both match and mismatch

conditions: In the match condition, the presence of a vowel hampers the identification of the corresponding word, by producing less accurate responses (Table 4). By contrast, in the mismatch condition, it is the absence of a vowel what yields worse accuracy results; in other words, it is when the competitor with a vowel is activated that Spanish L2 learners of English have difficulty determining whether the word contained a vowel.

4.6. Discussion

In Experiment 2, we used a word-monitoring task to examine the effect of perceptual difficulties on word recognition. We found that native English speakers could easily detect the target word in the auditory stimuli, independently of the presence or absence of a vowel. By contrast, native Spanish speakers had difficulty doing so and were less accurate in detecting the target words, particularly when there was a mismatch between the word they saw and the word they heard in the sentence. These results are consistent with two scenarios. First, L1 phonotactics constraints may be pervasive for L2 word recognition. Second, it is violations of L1 phonotactic constraints rather than violations of L1 phonotactics that cause difficulties in L2 word recognition. However, without a group comparison with another group of learners whose L1 allows for a similar syllable structure in word- and syllable-initial position, these two possibilities cannot be teased apart. A third possibility could be that these results can be, at least partly, attributed to L2 proficiency. However, considering the large range of proficiencies included in the study, and the lack of a proficiency effect (proficiency did not improve the statistical models), this possibility seems unlikely, although it should be further explored in future

studies.

As a result of this perception problems, Spanish L2 learners of English may inappropriately activate lexical competitors when they hear words that contain an /s/ cluster in near-word-initial position. The Spanish speakers also showed an effect of vowel that was in the opposite direction for the match and mismatch conditions: In the match condition, they were more accurate when listening to words that did not contain a vowel, whereas in the mismatch condition, they were more accurate when listening to words that contained a vowel. Another way to describe these results is that participants were more accurate whenever they saw a word without a vowel on the computer screen. Due to the difficulty in finding (near-)minimal word pairs, the words without vowels were in fact more frequent (based on a corpus search using the subtitle token corpus in EsPal (Duchon, Perea, Sebastián, Martí, & Carreiras, 2013)) than the words with a vowel. Hence, this frequency effect may be driving some of the Spanish speakers' results. In other words, although /s/-initial words may be more difficult to perceive accurately than words that contain a vowel, this difficulty may be somewhat offset by the higher frequency of these words, with Spanish listeners being perhaps more accurate in determining whether the auditory stimulus matches or mismatches /s/-initial words.

Similar activation of unintended lexical competitors has been reported in the word recognition literature on the effect of L1-L2 category assimilation (e.g., Broersma & Cutler, 2011; Escudero, 2007; Pallier, Colomée, and Sebastián-Gallés, 2001; Weber & Cutler, 2004). When perceiving spoken words, listeners must match the incoming auditory information with the lexical representations stored in memory. Word recognition models (e.g., Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994) stipulate that the

input activates multiple lexical candidates in parallel, which compete for selection. When a person hears speech segments, these segments activate words in the lexicon that overlap with them. When perceiving L2 words, L1 phonotactics may thus result in listeners activating competitor words that are in fact not present in the signal. These perceptual difficulties make L2 word recognition far less efficient. The perceived epenthetic vowel for the Spanish L2 learners of English in this study activates the lexical competitor with an actual vowel: When Spanish speakers saw a word spelled with a vowel (e.g., *especially*) and then heard a sentence with the other member of the minimal pair (e.g., *I prepared that specially for you*), their performance was not even at chance level. Even when hearing vowel-initial stimuli, the Spanish speakers in this study activated words with /s/-initial clusters. These results indicate that not just phonetic categories, but also phonotactics (or syllable structure), influence L2 word recognition (see Weber & Cutler, 2004).

All in all, the L1-induced lexical confusion and consequent lexical competition seems to explain the results obtained in Experiment 2. However, and although this explanation matches the results obtained in the study and is consistent with findings previously reported in the literature, the confusion shown by the Spanish learners could also be explained to be perceptual because the task could be performed without consulting the lexicon (as the carrier sentences were semantically neutral). Another possibility could be, thus, that L2 leaners did not rely on lexical representations to perform the task, and only paid attention to the acoustic properties of the acoustic input.

One important limitation of the current study is that, without comparison with a group of learners whose L1 allows for this type of

syllable structure, it becomes difficult to attribute the main findings of the study to the L1-specific phonological/syllable representation. At least in part these results could also be attributed to a general proficiency effect (although no proficiency effect was reported in the study) or to the effect of individual segments (rather than of syllable structure) on erroneous perception.

