When *blue* is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners

Isabelle Darcy and Trisha Thomas
Indiana University

* Acknowledgments
Funding for this study was provided by the Hutton Honors College via an undergraduate research grant to TT. The authors wish to thank Brandi Emerick, Eamon Anderson, Dr. Tom Busey and Dr. Natsuko Tsujimura for their support. We are indebted to Daniel Whyatt and Levi King for their help with recording the stimuli, as well as to Senyung Lee, Dr. Sun-Young Shin and Young Hwang for providing the loanword status evaluation of the experimental words. We also thank Abdullah Alotaibi for help with the acoustic measurements, and Michael Frisby and JangDong Seo from the Indiana University Statistical Consulting Center for help with the statistical analysis. We further gratefully acknowledge support through an Overseas Travel Grant from the Office of the Vice Provost for International Affairs and a 2016 Provost's Travel Award for Women in Science to ID. This work has further benefitted from excellent discussion and thought-provoking questions by audiences at New Sounds 2016 (Aarhus, Denmark) and at the phonetics and phonology reading group (PhLEGMe) at Indiana University. Finally, we also thank the Second Language Psycholinguistics Lab members, as always, for their feedback on earlier versions of this work.

Address for correspondence
Isabelle Darcy
Dept. of Second Language Studies
Indiana University—Bloomington
Morrison Hall 231
1165 E. 3rd St.
Bloomington, IN 47405
USA

Tel: +1 812 855 0033
E-mail : idarcy@indiana.edu
http://www.iub.edu/~psyling/

Abstract

Word-initial obstruent-liquid clusters, frequent in English (e.g. *blue*), are prohibited in Korean. Korean learners of English perceptually repair illicit word-initial consonant sequences with an epenthetic vowel [ʊ]. Thus they might perceive *blue* as b[ʊ]lue, and, at least initially, also represent it lexically as a disyllabic word. We ask whether the sound sequences permitted in one's L1 influence the way L2 words are represented in the mental lexicon. If they do, we predict that in a lexical decision task, Korean learners will accept nonwords containing epenthetic vowels ([bʊ'luː] for *blue*) as real English words more often than English listeners. These predictions were confirmed: we observed high error rates on test nonwords ([bʊ'luː]) by the Korean participants only, accompanied by few errors on control nonwords ([bɪ'luː]), suggesting that learners' lexical representations for familiar L2 words can be activated by nonwords that obey their L1 phonotactic grammar.

Keywords

Phonotactic grammar; perceptual epenthesis; phonolexical encoding; bilingual mental lexicon; Korean; L2 English

When *blue* is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners

1. Introduction

When we need to pronounce a foreign word containing sound sequences that do not exist in our native language (L1), we often modify these sequences in order to conform to those permitted in our L1 (Polivanov, 1931). For example, when English speakers need to talk about the capital city of Georgia (Tbilisi), or about a city on the Baltic coast in Poland (Gdańsk), they tend to insert a vowel and pronounce /tə'bilisi/ and /gə'dænsk/, because /tb/ and /gd/ are not permitted at the beginning of English words. Inserting vowels or simplifying consonant sequences are two common ways such adjustments take place. When learning words in a second language (L2), learners may experience these adjustments at the moment of articulating the right sequences of sounds, but also when perceiving them, and even while storing the pronunciation of the word in the mental lexicon. The present study is concerned with this latter process: how do people mentally store the phonological form of words of their L2? In particular, it examines how the sound sequences permitted in one's L1 influence the way words from the L2 are represented and accessed in the mental lexicon.

1.1 Phonotactic grammar

Each language allows for certain phoneme sequences but not others in specific syllabic positions. For example, liquids followed by obstruents in the syllable onset are not allowed in English, so that *lbog* is not a well-formed word, whereas *blog* is. Similarly, syllable structure constraints place restrictions on how many phonemes are allowed in specific positions such as onset, nucleus or coda. For example, Japanese only allows single-consonant onsets, whereas English can allow up to three consonants in the onset position (e.g. /stɹɪŋ/ 'string'). These sequencing and structural constraints are often referred to as PHONOTACTIC CONSTRAINTS.

Native speakers develop a tacit knowledge of the phonotactic constraints that govern the arrangement of phonemes in their first language early on, which we refer to as the PHONOTACTIC GRAMMAR. During their first year of life, at around 9 months, infants acquire the phonotactic regularities of their surrounding language, and demonstrate sensitivity to well-formedness patterns for their L1, but not for other unfamiliar languages (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993; see also Sebastián-Gallés & Bosch, 2002). This early attunement to the phonotactic patterns and syllable structure constraints of the native language appears critical for early word learning (Mattys & Jusczyk, 2001; Graf Estes, Edwards & Saffran, 2011). Reliance on the L1 phonotactic grammar continues to play an important role in adult speech production and processing, for instance during sound categorization, phoneme identification, and syllabification or segmentation decisions (Massaro & Cohen, 1983; Moreton, 2002; Pitt, 1998). For example, French listeners tend to perceive the sequences [dl] and [tl] as [gl] and [kl], respectively, in syllable-initial position, because the /dl/ and /tl/ clusters are phonotactically illegal in French (Hallé, Segui, Frauenfelder & Meunier, 1998).

Since phonotactic constraints are language-specific, the phonotactic grammar necessarily impacts speech processing in a language-specific way (see next section), and has consequences for non-native listening (Weber & Cutler, 2006). This commonly results in an array of adjustments (consonant deletion, vowel insertion, consonant mutation, among others) when perceiving or producing phonotactically illegal sequences (for a production example, see Carlisle, 1991). We refer to this kind of adjustment as PHONOTACTIC REPAIR, in the sense of Hallé, Dominguez, Cuetos, and Segui (2008). The nature and the scope of phonotactic repairs is not fully elucidated: The adjustments observed in production are not only the result of on-line production difficulties related to unfamiliar coordination of movements and insufficient articulatory overlap, even though this also plays a role (Davidson, 2006b). To a large extent, these repairs in production have their source in perception. The perceptual nature of phonotactic repairs has been confirmed in a number of cross-linguistic studies, where listeners have been shown to repair perceived structures if their L1 does not permit them (e.g. Hallé et al., 1998; Hallé & Best, 2007).

An additional major finding is that the *kind* of phonotactic repairs listeners will make is language-specific: listeners of one language will use one type of repair strategy, while those of a different language will use a different strategy (e.g. Dupoux, Kakehi, Hirose, Pallier & Mehler, 1999; Dupoux, Parlato, Frota, Hirose & Peperkamp, 2011). For instance, Dupoux and colleagues (1999) have shown that in Japanese, illicit consonant clusters are perceptually repaired by inserting a high back or central /u/-like vowel to break them up (or /o/ after coronal consonants). This phonotactic repair is referred to as perceptual vowel epenthesis. They found that if the L1 prohibits medial consonant clusters such as /bn/, the difference between two items such as /ebna/ and /ebuna/ evades Japanese listeners in a discrimination task—likely a result of the language-specific perceptual repair process (/u/-epenthesis) happening during the phonetic processing of the cluster stimuli (with a VCCV[1] structure) and making them sound very similar to the VCVCV stimuli. This repair effect was not observed when other vowels were inserted (e.g. /ebna/-/ebana/), nor was it observed in French listeners, whose L1 does not prohibit medial /bn/ clusters. In Brazilian Portuguese, for very similar environments, the epenthetic vowel of choice is a /i/-like high front vowel (e.g. Cardoso, John & French, 2008). In fact, Dupoux and colleagues (2011) show that Japanese and Brazilian listeners perceive different vowels (/i/ vs. /u/) in the same contexts, demonstrating that two languages that prohibit the same phonotactic sequence might resolve it in different ways.

Lastly, research also suggest that during non-native listening, phonotactic repairs arise very early in the time course of processing and presumably have a prelexical locus (Cuetos, Hallé, Domínguez & Segui, 2011; Dupoux, Pallier, Kakehi & Mehler, 2001; Dupoux et al., 2011; Dehaene-Lambertz, Dupoux & Gout, 2000; Hallé et al., 2008; Hallé et al.,1998). For instance, Dehaene-Lambertz and colleagues (2000) have shown that electrophysiological evidence of the repair is already evident 200 ms after the stimulus onset in a group of Japanese listeners who were listening to sequences of stimuli like /ebna/ vs. /ebuna/. The authors conclude that a fast and automatic coding of the input is shaped by L1 phonotactic constraints, and takes place early in speech processing, before word candidates are lexically activated. These findings are important, as they highlight the automaticity of L1-based phonotactic processing, making it hard to suppress during L2 processing.

Since phonotactic repairs are automatic, alter the perception of non-native speech very early during processing, and are language-specific, an important question to elucidate is to what extent L2 learners can overcome the influence of their L1 phonotactic grammar when learning a second language. Just like during non-native listening (that is, not involving learners), it appears to be generally the case that L1 phonotactic and syllable structure constraints are activated and trigger repairs during L2 listening or speaking (Abrahamsson 1999; Cardoso et al., 2008; Carlisle, 1991; Dupoux et al., 1999; Freeman, Blumenfeld & Marian, 2016; Gibson, 2012; Kabak & Idsardi, 2007; De Jong & Park, 2012; Parlato-Oliveira, Christophe, Hirose & Dupoux, 2010; Weber & Cutler, 2006). For instance, perceptual epenthesis has been observed in Korean learners of English, for whom discriminating a pair such as /pʰakma/-/pʰakʰʊma/ was more difficult than for American English listeners (Kabak, 2003), because the difference between both stimuli was less salient to them. In a same/different task, the learners' discrimination *d'* score was significantly lower (d' = 2.28) than the *d'* of American English listeners (d' = 3.58), and it can be concluded that to some significant extent, Koreans listeners perceptually repaired these illicit VCCV sequences with an epenthetic vowel (VCvCV), thus perceiving both stimuli in the pair as very similar. While most perceptual studies used nonwords, similar perceptual phonotactic repairs have been observed with real L2 words in bilinguals (pre-intermediate Brazilian Portuguese L1, English L2 learners: Bettoni-Techio, Rauber, & Koerich, 2007; highly proficient Spanish L1, English L2 bilinguals: Freeman, et al., 2016). Freeman and colleagues investigated the case of the word initial /s+C/ sequence which is prohibited in Spanish words, and show that a group of (highly proficient) Spanish-English bilinguals were faster to make a decision on /es+C/-initial nonwords (e.g. *estimagle*) when primed with /s+C/-initial English words (e.g., *strong*), compared to baseline nonwords that do not begin with /es+C/ (e.g. *atongside*). Such priming effect is expected to emerge if listeners perceptually repair the illegal word-initial /s+C/ sequence of the prime word into a legal sequence by inserting an initial /e/, therefore creating a form overlap (*estrong*) triggering facilitation. While this research suggests that L1-based phonotactic processing is difficult to fully suppress in a L2 even after substantial exposure, its influence may still diminish over time. Even though Dupoux et al. (1999) did not find any effect of L2 proficiency on the incidence of perceptual epenthesis in Japanese participants, Carlson and colleagues (2016) show with a gating task that English-dominant Spanish English bilinguals experienced fewer perceptual repairs in Spanish than the bilinguals who were Spanish-dominant (see also Cardoso, 2011). This study suggests that L2-specific phonotactic regularities can be acquired and can impact L1 perception. Generally, novel phonotactic constraints specific to the L2 appear difficult to learn, but seem acquirable when they are not incompatible with L1 constraints (Parlato-Oliveira et al., 2010; Weber & Cutler, 2006), even though there are still too few studies so far to be able to paint a complete picture of the L2 acquisition of phonotactic constraints.

