

The Role of Phrasal Phonology in Speech Perception:
What Perceptual Epenthesis Shows Us.

Karthik Durvasula^a, Jimin Kahng^b

^aDepartment of Linguistics and Languages, Michigan State University

B330 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824, USA, durvasul@msu.edu

^bTeaching English to Speakers of Other Languages, Northeastern Illinois University

3069 LWH, 5500 N St. Louis Avenue, Chicago, IL 60625, USA, j-kahng@neiu.edu

Corresponding author: Karthik Durvasula

Address: B330 Wells Hall, 619 Red Cedar Road, East Lansing, MI 48824, USA

Telephone: 1-517-432-0194

Fax: 1-517-432-2736

Abstract

Recent research using the phenomenon of illusory vowels has raised our awareness of the extent to which speech perception is modulated by the listener's native-language phonological knowledge. However, most of the focus has been limited to *word-level* phonological knowledge. In this article, we suggest that the perceptual system recruits segmental phonological knowledge that makes crucial reference to prosodic domains far beyond the word-level. We report the results from three identification experiments on Korean and American English participants. In accordance with their native-language phonotactic constraints at the level of the *Intonational Phrase*, Korean listeners unlike American English listeners hear more illusory vowels in stimuli containing the sequence of voiced stops followed by nasal consonants (e.g. [egma]) than those containing voiceless stops followed by nasal consonants (e.g. [ekma]). The results are interpreted as support for the view that speech perception makes crucial reference to the concept of *reverse inference*.

**The Role of Phrasal Phonology in Speech Perception:
What Perceptual Epenthesis Shows Us^{*}.**

Keywords: Speech Perception; Illusory Vowels; Prosodic Domains; Korean; American English.

Introduction

There has been growing awareness in the last few decades that speech perception is modulated by the listener's phonological knowledge. Listeners are more attuned to phonological categories that are contrastive in their native-language (C. Best, 1994; Dupoux, Sebastián-Gallés, Navarete, & Peperkamp, 2008). Listeners also seem to employ their knowledge of the phonotactic constraints in their native-language (Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Kabak & Idsardi, 2007). However, it is important to note that the evidence provided in support of phonotactic knowledge being recruited during speech perception is largely focused on *word-level* phonotactics. In this article, we suggest through the phenomenon of *illusory vowels* that, along with knowledge of word-level phonotactic constraints and phonological categories in the native language, the perceptual system recruits segmental phonological knowledge that makes crucial reference to prosodic domains far beyond the word-level.

* This article was made possible due to the help and support of many individuals. First and foremost, we would like to thank the associate editor and three anonymous reviewers for valuable criticism that helped make this article much better. Second, we would like to thank Bill Idsardi, Alan Beretta, Yen-Hwei Lin and the members of the Phonology-Phonetics group at Michigan State University for many helpful discussions. Third, we would like to thank Hongjun Seo and Boram Koo for helping us with experiment design. Fourth, we would like to thank Alan Munn, Cristina Schmitt, and Suzanne Wagner for help with experimental equipment. Finally, we would like to thank the audiences of CLS 49 and Manchester Phonology Meeting 21 for probing questions, and helpful discussion.

In the past decade and a half, the phenomenon of illusory vowels has been especially useful in allowing researchers to develop an enriched understanding of the nature of the phonological knowledge that is implicated during speech perception (Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Berent, Steriade, Lennertz, & Vaknin, 2007; Dehaene-Lambertz et al., 2000; Dupoux et al., 1999; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011; Kabak & Idsardi, 2007; Monahan, Takahashi, Nakao, & Idsardi, 2008; *inter alia*). The general finding with studies on illusory vowels is that when a native speaker is presented with a nonsense word that contains a consonant sequence that violates the phonotactic constraints in their language, an illusory vowel is often perceptually induced in between such sequences, thereby creating an illusory sequence that respects the phonotactic constraints of the language. For example, when a Japanese listener is auditorily presented with stimuli such as [ebzo], where the consonant sequence is either naturally produced or created by splicing out the medial vowel completely from productions such as [ebizo] or [ebazo], they may actually perceive [ebuzo] given that [bz] is an illicit consonant sequence in Japanese, as shown originally by Dupoux et al. (1999). In contrast, French speakers correctly identified far fewer (if any) illusory vowels for the relevant stimuli in the same experiment. Thereby suggesting that the illusory vowels heard by the Japanese speakers were a language specific effect and not simply due to fine phonetic detail. The characteristic that is most crucial to such studies is that the test stimuli have sequences of sounds that are phonotactically illicit, and therefore contain novel combinations of segments (and phonological features). This allows one to probe the types of generalizations that are crucially implicated in speech perception.

Previous research on the perception of illegal phonotactic sequences has led to a variety of new proposals and insights related to speech perception: (1) Listeners make use of abstract

featural co-occurrence constraints (Moreton, 2002), (2) Abstract principles such as the Sonority Sequencing Principle modulate speech perception (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008; Berent et al., 2009, 2007), (3) Syllable-structure phonotactics play a more important role than linear phonotactics¹ (Kabak & Idsardi, 2007), (4) More generally, phonotactically illegal consonant sequences are perceptually “repaired” in a variety of ways other than through illusory vowels (Davidson, 2006, 2007; Davidson & Shaw, 2012; Hallé, Segui, Frauenfelder, & Meunier, 1998; Wilson & Davidson, 2013).

It is important to note that a lot of the above work probing the perception of illegal phonotactics is dependent on the phenomenon of perceptual epenthesis. First, Berent et al (2008, et seq) show that both English and Korean listeners perceive more illusory vowel in onsets with decreasing sonority (e.g., [md] and [nb]) than in onsets with rising sonority (e.g., [ml] and [nw]). Second, Kabak & Idsardi (2007) argue that for Korean listeners, there are more illusory vowels perceived in stimuli that violate syllable-structure phonotactic restrictions (e.g., [ecma], where [c] is not possible in coda positions in Korean), than those that violate just linear phonotactic restrictions (e.g., [elma], where the sequence [lm] is not a possible sequence of sounds in Korean within a word, though [l] is an acceptable coda, and [m] is an acceptable onset). Third, the phenomenon of illusory vowels is also modulated by the low-level acoustic differences of the stimuli containing illegal phonotactic sequences (Davidson, 2006; Davidson & Shaw, 2012; Dupoux et al., 2011). For example, Dupoux et al (2001) show that illegal phonotactic sequences created by splicing out a medial vowel from stimuli trigger more perceptual illusions for the

¹ Following Kabak & Idsardi (2007), ‘linear phonotactics’ refers to phonotactic knowledge about word-internal segmental sequences, without reference to any prosodic information. For example, in the nonsense word [elna], the sequence [ln] violates linear phonotactic constraints in Korean, since the sequence of segments is not possible within words. However, the sequence [ln] does not violate syllable-structure phonotactics because [l] is a perfectly acceptable coda segment and [n] is a perfectly acceptable onset segment in Korean, i.e., since the syllable [el] and [na] are both acceptable syllables in Korean, the nonsense word [elna] does not violate any syllable-structure phonotactics.

spliced out vowel than those that are naturally produced with the same phonotactic violation.

Finally, there have also been reports of illusory vowels being modulated by the place of articulation of the preceding consonant (de Jong & Park, 2012). This last set of results might perhaps be attributed both to language specific (stochastic) phonotactic patterns or to language-specific phonetic patterns such as differences in stop burst releases for different places of articulation² (Byrd, 1992; Crystal & House, 1988; Kang, 2003; Rositzke, 1943).

While there is extensive work on the perception of stimuli containing illegal phonotactics and how such violations along with other factors influence perceptual illusions, the bulk of the research has focused on the perception of nonsense words containing word-level phonotactic violations (Berent et al., 2009, 2007; Davidson & Shaw, 2012; Davidson, 2007; Dehaene-Lambertz et al., 2000; Dupoux et al., 1999, 2011; Kabak & Idsardi, 2007; Monahan et al., 2008; *inter alia*).

In contrast, there is little research on whether phonological generalizations beyond the word-level are utilized in speech perception. In a study that probed the use of language specific prosodic knowledge by listeners, French participants were presented with identical segmental sequences of /selafis/ *C'est la fiche/l'affiche* ‘It's the sheet/poster’ that differed only in the $f0$ of the vowel /a/ (Spinelli, Grimault, Meunier, & Welby, 2010). In a two-way forced choice task, increasing the $f0$ of the vowel /a/ increased the identification rates for the word *affiche* ‘poster’ in line with the intonational patterns in French across the *Accentual Phrase*, a phonological domain beyond the word level (Jun & Fougeron, 2002). While the results were consistent with listeners using language specific prosodic knowledge, a potential confound stems from the fact that the

² This latter connection between the modulation of perceptual epenthesis by the place of articulation of the preceding consonant was first explored to our knowledge in the context of accounting for vowel epenthesis in loanword patterns by Kang (2003).

study lacked a control linguistic group that did not pattern like the French participants. One might argue that this control linguistic group is important to establish that language-general auditory processes to do with f_0 processing cannot explain the results.

Additionally, in very recent work probing the effect of prosodic boundaries on phonetic categorization, Kim & Cho (2013) show that native Korean and American English listeners needed a longer VOT to identify a sound as /p/ after an *Intonational Phrase* boundary than after a word boundary. They interpreted this as evidence of the use of general prosodic knowledge by both native Korean and American English listeners. However, as the authors acknowledge, it is not clear that the participants used language-specific prosodic knowledge during the task, given both language groups showed the same pattern. Furthermore, an alternative explanation for their results is the possibility that the VOT boundary of the listeners was modulated by the durational differences in the pre-stimulus carrier phrase. For example, for the English listeners, the pre-stimulus carrier phrase “*Let's hear*” was 355ms in the Intonational Phrase condition, and 147ms in the word boundary condition. It is possible that this durational difference in the two conditions was interpreted by the participants as a speech rate difference, thereby resulting in a shift of the listeners' VOT boundary (Nagao & de Jong, 2007). More generally, given that listeners are able to use distal and proximal speech rate cues available in the acoustic input during speech perception (Dilley & McAuley, 2008; Heffner, Dilley, McAuley, & Pitt, 2013), to show that listeners are necessarily using their knowledge of prosodic patterns in their language during speech perception, it is important to ensure that such durational/speech rate cues be absent.

There is of course related work on speech perception that argues that listeners are able to compensate for assimilatory changes at the edges of words (Coenen, Zwitserlood, & Bölte, 2001; Gaskell & Marslen-Wilson, 1996; Gow, 2003; Mitterer, Kim, & Cho, 2013). For example, the

phrase “garden bench” /ga:dn bɛntʃ/ is often pronounced as [ga:dm bɛntʃ]³, where the word-final nasal /n/ has assimilated to the place of articulation of the following segment. It has been shown that listeners are able to compensate for such coarticulatory changes, i.e., when presented with an assimilated variant (e.g., [ga:dm]), listeners are able to recognize the unassimilated word (e.g., “garden”), but only when the nasal consonant is followed by a word that begins with a bilabial sound (e.g., [ga:dm bɛntʃ]). Evidence that such compensations to assimilatory changes are (at least, partly) due to knowledge of language specific patterns comes from research looking at compensations to voicing and place of articulation assimilations by French and English listeners (Darcy, Ramus, Christophe, Kinzler, & Dupoux, 2009). They show that while English speakers compensate for place of articulation assimilation (which is present in the language), they do not compensate for voicing assimilation (which is not present in the language). In contrast, French speakers compensate for voicing assimilation (which is present in the language), but not for place of articulation assimilation (which are not present in the language).

One could interpret such results as evidence of the speech perception mechanism using phonological knowledge about segmental alternations beyond the word-level. However, there are some contradictory findings and confounds in interpreting such results as evidence for the use of the relevant phonological knowledge; some of these have been discussed in the relevant literature. First, at least for the case of place of articulation in English, it has been argued that listeners take advantage of phonetic cues present in the acoustic input of the relevant coarticulated segment, so perhaps the listeners are just paying attention to acoustic cues within

³ We use /.../ to represent phonemic or “underlying” representations, and [...] to represent pronunciations or “surface” representations.

the segment instead of the context *per se* (Gow, 2003)⁴. Second, some work on place assimilation in German compounds suggests that the unassimilated variant is recovered both in appropriate phonetic contexts and inappropriate phonetic contexts, thereby suggesting again that knowledge of the contextual generalization might not be the underlying cause for such compensations (Gumnior, Zwitserlood, & Bölte, 2005). Finally, the phonological changes that such studies focus on typically happen *word-internally* too. To return to the example from English, the same assimilation also happens *within* words, e.g., “unbelievable” /ʌnbɪlivəbl/ is often pronounced as [ʌmbɪlivəbl]⁵ (Coetzee & Pater, 2011; Gumnior et al., 2005; Kiparsky, 1985). Therefore, it is not obvious that listeners are recruiting their knowledge of word-boundary alternations, as opposed to word-internal alternations, in order to compensate for the coarticulation in such experiments.

In this article, we test if listeners recruit their knowledge about phonological generalizations beyond the word-level during speech perception. More specifically, we test if Korean and (American) English⁶ listeners use their native-language knowledge of segmental generalizations that lead to a certain set of distributional facts at the level of a high prosodic domain (the *Intonational Phrase*) during speech perception.

The sensitivity of phonological generalizations to prosodic domains beyond the word-level is not unique to Korean or English. In fact, based on studies on different languages, many

⁴ Rather tangentially to our present purposes, the possibility that Gow (2003) raises does not seem to us to be a possible explanation for at least the Gaskell & Marslen-Wilson (1996) results, since unlike Gow (2003) they used naturally-produced fully-assimilated tokens wherein there was no possibility of remnant acoustic cues of the underlying form (phoneme).