5. General Discussion and Conclusion

The present study is the first to systematically examine whether Spanish L2 learners of English misperceive segmental sequences that are not licensed by their L1 phonotactics, here /s/-initial clusters in syllable-onset (and thus word onset) position, and how these perceptual difficulties may affect L2 word recognition. Spanish and English share the phonemic segments investigated in the current study, but only English allows them to be combined in syllable-initial and thus word-initial position. /s/-initial clusters are known to be difficult (at least in production) for Spanish L2 learners of English, even at advanced levels of proficiency (e.g., Abrahamsson, 1999; Carlisle, 1997, 1998).

The results of our AXB task (Experiment 1) showed that only Spanish L2 learners of English had difficulties discriminating /s/-initial clusters preceded vs. not preceded by an epenthetic vowel, in line with other studies that argue that L1 phonotactic (or syllable structure) constraints have an important effect on speech perception (e.g., Kabak, & Idsardi, 2007; Matthews, & Brown, 2004).

Despite the extensive research on differences among phonemic repertoires and how the mismatch between two languages can

produce speech perception problems, to date, very little work had been dedicated explicitly to address the effect of these constraints on L2 word recognition. Experiment 2 investigated how L2 learners' perceptual difficulties affected their L2 word recognition. The results indicated that not just phonetic categories, but also L1 phonotactic constraints, influence L2 word recognition. These findings, together with previous evidence (e.g., Weber & Cutler, 2004), suggest that non-native listeners' perceptual difficulties have important consequences for L2 word recognition by potentially activating unintended (or "phantom") lexical competitors and creating lexical confusion for L2 learners (e.g., Broersma & Cutler, 2011). However, when processing input in a real-word situation, listeners could make use not only of bottom-up acoustic phonetic information, but also of top-down contextual cues. In the current study, stimuli were semantically ambiguous so that only bottom-up acoustic phonetic information could help disambiguate between the two forms tested (with vs. without a vowel). Future research should examine whether other types of cues (e.g., semantic or syntactic cues) constrain lexical activation in L2 speech processing.

Finally, in the present study, proficiency did not improve the statistical models run for either experiment. These findings may suggest that the effects of the L1 phonotactics on L2 speech perception exist across proficiency levels and survive even at advanced levels. Alternatively, the cloze test used in the current study to specify the proficiency of the participants may not have been an adequate measure of aural proficiency. This cloze test is conducted in the visual modality rather than in the aural modality; thus, it may not have been sensitive enough to capture variability in L2 learners' perceptual skills. Future studies should examine the extent to which native-like production and perception of /s/-initial

consonant clusters is indeed acquirable for Spanish L2 learners whose proficiency is assessed aurally.

References

Abrahamsson, N. (1999) Vowel epenthesis of/sC (C)/onsets in Spanish/Swedish interphonology: longitudinal case study. *Language learning 49*(3): 473-508.

Anderson, J. (1987) The markedness differential hypothesis and syllable structure difficulty. In G. Ioup, & S. Weinberger (eds.), *Interlanguage phonology: The acquisition of a second language sound system*. Cambridge, MA: Newbury House.

Baayen, R. H. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Best, C. T. (1995) Chapter 6: A Direct Realist View of Cross-Language Speech Perception. *Speech perception and linguistic experience: Issues in cross-language research*, 171-204.

Best, C. T. (1995) A direct realist view of cross-language speech perception, in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), 171 - 204.

Best, C. T., McRoberts, G. W., · Sithole, N. N. (1988) Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology 14*: 345 - 360.

Best, C. T., · Tyler, M. D. (2007) Nonnative and second-language speech perception: Commonalities and complementarities. Language experience in second language speech learning: In honor of James Emil Flege, 13-34.

Bohn, O. S., · Munro, M. (Eds.). (2007) *Language experience in second language speech learning in honor of James Emil*

Flege. Amsterdam: Benjamins.

Broersma, M., · Cutler, A. (2011) Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology 64*: 74-95.

Broselow, E. (2004) Unmarked structures and emergent rankings in second language phonology. *International Journal of Bilingualism 8*: 51-65.

Broselow, E., · Finer, D. (1991) Parameter setting in second language phonology and syntax. *Second Language Research 7*(1): 35 - 59.

Brown, J. D. (1980) Relative merits of four methods for scoring cloze tests. *Modern Language Journal 64*: 311 - 317.