The prelexical locus of the repairs also opens the possibility that they have consequences for learners who need to form lexical representations for the L2 words and access these representations during listening (see also Davidson, Shaw & Adams, 2007).

1.2 Word recognition and phonolexical encoding in L2

When recognizing spoken words, listeners continuously attempt to map the incoming speech signal onto lexical representations stored in memory (McClelland & Elman, 1986; Norris, 1994): lexical entries that overlap to some extent with the perceived speech input are activated until the lexical candidate that best matches the input wins over its competitors, a process known as lexical competition. However, models of spoken-word recognition mainly address native speakers' lexical behavior, assuming that their perception of the input is precise under most conditions, and that their lexical representations are stable and contain at least the phonological form of words in citation.[2]

Recognizing spoken words in a second language often presents a much more complicated picture. In the initial stages at least, L2 listeners' perception of the input is influenced by their L1 phonological grammar, and as a result, word recognition presents at least two types of related challenges for learners' developing lexicon, pertaining to either the form of words in their mental lexicon (lexical representations), the activation and recognition of words during listening (lexical access), or both. For instance, the influence of L1 during segmental perception can initially lead to under-differentiation of lexical entries, with learners storing two different words as homophones when these words differ in a phonemic contrast that is hard to perceive. It can also lead to over activation of lexical candidates (e.g. Broersma & Cutler, 2011; Darcy, Daidone & Kojima, 2013; Dupoux, Sebastián-Gallés, Navarrete & Peperkamp, 2008; Ota, Hartsuiker & Haywood, 2009; Pallier, Colomé & Sebastián-Gallés, 2001). Of note, both studies by Darcy et al. (2013) and Ota et al. (2009) show that imprecisions in lexical representations can persist even after segmental perception improved and after the contrast can be accurately distinguished. More generally, L2 form-to-meaning mappings can also be imprecise for items that do not pose particular perceptual difficulties (Cook, Pandža, Lancaster & Gor, 2016), indicating that phonolexical processing issues can persist independently from improvements in phonetic perception.

While the consequences of phonotactic repairs for lexical representations and word activation patterns are less well understood, they are possibly similar to those of phonemic distinctions, affecting either lexical representations, access, or both. In a longitudinal production study of a Spanish L2 learner of Swedish, Abrahamsson (1999) evokes a theoretical possibility to explain epenthesis in production, in which the native syllable structure constraints are responsible for the insertion of [e] in productions of initial /s+C/ clusters in L2 by Spanish L1 speakers. According to the author, "such an analysis would thus suggest that L2 production follows the same principles as loanword adjustment, of course with the exception that the epenthetic forms are not listed in the L2 lexicon" (p. 478-479). Here, the author seems to rule out the possibility that L2 words might be represented in the mental lexicon in a form that approximates the way they were first perceived: with epenthesis. Yet, a word learning study by Davidson et al. (2007) suggests that phonological representations established for novel words are influenced by the native language phonological prohibition on specific sequences, hampering "a detailed phonetic encoding of these items, at least at an early stage of exposure" (p. 3706). As this quote implies, the encoding or representation formed for words containing a prohibited sequence do not necessarily mirror the phonetic input – in fact, it leaves open the possibility that potential perceptual repairs could be encoded in lexical representations. Thus, in spite of Abrahamsson's doubts, L1-based phonotactic processing could have consequences for lexical representations or lexical access when learning a L2 that allows sequences that are illegal in the L1.

If the influence of L1 phonotactics is strong during L2 processing and learning, it is likely that for L2 learners, the L2 input they perceive—or their intake—will contain many words that underwent perceptual repairs, at least at the initial stages of learning (Matthews & Brown, 2004). As a result, building native-like, or 'repair-free' lexical representations of these L2 words may be compromised.

The goal of the present study is to examine how L2 words are represented in the mental lexicon of L2 learners, and whether or not their phonological form reflects L1-based phonotactic processing. That is, we ask whether the kind of phonotactic repairs observed in cross-linguistic and L2 studies while perceiving phonotactically illegal L2 words has long-term consequences at the lexical level. Specifically, we examine whether the phonolexical representations of English words reflect vowel epenthesis in initial obstruent-liquid sequences for L1 Korean learners of English.

2. The current study

To examine the presence of perceptual repairs in the mental lexicon, we chose Korean learners of English. This population is well-studied, and previous studies have documented the presence of perceptual repairs due to syllable structure/phonotactic constraints (Kabak, 2003; Kabak & Idsardi, 2007). Word-initial obstruent-liquid clusters are very common in English (e.g. *blue*) but prohibited in Korean, and as mentioned earlier, vowel epenthesis has been described as a mechanism to repair disallowed cluster words—in loanwords for instance (Dupoux et al., 1999; Kabak & Idsardi, 2007). Korean listeners perceptually repair illicit word-initial consonant sequences with an epenthetic vowel [ɯ], a high central unrounded vowel, which is perceived as similar to American English [ʊ][3] (Kabak, 2003, p. 56). Thus they might perceive *blue* as [bʊ'luː] or [bɯ'luː], and, at least at an early stage of exposure, also encode it lexically as a disyllabic word /bʊ'luː/[4].

We hypothesize that Korean learners of English store this type of cluster-initial English words with spurious vowels, as a result of perceptual epenthesis. If so, we predict that their representations can be activated by epenthesis-containing stimuli, and that Korean L2 learners will accept nonwords containing epenthetic vowels ([bʊ'luː]) as real English words more often than English listeners. The reason behind such activation is that decisions about real-word-status presuppose a successful word recognition involving a degree of overlap between incoming signal and stored word forms. Importantly, because the nonwords already contain the epenthetic vowel, the need for perceptual repair during the presentation of experimental stimuli is void. Hence, we do not examine whether or not listeners perceptually repair stimuli as they hear them (as was done for example in Dupoux et al., 2001): our design is in essence the mirror image of Dupoux et al., 2001, asking whether L2 learners' phonolexical representations for English words can be activated by stimuli that contain epenthesized vowels. One of the challenges inherent to auditory lexical decision tasks is presenting stimuli that are phonetically unfamiliar or phonotactically illicit for listeners. In those cases, the danger of perceptually compensating/repairing what one hears is serious, and may make conclusions about what listeners encode – given the possibility that they repaired the input – more difficult (see discussion). Therefore, our critical test condition does not contain any initial consonant sequences that would be illicit in Korean phonology.

However, given that our participants are fluent in English, are full-time students at a large Midwestern university, taking classes taught in English, and have been living in an English-speaking country for at least several months, an alternative hypothesis is possible. Our participants' substantial exposure to written forms could be sufficient to prevent perceptual epenthesis from "contaminating" the lexicon: Gibson (2012) reports that perceptual epenthesis effects were reduced during perception when listeners saw written forms. The orthography of English indicates the absence of a vowel in word-initial consonant clusters: <**bl**ue>, <**pr**oud>, <**pl**ay>. Taking Gibson's findings one step further could mean that orthographic knowledge of words may contribute to reduce perceptual epenthesis, and consequently it could prevent the establishment of representations with an epenthetic vowel, or it could help redefine representations that may have been initially stored with a vowel..

2.1 Method

We test these hypotheses with an auditory lexical decision task, where participants have to indicate by a button press whether an item is a real English word or not. To make such a decision about lexical status, one needs to compare the incoming input (stimulus) to stored phonological representations for words. The only way to correctly reject a nonword (which is always a potential word) is to have ‘repair-free’ phonolexical representations. If the nonword [bʊ'luː] matches a lexical representation sufficiently well, the word will be activated and recognized as an English word, and trigger a false positive (= *yes* answer). We will interpret a higher rate of false positives to indicate that learners’ representations for words like *blue* are activated by disyllabic stimuli containing the epenthetic vowel [ʊ] (e.g. [bʊ'luː]), possibly because the representations themselves contain the epenthetic vowel, as /bʊ'luː/ (see discussion; Lahiri & Reetz, 2002, or Darcy et al., 2013).

Participants

A total of 39 participants (L1 English: *N* = 21; L1 Korean learners of English: *N* = 18) completed a speeded auditory lexical decision task and a background/word-familiarity questionnaire. All participants reported normal (or corrected-to-normal) vision and hearing, and had a similar level of education. The native speakers of English did not know any Korean, but five participants knew a second language (Spanish), acquired after the age of 11 (*M* = 12.4).

The native speakers of Korean were proficient in English at an intermediate to advanced level (self-reported; see Table 1), and reported spending an average of 30 hours (*SD* = 20) per week using English. None of them reported being an early English-Korean bilingual. Table 1 presents an overview of demographic information and language learning profile for the participants.