⁵ While we are unaware of any corpus or production studies establishing this particular fact, we have referenced other linguists who have discussed the same observation. Furthermore, our discussions with other native speaker linguists (of American English) confirm this observation, at least impressionistically. Finally, we hasten to point out to the reader that if it turns out that there is no such word-internal optional assimilation, it would only support the general point we try to establish in this article that segmental patterns outside the word-level affect speech perception.

⁶ For the rest of this article, we use “English” to mean “American English”.

phonologists have argued that phonological generalizations in many languages are sensitive to domains beyond the word-level (Grijzenhout & Kabak, 2009; Nespor & Vogel, 1986; Selkirk, 1972, 1986; *inter alia*). Furthermore, it has been suggested that these domains form a *Prosodic Hierarchy*, and that there are at least three levels of prosodic organization above the (phonological) word-level – the *Phonological Phrase*, the *Intonational Phrase* and the *Utterance Phrase*. While others have argued for a more elaborate *Prosodic Hierarchy* (Jun, 1993; Jun 1996; Vogel, 2009; *inter alia*), such distinctions are not relevant for present purposes.

The arguments that motivate such a hierarchy involve the observation that phonological generalizations are typically bounded within one of these prosodic domains (Grijzenhout & Kabak, 2009; Nespor & Vogel, 1986; Selkirk, 1972, 1986; *inter alia*). Further evidence for the hierarchy comes from articulatory studies that have shown that the domain-edge segments get progressively longer adjacent to higher prosodic domains (Taehong Cho, Keating, Fougeron, & Hsu, 2003; Fougeron & Keating, 1997; Keating, 2003). For example, an [n] at a syllable boundary is shorter than one at a word boundary, which in turn is shorter than one at a *Phonological Phrase* boundary. Such incremental domain-edge lengthening differences can be observed both at the beginning and at the end of each domain. Dilley, Shattuck-Hufnagel, & Osterdorf (1996) provide additional support for the prosodic hierarchy through a study of glottalization in vowel-initial words from a radio-news corpus. They showed that the likelihood of glottalization depended on the prosodic position of the word, with the highest likelihood at the beginning of *Intonational Phrases*, a lower likelihood at the beginning of *Phonological Phrases* (labeled *Intermediate Phrases* in Dilley et al. (1996)), and the lowest likelihood at the beginning of words that are phrase-medial.

In what follows, we briefly present the phonological facts from Korean and English that are relevant for this article. In Korean, there is no lexical contrast between voiced and voiceless (unaspirated) oral stops (Silva, 1992; Sohn, 1999). The same stop phoneme manifests as a voiced stop in some phonological contexts, and as a voiceless (unaspirated) stop in other phonological contexts; thus, they are allophones of the same phoneme. In English, while the traditional viewpoint has been that there is a contrast between voiced and voiceless stops, more recently, laryngeal contrast in obstruents has been argued to be an aspiration or spread glottis contrast, where the unaspirated obstruent segments surface as voiced through passive voicing (Avery & Idsardi, 2001; Iverson & Salmons, 2003b; Jessen & Ringen, 2002). Having said this, we will assume the traditional viewpoint for expository reasons. This however does affect the argument presented in this article. As will become clear later, what is important for the current article is that phonetically voiced stops contrast with phonetically voiceless stops in coda positions in English.

The descriptive facts that are crucial to appreciate are as follows. The domain-sensitive phonological generalizations in Korean ensure the following distributional facts: (a) *Within* words neither voiced stops nor voiceless stops are licit adjacent to nasal consonants; word-medial sequences such as [...*km...]⁷ or [...*gm...] do not exist. In fact, /km/ and /gm/ sequences at the phonemic level surface as [ŋm] within the *Intonational Phrase*, which leads to divergent non-native pronunciations by Korean speakers for English loanword such as [bɪŋmɛk] ‘Big Mac’ (Kabak & Idsardi, 2007). (b) *Across* the much larger prosodic domain of the *Intonational Phrase* (IP), but not across any smaller domain, voiceless stops followed by nasals

⁷ We use the asterisk to represent impossible segmental sequences.

are licit [....k #_{IP} m...]⁸, but voiced stops followed by nasals are still not licit sequences [....*g #_{IP} m...] (Y. Y. Cho, 1990; Jun, 1993, 1996; Yongsoon Kang, 1992; Silva, 1992). We refer the reader to *Appendix A* for more elaborate linguistic evidence in support of this generalization from Korean. Contrastingly in English, voiced and voiceless stops are licit before nasal consonants both within words and across higher prosodic domains (for example, “batman” [bætmæn], and “The bat matters” [ðə bæt mærəz]; “Bradman” [brædmæn], and “The bad matters” [ðə bæd mærəz]). Consistent with the above asymmetry across larger prosodic domains in Korean, we show in this article that Korean listeners appear to perceive far more illusory vowels in nonsense stimuli with sequences of voiced stops followed by nasal consonants than stimuli with sequences containing voiceless stops followed by nasal consonants. However, since both sequences are prohibited within domains smaller than the *Intonational Phrase* (including words), the asymmetrical perceptual epenthesis cannot be a result of phonotactics at any domains smaller than the *Intonational Phrase* (including word-level phonotactics). Furthermore, consistent with the absence of a similar asymmetry in English phonotactic generalizations, English listeners show no such asymmetry in illusory vowel perception in similar contexts.

Before laying out the expectations for the series of experiments discussed in this article, it is important to present our conception of the nature of the problem that is being solved during speech perception (following Marr (1982)). We suggest that the task of the listener in speech perception is primarily a task of *reverse inference* - it is to identify the phonological representation that best maps to the given acoustic input. Therefore, when a Korean listener hears a nonsense sequence with a voiceless stop followed by a nasal consonant such as [ekma], it is not possible (given the word-phonotactics of Korean) to infer that it is a single word, but it is

⁸ We use #_{IP} to represent an *Intonational Phrase* boundary.

possible for them to parse the sequence as having an *Intonational Phrase* boundary between [k] and [m] ([ek #_{IP} ma]). Therefore, there is no need to perceptually “repair” the sequence [ekma] through an illusory vowel. Contrastingly, when a Korean listener hears a nonsense sequence like [egma], since the sequence [...gm...] is phonotactically illicit across all prosodic domains in Korean, there is no domain boundary (for example, a syllable boundary, or a word boundary, or a higher prosodic domain boundary) that a Korean listener could infer that would allow the listener to accept the sequence as a licit phonotactic sequence. Furthermore, as a result of the domain-sensitive generalizations in Korean, voiced stops in Korean can only surface before vowels; therefore the listener necessarily infers an illusory vowel after the voiced stops in [...gm...] contexts. So, we predict an asymmetry in the likelihood of illusory vowels induced after voiced and voiceless stops adjacent to nasal consonants for Korean listeners: voiced stop-nasal consonant sequences (e.g., [...gm...]) should induce more illusory vowels than voiceless stop-nasal consonant sequences (e.g., [...km...]) for Korean listeners. In contrast, since both voiced or voiceless stops are allowed before nasal consonants in English (within words and across higher prosodic domains), English listeners should be able to parse all the relevant sequences into licit phonological sequences without perceiving any illusory vowels. Therefore, there should crucially be no asymmetry in the perception of illusory vowels for the English listeners.

With respect to the quality of the illusory vowel during perceptual illusions, Dupoux et al. (2011) argue that it is the phonetically minimal element, or the shortest vowel, in the language ([u] in Japanese, and [i] in Brazilian Portuguese). They reason that if a vowel is necessary in a particular position for language-specific phonotactic reasons but is not present in the acoustic input, then the least egregious perceptual repair would be one where the shortest vowel in the language is inserted to fix the phonotactic violation. Acoustic studies of Korean vowels have

shown that the vowel [ɯ]⁹ is the shortest vowel in the language (Chung, Kim, & Huckvale, 1999; Han, 1964; K.-O. Kim, 1974). The typical duration of the vowels [ɯ] in phrase-initial contexts is around 144ms; the duration of the vowel [i] and [u] in a similar position is around 160ms and 165ms respectively (Chung et al., 1999). Given Dupoux et al.'s (2011) claim that the phonetically minimal element or shortest vowel is the illusory vowel, one would expect the vowel [ɯ] to be the illusory vowel. Another reason to expect the vowel [ɯ] to be the illusory vowel is that it is often deleted in Korean, especially in weak non-initial open syllables (Yoonjung Kang, 2003; Kim-Renaud, 1987). Therefore, following the logic of reverse inference laid out earlier, it is a good vowel to infer for a listener in phonological environments where a vowel is not present but is expected based on the phonological patterns of the language.

There are some seemingly inconsistent results from prior research that have a direct bearing on the above predictions (de Jong & Park, 2012; Hwang, 2011; Kabak & Idsardi, 2007). Hwang (2011) manipulated the duration of the medial vowel in [VC₁VC₂V] sequences where C₁ was either a voiced or voiceless stop, and tested these sequences on Korean and American English listeners. She observed that when C₁ was a voiced stop and the medial vowel was completely spliced out, Korean listeners perceived illusory vowels to a substantial degree, while English listeners did not. In contrast, when C₁ was a voiceless stop and the medial vowel was completely spliced out, Korean listeners didn't hear any appreciable number of illusory vowels just like the English listeners. In contrast, Kabak & Idsardi (2007) presented Korean and English listeners with naturally produced [VC₁C₂V] sequences where C₁ was either a voiced or voiceless

⁹ There is some debate in the phonological literature on the use of the *unrounded high back vowel* [ɯ] for the Korean letter ㅗ. Some have suggested that the *unrounded high central vowel* [ɨ] is perhaps more appropriate. Since the focus of the current article is not directly related to this issue, we use [ɯ] throughout.

stop (amongst other types of consonants), and found only a nonsignificant increase in the illusory vowel rates when C_1 was a voiced stop. Finally, de Jong & Park (2012) looked at the effect of word-final consonants on syllable perception of Korean listeners through a syllable counting task, and did not find a significant difference between the syllable counting errors associated with word-final voiced and voiceless consonants. Their results suggest that there is no appreciable difference between the illusory vowel rates for voiced and voiceless consonants in word-final position.

While the above results from the three papers appear inconsistent, it is important to point out that both the experiments that found nonsignificant results had stimuli that were spoken by native American English speakers (de Jong & Park, 2012; Kabak & Idsardi, 2007), while the experiment that did show a strong effect in the expected direction used a Korean-English bilingual (Hwang, 2011). As mentioned earlier, the voiced stops in American English have been argued to not be phonologically voiced, and most probably surface as voiced through *passive voicing*¹⁰ (Avery & Idsardi, 2001; Iverson & Salmons, 2003b; Jessen & Ringen, 2002). Furthermore, passive voicing into consonants in coda positions is typically very weak and is also phonetically conditioned by many prosodic and contextual factors (Iverson & Ahn, 2007; Jessen & Ringen, 2002). In both the experiments where there was no clearly observed increase in illusory vowels after voiced obstruents, the crucial segments were naturally produced in coda positions by American English speakers. Therefore, the lack of a clear voicing effect on illusory vowels could possibly have been due to insufficient voicing in the stimuli during the closure period of the relevant consonants. In contrast, in Hwang's (2011) study, the crucial segments were produced by a Korean-English bilingual in an intervocalic position, which is associated

¹⁰ *Passive voicing* is defined as phonetic voicing for which “the aerodynamic conditions for voicing are not actively controlled; voicing only continues during a stop closure as long as the articulatory context allows a pressure drop across the glottis” (Kohler, 1984, p. 162).

with much more robust voicing even if the voicing is through passive voicing. Therefore, the fact that there was a clearly observed voicing effect on illusory vowels in Hwang (2011) compared to the other two studies is not completely surprising. However, the fact that the stimuli in Hwang (2011) were created by splicing out a medial vowel raises the possibility of there being remnant vowel cues that caused the observed patterns. In fact, Dupoux et al. (2011) discuss results that support this possibility. They show that there are language-specific differences for stimuli with remnant vowel cues, and that participants treat stimuli with vowels spliced out differently from naturally produced stimuli with no vowels.

In the following sections, we discuss three identification experiments that support the predictions laid out above. While the focus of each of the experiments is on the performance of the Korean listeners, we specifically recruited American English participants in Experiments 1 and 2 as a control group to further ensure that the effects exhibited by the Korean listeners could not be attributed to general auditory processes, or stimulus artifacts or to response bias as a result of the experiment design. In Experiment 1, we test Korean and American English listeners on test items that have either a voiced or voiceless stop followed by a nasal consonant (for example, [egma] and [ekma]), at three places of articulation (labial, alveolar, and velar). As expected the Korean listeners perceive more illusory vowels in the voiced stop contexts than the voiceless stop contexts, but crucially the American English listeners show no such differences. In doing so, we replicate Hwang's (2011) results with naturally produced stimuli and thereby provide more evidence that her results were not spurious. In Experiment 2, we addressed the possibility that the Korean listeners perceive more illusory vowels in the voiced contexts because they were attending to phonetic cues in the release bursts of the voiced stops that they interpreted as vocalic cues, while the American English listeners were not. We gave Korean and American English

participants test stimuli at two places of articulation (velar and labial). The test stimuli for both places of articulation formed a voicing continuum, where the closure voicing duration of the crucial stop consonants was manipulated. Crucially, the stimuli for each place of articulation all had the same release burst, therefore, an increase in the illusory vowel rates with increasing levels of voicing could not be because of differences in release bursts. In Experiment 2, we show that the Korean listeners, but not the American English listeners, perceived more illusory vowels with increasing amounts of closure voicing. In Experiment 3, we addressed the possibility that perhaps the Korean listeners in Experiment 2 showed an increase in the perception of illusory vowels with increasing amounts of voicing because there is an interaction between voicing and the release cues, i.e., perhaps the release cues are perceived as more vocalic adjacent to increased amounts of closure voicing. To address this concern, we present Korean listeners with test stimuli similar to those in Experiment 2 with and without release burst, and show that the increase in illusory vowel perception with increasing voicing is present with both the stimuli with consonantal release bursts and the stimuli without consonant release bursts.