Carlisle, R. (1997) The modification of onsets in a markedness relationship: Testing the interlanguage structural conformity hypothesis. *Language Learning 47*: 327-361.

Carlisle, R. S. (1998) The acquisition of onsets in a markedness relationship. A longitudinal study. *Studies in Second Language Acquisition 20*: 245 - 260.

Davidson, L. (2006) Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics 34*(1): 104 - 137.

Davidson, L., Jusczyk, P., · Smolensky, P. (2004) The initial and final states: Theoretical implications and experimental explorations of richness of the base. In R. Kager, W. Zonneveld, & J. Pater (eds.), *Fixing priorities: Constraints in phonological acquisition*. Cambridge: Cambridge University Press.

Dehaene-Lambertz, G., Dupoux, E., · Gout, A. (2000) Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience 12*:

635-647.

Duchon, A., Perea, M., Sebastián, N., Martí, M. A., · Carreiras, M. (2013) EsPal: one-stop shopping for Spanish word properties. *Behavior Research Methods 45*(4): 1246 - 58. http://doi.org/10.3758/s13428-013-0326-1

Dupoux, E., Kaheki, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion. *Journal of Experimental Psychology: Human Perception and Performance 25*, 1568 - 1578.

Dupoux, E., Pallier, C., Kakehi, K. · Mehler, J. (2001) New evidence for prelexical phonological processing in word recognition. Language and Cognitive Processes 16(5-6): 491-505.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*(1), 23-41.

Eckman, F., & Iverson, G. (1993) Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research 9*: 234 - 252.

Escudero, P. (2007). Multilingual sound perception and word recognition. Stem, *Spraak en Taalpathologie*, 93-103.

Escudero, P., · Boersma, P. (2004) Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition 26*: 551 585.

Everitt, B., · Hothorn, T. (2006) *A Handbook of Statistical Analyses Using R*. Chapman & Hall/CRC, Boca Raton, FL.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems, in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Baltimore): 233 -

273.

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.

Flege, J. E., MacKay, I. R., & Meador, D. (1999) Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America 106*: 2973.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5), 3125-3134.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of memory and language, 41*(1), 78-104.

Goto, H. (1971) Auditory perception by normal Japanese adults of the sounds ″l″ and ″r″. *Neuropsychologia 9*(3): 317 - 323.

Hancin-Bhatt, B., · Bhatt, R. (1998) Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition 19*: 331 - 378.

Kabak, B., & Idsardi, W.J. (2007) Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech 50*: 23-52.

Kilborn, K., & Moss, H. (1996) Word monitoring. *Language and Cognitive Processes 11*(6): 689-694.

Lee, G. (2016) Use of acoustic cues in the perception of complex syllable structure. *Language and Linguistics 70*: 393-422.

Lenneberg, E. H. (1967) The biological foundations of language. *Hospital Practice 2*(12): 59-67.

Marslen-Wilson, W.D. (1987) Functional parallelism in spoken

word-recognition. *Cognition 25*: 71-102

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

McClelland, J. L., & Elman, J. L. (1986) The TRACE Model of Speech Perception. *Cognitive Psychology 18*: 1-86.

Norris, D. (1994) Shortlist: A connectionist model of continuous speech recognition. *Cognition 52*: 189-234.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001) The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science 12*: 445 - 449.

Patkowski, M. S. (1990) Age and accent in a second language: A reply to James Emil Flege. *Applied linguistics, 11*(1), 73-89.

R Development Core Team (2009) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. ISBN 3-900051-07-0. Available at http://www.R-project.org.

Rose, Y., · Demuth, K. (2006) Vowel epenthesis in loanword adaptation: Representational and phonetic considerations. *Lingua 116*: 1112-1139.

Schmidt, A. M., & Flege, J. E. (1995) Effects of speaking rate changes on native and nonnative speech production. *Phonetica 52*(1): 41-54.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language learning, 19*(3 4), 245-253.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech.* Newbury House Publishers.

Tagliaferri, B. (2005) *Paradigm*. Perception Research Systems, Inc.

Available at www.perceptionresearchsystems.com

Weber, A., · Cutler, A. (2004) Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language 50*(1): 1-25.

Faculty of Spanish
Hankuk University of Foreign Studies
107 Imun-ro, Dongdaemun-gu, Seoul 02450, Rep. of Korea
e-mail: mtmg87@gmail.com

Received: Apr. 5, 2018
Revised: May. 23. 2018
Accepted: May. 25, 2018