Table 1. *Participant background information.*

Group	Females / Males	Mean age (SD; Range; Median)	Median age of arrival in USA (SD; range)	Mean length of stay in USA in years (SD; range)	Median age of learning L2 (SD; range)	Mean self-estimated proficiency (L1; L2)	Mean length of Eng. learning in years (SD; range)
English	13 / 8	28.2 (10.6; 18-50; 22)	born in USA	28.2 (10.6; 18-50)	12 (1.14; 11-14)	10; 6^	--
Korean	11 / 7	22.7 (3.2; 19-29; 22)	18* (3.5; 15-29)	3.5 (2.2; 0.25-7)	11 (2.24; 7-15)	10; 7	10.8 (2.39; 6-15)

Note: *Age indication missing for one Korean participant; ^ This L2 estimate does not include the two early bilinguals.

All but two participants were born in monolingual families and grew up speaking only English or Korean respectively. The two early bilinguals—from the English group—grew up speaking English/Mandarin and English/Hindi; both were born and grew up in the USA. Proficiency estimates (self-reported) for L1 and L2 were obtained by averaging self-rating scores of the participants' ability to speak spontaneously, write, understand, and read their first/second language, as well as a rating of their accentedness, on a scale of 1 (none/very heavy accent) to 10 (perfectly/little to no accent). No formal proficiency test was administered.

Materials

Experimental stimuli consisted of 30 common English words containing onset clusters (*blue*, *play*), which were modified to create 30 pairs of experimental nonwords: Thirty test items containing [ʊ], which is similar to the Korean epenthetic vowel for onset clusters (for instance [bʊ'luː] or [pʊ'leɪ], based on *blue* and *play*, respectively), and thirty control items, which contained the vowel [ɪ] (for instance [bɪ'luː] or [pɪ'leɪ]). This resulted in 60 recorded nonwords. An additional 122 words (for instance *relax* [rɪ'læks], *well* [wɛl]) and nonwords (for instance *peef* [piːf], *altron* ['ɑltrən]) made up the distractor condition. The distractors contained 92 words and 30 nonwords, such that a nearly equal proportion of words and nonwords was represented in the experiment. These distractor items did not contain word-initial obstruent-liquid clusters. Among the 122 distractors, 74 were monosyllabic and 48 were disyllabic items (38 words, 10 nonwords). Eleven out of the 48 disyllabic distractors (8 words, 3 nonwords) were stressed on the second syllable, roughly equivalent to 22% the stimuli (a value similar to the 25% proportion of English polysyllabic words that receive stress on the second syllable, Cutler, 2005, p. 271). This was done to ensure that stress placement could not be a perfect indicator of lexical status. Since the test and control nonwords were stressed on the second syllable due to the vowel insertion manipulation, the iambic stress pattern might have cued participants to systematically respond *no*. By including a number of word items with second-syllable stress, we reduced this potential confound.

The control condition using the vowel [ɪ] allows to verify that the effect is driven by the influence of L1-based phonotactic processing, and not due to a possible bias for learners to answer *yes* and experience false positives more often generally to items that contain any inserted vowel. In this condition, the vowel is [ɪ], which is not the epenthetic vowel for obstruent-liquid onset clusters in Korean. However, [ɪ] is a possible epenthetic vowel in Korean in the context of palatal consonants (Kabak, 2003, p. 56).

Stimuli were recorded naturally several times by two phonetically-trained male English native speakers from the Midwest in a sound-isolated recording booth at a sampling rate of 44,100 Hz with a 16-bit resolution, on a mono channel through a Sennheiser e835 microphone. From these recordings, one male voice was selected due to the heightened clarity of his inserted vowels. Among all accurately recorded items (as judged by auditory and visual examination of the recordings by the authors), thirty pairs of nonwords were selected as experimental stimuli (see appendix for a list of stimuli; all stimuli and instruments are available on the IRIS database, www.iris.org). The inserted epenthetic vowels were not stressed, but not fully reduced, so that they remained distinguishable from each other. This type of realization is also naturally present in English and can be found in words such as *police* [pəˈliːs], or *tomorrow* [təˈmɒɹəʊ], where [ə] can have a [ʊ] coloring. Similarly, the unstressed vowel in *because* [bɪˈkəz] and *guitar* [gɪˈtɑɹ] can have a /ɪ/-like quality.

The duration and spectral properties of the inserted vowels in the experimental stimuli were measured by the first author and an additional researcher, to verity that acoustic differences were implemented as intended. Table 2 summarizes the duration of each kind of epenthetic vowel (front/back), as well as their formant values (F1, F2 and F3). Measurements were taken in the vowel steady-state portion using Praat (Boersma & Weenink, 2016) with a 0.015 s window length and a 70dB dynamic range. The beginning of the vowel portion was manually marked at the point of a sharp increase in intensity coinciding with the onset of a periodic waveform with regular formant structure. The end of the vowels was more difficult to establish visually because of the following liquid (/r/ or /l/). It was marked conservatively, using both auditory information and visual cues from the spectrogram, at a point where the following liquid was not yet clearly identifiable and a discernible change in the formant structure was visible.

Table 2. *Descriptive statistics of acoustic measurements for the inserted vowel portion in the experimental stimuli.*

	Mean Duration			Mean F1			Mean F2			Mean F3		
Stimulus	ms SD	95% CI Lower	Upper	Hz SD	95% CI Lower	Upper	Hz SD	95% CI Lower	Upper	Hz SD	95% CI Lower	Upper
Control (i)	**89.2** 24.5	80.1	98.4	**272** 40	257	287	**1926** 120	1881	1970	**2440** 288	2333	2547
Test (u)	**103.8** 26.9	93.7	113.8	**367** 63	344	391	**909** 88	876	941	**2070** 186	2000	2139
Average	**96.5** 25.7			**319.9** 51.3			**1417.1** 103.8			**2254.8** 236.8		

Note. 95% CI = 95% Confidence Interval for Mean; SD = Standard Deviation

The measured values suggest three main conclusions: First, the spectral information clearly indicates that the two vowels were realized differently in our speaker's recordings, a difference mostly noticeable in the large difference in F2 that corresponds to the front/back distinction ($t(53.2) = 37.5$, $p < .001$, two-tailed). Second, the average duration of the inserted vowels is not particularly "subtle", at 96 ms (range: 49-159 ms). The test (u) vowels are minimally longer than the test (i) vowels ($t(58) = -2.2$, $p = .03$, two-tailed). Finally, this range of duration has been found in other acoustic studies of epenthetic vowels: for instance, they are comparable to the ones obtained for L2 English learners in Shin (2014), who observed a mean duration of 85 ms (SD = 26.9), and high variability across speakers. The range of duration is also comparable or higher to the ones obtained for lexical vowels in Davidson (2006a), suggesting that they were realized with a clear articulatory target (and not as a transitional vowels) by our speaker. Of note, the vowel duration in our stimuli is also longer on average than the last two steps (72 ms and full vowel) of the vowel continuum in Dupoux et al. (1999, Figure 1), at which all participants reported perceiving a vowel in more than 90% of the stimuli. Taken together, these data suggest that our stimuli contained clearly perceptible, unambiguous vowels.

Procedure

All participants were tested individually in a quiet room. The lexical decision task and the questionnaire were administered in a single session using a portable computer and over the ear headphones. The 182 total items were split into two lists of 91 items, which were presented as blocks separated by a break during the experiment. Each block contained 46 real and 45 nonwords. The nonwords were comprised of 15 distractors, 15 [ʊ] and 15 [ɪ] items. The [ʊ] and [ɪ] nonwords generated from the same base word (e.g. [bʊ'luː] and [bɪ'luː] both generated from *blue*) were never presented in the same block. The task was programmed in Python with the aid of the Psychopy builder (Peirce, 2007).

The blocks were randomized, as were the items within each block. The task was speeded and participants were instructed to make their answer within 2500 ms. Participants were asked to indicate, by pressing one of two keys on the computer keyboard labeled “yes” and “no”, whether each item they heard was a real English word or not. The experimental portion of the task was preceded by a short practice session with feedback, using examples such as *shrimp* (real word) vs. **sharimp* (nonword), to favor a phonological decoding interpretation of the task rather than an acoustic one. For the same reason, a different voice was used for these trials. The lexical decision task took about 20 minutes to complete. Error rates were tallied and RTs were measured from the onset of the stimulus.

After the lexical decision task, participants completed a background questionnaire and a word familiarity questionnaire. We decided to use familiarity rather than frequency, given that frequency estimates are likely different for L2 learners and native speakers. Familiarity ensures comparability across groups. The questionnaire intended to verify their familiarity with the written form of the distractor words and base words from the lexical decision task. Each word was to be ranked on a scale adjusted from Paribakht and Wesche (1993; Wesche & Paribakht, 1996). There were three rankings as follows: 1) “Very familiar, this is a word I know and am able to use in an English sentence”, 2) “I am not sure about what it means, but I know it’s an English word” and 3) “I have never seen it, or I am not sure this is a real English word”. Participants were asked to put a check in the box that best fit their familiarity with each word. A few foil nonwords were included in the questionnaire to check that participants were providing answers corresponding to their actual knowledge of each word.

Overall, completion of the full testing session took about 45 minutes. Participants received a small payment in return for their time and were debriefed after completing the experiment. All procedures described in this study have been approved by the Indiana University Institutional Review Board.

Predictions

There are two expected possible patterns for phonolexical representations: onset cluster (CCV) vs. epenthesis (CvCV). In the first case where a given participant successfully encodes onset clusters in their mental lexicon, the word *blue* will be represented as /bluː/. If a participant follows the epenthesis pattern in their phonolexical representations, the word *blue* might be represented as /bʊ'luː/ according to Korean phonotactic constraints. The extent to which a participant represents words via onset cluster or epenthesis will be indexed in terms of error proportions in each condition.

For the test nonwords condition ([bʊ'luː]), if a word such as *blue* is encoded as CCV, hearing a nonword such as [bʊ'luː] will result in a mismatch with contacted representations, and trigger no lexical activation. This would accordingly generate low rates of *yes* responses, that is, low error rates, in the test(u) condition. But—this is our test case—in case of epenthesis, the word *blue* may be encoded as CvCV, as /bʊ'luː/. Upon hearing the stimulus nonword [bʊ'luː], there will be a sufficient overlap (match) with the stored representation, *blue* will be activated, and this nonword will be recognized as a real word. This would lead to more *yes* responses in the test(u) condition, and thus higher error rates. These two diverging response patterns in the test condition correspond to what we expect for the native English speakers and for the Korean L2-English learners, respectively.