Experiment 1

Method

Participants

17 native Seoul Korean speakers (mean age 28, 10 men and 7 women) and 22 native (Midwestern) American English speakers (mean age 21, 4 men and 18 women) participated in

the experiment. All the subjects were recruited at Michigan State University, East Lansing, United States. None of the Korean speakers started to learn English or came to the US before the age of twelve. The Korean participants were compensated for their participation (\$10) and the English participants received extra credit for their participation.

Stimuli

Fifteen nonce words with the template eC_1V_1ma were used, in which C_1 was a labial, alveolar, or velar consonant [p, b, t, d, k, g]; V_1 was [i, u, Ø (null)]; and three distractors were also included [emma, enma, ejma]. Table 1 shows the test items used in the experiment. The crucial test items without intervening vowels had either a voiced stop followed by a nasal consonant, i.e., [ebma, edma, egma], or a voiceless stop followed by a nasal consonant, i.e., [epma, etma, ekma], at three places of articulation (labial, alveolar, and velar). For the items with vowels, voiced stops were used for C_1 , as voiceless (unaspirated) stops are pronounced as voiced in intervocalic positions in Korean.

None of the stimuli were words in either Korean or in English. They had stress on the first vowel and were natural recordings by a trained male phonetician (one of the authors), who is a native speaker of Indian English and Telugu, and a near-native speaker of standard Hindi. The speaker was chosen for two reasons: (a) to avoid providing either linguistic group with stimuli from their native dialect/language, (b) unlike many American English speakers, the speaker of this dialect of Indian English has clear and consistent closure voicing in coda position. Each item was recorded several times using the software Praat (Boersma & Weenink, 2012) with a microphone (Logitech USB Desktop Microphone; Frequency Response – 100Hz-16KHz) at a 44KHz sampling rate (16-bit resolution; 1-channel). From these recordings, two tokens were

selected for each item and they were each presented twice; therefore, there were 60 tokens in the experiment. The medial vowels of [ebuma, eduma, eguma] were shortened by splicing at zero-crossings in the middle of the vowel to be approximately 50ms in order to match the medial vowel length of the other tokens. The rest of the items were shortened by one pitch period, approximately 9ms, in the [m] at zero crossings in order to have all the items modified and they do not sound too different from the other tokens. The modification did not affect the naturalness of the stimuli. The stimuli were normalized in Praat to have a mean intensity of 60dB.

For native Korean listeners, we predict that the voiced stops before nasal consonants trigger illusory vowels to a much higher degree than voiceless stops in the same context. The expected illusory vowel in such contexts is the vowel [u], as discussed earlier. In contrast, for English speakers, since both voiced and voiceless stops are phonotactically licit before nasal consonants, we predict little to no perceptual epenthesis in such contexts.

Table 1

Test Items for Experiment 1

Medial Vowels				
	[u]	[i]	None	
			Voiced stop	Voiceless stop
Labial	ebuma	ebima	ebma	epma
Alveolar	eduma	edima	edma	etma
Velar	eguma	egima	egma	ekma

Procedure

An identification task was used to investigate a perceptual epenthesis effect. The experiment was conducted individually in a quiet room using a laptop computer. The stimuli were presented to each participant through an identification task scripted in Praat with a low-noise headset (Koss R80 headphones). All the instructions were in English for the English speakers ('Choose the vowel between the two consonants') and in Korean for the Korean speakers ('두 자음 사이의 모음을 고르세요'). The participants were asked to listen to each stimulus and determine whether the medial vowel was [u], [i], or nothing and click on the corresponding box (The actual choices were "u", "i", "nothing" for English listeners, and "으", "이", "없음" for Korean listeners) on the screen with a mouse. Before the actual experiment, each participant completed a practice session to ensure familiarity with the task. The practice session had 12 trials with another set of nonce words. In the actual experiment the inter-trial interval was 1000ms and the test tokens were presented in random order for each participant.

Results

Responses to all the stimuli can be found in *Appendix B* (Table B.1). When there was a medial vowel in the stimuli, both Korean and English participants showed high rates of correctly identifying the medial vowel. Furthermore, as expected, the no-vowel contexts were rarely misidentified as [i] by both Korean and English listeners, ranging 0% to 2.3%, except for [egma]. Korean listeners misidentified [i] 11.8% when presented with [egma]; however, the 3 out of a total 8 misidentifications came from a single participant. Figure 1 shows the percentages of

misidentification of the no-vowel contexts as [u]. As can be seen in the figure, the Korean subjects are much more likely to perceive an [u] in the voiced stop contexts than after the voiceless stop context. On the other hand, the English speakers are at ceiling with all the relevant stimuli. The percentages and error bars are mostly zeroes for the English speakers because of the floor effects.

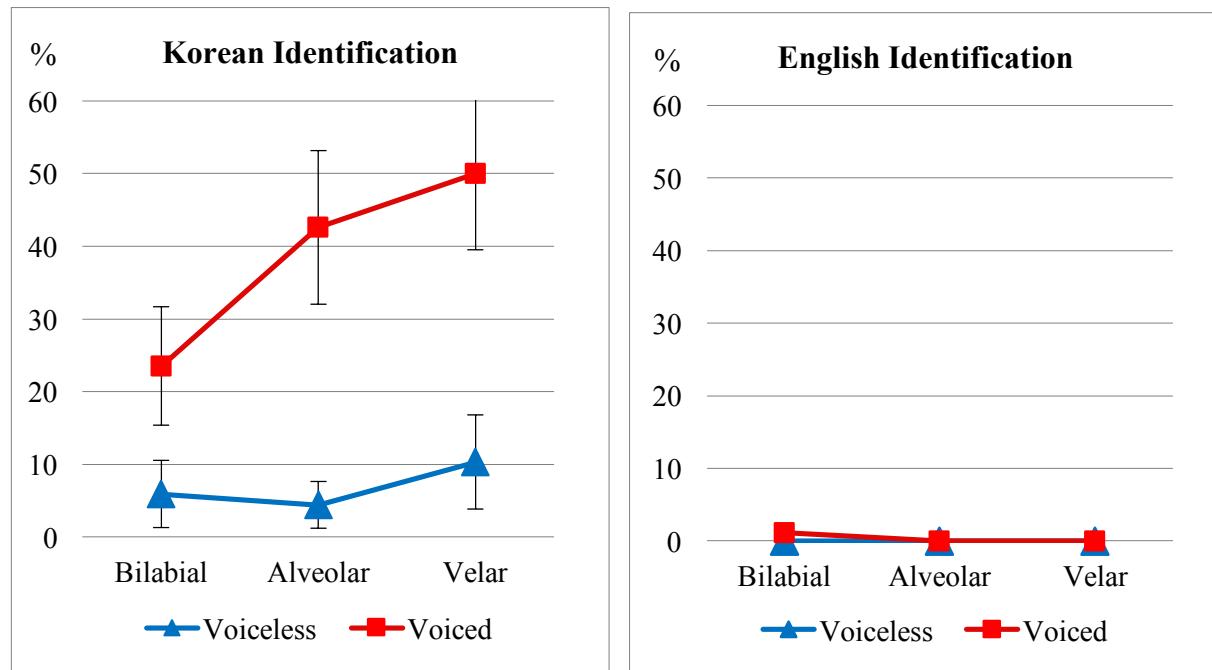


Figure 1. Percentages of erroneous [u] vowel identification for eCma stimuli, in Korean (left) and English (right) groups. Error bars represent standard errors. [Note: C = consonant]

In this article, wherever possible, subject responses were analysed using mixed-effects logistic regression models in the statistical software package R (R Development Core Team, 2014). The models were fitted using the *glmer* function available through the *lme4* package (Bates, Maechler, Bolker, & Walker, 2014). We attempted to obtain the maximal random effects structure that was justified by the data (Barr, Levy, Scheepers, & Tily, 2013). However, as is

typical in psycholinguistic data (and in our own experience), the models with the most complex random effects structures did not converge. It is important to note that the field and the statistical literature in general have not come to a consensus on how to best proceed in identifying the best random effects structure, especially when a model with a particular random effects structure does not converge (Bolker, 2014). In what follows, we describe the random effects structure selection process that we used for our experiments by following other experienced linear mixed effects modelers in psycholinguistics (Barr et al., 2013; Jaeger, 2009, 2011). We identified the appropriate random effects structure, by keeping the fixed effects constant – we used the full fixed effects model for the experiment (i.e., with interactions for all the fixed effects). We started with the most complex random effects structure. In the case of non-convergence of the complex random effects model, we systematically pared down the random effects structure till convergence was reached¹¹. The least complex random effects structure we entertained was one with a varying intercept for both subjects and items. When convergence was reached, the corresponding random effects model was identified to be the maximal random effects structure justified by the data.

Using the maximal random effects structure justified for the data that was identified as detailed above, model comparison to identify the best combination of fixed effects was performed through backwards elimination of non-significant terms, beginning with the interactions, through a Chi-squared test of the log likelihood ratios. The most complex fixed

¹¹ This is the same strategy used by Barr et al (2013) for their simulations: “Nonconverging LMEMs were dealt with by progressively simplifying the random effects structure until convergence was reached. Data from these simpler models contributed to the performance metrics for the more complex models. For example, in testing maximal models, if a particular model did not converge and was simplified down to a random-intercepts-only model, the p values from that model would contribute to the performance metrics for maximal models. This reflects the assumption that researchers who encountered nonconvergence would not just give up but would consider simpler models.” (Barr et al., 2013, pg. 266)

effects model entertained was the full model with all interaction terms, and the least complex model entertained was the model with only an intercept term and no fixed effects.

We attempted to fit mixed-effects models for all the data from both the language groups. Whereas the experiment had three responses, our crucial predictions were about increasing levels of [w] responses with voicing; therefore, the dependent variable was coerced into a binary variable that codes for whether listener responded with an [w] or not to the test items (i.e., [w] vs. everything else). This in effect models the percentage of [w] responses amongst all responses given by a subject for a particular stimulus. The independent variables were *Language*, *Place of Articulation* (POA) and *Voicing*. *Language* has two levels (English, and Korean). *POA* has three levels (Bilabial, Alveolar, and Velar), and *Voicing* has two levels (Voiced and Voiceless). The baseline was the alveolar voiced consonant context [d] for English listeners. Unfortunately, many of the more complex models did not converge even for the simplest random effects structure considered, i.e., varying intercepts for subjects and items (perhaps, due to the floor levels of the English responses), so the results of any subsequent model comparisons are not meaningful.

However, from visual inspection it is extremely clear that the results for the English and Korean listeners were substantially different, as observable in Figure 1, so we proceeded to fit two separate sets of mixed-effects logistic-regression models. The dependent variable was a binary variable that codes for listener responses to the test items ([w] vs. everything else), and the independent variables were *POA* and *Voicing*. As before, *POA* has three levels (Bilabial, Alveolar, and Velar), and *Voicing* has two levels (Voiced and Voiceless). The baseline was the alveolar voiced consonant context [d].

For the Korean listeners, the maximal possible random effects structure justified by the data was one with a varying intercepts for both subjects and items. The model with two independent (non-interacting) factors for *Voicing* and *POA* was the best model for the above random effects structure. Table 2 shows this model in more detail. As compared to the baseline, there was a statistically significant reduction in [w] identification rates for bilabial contexts (and trend towards an increase in [w] identification rates for the velar contexts). Since the model with the interaction term was not significantly different from the model without the interaction term, this suggests that there was a statistically significant result for all voiceless consonants compared to their voiced counterparts, thereby suggesting that the Korean listeners were indeed perceiving more illusory vowels in the context of voiced consonants than in the context of voiceless consonants.

Table 2

Logistic mixed effects model for Korean listeners (baseline = alveolar voiced consonant context [d])

Fixed Effects	Estimate	Std. Error	Z value	P-value
Intercept	-0.908	0.704	-1.290	0.2
POA: Bilabial	-1.118	0.446	-2.508	0.01 *
POA: Velar	0.728	0.409	-1.778	0.08
Voicing: Voiceless	-3.504	0.469	-7.471	<0.001 ***

The logistic mixed effects model for the English listeners did not converge even for the simplest random effects structure of a varying intercept for subjects and items, and perhaps it is even meaningless to fit such a model to the English data given all the participants were at floor levels with [w] identification rates – there was just one [w] identification in the results for the

test items; i.e., the English participants uniformly and categorically identified no vowels in the test items with no medial vowels.

We also tested if the English and the Korean listeners differed in their [u] identification rates for the voiceless consonant contexts. The resulting mixed effects logistic regression model did not converge even with the simplest random effect structure, and therefore we attempted to model the data with a regular logistic regression model with *POA* and *Language* as interacting factors¹² (Table 3). There were no statistically significant differences between the English and Korean listeners for the voiceless consonants, thereby suggesting that the [u] identification rates for the voiceless consonants for the Korean listeners did not differ statistically from the English listeners. However, a closer look at the very large standard errors in the model suggests that though it has converged, the estimates are not believable in our opinion; therefore, the above inference is far from secure.

Table 3

Logistic regression model for only the voiceless consonant contexts. (baseline = alveolar voiced consonant context [d] for the English listeners)

Fixed Effects	Estimate	Std. Error	Z value	P-value
Intercept	-21.57	3.11×10^3	-0.007	0.994
POA: Bilabial	-3×10^{-8}	4.407×10^3	0.0	1
POA: Velar	-4×10^{-8}	4.407×10^3	0.0	1
Language: Korean	18.49	3.12×10^3	0.006	0.995

¹² One could perhaps object that modeling the raw responses using logistic regression leads to a strong violation of the independence assumption made for each of the data points. For this reason, we also calculated the proportion of responses ([u] vs. everything else) for each subject, and then used the proportions to model a logistic regression model with the same predictor variables (which is perhaps not possible in other statistical software but can be done using the statistical software R as discussed by Greg Snow in this forum thread (Snow, 2012)). The consequent model had nearly identical estimates, standard error, z-values and p-values.

POA(Bil):Lng.(K)	0.30	4.407×10^3	0.00	1
POA(Vel):Lng.(K)	0.91	4.407×10^3	0.00	1

Given that the results for the voiceless consonant contexts, in our opinion, could not be modeled with a reasonable statistical model, we chose to look directly at the [w] identification rates in voiceless contexts. For the Korean listeners, 9 out of the total 14 [w] identifications in the voiceless contexts were from the same Korean listener. This presumably contributed to the rather large standard error estimates¹³. Furthermore, thirteen of the Korean listeners had no [w] identifications for the voiceless contexts, i.e., they were at floor levels just like the English listeners. Compare this to the voiced contexts where there were a total of 79 [w] identifications. This strongly suggests that, in general, the Korean listeners did not differ from the English listeners with respect to the voiceless consonant contexts.

While the data are completely consistent with the predictions we laid out earlier, there are three aspects that do require a comment. First, as mentioned earlier, there is an effect of POA on the [w] identifications rates. It is important to note that the experimental stimuli were all naturally produced. This raises the possibility of at least two potential sources of explanation for the POA effect we found. First, since the stimuli were naturally produced, they were not controlled for levels of voicing in that the consonants at different places of articulation could potentially have different degrees of acoustic cues associated with voicing¹⁴, and therefore the

¹³ Thanks to an anonymous reviewer for bringing this point to our attention.

¹⁴ Note: it is not enough to measure the VOTs of the different tokens to ascertain this fact. Since these are naturally produced stimuli and voicing is known to have many acoustic cues, assessing the aggregate effect of these cues on perception, and ascertaining different degrees of voicing does not have an obvious solution. This is one of the reasons we conducted Experiments 2 and 3, where the stimuli are more controlled to vary in a particular acoustic dimension only.

only prediction that can be directly tested is that of a difference between voiced and voiceless contexts in general. Second, again since the stimuli were all naturally produced, it is possible that there were subtle differences in consonant release associated with the coda consonants at different POA. Such differences in consonantal release at different POA have been observed for American English speech (Byrd, 1992; Crystal & House, 1988; Kang, 2003; Rositzke, 1943). Furthermore, different levels of consonantal release have been linked to different rates of vowel epenthesis in loanwords (Kang, 2003). Finally, epenthetic vowels in loanwords have been attributed to illusory vowels (Peperkamp, 2005; Peperkamp & Dupoux, 2003). It is therefore possible that there were subtle differences in consonantal release in the stimuli that might have been reinterpreted as vocalic cues by the Korean speakers (as Korean coda stops are unreleased). However, the English listeners might have just treated them as consonantal releases (as such variation is possible for coda stops in English).

Second, the floor effects observed for the English listeners in Experiment 1 need to be interpreted with caution. While the absence of difference between the voiced and the voiceless contexts is very much in line with our predictions for the English listeners, it is possible that any underlying differences are actually masked by the floor effects observed. Although the experiment might be interpreted as being “easier” for the English listeners, it is unclear what fact about the design or stimuli can lead to such ease for only the English speakers. Therefore, given the floor effects, we suggest that perhaps future experiments probing similar questions attempt to make the task harder, perhaps through more attentionally demanding designs, to move away from such floor effects for the control group.

Third, a closer look at the [w] identification rates in the case of voiceless stimuli revealed that while most Korean listeners patterned as the English listeners, one listener in particular

behaved very differently. The individual differences observed could either result from generic random response variation, or from response error on the part of the subject, or from systematic differences in the knowledge possessed. The issue of individual differences is a difficult one to address, especially in the context of our experiments, because it is not clear to us how to separate random variation in subject responses from systematic knowledge-driven differences in subjects. Furthermore, it is not clear to us that any of the models that are out there do a good job of separating these two aspects. However, we acknowledge the importance of the issue and leave it as a topic for future work.

Returning to the main prediction tested in this experiment, despite the observed POA effect, it is clear that there is a clear effect of voicing on illusory vowel rates for the coda consonants at all POA in the case of the Korean listeners, but not in the case of the English listeners.

Experiment 2

Experiment 1 showed that phonotactically illicit voiced stops in Korean triggered more illusory vowels than their voiceless counterparts in the same environment, whereas English listeners showed no such difference between voiced and voiceless stops. Furthermore, as expected English listeners showed no evidence of perceiving illusory vowels in any context. However, it is possible that there are some asymmetric phonetic cues for vowels in the release portions of the voiced and voiceless consonants – the release portions of the voiced stops might have cues more consistent with reduced vowels than the release portions of the voiceless stops

(Wilson & Davidson, 2013). Furthermore, different language-specific perceptual strategies related to the presence/absence of release cues have been observed for French and English listeners (Flege & Hillenbrand, 1987).

More directly relevant to our present experiments is that fact in Korean, syllable-final (coda) stops are typically unreleased (except perhaps in word-final positions where there is a brief audible release (Kim & Jongman, 1996)), while in English coda stops may be released. Furthermore, there are differences in release cues of coda stops in English due to both place of articulation and voicing (Byrd, 1992; Crystal & House, 1988; Kang, 2003; Rositzke, 1943). Therefore, it is possible that the English listeners due to their experience with differences in release cues of coda voiced/voiceless stops are able to accommodate for such differences during perception. In contrast, since Korean listeners do not have such relevant experience, they might not be able to accommodate for different amounts of release, and might treat them as vocalic cues with different associated strengths. As a consequence, it is possible in Experiment 1, that the asymmetry between the English and the Korean listeners is due to the release cues directly, and not voicing per se. Therefore, if such asymmetric cues are present in the stimuli in Experiment 1, the two language groups may just be treating them differently; that is, the Korean listeners might be attending to the cues, while the English listeners might be ignoring such cues. In order to test the effect of the voicing of the stop consonant on perceptual epenthesis, it is important to show that the Korean listeners behave differently with different levels of voicing of the stop. Therefore, Experiment 2 presented different levels of voicing to the two language groups, while keeping the release bursts constant, and compared their identification of illusory vowels. The prediction was that the more voiced the stop consonant before the nasal consonant is, the more that context should trigger perceptual epenthesis for Korean listeners, but not for

English listeners. As mentioned earlier, another reason for including the American English participants was to ensure that the behavior of the Korean listeners could not be attributed to general auditory processes, or stimulus artifacts or to response bias as a result of the experiment design.

Method

Participants

16 native Seoul Korean speakers (mean age 28, 10 men and 6 women) and 18 native (Midwestern) American English speakers (mean age 21, 4 men and 14 women) participated in the experiment. They were the participants who took part in the first experiment. All the subjects were recruited at Michigan State University, East Lansing, USA. None of the Korean speakers started to learn English or came to the US before the age of twelve.

Stimuli

In order to test the effects of degree of voicing on perceptual epenthesis, [ebma, ebwma, egma, eguma] from Experiment 1 were used as base tokens for creating the stimuli in Experiment 2. To build the two continua of [epma-ebma] and [ekma-egma], the base token with minimal voicing was first created by removing the closure portion completely from [ebma] and [egma] and replacing it with silence (100ms) at zero crossings. Voicing was extracted from the voiced token from the beginning of closure. A continuum of minimal closure voicing-maximal closure voicing tokens was created by splicing in an increasing number of pitch periods to the

base voiceless token. Each step of the continuum contained approximately an additional 10ms of voicing for bilabials and 8ms of voicing for velars. For every pitch period of voicing added, a certain amount of silence was removed from the base (12.5ms). Therefore, the closure portion got shorter as voicing was added. This created a continuum of 9 stimuli for [ebma] and [egma]. Out of this, five equally spaced stimuli representing approximately 0%, 25%, 50%, 75%, and 100% of voicing were chosen for both [ebma] and [egma]. Figure 2 illustrates the two continua, each with five levels of voicing.

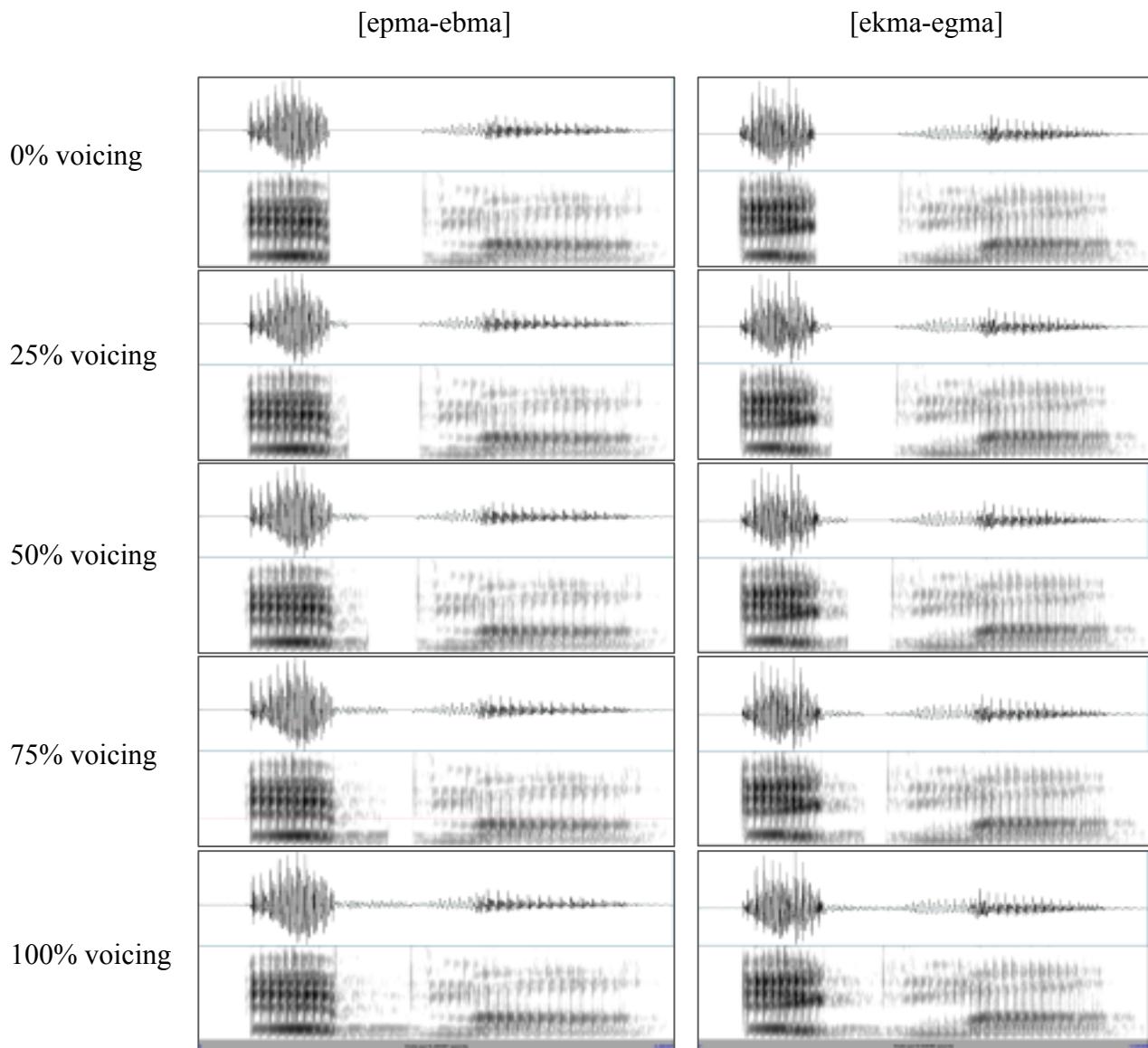


Figure 2. Stimuli for Experiment 2

In addition, [emma] and [enma] were included as distractors. They were shortened by two pitch periods in the first consonant at zero crossings so that all the stimuli are modified and they do not sound too different from the other tokens. The items with a medial vowel [ebuma, egwuma] and the two additional items [emma, enma] served as distractors and were expected to be perceived by all participants successfully. All the stimuli were normalized in Praat to have a mean intensity of 60dB.

Procedure

As in Experiment 1, an identification task was used to investigate the perception of illusory vowels. The ten test items in Figure 2 and five tokens of each of the distractors [ebuma, egwuma, emma, enma] were repeated seven times; therefore, there were 210 tokens in each session.

The experiment was conducted individually in a quiet room using a laptop computer. The stimuli were presented to each participant through an identification task scripted in Praat with a low-noise headset (Koss R80 headphones). All the instructions were in English for the English speakers and in Korean for the Korean speakers as in Experiment 1. The participants were asked to listen to each stimulus and determine whether the medial vowel was [u] or nothing and click on the corresponding box (The actual choices were “u” and “nothing” for English listeners, and “으” and “없음” for Korean listeners) on the screen with a mouse. In Experiment 2, the option

“i” was dropped to make the task more straightforward. As mentioned in the discussion for Experiment 1, the option “i” was rarely chosen by either Korean or English listeners; therefore, this alteration does not affect the interpretation of the results for Experiment 2. After the participants completed Experiment 1, they took a short break. Before the second experiment, each participant completed another practice session to ensure familiarity with the task of Experiment 2. The practice session had 12 trials with another set of nonce words. In the actual experiment there were 210 trials in total. The inter-trial interval was 1000ms. All the trials were randomized for each participant.