In the control condition, the vowel is [ɪ] ([bɪ'luː]); Regardless of which pattern is represented in the mental lexicon (cluster or epenthesis), the stimulus nonword should activate neither /bluː/ nor /bʊ'luː/, and should result in a low proportion of *yes* responses for the control items, that is, a low error rate for both groups.

Finally, regardless of which pattern is encountered, we expect mostly *yes* answers for real words and *no* answers for nonwords in the distractor condition, with a low error rate overall, since these items do not contain initial consonant clusters.

2.2 Results

Familiarity data

The familiarity questionnaire revealed that the base words used to create the experimental items (test and control) were highly familiar to all participants, as indicated by the proportion of "very familiar" responses (M = 100%, SD = 0). It can be concluded that these words were thus highly familiar to our participants in both orthography and meaning.

Lexical decision: Sample characteristics and data screening

RT data were screened for extreme values, which were defined as being above 2.5 SD from each group's mean RT over all conditions, or below 250 msec.; For Koreans, 107 RT data points (1.5% of the data) were removed; for English, 100 data points were removed (1.4 % of the data). The resulting mean error rate and RT were computed for each item and each participant, and were screened for outliers. Participants with an error rate above 2 SD from the group's mean on the distractor condition were excluded. One English participant was excluded from analyses because of a mean error rate of 94% on this condition. High error rates for this participant in the other conditions as well (M = 95%) resulted in unanalyzable RT data with fewer than 10 observations. While this behavior might indicate a button reversal, it was deemed safer to exclude this participant. No Korean participant was excluded.

In addition, two items (both were distractors, one word and one nonword) were excluded because they triggered very high error rates among the native speakers after the outlier participant was removed (45% and 35% respectively, beyond 2 SD from the mean for this group). RTs were examined for normality and for homogeneity of residuals. The RT, which were positively skewed, required a natural log transformation in order to improve the normality of residual distribution.

Lexical decision: Error and RT data

Descriptive statistics for mean error rates and RTs (computed over correct answers only) for each group and each condition are presented in Table 3. A first inspection of the error data suggests that our predictions are confirmed: listeners in both groups are similarly accurate on two out of three conditions (control and distractors) but differ clearly on the test condition.

Table 3. *Mean error (%), SD, 95% CI, Mean RT (ms), SD, 95% CI, Median RT and interquartile range in each condition for each group.*

Group	Condition	Mean error (%)	SD	95% CI	Mean RT (ms)	SD	95% CI	Median RT (ms)	RT Interquartile range (ms)
English	Test (u)	14.0	20.8	07.0–21.0	1237	447	1197–1276	1114	567
	Control (i)	11.5	16.2	04.5–18.5	1296	469	1255–1336	1163	600
	Distractor	04.2	3.4	-02.8–11.3	1161	439	1143–1179	1047	517
Korean	Test (u)	38.7	25.2	31.2–46.1	1436	586	1370–1501	1232	703
	Control (i)	16.5	11.9	09.1–23.9	1418	534	1368–1468	1262	602
	Distractor	09.1	5.1	02.0–16.5	1272	516	1249–1295	1115	615

Note. The RT values are based on correct answers.

Figure 1 clearly shows the differences in error proportions among groups. As predicted, Korean listeners experienced a sharp increase in false positives for the test (u-nonwords) condition (*M* = 39 % error), but not the English listeners (*M* = 14 % error). We also descriptively observe that the Korean group's mean RT (*M* = 1436 ms) in this condition is about 200 ms slower than the mean RT of the English group (*M* = 1237 ms). In both other conditions, the Korean group's mean RTs were also slower than the English group but the difference was smaller, at about 110 ms.

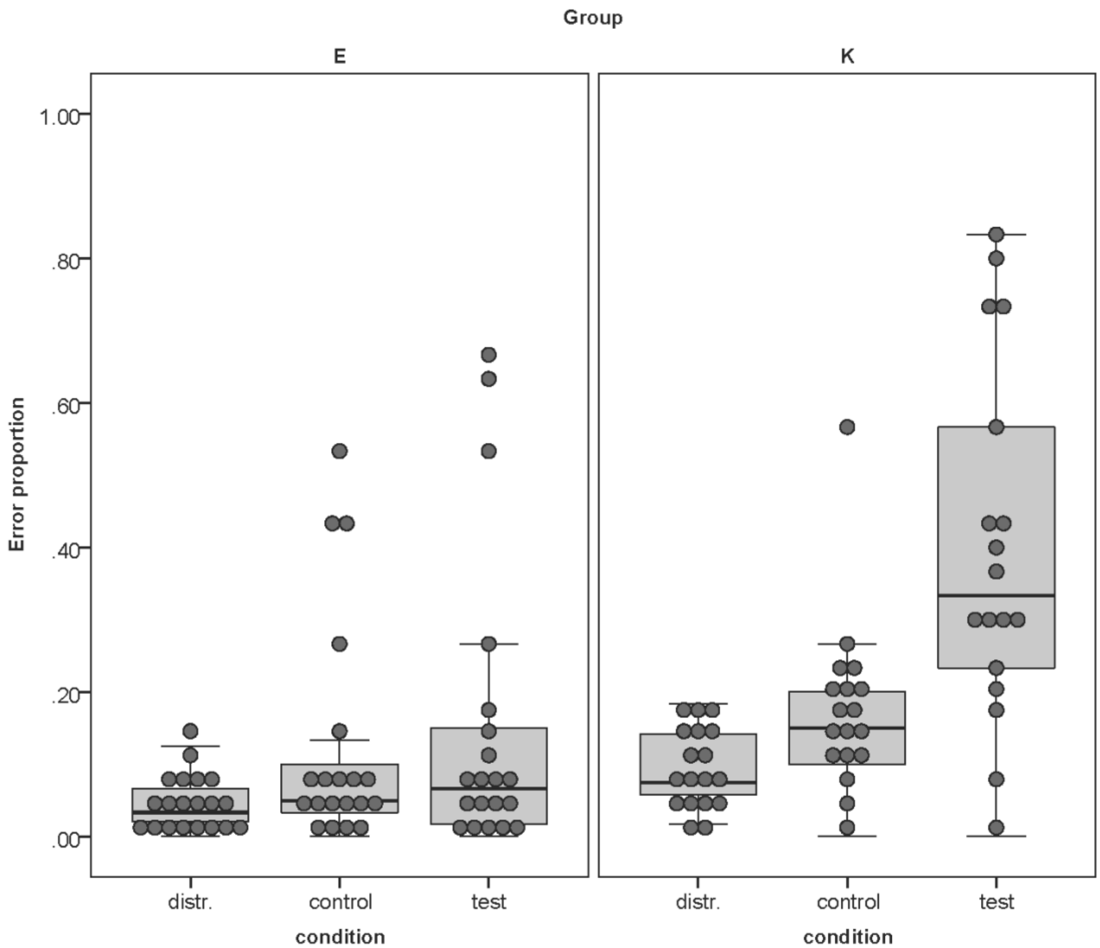

Figure 1. Box plot with overlaid dot plot for each condition and group. Each dot is one participant. Horizontal lines are medians, boxes show the interquartile range (IQR) representing 50% of the cases, whisker bars extend to 1.5 times the IQR. Outliers (circles) are cases with values between 1.5 and 3 times the IQ range, i.e., beyond the whiskers.

Planned analysis: Error rates. The observations from Table 3 were confirmed by inferential statistics. For all analyses, alpha is set at p = .05. We analyzed the lexical decision data by modeling the likelihood of a response type (error or correct answer) in each group and condition, using the *lme4* package (https://CRAN.R-project.org/package=lme4; version 1.1-14, Bates, Maechler, Bolker &

Walker, 2015) in the R statistical language (http://www.r-project.org/, version 3.4.2). We used a generalized linear mixed model (GLMM) with binomial logit link function fit by maximum likelihood (Laplace approximation) for the categorical correct/incorrect dependent variable, and included random effects (intercepts) for both item and participant[5]. The random structure had to be simplified because of non-convergence, and only includes intercepts. Random slopes for participants are not included and thus do not allow to account for the extent to which differences between condition might differentially affect participants. However, since this difference is also implemented through specific items in each condition, the random effects for items will give a sense of this dimension of the variance. Table 4 summarizes the random and fixed effects estimates.

Table 4. *Overview of random and fixed effects estimates for the GLMM model, for the categorical error variable.*

Random effects:

Groups	Name	Variance	SD
Item	(Intercept)	1.0637	1.031
Participant	(Intercept)	0.7157	0.846

Number of observations: 6657, groups: item, 180; participant, 38

Fixed effects:

	Estimate	Std. Error	z value	Pr (> |z|)
(Intercept)	-2.646	0.308	-8.605	< 0.001 ***
Group K	0.739	0.336	2.198	0.028 *
Condition distractor	-1.486	0.281	-5.289	< 0.001 ***
Condition test(u)	0.281	0.331	0.849	0.396
Group K: condition distractor	0.497	0.229	2.173	0.030 *
Group K: condition test(u)	1.090	0.250	4.356	< 0.001 ***

Note. K = Korean (English is the reference group). Control is the reference condition

In addition, *p*-values were obtained using the *lmerTest* package (https://CRAN.R-project.org/package=lmerTest) for the fixed effects of condition (distractor, test and control) and group (English vs. Korean), as well as the interaction between these. This package uses Satterwaithe's approximation to degrees of freedom for the *p* values. We report the chi-square tests from the GLMM results here (Type III Wald chi-square tests). There was a significant effect of Group ($X^2(1) = 4.83$, p =

.028), of Condition ($X^2(2) = 53.1$, $p < .001$), and a significant interaction between these factors ($X^2(2) = 19.3$, $p < .001$).

Finally, an analysis of odds-ratio was used to unpack this interaction and to estimate the likelihood of each group to make errors in each condition, using the package *lsmeans* (https://CRAN.R-project.org/package=lsmeans). Tables 5a and 5b summarize the odds ratio by condition and by group.

Table 5a. *Odds ratio of the two groups making errors on the lexical decision task in each condition (K = Korean; E = English).*

Condition	Contrast	Odds ratio	SE	*z* ratio	*p* value
Test (u)	K – E	6.22	2.025	5.622	< .0001
Control (i)	K – E	2.09	0.703	2.198	0.0279
Distractor	K – E	3.44	1.071	3.966	0.0001

Note. Tests are performed on the log odds ratio scale. *p*-value adjustment through Tukey's method.

Table 5b. *Odds ratio of making errors on the test vs. the control condition in the lexical decision task, within each group (critical comparison in boldface).*

Group	Contrast	Odds ratio	SE	*z* ratio	*p* value
English	Distractor - Control (i)	0.23	0.064	-5.289	< .0001
	Test (u) - Distractor	5.85	1.626	6.364	< .0001
	Test (u) - Control (i)	**1.32**	**0.438**	**0.849**	**0.6725**
Korean	Distractor - Control (i)	0.37	0.098	-3.764	0.0005
	Test (u) - Distractor	10.6	2.676	9.351	< .0001
	Test (u) - Control (i)	**3.94**	**1.229**	**4.398**	**< .0001**

Note. Tests are performed on the log odds ratio scale. *p*-value adjustment through Tukey's method.

This analysis indeed revealed that in the test condition, the Korean group was much more likely to make errors than the English group (Odds ratio = 6.22), whereas the likelihood of each group to make an error in the control condition was more comparable (Odds ratio = 2.09). This difference in the two odds ratios was significant, $F(1, 6457) = 19.39$, $p < .0001$. Another way to look at this interaction is to consider the differences between conditions within each group of participants. English participants show similar odds of making errors in the test and control conditions ($p > .6$), whereas Korean participants have significantly higher odds of making an error in the test than in the control condition ($p < .0001$, see Table 5b), confirming our predictions that Korean participants store /u/ forms differently from /i/ forms whereas English participants do not process the two differently. The pattern suggested in Table 3 is thus corroborated showing that the largest difference between Korean and English groups lies in the test condition (error rate of the Korean group in the test condition, M = 39%, vs. M = 14% for the English group). Figure 2 visualizes this relationship by condition, showing that Koreans are more likely to make a mistake than the English participants generally speaking, with this effect being more pronounced in the test condition.

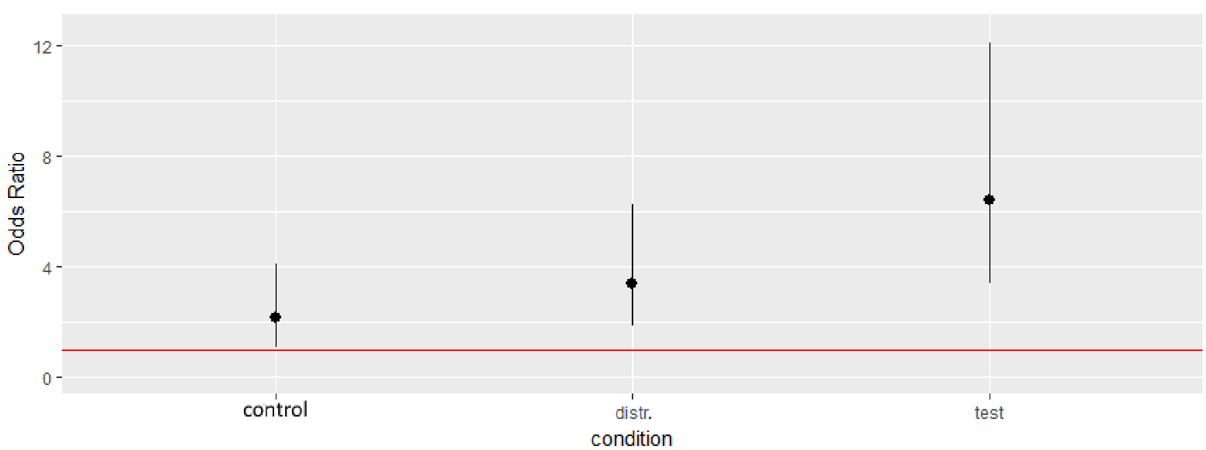

Figure 2. Odds-ratios of the Korean group to make a mistake relative to the English group. Bars represent 95% confidence interval. The solid horizontal line (1) shows the level at which both groups have an equal likelihood (1) of making errors.

Exploratory analyses: RT. In terms of RTs, it is important to rule out the presence of speed-accuracy tradeoffs (Sternberg, 1998). In addition, if Korean participants experience a higher processing load, we predict that their RT would be slowest to correctly answer on the test (u) condition, and possibly much slower than the English participants on this condition, whereas the difference between groups on the other two conditions might be smaller. We conducted the analysis on the log-transformed RT data (for descriptives, see Table 3), using a linear mixed effects model (in SPSS 24) declaring subject and item as random effect. The model examined the fixed effects of condition (distractor, test and control) and group (English vs. Korean), as well as the interaction between these. For convenience, the untransformed mean RTs and their corresponding CIs are reported below. The *p*-values are for the logRT. The effects do not change in the model using log values, compared to the untransformed RT model.

The Type III tests of fixed effects revealed a significant main effect of condition on the logRT ($F(2, 192.0) = 33.8$, $p < .001$), a marginal main effect of group ($F(1, 37.2) = 3.65$, $p = .064$), and a significant interaction ($F(2, 5737.2) = 3.16$, $p = .042$). Across both groups, the distractor condition was responded to faster ($M = 1222$ ms, CI = 1153–1290) than both other conditions (control $M = 1368$ ms, CI = 1296–1440; test $M = 1364$ ms, CI = 1290–1437), which were not different from each other ($p = .9$). While Korean participants were slower in all three conditions compared to the English participants (see Table 3), Sidak-corrected post-hoc tests show that the interaction is mainly driven by the group difference in latency on the test condition (Koreans were on average 211 ms slower, CI = 63–358, $p = .015$), a difference about twice as large as the group difference on the other two conditions (mean difference for distractor items = 111 ms, CI = -26–247, $p > .1$; mean difference for control items = 107 ms, CI = -38–251, $p > .1$). However, this effect is likely not a very large one as indicated by the overlapping confidence intervals. It is noticeable that the CI is crossing 0 for the distractor and control, but not for the test condition. The parameter estimates are shown in Table 6.

Table 6. *Parameter estimate, standard error, t-value, p-value, and 95% confidence interval of the predictors for log-transformed RT.*

Fixed Effects	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower	Upper
Intercept	3.138	0.019	60.29	168.6	.000	3.100	3.175
[condition=ctrl(i)]	-0.011	0.012	478.2	-0.90	.370	-0.035	0.013
[condition=distractor]	-0.059	0.010	537.1	-5.87	.000	-0.079	-0.040
[Group=E]	-0.061	0.024	48.60	-2.53	.015	-0.110	-0.013
[condition=ctrl(i)] * [Group=E]	0.029	0.013	5746.9	2.27	.023	0.004	0.055
[condition=distractor] * [Group=E]	0.025	0.011	5772.4	2.35	.019	0.004	0.045
Parameter	Estimate	Std. Error					
Residual	0.017	0.000					
Item	0.001	0.000					
Participant	0.005	0.001					

Note. E = English; (Korean is the reference group). Test is the reference condition

However, given that the error proportions are higher in the Korean group, there are also fewer RT data points to include in the analysis. In addition, one could argue that to this participant group, the lexical status (word vs. nonword) of these items is ambiguous, such that it is not clear, for this group, which answer should be considered “correct”. Thus, it can prove interesting[6] to compare the RTs of the Korean group for both ‘word’ and ‘nonword’ responses in the test vs. control condition. We thus analyzed the responses according to the predicted lexical status of the items, that is, responses as ‘word’ for items in the test (u) condition and responses as ‘nonword’ for items in the control (i) condition (see Dupoux et al., 2001, for a similar reasoning). These latencies are shown in Table 7. The comparison of least square means (of the log RTs) shows that for test (u) items, ‘word’ responses (that is, incorrect answers) were given as rapidly as ‘nonword’ ones (*t* ratio = 0.83; $p = .4$). In contrast, items in the control (i) condition yielded slower RTs for ‘words’ responses compared to the responses as ‘nonwords’ (*t* ratio = 3.98; $p < .001$). This pattern might be taken as indirect evidence that in the test (u) condition, participants were more confident in their incorrect ‘word’ answers than they were in the control (i) condition, where incorrect responses were made more slowly than correct answers. However, strong conclusions should not be drawn from this analysis, given that responding correctly may involve different processes than responding incorrectly (see Sebastián-Gallés et al., 2006).

Table 7. *Reaction times (ms) and standard error to test and control items by response for the Korean group.*

	Response: Nonword		Response: Word	
Condition	RT	SE	RT	SE
Test (u)	1449	34.32	1551	43.02
Control (i)	1423	25.78	1699	65.32

Exploratory analyses: Items. An analysis of individual patterns for items revealed (see Figure 3) that two items triggered high error rates in both groups. These items turned out to be the same for both groups: they were items that unintentionally resulted in near-words or words through the insertion of the [ɪ]-vowel: *drive* and *train*, which resulted in items very close to *derive* and *terrain*. Another noticeable pattern in Figure 3 is the wide dispersion of error rates for the test items in the Korean group. Some items triggered an error rate of 80% across all participants, while others were accepted as words much less often.

An immediate explanation for this variability does not present itself. Korean uses many loanwords from English, which might explain why certain nonwords trigger higher error rates than others (see Figure 4). A follow up survey was conducted with three trained linguists who spoke Korean as their first language. They were asked to mark which of the experimental items they thought might be common loanwords in Korean. They confirmed that loan words are almost always realized with an epenthetic vowel in the Korean form. English words that are used as loanwords in Korean without translation, in the same or slightly adapted meaning (e.g., *cream*, or *drive*) and having no alternative in the Korean language, were given a score of 2.