Results

In Experiment 2 participants heard stimuli that contained a stop consonant with different levels of voicing and determined whether the medial vowel was [u] or nothing. Responses to all the stimuli can be found in *Appendix B* (Table B.2). When there was a medial vowel in the stimuli, both Korean and English participants correctly identified the medial vowel. Figure 3 shows the percentages of misidentification of the no-vowel contexts as [u]. As can be seen, the Korean subjects are more likely to perceive an [u] as the level of voicing increases. On the other hand, the percentages and error bars are mostly zeroes for the English speakers.

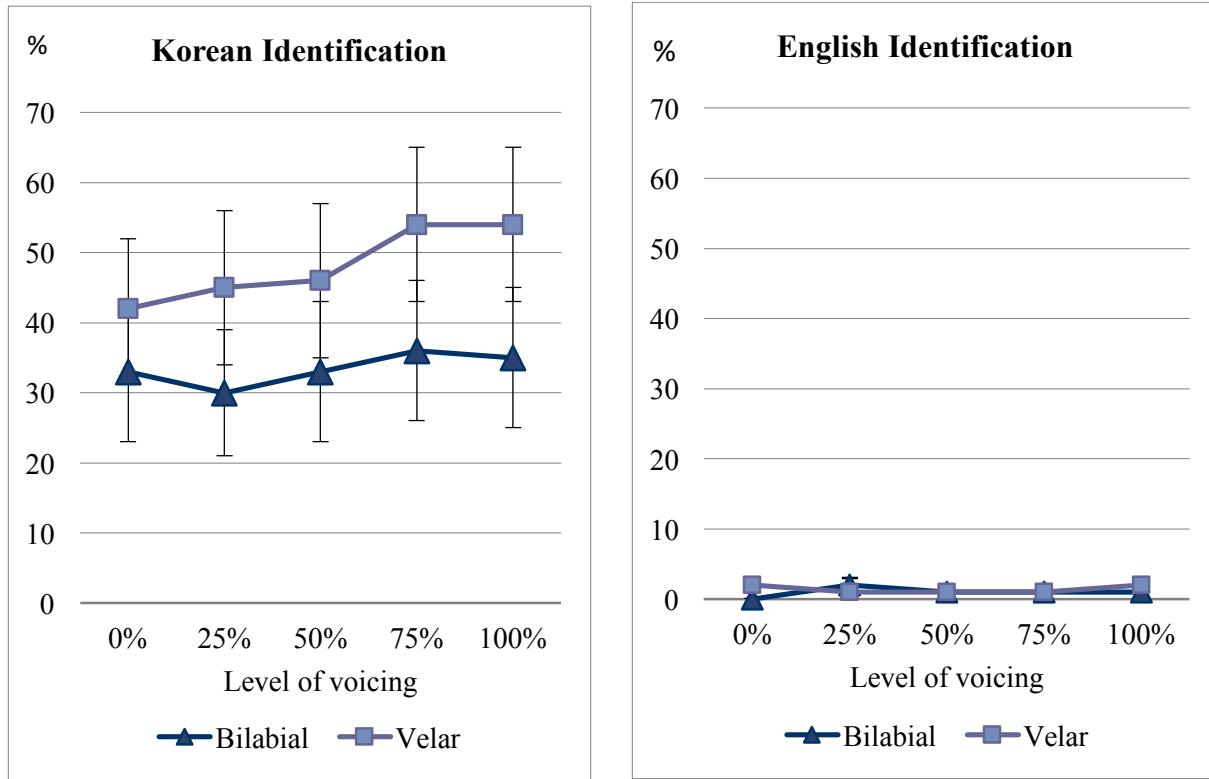


Figure 3. Percentages of erroneous [u] vowel identification when presented with different levels of voicing, in Korean (left) and English (right) groups. Error bars represent standard errors.

As in the first experiment, the results for the English and Korean listeners were substantially different, as observable in Figure 3, so we decided to fit mixed-effects logistic-regression models. The models were fitted and compared in the same manner as described for Experiment 1. The dependent variable was a binary variable that codes for listener responses to the test items ([u] vs. nothing), and the independent variables were *Language*, *Place of Articulation* (POA) and *Voicing*. *Language* has two levels (Korean and English). *POA* has two levels (Bilabial, and Velar), and *Voicing* was treated as a continuous variable. The baseline was the bilabial consonant context [b] with minimal voicing for English listeners.

The fully saturated model with three-way interaction did not converge for even the simplest random effects structure considered, i.e., varying intercepts for subject and items.

However, all other models except the fully saturated model converged. So, in Table 4 below, we present the results of the best model amongst the remaining models. The best model was the one with an interaction term for *POA* and *Language*, and a main effect term for *Voicing*¹⁵. As can be observed from the table, there was a statistically significant increase in [w] identification rates with increasing levels of *Voicing*. There was also a main effect of *Language*. Finally, there was also an interaction of *POA* and *Language*.

Table 4

Logistic mixed effects model for both language groups in Experiment 2 (baseline = bilabial consonant context [b] with minimal voicing)

Fixed Effects	Estimate	Std. Error	Z value	P-value
Intercept	-7.61	1.01	-7.54	<0.0001
POA (Velar)	0.35	0.60	0.6	0.56
Language (Korean)	5.25	1.20	4.38	<0.0001
Voicing	0.21	0.07	3.08	<0.01
POA (Velar): Language (Korean)	1.24	0.64	1.95	0.05

However, it should be strongly emphasized that, given that the fully saturated fixed effects model did not converge and was not part of the model comparisons, the results of the above model are very difficult to interpret. So, we suggest caution in deriving any inferences from it, and in what follows, we look at each language separately for a more accurate picture of the results of this experiment.

For the individual languages, the dependent variable was a binary variable that codes for listener responses to the test items ([w] vs. nothing), and the independent variables were POA

¹⁵ On the request of an anonymous reviewer, we included the next largest fixed effects model with an interaction term *Voicing* * *Language* in Appendix C.

and *Voicing*. *POA* has two levels (Bilabial, and Velar), and *Voicing* was treated as a continuous variable. The baseline was the bilabial consonant context [b] with minimal voicing.

For the Korean listeners, the maximal possible random effects structure justified by the data was one with a random intercept and a random slope of *Voicing* for subjects and a random intercept for items. The best model with this random effects structure was the model with non-interacting factors for *Voicing* and *POA*, thereby suggesting that the effect of increasing levels of voicing on both velars and bilabial consonant contexts was not significantly different. Table 5 shows this model in more detail. As can be gleaned from the table, there was a statistically significant increase in [w] identification rates with increasing levels of *Voicing*. Furthermore, the velar consonant contexts in general triggered more illusory [w] identifications than the bilabial consonant contexts.

Table 5

Logistic mixed effects model for the Korean listeners in Experiment 2 (baseline = bilabial consonant context [b] with minimal voicing)

Fixed Effects	Estimate	Std. Error	Z value	P-value)
Intercept	-2.715	0.948	-2.865	<0.01 **
Voicing	0.296	0.107	2.766	<0.01 **
POA: Velar	1.618	0.218	7.437	<0.001 ***

For the English listeners, the maximal possible random effects structure justified by the data was one with a varying intercept and varying slope of *Voicing* for subjects and a varying intercept for items. However, with that random effects structure, there were no significant effects, i.e., the simplest model with no factors best represented the data. This confirms the interpretation that is available through Figure 3 that there are no observable effects of *Voicing* or

POA, i.e., as in Experiment 1, the English speakers were close to floor levels with respect to [u] identification rates in the test items.

Again, the results confirm the expectations laid out earlier. However, there are two aspects of the results that can benefit from further discussion. First, even the lowest levels of voicing triggered higher levels of [u] identification for both bilabial and velar contexts compared to the voiceless contexts in Experiment 1. This seems reasonable given the fact that the base stimulus for each POA was the voiced token from Experiment 1, from which all voicing during the closure was manipulated. This means, that the other acoustic cues that correlate with voicing were still present in all the stimuli. Therefore, even the minimally voiced stops are more likely to be categorized as voiced stops in Experiment 2 than the voiceless stops in Experiment 1. As a consequence, even the minimally voiced stops in Experiment 2 should trigger higher rates of [u] identification than the voiceless contexts in Experiment 1. Second, just as in Experiment 1, the [u] identification rates related to the bilabial contexts were observed to be lower for Experiment 2; however, this is also to be expected given the stimuli for Experiment 2 were constructed using those of Experiment 1; therefore, any stimuli differences in the way of different levels of voicing or different levels of release cues would carry over Experiment 2. As a consequence, the reasons mentioned to account for the effect of POA in the results section of Experiment 1 are equally pertinent for the current experiment.

Finally, a potential confound in both Experiment 1 and Experiment 2 is the fact that English does not have the vowel [u] in it, therefore, the speakers might have been reluctant to identify the vowel in the crucial test items with no medial vowels. However, though the vowel [u] does not exist in English, the vowels [i] and [u] do exist in the language. We expected, given

how close [u] is to these two vowels, especially [u], that English listeners would largely choose [u] or “u”, and perhaps sometimes choose [i] or “i” in their responses. Furthermore, the vowel [u] is only very slightly rounded in casual American speech, and therefore, phonetically it is extremely close (in fact, overlapping in F1-F2 space) to the vowel [u] (Kabak & Idsardi, 2007). It is well documented that when a vowel is not in the language, but there are phonetically close vowels, the listeners of the language classify the non-native vowel as the closest vowel category that exists in their language (C. T. Best, Halle, Bohn, & Faber, 2003). This expectation was validated both in Experiment 1 and Experiment 2. Crucially, in both experiments, the choice of *no vowel* was minimal when the stimuli did have a full medial vowel. This is especially clear in the case of Experiment 2, where the only options were “u” and *no vowel*. In this case, the English listeners overwhelmingly chose “u” (about 95% of the time) as their preferred choice when presented with the full vowel control stimuli¹⁶. So, it is very unlikely that the English listeners chose the *no vowel* responses for the crucial test items without medial vowels because the vowel [u] does not exist in the language. Finally, it needs to be mentioned that Kabak & Idsardi (2007) also used English speakers as a control population to probe the illusory vowel effect in Korean (but testing a different set of hypotheses), and they reported that the English listeners had no trouble interpreting the [u] as [u], which is a vowel in their language. However, while we maintain that the different vowel choices given to the English listeners are unlikely to be the source of the observed language-specific differences for the reasons discussed above, we would like to caution the reader that they keep this potential confound in mind while interpreting the results of Experiment 1 and Experiment 2.

¹⁶ See *Appendix B* for details.

Returning to the main hypothesis tested through this experiment, as predicted for the Korean listeners, there is a clear effect of increasing levels of closure voicing on illusory vowel rates for the coda consonants.

Experiment 3

In Experiment 2 only Korean listeners but not the American English listeners perceived more illusory vowels with increasing amounts of closure voicing. Crucially, all the test items in Experiment 2 had natural release bursts after the crucial prenasal consonants. Flege & Hillenbrand (1987) showed that the perception of native French listeners, but not native English listeners, was influenced by the presence/absence of final stop release cues. They ascribed this difference in perception to the patterns of final stop releases in native French and English productions. Therefore, it is important to rule out the possibility that the Korean and English listeners in Experiment 1 and Experiment 2 were simply attending differentially to the consonantal release cues based on the native language syllable-final stop release patterns.

Korean syllable final stops are typically described as unreleased; however, Kim & Jongman (1996) document in a production experiment that up to 83% of the word-final stops had a brief (inaudible) release burst. On the other hand, English syllable final stops can be realized as either released or unreleased in word-final position (Lisker, 1999), and Henderson & Repp (1982) observed that in preconsonantal positions, the release rates of syllable final stops in English are on average about 58% (Henderson & Repp, 1982). Furthermore, it has been observed that differences in consonantal release of stimuli do affect Korean listeners during segment

identification (T Cho & McQueen, 2006). With respect to English listeners, the observed effect of consonantal release on segmental identification has been a bit more inconsistent in that some studies show that such cues do not appear to be attended to, while others show that the consonantal release cues appear to modulate the perception of segmental identification at some places of articulation but not others (Flege & Hillenbrand, 1987; Lisker, 1999).

As mentioned earlier, listeners sometimes interpret consonantal releases as (reduced) vowels (Wilson & Davidson, 2013). Taking this fact into account along with the cross-linguistic differences in the perceptual effects of consonantal releases discussed in the preceding paragraph, it is possible to ascribe the differences between the Korean and English listeners in Experiments 1 and 2 to differences in the perceptual strategies of the two language groups in dealing with the consonantal release cues in syllable final stops, i.e., perhaps the release cues are perceived as more vocalic adjacent to increased amounts of closure voicing for the Korean listeners, but not the English listeners. Therefore, Experiment 3 addressed this concern and we tested Korean listeners with test stimuli similar to those in Experiment 2 with and without release burst. We show that the increase in illusory vowel perception with increasing voicing is present with both the stimuli with consonant release bursts and the stimuli without consonant release bursts.

Method

Participants

20 native Seoul Korean speakers (mean age 26, 13 men and 7 women) participated in the experiment. All the subjects were undergraduate or graduate students at a university in Seoul,

South Korea. All participants had at least several years of formal school experience with English. The participants were compensated for their participation (₩5,000).

Stimuli

In order to test the effects of degree of release on perceptual epenthesis, we used the closure-voicing continuum for [egma] from Experiment 2. Furthermore, a closure-voicing continuum for [egma] without release was created by splicing out the release burst at zero crossings from the base token with minimal closure voicing from Experiment 2 and then adding increasing amounts of closure voicing as described in Experiment 2. This created a continuum of 9 stimuli each for [egma] with and without release. Out of this, five equally spaced stimuli representing approximately 0%, 25%, 50%, 75%, and 100% of voicing were chosen for both [egma] with and without release.

In addition, [eguma, elma, eruma, ejma] were included as distractors. They were shortened by two pitch periods in the first consonant at zero crossings so that all the stimuli are modified and they do not sound too different from the other tokens. All the stimuli were normalized in Praat to have a mean intensity of 60dB.

Procedure

As in Experiments 1 and 2, an identification task was used to investigate the perception of illusory vowels. The ten test items and five tokens of each of the distractors [eguma, elma, eruma, ejma] were repeated six times; therefore, there were 180 tokens in each session.

The experiment was conducted individually in a quiet room using a laptop computer. The stimuli were presented to each participant through an identification task scripted in Praat with a low-noise headset (Sony MDR-ZX100 ZX). All the instructions were in Korean (‘뭐라고 들렸습니까?’, Translation: What did you hear?). The participants were asked to listen to each stimulus and choose the sound they heard written in Korean by clicking on the corresponding option on the screen with a mouse. In order to make the task as non-metalinguistic as possible, the participants were given options containing the whole sequence of sounds written in Korean orthography. There were four options—one with the medial vowel [ɯ], one without the medial vowel, one with a nasal coda, and “none of the above.” For instance, the options for [egma, eguma, eŋma] were “이 그 마” ([eguma]), “에 마” ([ekma]), “엥 마” ([eŋma]), and “모두 아님” (none of the above). Before the actual experiment, each participant completed a practice session to ensure familiarity with the task. The practice session had 9 trials with another set of nonce words. In the actual experiment there were 180 trials in total. The inter-trial interval was 1000ms. All the trials were randomized for each participant.

Results

In Experiment 3 participants heard stimuli that contained stop consonants with different levels of voicing either with or without release and identified what they heard among the options written in Korean orthography. The options were 1) sound with the medial vowel [ɯ], e.g., “이 그 마” ([eguma]), 2) sound without the medial vowel, e.g., “에 마” ([ekma]), 3) sound with a nasal coda, e.g., “엥 마” ([eŋma]), and 4) “모두 아님” (none of the above). Responses to all the

test stimuli can be found in *Appendix B* (Table B.3). Stimuli with no vowel were rarely misidentified as a sound with a nasal coda or “none of the above” by the Korean listeners, ranging 0% to 6%. Figure 4 shows the percentages of misidentification of the no-vowel contexts as [w]. As can be seen in the figure, for both items with and without release, the Korean subjects are more likely to perceive an [w] as the level of voicing increases.

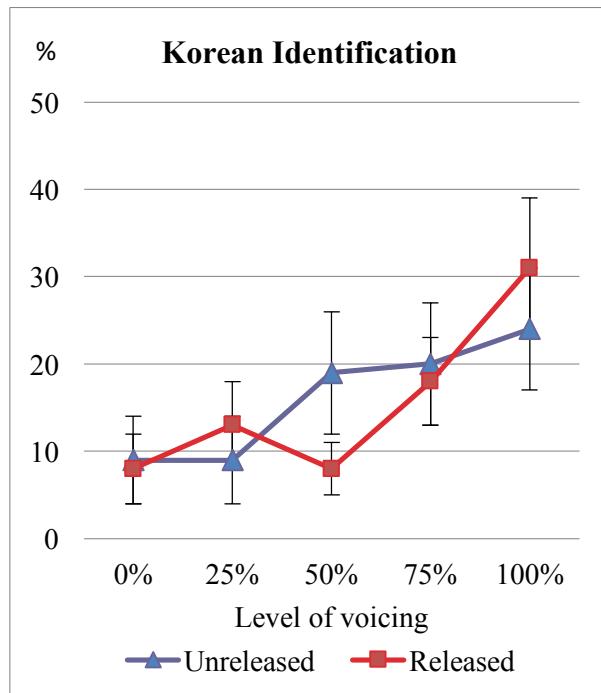


Figure 4. Percentages of erroneous [w] vowel identification when presented with different levels of voicing using items with and without release. Error bars represent standard errors.

Following the methodology laid out in Experiment 1, we fitted mixed-effects logistic-regression models to the results of the Korean listeners. As in Experiment 1, the dependent variable was coerced into a binary variable that codes for whether listener responded with an [w] or not to the test items (i.e., [w] vs. everything else), and the independent variables were *Release*

and *Voicing*. *Release* has two levels (Released, and Unreleased), and *Voicing* was treated as a continuous variable. The maximal possible random effects structure justified by the data was one with a varying intercept and a varying slope of *Voicing* for subjects, and a varying intercept for items. The baseline was the released consonant with minimum voicing.

The best model was the one with only one factor, *Voicing*, thereby suggesting that there were no statistically significant differences between the [w] identification rates of released and unreleased stimuli. Table 6 shows this model in more detail. As can be seen, there was a statistically significant increase in [w] identification rates with increasing levels of voicing.

Table 6

Logistic mixed effects model for Korean listeners in Experiment 3

Fixed Effects	Estimate	Std. Error	Z value	P-value)
Intercept	-6.10	1.38	-4.41	<0.001
Voicing	0.79	0.24	3.24	<0.01

The results of Experiment 3 show that the effect of voicing on illusory vowels is present regardless of the presence/absence of release cues. Therefore, the results show that the results in Experiments 1 and 2 are unlikely to be purely due to a complex interaction between voicing and release cues. However, we suggest more circumspection in the interpretation of the lack of observed difference between the stimuli with release cues and those without release cues as us claiming that there is no difference. Kang (2003) argues that differences in release cues in English source words do affect epenthetic vowels in Korean loan words. To the extent that such epenthetic vowels can be attributed to illusory vowels, it might seem reasonable to infer that

release cues do affect the perception of illusory vowels for Korean listeners. In fact, visual inspection of the highest voicing levels of both continua does impressionistically suggest that release cues show a tendency to increase illusory vowel percepts even in our experiment, since the highest amount of incorrect illusory vowel identifications do seem to be related to the released continuum. However, given that this is not statistically significant in our experimental results, it is unclear if Kang's (2003) results regarding epenthetic vowels in Korean loanwords can be reduced to the effect of release cues on illusory vowels.

Returning to the main hypothesis tested through this experiment, it is clear that there is an effect of increasing levels of closure voicing on illusory vowel rates for both released and unreleased coda consonants.

Discussion

In this article, we have shown that listeners recruit knowledge about segmental generalizations in their native-language beyond the word-level during speech perception. More specifically, we have shown that Korean and American English listeners recruited their knowledge of domain-sensitive segmental generalizations that result in a certain set of distributional facts at the level of the *Intonational Phrase* (IP). As discussed above, Korean, but not English, phonological (segmental) generalizations ensure that there is an asymmetry with respect to voiced and voiceless stops before nasal consonants, i.e., [...C_[voiced]N...]¹⁷ and [...C_[voiceless]N...] respectively. Consequently, while voiced stops followed by nasal consonants, i.e. [...C_[voiced]N...] sequences, are prohibited across all prosodic domains in Korean, voiceless

¹⁷ Note: we use 'C' for oral consonants and 'N' for nasal consonants.

stops followed by nasal consonants, i.e. [...C_[voiceless]N...] sequences, are possible (only) across *Intonational Phrases*. Following previous work on perceptual epenthesis that shows that phonotactically illicit sequences trigger illusory vowels, we show that Korean listeners (unlike the English listeners) perceive more illusory vowels in the voiced stop contexts than in the voiceless stop contexts (Experiment 1). It is important to note that since both sequences are prohibited within words, the asymmetrical perceptual epenthesis cannot be a result of word-level phonotactics. Furthermore, we also show that the asymmetry in the illusory vowels perceived cannot be due to phonetic differences in the release characteristics of the voiced and voiceless stops (Experiments 2 & 3). In Experiment 2, we controlled for the possibility that the Korean listeners were interpreting the release portions of the voiced and voiceless stops before nasal consonants differently from the English listeners, and therefore performed differently from the English listeners in Experiment 1. In Experiment 1, there was a possibility that the English listeners completely ignored the consonant release bursts, while the Korean listeners misinterpreted the different release characteristics of voiced and voiceless stops as either more or less vowel-like. This would also account for the asymmetry observed in Experiment 1. Therefore, in Experiment 2 we presented the participants (Korean and English listeners) with phonetic voicing continua synthesized from the same tokens; thus ensuring that the release characteristics of all the tokens in each continuum were identical. Consistent with our expectations, Korean listeners, but not English listeners, showed an increase in perceptual epenthesis with increasing amounts of voicing. Furthermore, given that the English listeners patterned very differently from the Korean listeners in both Experiment 1 and Experiment 2, the behavior of the Korean listeners cannot be attributed to general auditory processes, or stimulus artifacts or to response bias as a result of the experiment design. Finally, in Experiment 3, we

controlled for the possibility that the Korean listeners in Experiment 2 were interpreting the release portions of the stops as increasingly vowel-like in the context of increasing amounts of closure voicing of the preceding stop, and therefore perceived increasing amounts of illusory vowels. We tested Korean listeners on another set of continua where release portions of the base voiceless tokens were completely spliced out. The Korean listeners heard an increasing number of illusory vowels with increasing voicing levels of the stop consonant even when the release portion was spliced out. The results of the three experiments are consistent with the view that the Korean listeners were perceiving more illusory vowels in [...]C_[voiced]N...] contexts than in [...]C_[voiceless]N...] contexts. The observed asymmetry in the perceptual illusions strongly suggests that the Korean listeners used their knowledge of Korean segmental phonology at prosodic domains far above the word during speech perception.

One could attempt to account for the facts presented in this article by saying that the local context of voiced stops is intervocalic even within the word; therefore, a Korean listener may choose the vowel option in the above experiments after voiced stops because voiced stops in Korean are likely to be followed by a vowel¹⁸. However, this potential explanation cannot account for why the illusory vowel is [u] and not [i] for Korean listeners in Experiment 1, though both vowels trigger intervocalic voicing in Korean¹⁹. Furthermore, this potential account also makes what seem to us to be very unlikely predictions for the voiceless stop contexts in Korean. Crucially, unreleased voiceless stops in Korean cannot be followed by nasal consonants within a word in Korean; they can however be followed by other voiceless obstruents within a word. Therefore, such unreleased voiceless stop segments in the input, as per the above account,

¹⁸ Thanks to an anonymous reviewer for bringing it to our notice.

¹⁹ In fact, our experience with such experiments strongly suggests that the illusory vowel always seems to be [u] in these contexts, and no other vowel. Therefore, the choice of illusory vowels is a bigger problem for this view than is alluded to in the text.

should trigger the perception of another obstruent consonant following the voiceless stop because these are highly likely based on surface phonotactic patterns within words. This strikes us as an extremely unlikely result, since native speakers in informal discussions (including one of the authors) reject this as impossible. While our experiment does not directly test this possibility, given the extremely unlikely nature of this particular repair and the indeterminacy of the actual illusory vowel in voiced contexts alluded to earlier in the paragraph, we suggest that the above account cannot be a reasonable explanation of our results.

A second issue revolving around the proposal in the current article is that typically pitch (F0) contour, or segment duration (and perhaps syntactic structure) have been thought to be relevant to infer an IP boundary, while we claim that segmental alternations can also be used by listeners to infer an IP boundary²⁰. This might at first sight seem quite surprising. However, to our knowledge, there has never been an account that argues that pitch contours and segmental duration are the only such predictors of higher prosodic domains such as the IP. In fact, the viewpoint developed in this article naturally extends to patterns beyond pitch contours and segmental durations. Pitch contours and segmental duration are both aspects that are systematically manipulated in the phonological (or phonetic) grammars of languages; that is, these are aspects that can be systematically different across (or at the end of) different prosodic domains in a language. Therefore, from our perspective, it is only natural that these phonetic cues can be very good predictors of higher prosodic domains during speech perception. Following this line of reasoning, there is in fact nothing surprising about a segmental alternation being used to infer a higher prosodic domain as long as that alternation is systematically used to mark that particular domain boundary in the phonological or phonetic patterns of the language, which is exactly the case with the voicing and nasalization patterns in Korean.

²⁰ Thanks to an anonymous reviewer for bringing it to our notice.