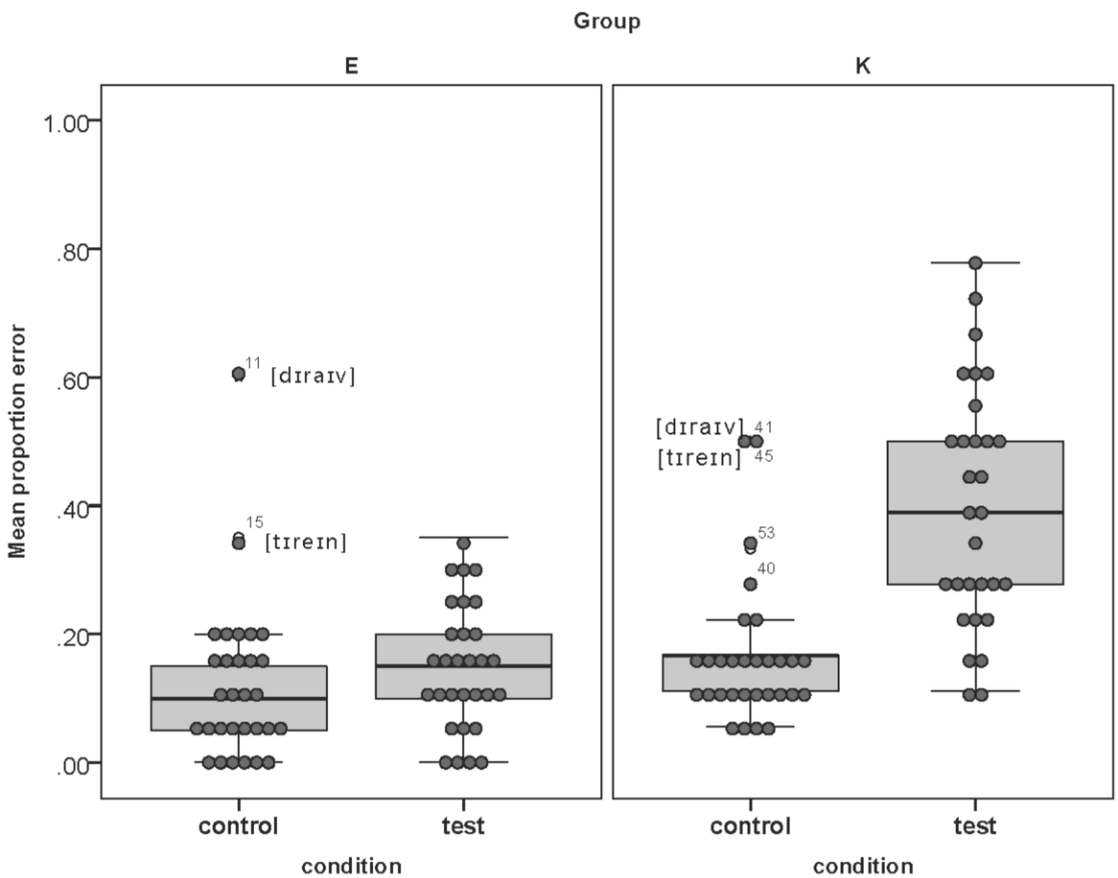

Figure 3. Box plot with overlaid dot plot for each condition and group (E = English; K = Korean). Each dot is one item. Horizontal lines are medians, boxes show the IQR representing 50% of the cases, whisker bars extend to 1.5 times the IQR. Outliers (circles) are cases with values between 1.5 and 3 times the IQ range, i.e., beyond the whiskers.

Words that are simply alternatives to the equivalent Korean forms (e.g. *blue*, *black*, *clean* and *play*) and that are sometimes used as loanwords (i.e., *clean* as [kʊli:n]), but whose Korean equivalents are much more frequently used, were coded as 1. Other words that they deemed were not or rarely used as loanwords in Korean were coded as 0. Finally, the experimental word list was checked against a list of attested English loanwords provided by the National Institute of the Korean Language. A match was coded as 1. The ratings given for each word were added. The resulting loanword scores ranged from 0 to 6, where 6 is a very frequent and common loanword.

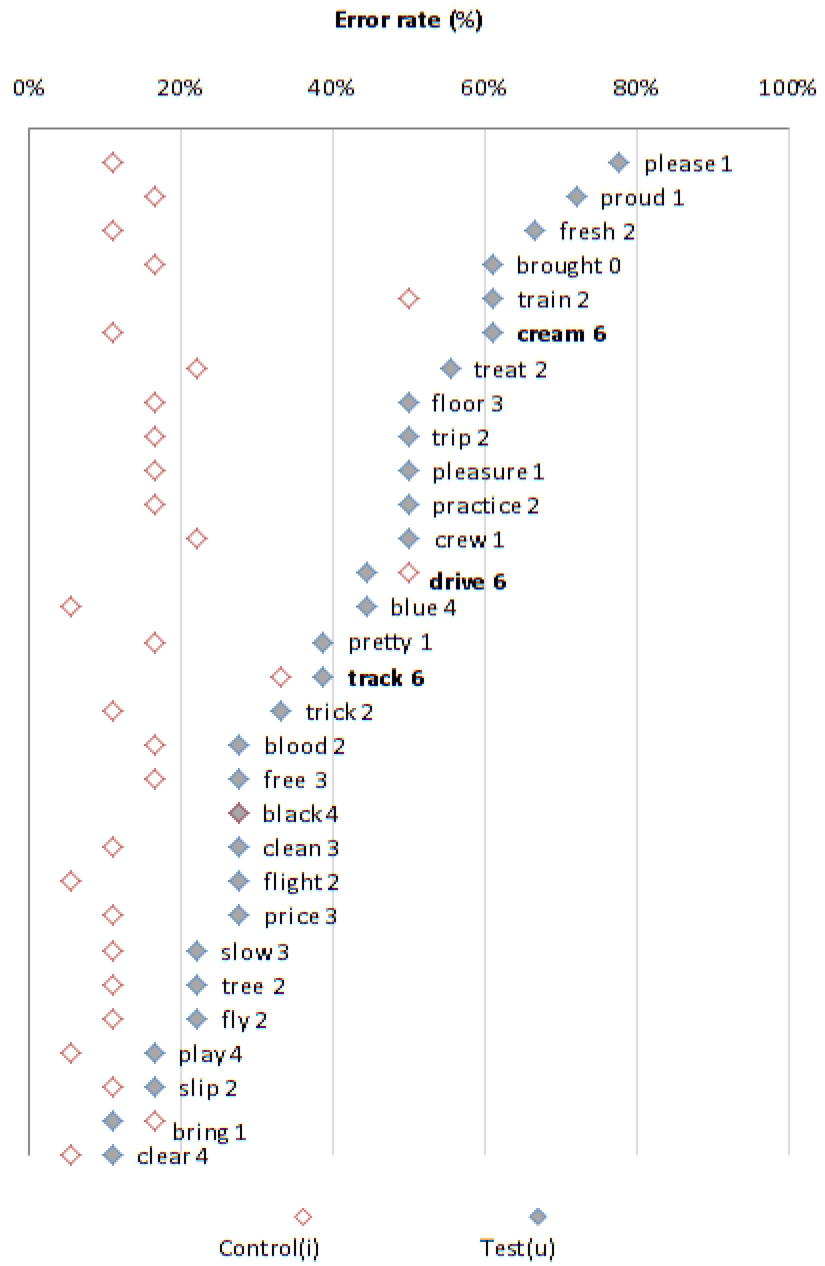

Figure 4. Individual error proportion by item for the Korean group in the control and test condition only. Empty diamonds represent an item in the control condition, full diamonds represent the corresponding item in the test condition.

If a status as common loanword were to explain the variance in error rates, one would expect that the error rates might be highest for words that are very commonly used as loanwords in Korean (i.e. with a high loanword score). As shown in Figure 4, where each word label is followed by its loanword score, three words received a definite status as loanword (*cream*, *drive*, and *track*) with a score of 6. Their respective error rates were 61%, 44% and 39%. Interestingly, the one word with a loanword score of 0 (*brought*) also triggered 61% errors. Loanword status in our stimuli does not appear to yield a clear effect, but this would be an interesting direction for future research.

There was no clear trend for clusters containing a voiced over voiceless consonant to trigger more error rates. However, the second consonant of the cluster ([ɹ] or [l]) was linked to different patterns of errors. The experimental word list contained 13 items out of 30 with a cluster containing [l], and 17 items out of 30 with a cluster containing [ɹ]. The effect of cluster type (/r/ vs. /l/) was examined by fitting a generalized linear mixed model with logit link function (with the *lme4* package in R 3.4.2, fit by maximum likelihood Laplace approximation) to the data with a categorical correct/incorrect dependent variable, as well as random effects (intercepts) for both item and participant. Cluster type had no main effect on error rates (b = -0.027, SE = 0.29, z = -0.092, p = .93) but the cluster-by-group interaction was significant (b = 0.76, SE = 0.27, z = 2.81, p < .005); In the Korean group, the mean error rate was 32% (SD = 18) for [l]-clusters and 45% (SD = 17) for [ɹ]-clusters, whereas the English group did not show any difference by cluster consonant. In summary, the effect of cluster type in our data—while not sufficient to draw robust conclusions given that the stimuli were not matched along this dimension—indicates that [ɹ]-words trigger more errors for Korean listeners. Further research is needed to possibly uncover why the difference between [ɹ] and [l] in the cluster might affect lexical access patterns in the Korean group.

Summary. Taken together, the analyses of error patterns and response latencies indicate that compared to English participants, Korean listeners experienced more difficulties—expressed both in higher error rates and slower RTs—when they had to reject nonwords containing the epenthetic vowel (e.g. [bʊ'luː]), but were overall more similar to native listeners when rejecting nonwords containing the control vowel (e.g. [bɪ'luː]). Therefore, it appears that our L2 participants did not globally accept nonwords containing any additional vowel: the effect we observe suggests that hearing a nonword with [ʊ] successfully activates a phonolexical representation, but this does not happen when the nonword contains a different vowel.

One possible interpretation of this effect is that the learners' phonolexical representation for this type of English words in fact contain this epenthetic vowel [ʊ], thus presenting a match to the percept. A slightly different interpretation – to which we return below in the general discussion – assumes that the stored representation is underspecified for the presence or absence of the vowel element (Lahiri & Marslen-Wilson, 1991). Accordingly, even if a vowel is determined to be present in the percept, it will not create a conflict with the underspecified representation during the matching process. Therefore, the percept will not effectively mismatch with stored representations (no-mismatch), and will not inhibit the activation of a lexical representation. In either scenario, a lexical representation can be activated by a nonword containing [ʊ].

3. General discussion

The main purpose of this study was to investigate the potential consequences of perceptual repairs in the mental lexicon for Korean learners of English. We hypothesized that Korean learners of

English might store English words with spurious vowels in their mental lexicon, as a result of perceptual epenthesis during word learning. We predicted that if epenthesized vowels were indeed lexically represented, Korean L2 learners would accept nonwords containing epenthetic vowels ([bʊ'luː] for *blue*) as real English words more often than English listeners. Our predictions were largely confirmed in a lexical decision task, where we observed a high error rate on test nonwords ([bʊ'luː]) by the Korean participants, accompanied by a low error rate on control nonwords ([bɪ'luː]). This pattern suggests a match between incoming signal and stored word forms in the case of the [ʊ] vowel. Notably, we demonstrate that Korean learners did not accept just any vowel as a word match, but only the [ʊ] nonwords, whereas the English native listeners mostly rejected all nonwords containing either [ʊ] or [ɪ]. Based on these results, we conclude that perceptual epenthesis can be lexically encoded in the mental lexicon of at least some of the Korean learners in our study, and that L1-based phonotactic and syllable structure restrictions can create pervasive phonolexical encoding issues for learners, underlining the challenging nature of L2 word learning.

An important aspect of our design was to circumvent a longstanding issue inherent to auditory lexical tasks in L2, in which the possible presence of perceptual difficulties (e.g. in case of phonotactically illegal stimuli) make it difficult to unambiguously conclude that the *representations* themselves are not target like (it could be an issue of input misperception, for instance). By avoiding the need for perceptual repairs during the presentation of experimental stimuli (all experimental items contained phonotactically legal word onsets for Korean), we sought to isolate the locus of the difficulty at the level of phonolexical representations themselves. Thus, we did not examine explicitly whether or not the stimuli were perceptually repaired during our experiment. This has been shown elsewhere (e.g. Kabak, 2003; see also Dupoux et al., 1999, among others). However, our design cannot completely rule out another, slightly different explanation. As we have shown, it is likely that all participants perceived the presence of inserted vowels clearly enough, and neither the acoustic analyses presented in Table 2 nor the pattern in Figure 3 suggest an exaggerated difference in acoustic salience between the two vowel conditions (which would have been picked up by both groups equally). Nevertheless, it is still possible that the Korean and the English participants *interpret* the vowel they perceive differently: For example, the inserted vowel might be perceptually less salient to Koreans, leading them to compensate for the presence of this vowel and accept these items as words more often (unlike the English participants), without necessarily implying that their representations contain this specific vowel. The reasons for reduced perceptual salience can be multiple. One possibility, related to variability in the input, is that the percept could be close enough, a "good enough match", or in lexical decision terms, not an effective mismatch to stored representations (Lahiri & Marslen-Wilson, 1991, Lahiri & Reetz, 2002). In other words, the Koreans' perceptual system might not allow the perceived stimulus to successfully inhibit the activation of words. This interpretation presupposes that lexical representations are not fully specified (see Paradis & Prunet, 2000; Marquez-Martinez, 2016) in some aspects of the CV skeleton, timing slots or syllabic structure—but exploring the various potential L2 representations in detail is beyond the scope of a single experiment. Further research will clearly be needed to elucidate these questions. Crucially however, while pointing to two different mechanisms and representations (specified with the epenthetic vowel vs. underspecified regarding the presence of the vowel), these two approaches essentially also lead to the same effect of L1 phonotactic grammar (and initial perceptual epenthesis) onto L2 lexical access.

We found evidence that a CVCV nonword ([bʊ'luː]) activates a representation in the learners' mental lexicon. Not all of the words, and not all the learners experienced this effect, but as a group, they were much more likely than the English participants to accept a CVCV stimulus as a good enough match for a word like *blue*. Taken together, our findings lead to the conclusion that the English phonotactic constraint licensing initial consonant clusters is not yet fully acquired by the majority of the Korean participants, and that to a substantial extent, the L1 phonotactic grammar shapes phonolexical representations for well-known English words. Yet, a few participants reached scores comparable to the native speakers' average of 14.5% error (SD = 21) in the test condition: two participants were below this "threshold", and 7 more were within 1 SD of this average (= below 36% error). In view of the large inter-individual variation we observed in our Korean participants (see Figure 1), it is important to address possible reasons for this variability. These include proficiency, perception, item familiarity, as well as the possibility of multiple representations.

It is possible that the learners, who have been learning English for an average of 11 years (range 6–15), differed in proficiency, and in how efficiently their L2 phonological grammar resisted the influence from the L1. We obtained a range of self-reported proficiency scores as well as data regarding learners' age, length of residence/exposure, length of learning, and weekly use, but no obvious factors appeared to easily explain these differences. None of these variables was significantly correlated with the error rate obtained in the test condition. Follow-up studies using standardized ways to measure proficiency would be needed to evaluate the lexical behavior of learners at different proficiency levels.

Another interesting question is whether learners' perception also relates to the quality of their lexical representations. We made the claim above that the form of learners' lexical representations might be a consequence of their initial tendency to perceptually repair words with onset clusters such as *blue* into /bʊ'luː/. While our study did not test for the presence of perceptual repairs in the learners, it is possible to imagine that once learners start being able to perceive onset clusters without perceptually repairing them, the lexical representations they form also might begin to encode words with CC clusters such as /bl/. It appears therefore useful to add a test of perceptual discrimination accuracy for nonword stimuli pairs such as [ˈplɔŋkət] vs. [pʊˈlɔŋkət], and [ˈplɔŋkət] vs. [pɪˈlɔŋkət], perhaps in combination with a novel word learning task.

Since it was important for the validity of the lexical decision task to establish that the participants knew the base words we used to create the nonwords, we only used known and highly familiar words for the study. In our data, given the variability in error rates we observed among these familiar items, it appears that familiarity with the words is unrelated to the error rates in the test condition. By the same token, it is unlikely that the orthographic familiarity with the base words had any strong impact on the error rates. Recall that familiarity with the written forms of the base words was extremely high. This possibly suggests that during the learning of these words, orthographic information was not sufficient to suppress phonotactic repairs. However, we did not specifically manipulate orthographic information, and our familiarity scale only contained three steps (see Paribakht & Wesche, 1993, and Wesche & Paribakht, 1996, for the full 5-step scale). It would be interesting to examine the potential impact of orthographic knowledge and word familiarity on error rates by obtaining more precise familiarity estimations, and by manipulating these factors explicitly.

Finally, it is clearly not the case that the phenomenon we observed—perceptual epenthesis in the L2 mental lexicon—unfolds in an all-or-none fashion for all L2 learners. Most likely, the development of lexical representations containing new L2-specific structures, as well as the suppression of those containing repairs, should be thought of as a gradual process, and the bilingual mental lexicon might contain at some point multiple variants for the same word.

Indeed, our data suggest that for the experimental words, most of the L2 learners have at least one phonolexical representation which contains an epenthetic vowel, but it might not be their only representation for these words. Target-like lexical representations may have been acquired as well, and could currently co-exist in the learners' mental lexicon—at least temporarily. From this perspective, in order to make accurate lexical decisions, the learners' task is likely twofold: acquire and inhibit. Once 'repair-free' lexical representations are established, the learner must still find a way to suppress or inhibit the activation of the co-existing, perhaps more entrenched non-native representations (with epenthesis), and learn to treat them as nonwords. In our dataset, participants probably also differed in how successful they were in doing so, for various reasons, ranging from perception accuracy to exposure or individual differences. It is possible that the learner who accepted 83% of the [ʊ]-nonwords had not yet built virtually any lexical representations containing initial clusters. But the majority of the other learners might already have co-existing, alternate lexical representations containing clusters for these words, since they rejected the items with epenthetic [ʊ] at least in some of the trials, which also suggests that they were able to inhibit or even fully suppress their initial representations with spurious vowels.

This proposal parallels the way words with two pronunciation variants may be stored in the lexicon with two lexemes (Bürki & Gaskell, 2012; see also Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer and Díaz, 2006). As the present data also show, some of these variants might be non-native, perceptually repaired forms, suggesting that encoding multiple lexemes may also extend to forms that result from interlanguage phonological processing, reflecting non-native phonotactic structure, such as /bʊ'luː/. This means that the bilingual mental lexicon may represent word forms that are never actually part of the speech signal, with a frequency of 0, unlike the attested pronunciation variants examined by Bürki and Gaskell (2012). Of course, it is possible that in our learners' linguistic environment, the frequency of forms with epenthetic vowels is not zero, but instead constitutes sizeable input. Without a detailed quantification of their input, it is not possible to fully answer these questions.

This explanation also echoes recent findings of learner variability in L2 productions by John and Cardoso (to appear), in which they also surmise that for L2 learners, more than one lexical representation can be temporarily co-existing. Gradually, learners become more adept at selecting the accurate form (without the inserted vowel), which initially happens less frequently, and mostly in formal or careful speech. Future research would be needed to examine lexical behavior using variationist methods.

Given the possibility of dual or multiple lexical representations and of less effective lexical candidate inhibition, future research will need to solve the fundamental questions of whether and how phonolexical representations can be corrected, overwritten, or their activation thresholds “selectively reduced” over time, and how this intersects with the acquisition of L2 phonotactic constraints in perception (eliminating the occurrence of perceptual repairs). Is improvement in a task such as ours dependent on reducing instances of perceptual epenthesis at the phonetic perception level? Or is lexical acquisition independent? (cf. Darcy et al., 2012; Darcy, Daidone & Kojima, 2013). If epenthetic representations indeed are first built because of the influence of the L1 phonotactic grammar, how can their impact be reduced? It would be important to triangulate these findings with perception, as said above, but also with production. Reducing the impact of such lexical representations containing spurious vowels might be a direct way to reduce their incidence during production (John & Cardoso, to appear). Variability should be taken into account, both in perception and in production, to determine which factors condition the likelihood of repairs in lexical representations, and their subsequent suppression.

From the learner’s perspective, however, for quite a long time *bʊlue* or *pʊroud* are perfectly good English words. As we have argued, this effect arises in large part because learners’ lexical representations for L2 words reflect the phonotactic grammar of their L1. While it is certainly possible to suppress non-native lexical representations, this process is gradual, and implies that perceptual epenthesis during L2 word learning can have long-term lexical consequences. To echo the words of Matthews and Brown (2004), the case at hand offers noteworthy evidence that intake exceeds input at the level of lexical representations too.