The view developed in this article that prosodic phonological knowledge is recruited during speech perception allows for a reinterpretation of the very interesting results by Kabak and Idsardi (2007). They suggested based on a perception experiment on Korean listeners (with English listeners as controls) that perceptual illusions are driven by listeners' knowledge of syllable-structure restrictions, but not by their knowledge of linear phonotactic restrictions. Given that speakers of a language possess knowledge about generalization that involves linear-phonotactic restrictions in their language (Albright, 2009; Moreton, 2002; Scholes, 1966) and that people are able to learn such linear phonotactic patterns in artificial language experiments (Sharon Peperkamp & Dupoux, 2007; Seidl & Buckley, 2005), it would be puzzling if listeners recruited their knowledge of syllable-structure restrictions but not linear-phonotactic restrictions during speech perception. In their experiment, Kabak and Idsardi (2007) used nonsense words with three different kinds of medial consonant sequences: (a) stimuli that have consonant sequences that are allowed in the language [kt^h , lt^h], (b) stimuli that have consonant sequences that violate the "linear" phonotactic restrictions of Korean [ln , gt^h , km , gm], (c) stimuli that have consonant sequences that violate the "syllable-structure" phonotactic restrictions of Korean that disallow palatal stops in coda positions [ct^h , jt^h , cm , jm]. They argue that Koreans appear to treat stimuli with the "linear" phonotactic violations on par with the stimuli with licit consonantal sequences in that they do not perceive substantially more illusory vowels compared to licit sequences. With respect to their stimuli containing consonant sequences that violate the "linear" phonotactic restrictions of Korean [ln , gt^h , km , gm] in their article, we have shown in this article that nonsense stimuli with voiced stops (at three different places of articulation) all consistently trigger more illusory vowels than corresponding voiceless stops in the same position. As mentioned in the introduction, Kabak and Idsardi themselves observed non-significant

differences for similar sequences; however, it is important to note that their results were clearly trending in the same direction as our results here. Again, as discussed in the introduction, the differences in the results could stem from a variety of factors such as experiment design and stimulus construction. The other two sequences that they labeled as linear phonotactic violations [ln, km] are both sequences that violate just linear *word-level* phonotactic restrictions; however, they are both well-formed linear sequences across the *Intonational Phrase* in Korean. On the other hand, the sequences they label as “syllable-structure” violations [ct^h, jt^h, cm, jm] are illicit across all prosodic domains, and are thus impossible within any prosodic domain in Korean. Therefore, it is straight-forwardly possible to reinterpret Kabak and Idsardi’s (2007) results as support for the main thesis of the current article that sequences in nonsense stimuli that are not possible across some domain in the listener’s language (be it within words or across larger prosodic domains) induce more illusory vowels than those that are possible across some domain. This in turn strongly suggests again that listeners recruit their knowledge of phonological generalizations beyond the word-domain during speech perception. It is also important to highlight in this context that though we provide a reinterpretation of Kabak and Idsardi’s results, we agree with their general position that the listener uses abstract phonological knowledge during the task of speech perception.

Our results are consistent with recent research in theoretical phonology. In recent years, there has been a growing awareness in theoretical work related to loan-word phonology that speech perception heavily recruits phonological knowledge (Boersma & Hamann, 2009; Kabak & Idsardi, 2007; Yoonjung Kang, 2003). Our results are consistent with such work in claiming that, perhaps, speech perception recruits phonological knowledge so heavily that it might be

difficult, or even impossible, to separate the effects of phonological knowledge and speech perception on loanword phenomena.

In conclusion, we have argued here through the phenomenon of illusory vowels that, along with knowledge of word-level phonotactic constraints and phonological categories in the native language, the perceptual system recruits segmental phonological knowledge that makes crucial reference to prosodic domains far beyond the word-level. In doing so, we hope to have increased the interest in studying the extent to which the native-language phonological grammar is utilized during speech perception.

Appendix A

A variety of researchers have argued that phonological domains are necessary to explain the phonological processes of Korean (Y. Y. Cho, 1990; Jun, 1993, 1996; Yongsoon Kang, 1992; Silva, 1992). There are two domain-sensitive processes in Korean that are relevant²¹ (Table A.1). First, Korean has a phonological process of *Voicing* that targets voiceless stops in inter-vocalic positions²². Second, Korean also has a process of *Nasalization* that targets stops before nasal consonants.

Table A.1

Relevant Phonological Processes in Korean (Ahn, 1985; Iverson, 1993; Sohn, 1999)

Process	UR	SR
Voicing		
Rule:	/i + kuk + i/	[igugi] ‘this soup (NOM)’
$/C_{\text{lax stop}}/ \rightarrow [\text{voiced}] / V - V$	/i + pap + i/	[ibabi] ‘this rice (NOM)’
Nasalization		
Rule:	/kuk + man/	[kuŋman] ‘soup only’
$/C_{\text{stop}}/ \rightarrow [\text{nasal}] / - [\text{nasal}]$	/pap + man/	[pamman] ‘rice only’

Furthermore, as previously recognized by many linguists, the two processes in Korean are bounded within certain phonological domains (Y. Y. Cho, 1990; Jun, 1993, 1996; Yongsoon Kang, 1992; Silva, 1992). For *Voicing*, the relevant phonological domain of application is the

²¹ We refer the reader to Jun (1993, 1996) for a more complete picture of how to identify different domains in Korean.

²² The actual generalization is perhaps more accurately described as inter-sonorant instead of inter-vocalic; however, we use inter-vocalic throughout the article as a simplification. This simplification does not affect the relevant facts.

*Phonological Phrase*²³ (Y. Y. Cho, 1990). For example, the Korean morpheme /kε/ 'dog' is pronounced with a word-initial [k] when pronounced in isolation (Table A.2). The morpheme-initial /k/ phoneme is pronounced as its voiced counterpart [g] when it appears in an inter-vocalic position within a word or within a single *Phonological Phrase* (Table A.2); however, the voicing process is blocked across *Phonological Phrases* and across *Intonational Phrases*. Therefore, the relevant phonological domain for *Voicing* in Korean is the *Phonological Phrase*.

Table A.2

The phonological domains where Voicing is active in Korean

Phonological Process	$/C_{\text{lax stop}}/ \rightarrow [\text{voiced}] / V - V$ [Domain: within <i>Phonological Phrases</i>]		
Example			
morpheme produced in isolation	[k]ε 'dog'		
Domains: process is active	across <i>Morphemes</i> s'əlmə-[g]ε 'sled dog' i [g]ε balmok ... across <i>Words</i> this dog ankle 'this dog's ankle...'		
Domains: process is blocked	across <i>Phonological Phrase</i> kojanji-ga [k]ε-lul sirəh-ɛjo cat-NOM dog-ACC hate-PRESENT 'The cat hates the dog' across <i>Intonational Phrase</i> i gojanji, [k]ε-lul sirəh-ɛjo this cat, dog-ACC hate-PRESENT 'This cat, hates the dog'		

²³ Jun (1993) shows evidence that the relevant domain for *Voicing* is in fact another domain that is intermediate between *Intonational Phrase* and the *Phonological Phrase* called the *Accentual Phrase*. However, for reasons of expository simplicity we present the domain as the *Phonological Phrase*, a domain more commonly acknowledged by phonologists. Again, this simplification is tangential to the focus of the article and therefore does not affect the argument made herein.

For Nasalization, the relevant phonological domain is the *Intonational Phrase* (Jun, 1993, 1996). For example, the Korean morpheme /pak/ ‘gourd’ is pronounced with a morpheme-final [k] when pronounced in isolation (Table A.3). The morpheme-final [k] is pronounced as its nasal counterpart [ŋ], a nasal consonant with the same place of articulation, when it appears adjacent to a nasal consonant within a word, or within a single *Phonological Phrase*, or across *Phonological Phrases*; however, the nasalization process is blocked across *Intonational Phrases*. Therefore, the relevant phonological domain for *Nasalization* in Korean is the *Intonational Phrase*.

Table A.3

The phonological domain where Nasalization is active in Korean

Phonological Process	/C _{stop} / → [nasal] / _ [nasal] [Domain: within <i>Intonational Phrases</i>]																
Example																	
morpheme produced in isolation	pa[k] ‘gourd’																
across Morphemes	pa[ŋ]-namul ‘gourd vegetable’																
Domains: process is active	<table border="1"> <tr> <td>across Words</td> <td>kui</td> <td>pa[ŋ]</td> <td>nemse-ga</td> </tr> <tr> <td></td> <td>That</td> <td>gourd</td> <td>smell-NOM</td> </tr> </table>			across Words	kui	pa[ŋ]	nemse-ga		That	gourd	smell-NOM						
across Words	kui	pa[ŋ]	nemse-ga														
	That	gourd	smell-NOM														
across Phonological Phrase	<table border="1"> <tr> <td></td> <td>kui</td> <td>pa[ŋ]</td> <td>məg-əs'əjo</td> </tr> <tr> <td></td> <td>That</td> <td>gourd</td> <td>eat-PAST</td> </tr> <tr> <td></td> <td colspan="3">‘The gourd was eaten’</td></tr> </table>				kui	pa[ŋ]	məg-əs'əjo		That	gourd	eat-PAST		‘The gourd was eaten’				
	kui	pa[ŋ]	məg-əs'əjo														
	That	gourd	eat-PAST														
	‘The gourd was eaten’																
Domains: process is blocked	<table border="1"> <tr> <td>across Intonational Phrase</td> <td>kui</td> <td>pa[k]</td> <td>molle</td> <td>məg-əs'əjo</td> </tr> <tr> <td></td> <td>That</td> <td>gourd</td> <td>secretly</td> <td>eat-PAST</td> </tr> <tr> <td></td> <td colspan="3">‘The gourd was eaten secretly’</td></tr> </table>			across Intonational Phrase	kui	pa[k]	molle	məg-əs'əjo		That	gourd	secretly	eat-PAST		‘The gourd was eaten secretly’		
across Intonational Phrase	kui	pa[k]	molle	məg-əs'əjo													
	That	gourd	secretly	eat-PAST													
	‘The gourd was eaten secretly’																

As can be observed in the above examples, the phonological domain for the *Nasalization* process is larger than that for the *Voicing* process. Furthermore, in domains smaller than the

Intonational Phrase, where both processes are applicable, a stop consonant before a nasal consonant is always realized as a nasal consonant. For example, the morphemes /kuk/ ‘soup’ and /man/ ‘only’ are pronounced together as [kuŋman] ‘soup only’. Contrastingly, in domains outside the *Intonational Phrase*, neither *Nasalization* nor *Voicing* is applicable, so stop consonants are always realized as voiceless. In other words, if a stop consonant and a nasal consonant are separated by an *Intonational phrase*, the stop surfaces as a voiceless stop. Therefore, the two processes conspire to ensure that there are no voiced stops adjacent to nasal consonants in the language, i.e., there are no sequences of [...C_[voiced] # M...] no matter what the domain boundary is between the two segments (# = domain boundary; C = consonant; M = Nasal consonant). However, there can be voiceless stops adjacent to nasal consonants as long as the two segments straddle an *Intonational Phrase* boundary, i.e., [...C_[voiceless] # M...] is licit in Korean if the domain boundary is the *Intonational Phrase*.

Appendix B

Table B.1

Means and standard errors of responses in Experiment 1

	Percentages of responses					
	Korean subjects (N = 17)			English subjects (N = 22)		
	[i]	nothing	[w]	[i]	nothing	[w]
ebima	97.06 (2.94)	0 (0)	2.94 (2.94)	96.59 (2.49)	1.14 (1.14)	2.27 (1.57)
ebma	0 (0)	76.47 (8.16)	23.53 (8.16)	1.14 (1.15)	97.73 (1.57)	1.14 (1.14)
ebuima	0 (0)	0 (0)	100 (0)	7.95 (4.47)	2.27 (1.57)	89.77 (5.11)
epma	0 (0)	94.12 (4.56)	5.88 (4.56)	2.27 (2.27)	97.73 (2.27)	0 (0)
edima	100 (0)	0 (0)	0 (0)	95.45 (2.67)	4.55 (2.67)	0 (0)
edma	0 (0)	57.35 (10.68)	42.65 (10.67)	0 (0)	100 (0)	0 (0)
eduuma	0 (0)	0 (0)	100 (0)	10.23 (3.91)	9.09 (4.52)	80.68 (5.68)
etma	1.47 (1.47)	94.12 (3.41)	4.41 (3.21)	1.14 (1.14)	98.86 (1.14)	0 (0)
egima	100 (0)	0 (0)	0 (0)	97.73 (1.57)	1.14 (1.14)	1.14 (1.14)
egma	11.76 (5.30)	38.24 (10.08)	50 (10.50)	1.14 (1.14)	98.86 (1.14)	0 (0)
eguuma	0 (0)	2.94 (2.01)	97.06 (2.01)	18.18 (5.74)	11.36 (5.87)	70.45 (7.65)
ekma	0 (0)	89.71 (6.45)	10.29 (6.45)	0 (0)	100 (0)	0 (0)
emma	0 (0)	100 (0)	0 (0)	0 (0)	100 (0)	0 (0)
enjma	0 (0)	100 (0)	0 (0)	0 (0)	98.86 (1.14)	1.14 (1.14)
enma	0 (0)	100 (0)	0 (0)	0 (0)	100 (0)	0 (0)

Table B.2

Means and standard errors of responses in Experiment 2

	Percentages of responses			
	Korean subjects (<i>N</i> = 16)		English subjects (<i>N</i> = 18)	
	nothing	[w]	nothing	[w]
ebma_00	67 (9.735)	33 (9.735)	100 (0)	0 (0)
ebma_25	69.69 (9.02)	30.31 (9.02)	98.44 (1.067)	1.56 (1.067)
ebma_50	67 (10.076)	33 (10.076)	99.22 (0.778)	0.78 (0.778)
ebma_75	64.31 (10.438)	35.69 (10.438)	99.22 (0.778)	0.78 (0.778)
ebma_100	65.19 (10.029)	34.81 (10.029)	99.22 (0.778)	0.78 (0.778)
ebuuma	1.13 (0.464)	98.88 (0.464)	2.11 (0.718)	97.89 (0.718)
egma_00	58.06 (9.892)	41.94 (9.892)	98.44 (1.067)	1.56 (1.067)
egma_25	55.44 (11.355)	44.56 (11.355)	99.22 (0.778)	0.78 (0.778)
egma_50	54.56 (10.699)	45.44 (10.699)	99.22 (0.778)	0.78 (0.778)
egma_75	46.44 (10.647)	53.56 (10.647)	99.22 (0.778)	0.78 (0.778)
egma_100	46.56 (10.883)	53.44 (10.883)	98.44 (1.067)	1.56 (1.067)
eguuma	0.75 (0.335)	99.25 (0.335)	5.78 (3.78)	94.22 (3.78)
emma	99.63 (0.256)	0.37 (0.256)	99.5 (0.364)	0.5 (0.364)
enma	100 (0)	0 (0)	99.39 (0.611)	0.61 (0.611)