References

Abrahamsson, N. (1999). Vowel Epenthesis of /sC(C)/ Onsets in Spanish/Swedish Interphonology: A Longitudinal Case Study. *Language Learning, 49,* 473–508

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67,* 1–48. doi:10.18637/jss.v067.i01

Bettoni-Techio, M., Rauber, A. S., & Koerich, R. D. (2007). Perception and production of word-final alveolar stops by Brazilian Portuguese learners of English. In *Proceedings of Interspeech 2007*, pp. 2293 –2296. https://www.isca-speech.org/archive/interspeech_2007/i07_2293.html (retrieved October 8, 2018).

Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer* (Version 6.0.13). Retrieved from http://www.praat.org/ (retrieved February 1, 2016)

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

Bürki, A., & Gaskell, M. G. (2012). Lexical representation of schwa words: Two mackerels, but only one salami. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 617–631. doi: 10.1037/a0026167

Cardoso, W., John, P., & French, L. (2008). The variable perception of /s/ + coronal onset clusters in Brazilian Portuguese English. In M. A. Watkins, A. Rauber & B. O. Baptista (eds.), *Recent research in second language phonetics/ phonology: Perception and production,* pp. 203–231. Newcastle upon Tyne: Cambridge Scholars

Cardoso, W. (2011). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research, 27,* 433–465. doi:10.1177/0267658311413540

Carlisle, R. S. (1991). the influence of environment on vowel epenthesis in Spanish/English interphonology. *Applied Linguistics, 12,* 76–95. doi: 10.1093/applin/12.1.76

Carlson, M. T., Goldrick, M., Blasingame, M., & Fink, A. (2016). Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism: Language and Cognition, 19,* 939–954. doi:10.1017/S1366728915000334

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-to-Meaning Mappings. *Frontiers in Psychology, 7,* 1345. doi: 10.3389/fpsyg.2016.01345

Cuetos, F., Hallé, P. A., Domínguez, A., & Segui, J. (2011). Perception of prothetic /e/ in #sC utterances: Gating data. In W. S. Lee & E. Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII),* pp. 540–543. Hong Kong: City University of Hong Kong.

Cutler, A. (2005). Lexical stress. In D.B. Pisoni & R.E. Remez (eds.), *The Handbook of Speech Perception,* pp. 264–289. Oxford: Blackwell.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon, 8,* 372–420.

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research, 28,* 5–40. doi: 10.1177/0267658311423455

Davidson, L. (2006a). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics, 34,* 104–137. doi:https://doi.org/10.1016/j.wocn.2005.03.004

Davidson, L. (2006b). Phonotactics and articulatory coordination interact in phonology: evidence from nonnative production. *Cognitive Science, 30,* 837–862. doi: 10.1207/s15516709cog0000_73

Davidson, L., Shaw, J., & Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *The Journal of the Acoustical Society of America*, *122*, 3697–3709.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience, 12,* 635–647. doi: doi:10.1162/089892900562390

De Jong, K., & Park, H. (2012). Vowel epenthesis and segment identity in Korean learners of English. *Studies in Second Language Acquisition, 34,* 127– 155. doi:10.1017/S0272263111000520

Dupoux, E., Kakehi, 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,* 491–505.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language, 64,* 199–210. doi: http://dx.doi.org/10.1016/j.jml.2010.12.004

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*, 682–706.

Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2016). Phonotactic constraints are activated across languages in bilinguals. *Frontiers in Psychology, 7,* 702. http://doi.org/10.3389/fpsyg.2016.00702

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech-perception. *Perception & Psychophysics, 54*, 287–295

Gibson, M. (2012) Perception-based vowel insertion by native Spanish-speaking learners of English. *TIPA. Travaux interdisciplinaires sur la parole et le langage [Online], 28.* http://tipa.revues.org/205 (retrieved September 2, 2013). doi: 10.4000/tipa.205

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279. doi:10.1037/0033-295X.105.2.251

Graf Estes, K., Edwards, J., & Saffran, J. R. (2011). Phonotactic constraints on infant word learning. *Infancy, 16,* 180–197. doi: 10.1111/j.1532-7078.2010.00046.x

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *The Journal of the Acoustical Society of America, 121,* 2899–2914.

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

Hallé, P. A., Segui, J., Frauenfelder, U. H., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 592–608.

John, P., & Cardoso, W. (to appear). Are word-final consonants codas? Evidence from Brazilian Portuguese ESL/EFL learners. In J. Volin & R. Skarnitzl (eds.). *Pronunciation of English by speakers of other languages.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language, 32,* 402–420

Kabak, B. (2003). *The perceptual processing of second language consonant clusters*. Unpublished Ph.D. Dissertation, University of Delaware, USA.

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

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38,* 245–294. doi:10.1016/0010-0277(91)90008-R

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7* (pp. 636–675). New York, Berlin: Mouton De Gruyter.

Marquez-Martinez, M.-A. (2016). *The Acquisition of French Nasal Vowels: From First Language Allophony To Second Language Phonological Contrast*. Unpublished Ph.D. Dissertation, Indiana University, USA.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34,* 338–348.

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. doi:10.1177/13670069040080010201

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, *78,* 91–121. doi: https://doi.org/10.1016/S0010-0277(00)00109-8

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30,* 1113–1126.

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition, 84,* 55–71. doi: https://doi.org/10.1016/S0010-0277(02)00014-8

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

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition, 111,* 263–269.

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.

Paradis, C., & Prunet, J.-F. (2000). Nasal vowels as two segments: Evidence from borrowings. *Language, 76,* 324–357. doi:10.2307/417659

Paribakht, T. S., & Wesche, M. B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11,* 09–29. doi: doi.org/10.18806/tesl.v11i1.623.

Parlato-Oliveira, E., Christophe, A., Hirose, Y., & Dupoux, E. (2010). Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. *The Journal of the Acoustical Society of America, 127,* 3738–3748. doi: http://dx.doi.org/10.1121/1.3327792

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162,* 8–13. doi: https://doi.org/10.1016/j.jneumeth.2006.11.017

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics, 60,* 941–951. doi: 10.3758/bf03211930

Polivanov, E. (1931). La perception des sons d'une langue étrangère, *Travaux du Cercle Linguistique de Prague, 4,* 79–96.

Sebastián-Gallés, N., & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 974–989.

Sebastián-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience, 18,* 1277–1291. doi:10.1162/jocn.2006.18.8.1277

Shin, D.-J. (2014). *Training Korean speakers on English vowels and prosody: Individual differences in perception, production and vowel epenthesis*. Unpublished Ph.D. Dissertation, University College London, UK.

Sternberg, S. (1998). Inferring mental operations from reaction time data: how we compare objects. In D. Scarborough & S. Sternberg (eds.). *Methods, models and conceptual issues. An invitation to cognitive science, Vol 4.* (pp. 365–454). Cambridge, MA: MIT Press

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America, 119*, 597–607.

Wesche, M. B., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: depth versus breadth. *Canadian Modern Language Review, 53*, 13–40.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58*, 475–482.

[1] We refer to the timing units for consonants and vowels as C and V (see Lahiri & Marslen-Wilson, 1991). This level of timing units is called the CV skeleton, upon which the syllable is built.

[2] In addition to the citation form of words, lexical representations likely also contain other forms, for example the reduced forms found in conversational speech. However, it is still a matter of debate whether native listeners represent spoken words exclusively as phonetically detailed exemplars (Johnson, 1997; Goldinger, 1998) or whether phonological abstraction also takes place (McQueen, Cutler & Norris, 2006).

[3] The Korean epenthetic vowel is [ɯ] in most contexts except after palatal consonants where it is [i] or [ɨ], which are similar to American English [ɪ] (see Kabak, 2003).

[4] In this paper, we mainly use the IPA symbol [ʊ] for the American English high back rounded lax vowel in our examples, for the two reasons that 1) the Korean epenthetic vowel is perceptually similar to it, despite being unrounded itself, and 2) the American English [ʊ] is what was produced in our stimuli. However, we acknowledge that this representation is only an approximation of what learners might in fact encode in their lexical representations. The encoded vowel might even be intermediate between [ʊ] and [ɯ], yet it might still be activated by perceiving [ʊ] in our stimuli.

[5] A parallel model examined potential effects of block on the conditions control and test. There was no main effect of block and no interaction with condition. Therefore, we report the simpler model without this variable.

[6] We thank an anonymous reviewer for suggesting this analysis

Appendix

Experimental items (test /ʊ/ items and control /ɪ/ items), IPA and loanword score. All items are stressed on the second syllable.

Base word	/ʊ/ items	/ɪ/ items	Loanword score
black	bʊlæk	bɪlæk	4
blood	bʊlʌd	bɪlʌd	2
blue	bʊluː	bɪluː	4
bring	bʊɹɪŋ	bɪɹɪŋ	1
brought	bʊɹɔt	bɪɹɔt	0
clean	kʊliːn	kɪliːn	3
clear	kʊliːɹ	kɪliːɹ	4
cream	kʊɹiːm	kɪɹiːm	6
crew	kʊɹuː	kɪɹuː	1
drive	dʊɹaɪv	dɪɹaɪv	6
flight	fʊlaɪt	fɪlaɪt	2
floor	fʊlɔːɹ	fɪlɔːɹ	3
fly	fʊlaɪ	fɪlaɪ	2
free	fʊɹiː	fɪɹiː	3
fresh	fʊɹɛʃ	fɪɹɛʃ	2
play	pʊleɪ	pɪleɪ	4
please	pʊliːz	pɪliːz	1
pleasure	pʊlɛʒəɹ	pɪlɛʒəɹ	1
practice	pʊɹæktəs	pɪɹæktəs	2
pretty	pʊɹɪɾi	pɪɹɪɾi	1
price	pʊɹaɪs	pɪɹaɪs	3
proud	pʊɹaʊd	pɪɹaʊd	1
sleep	sʊliːp	sɪliːp	2
slow	sʊloʊ	sɪloʊ	3
track	tʊɹæk	tɪɹæk	6
train	tʊɹeɪn	tɪɹeɪn	2
treat	tʊɹiːt	tɪɹiːt	2
tree	tʊɹiː	tɪɹiː	2
trick	tʊɹɪk	tɪɹɪk	2
trip	tʊɹɪp	tɪɹɪp	2