Table B.3

Means and standard errors of responses in Experiment 3

	[ŋ]	none of them	nothing	[w]
egma_00	0 (0)	0 (0)	90.85 (4.58)	9.15 (4.58)
egma_25	2.55 (1.393)	2.5 (1.816)	85.8 (5.573)	9.2 (5.197)
egma_50	1.7 (1.17)	0.85 (0.85)	78.25 (6.642)	19.2 (6.661)
egma_75	4.2 (2.054)	0.85 (0.85)	75 (7.096)	20 (7.003)
egma_100	4.2 (2.054)	1.7 (1.17)	70 (6.567)	24.2 (6.664)
egma_r_00	0 (0)	0 (0)	91.6 (3.743)	8.4 (3.743)
egma_r_25	2.5 (1.816)	0.85 (0.85)	83.25 (5.669)	13.45 (4.948)
egma_r_50	1.7 (1.17)	0.85 (0.85)	89.15 (3.467)	8.3 (3.059)
egma_r_75	6.7 (2.532)	0 (0)	75.9 (5.717)	17.45 (5.315)
egma_r_100	1.7 (1.17)	2.55 (1.393)	65.1 (7.527)	30.85 (7.551)
eguuma	0 (0)	0.15 (0.15)	0.15 (0.15)	99.7 (0.206)

Appendix C

For Experiment 2, when looking at both languages simultaneously, the fully saturated fixed effects model did not converge even for the simplest random effects considered in this article. The best model, amongst the rest, had an interaction term for *POA* and *Language*, and a main effect term for *Voicing* (Table 4). In what follows, we also present the results of the next largest fixed effects model that included an interaction term for *Voicing* and *Language* (Table C.1). It is important to note that this is not the best model amongst the remaining models; however, an anonymous reviewer suggested that presenting the model with an interaction term for *Voicing* and *Language* might be useful to the reader. As with the model presented in the main article (Table 4), the random effects structure was a varying intercept for subjects and items. This model suggests that there was a statistically significant increase in [w] identification rates for Korean listeners, and marginally statistically significant increase in [w] identification rates for velar sounds. Finally, there was also a marginal interaction of *POA* and *Language*.

Table C.1

Logistic mixed effects model for both language groups in Experiment 2 with an additional interaction term (*Voicing* * *Language*) over the model presented in Table 4 (baseline = bilabial consonant context [b] with minimal voicing)

Fixed Effects	Estimate	Std. Error	Z value	P-value
Intercept	-7.30	1.17	-6.23	<0.0001
POA (Velar)	0.75	0.44	1.69	0.09
Language (Korean)	5.28	1.34	3.95	<0.0001
Voicing	-0.15	0.24	-0.64	0.522
POA (Velar): Language (Korean)	0.25	0.14	1.80	0.07
Voicing: Language (Korean)	0.26	0.23	1.15	0.25

References

Ahn, S.-C. (1985). *The Interplay of Phonology and Morphology in Korean* (Ph.D. Dissertation). University of Illinois, Urbana-Champaign.

Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology*, 26, 9–41.

Avery, P., & Idsardi, W. J. (2001). Laryngeal Dimensions, Completion and Enhancement. In T. A. Hall (Ed.), *Distinctive Feature Theory* (pp. 41–70). de Gruyter, Berlin.

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

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 (Version R package version 1.1-7). Retrieved from <http://CRAN.R-Project.org/package=lme4>.

Berent, I., Lennertz, T., Jun, J., Moreno, M., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, 105, 5321–5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology*, 26, 75–108.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104, 591–630.

Best, C. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Best, C. T., Halle, P., Bohn, O.-S., & Faber, A. (2003). Cross-language perception of nonnative vowels: Phonological and phonetic effects of listeners' native languages. In *15th International Congress of Phonetic Sciences* (pp. 2889–2892). Barcelona.

Boersma, P., & Hamann, S. (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese & W. L. Wetzel (Eds.), *Loanword Phonology* (pp. 11–58). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer [Computer program]. Version 5.3.20, retrieved 28 June 2012 from <http://www.praat.org/>.

Bolker, B. M. (2014). Random- and fixed-effects structure in linear-mixed models. *Cross Validated (StackExchange)*. Retrieved from <http://stats.stackexchange.com/questions/130714/random-and-fixed-effects-structure-in-linear-mixed-models>

Byrd, D. (1992). A note on English sentence-final stops. *UCLA Working Papers in Phonetics*, 81, 37–38.

Cho, T., Keating, P., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In *Laboratory Phonology 6* (pp. 145–163). Cambridge University Press.

Cho, T., & McQueen, J. M. (2006). Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants. *Journal of Acoustical Society of America*, 119(5 Pt 1), 3085–3096.

Cho, Y. Y. (1990). Syntax and Phrasing in Korean. In S. Inkelas & D. Zec (Eds.), *The Phonology-Syntax Connection* (pp. 47–62). Chicago: University of Chicago Press.

Chung, H., Kim, K., & Huckvale, M. (1999). Consonantal and Prosodic Influences on Korean Vowel Duration. In *Proceedings of EuroSpeech99*. Budapest, Hungary.

Coenen, E., Zwitserlood, P., & Bölte, J. (2001). Consequences of assimilation for word recognition and lexical representation. *Language & Cognitive Processes*, 16, 535–564.

Coetzee, A. W., & Pater, J. (2011). The Place of Variation in Phonological Theory. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The Handbook of Phonological Theory* (pp. 401–434). Wiley-

Blackwell. Retrieved from
<http://onlinelibrary.wiley.com/doi/10.1002/9781444343069.ch13/summary>

Crystal, T. H., & House, A. S. (1988). The duration of American-English stop consonants: an overview. *Journal of Phonetics*, 16, 285–294.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. Vijver (Eds.), *Variation and Gradience in Phonetics and Phonology* (pp. 265–309). Berlin, New York: Mouton de Gruyter.

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

Davidson, L. (2007). The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology*, 24(2), 261–286.

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics*, 40(2), 234–248.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study. *Journal of Cognitive Neuroscience*, 12(4), 635–647. <http://doi.org/10.1162/089892900562390>

De Jong, K., & Park, H. (2012). Vowel epenthesis and segment identify in Korean learners of English. *Studies in Second Language Acquisition*, 34(1), 127–155.

Dilley, L., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59, 294–311.

Dilley, L., Shattuck-Hufnagel, S., & Osterdorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423–444.

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(6), 1568–1578.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, 64(3), 199–210.

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

Flege, J. E., & Hillenbrand, J. (1987). A differential effect of release bursts on the stop voicing judgments of native French and English listeners. *Journal of Phonetics*, 15, 203–208.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of Acoustical Society of America*, 101(6), 3728–3740.

Gaskell, G., & Marslen-Wilson, W. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144–158.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65(4), 575–590.

Grijzenhout, J., & Kabak, B. (Eds.). (2009). *Phonological Domains: Universals and Deviations*. Mouton de Gruyter.

Gumnior, H., Zwitserlood, P., & Bölte, J. (2005). Assimilation in existing and novel German compounds. *Language and Cognitive Processes*, 20(3), 465–488.
<http://doi.org/10.1080/01690960444002174>

Hallé, P., 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.

Han, M. S. (1964). Duration of Korean Vowels. In *Studies in the Phonology of Asian Languages 2*. University of Southern California, Los Angeles: Acoustics Phonetics Research Laboratory.

Heffner, C. C., Dilley, L., McAuley, J. D., & Pitt, M. (2013). When cues combine: How distal and proximal cues are integrated in word segmentation. *Language & Cognitive Processes*, 28(9), 1275–1302.

Henderson, J. B., & Repp, B. H. (1982). Is a stop consonant released when followed by another stop consonant? *Phonetica*, 39, 71–82.

Hwang, J. (2011). *Non-native Perception and Production of Foreign Sequences* (PhD Dissertation). State University of New York at Stony Brook, Stony Brook.

Iverson, G. K. (1993). (Post) Lexical Rule Application. In *Studies in Lexical Phonology* (pp. 255–275). Seoul: Academic Press.

Iverson, G. K., & Ahn, S.-C. (2007). English voicing in dimensional theory. *Language Sciences*, 29(2-3), 247–269.

Iverson, G. K., & Salmons, J. C. (2003b). Laryngeal enhancement in Germanic. *Phonology*, 20, 43–74.

Jaeger, T. F. (2009). Random effect: Should I stay or should I go? Retrieved from <https://hlplab.wordpress.com/2009/05/14/random-effect-structure/>

Jaeger, T. F. (2011). More on random slopes and what it means if your effect is not longer significant after the inclusion of random slopes. Retrieved from <https://hlplab.wordpress.com/2011/06/25/more-on-random-slopes/>

Jessen, M., & Ringen, C. (2002). Laryngeal features in German. *Phonology*, 19(2), 189–218.

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody* (Ph.D. Dissertation). The Ohio State University, Columbus, Ohio.

Jun, S.-A. (1996). *The Phonetics and Phonology of Korean Prosody: intonational phonology and prosodic structure*. New York, NY: Garland Publishing Inc.

Jun, S.-A., & Fougeron, C. (2002). The realizations of the accentual phrase in French intonation. *Probus*, 14, 147–172.

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.

Kang, Y. (1992). Prosodic Structure of Korean. In *Proceedings of the 1992 Seoul International Conference on Linguistics* (pp. 199–208). Seoul, Korea: Hanguk Munhwasa, Seoul.

Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology*, 20, 219–273.

Keating, P. (2003). Phonetic Encoding of Prosodic Structure. In *6th International Seminar on Speech Production*. Syndey.

Kim, H., & Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics*, 24, 295–312.

Kim, K.-O. (1974). *Temporal Structure of Spoken Korean: an Acoustic Phonetic Study* (Ph.D. Dissertation). University of Southern California.

Kim-Renaud, Y.-K. (1987). Fast speech, casual speech and restructuring. *Harvard Studies in Korean Linguistics*, 2, 341–359.

Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *J Acoust Soc Am*, 134(1), EL19–EL25. <http://doi.org/10.1121/1.4807431>

Kiparsky, P. (1985). Some Consequences of Lexical Phonology. *Phonology*, 2, 85–138.

Kohler, K. J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica*, 41, 150–174.

Lisker, L. (1999). Perceiving Final Voiceless Stops without Release: Effects of Preceding Monophthongs versus Nonmonophthongs. *Phonetica*, 56, 44–55.

Marr, D. (1982). *Vision: A Computational Approach*. San Francisco: Freeman & Co.

Mitterer, H., Kim, S., & Cho, T. (2013). Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language*, 69, 59–83.

Monahan, P. J., Takahashi, E., Nakao, C., & Idsardi, W. J. (2009). Not All Epenthetic Contexts are Equal: Differential Effects in Japanese Illusory Vowel Perception. In I. Shoichi, H. Hoji, P. M. Clancy, & S.-O. Sohn (Eds.), (pp. 391–405). Presented at the Japanese/Korean Linguistics, 17, Stanford, CA: CSLI Publications.

Moreton, E. (2002). Structural constraints in the perception of English stop–sonorant clusters. *Cognition*, 84, 55–71.

Nagao, K., & de Jong, K. (2007). Perceptual rate normalization in naturally produced rate-varied speech. *J Acoust Soc Am*, 121(5 Pt1), 2882–98.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Peperkamp, S. (2005). A psycholinguistic theory of loanword adaptations (pp. 341–352). Presented at the Berkeley Linguistics Society 30.

Peperkamp, S., & Dupoux, E. (2003). Reinterpreting loanword adaptations: the role of perception. In *Proceedings of the 15th International Congress of Phonetic Sciences*. 367-370.

Peperkamp, S., & Dupoux, E. (2007). Learning the mapping from surface to underlying representations in artificial language learning. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 315–338). Berlin: Mouton de Gruyter.

R Development Core Team. (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>

Rositzke, H. A. (1943). The articulation of final stops in General American speech. *American Speech*, 18, 39–42.

Scholes, R. (1966). *Phonotactic grammaticality*. The Hague: Mouton.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, 1, 289–316.

Selkirk, E. (1972). *The phrase phonology of English and French* (Ph.D. Dissertation). MIT.

Selkirk, E. (1986). On derived domains in sentence phonology. In *Phonology Yearbook* (Vol. 3, pp. 371–405). Cambridge, MA: MIT Press.

Silva, D. J. (1992). *The Phonetics and Phonology of Stop Lenition in Korean* (Ph.D. Dissertation). Cornell University, Ithaca, New York.

Snow, G. (2012). How to do logistic regression in R when outcome is fractional? *Cross Validated (StackExchange)*. Retrieved from <http://stats.stackexchange.com/questions/26762/how-to-do-logistic-regression-in-r-when-outcome-is-fractional>

Sohn, H.-M. (1999). *The Korean Language*. Cambridge, UK: The Cambridge University Press.

Spinelli, E., Grimault, N., Meunier, F., & Welby, P. (2010). An intonational cue to word segmentation in phonemically identical sequences. *Attention, Perception, & Psychophysics*, 72(3), 775–787. <http://doi.org/10.3758/APP.72.3.775>

Vogel, I. (2009). The Status of the Clitic Group. In J. Grijzenhout & B. Kabak (Eds.), *Phonological Domains: Universals and Deviations* (pp. 15–46). Mouton de Gruyter.

Wilson, C., & Davidson, L. (2013). Bayesian Analysis of Non-native Cluster Production. In *Proceedings of NELS 40* (Vol. 2, pp. 265–278). MIT, Cambridge, MA.