Phonological Alternations Modulate Illusory Vowels in Perceptual Epenthesis

Acknowledgements

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 NELS 43 and J/K Conference 22 for probing questions, and helpful discussion.

Abstract

Native speakers perceive illusory vowels when presented with consonant sequences that violate phonotactic constraints in their language. Previous research suggests that the phenomenon motivates speech perception models that include surface phonotactic information and are sensitive to the acoustics of the speech tokens. In this article, inspired by Bayesian models of speech perception, we claim that the task of the listener in speech perception is to identify the target underlying representations. This predicts that the phenomenon of perceptual illusions will be modulated, not only by surface phonotactics and the acoustics of the speech tokens, but also by the *phonological alternations* of a language. We present the results of three experiments (an AX task, an ABX task, and an identification task) on native Korean listeners, with native English listeners as controls, showing that they perceive different sets of illusory vowels in different phonological contexts, in accordance with the phonological processes of *Vowel Deletion* and *Palatalization* in the language.

Keywords: speech perception, perceptual epenthesis, illusory vowels, phonotactic constraints, phonological alternations, Korean phonology.

1. Introduction

The phenomenon of illusory vowels has received a lot of attention in the recent literature (Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Berent, Steriade, Lennertz, & Vaknin, 2007; Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011; Kabak & Idsardi, 2007; Monahan, Takahashi, Nakao, & Idsardi, 2009; inter alia). The general finding with these studies is that listeners sometimes perceive illusory vowels in stimuli that contain consonant sequences that are phonotactically illicit in their native languages. When a native speaker is presented with a nonsense word containing a consonant sequence that violates the phonotactic constraints in their language, an illusory vowel is perceptually induced in between such a sequence thereby creating an illusory sequence that respects the phonotactic constraints of the language. For example, when a Japanese listener is auditorily presented with [ebzo], they may actually perceive [ebɯzo] given that [bz] is an illicit consonant sequence in Japanese, as shown originally by Dupoux et al. (1999).

As discussed by Dupoux et al. (2011), the contextual and phonetic effects observed with illusory vowels are difficult to account for through most current psycholinguistic models of speech recognition, where the primary units are segments and phonological/phonetic features (Best, 1994; Kuhl, 1993; Lahiri & Reetz, 2002, 2010; McClelland & Elman, 1986; Norris & McQueen, 2008). They suggest that this can be remedied by having phonotactic constraints that refer to surface sequences of segments interact with categorization in a single processing step. We argue in this article that the phenomenon of illusory vowels shows us that, along with surface phonotactic constraints and phonetic representations, there is also a need to take into account the phonological alternations present in a language in understanding speech perception. Inspired by

Bayesian models of speech perception[1] (Bever & Poeppel, 2010; Feldman & Griffiths, 2007; Poeppel & Monahan, 2011; Sonderegger & Yu, 2010; Wilson & Davidson, in press; Yu, 2011), we claim that the task of the listener in speech perception is primarily a task of reverse inference: it is to identify the best estimate of the intended underlying categories of the utterance for the incoming acoustic token. In this case, the underlying category information we make reference to is the phonemic/underlying representations. The knowledge about what underlying categories map to what surface categories must include information about both phonological alternations and phonotactic constraints. Therefore, both phonological alternations and phonotactic constraints are expected to play a role in speech perception, along with the phonetic characteristics of the language. As we show below, the actual *quality* of the illusory vowels in different contexts is modulated by the phonological processes of the language.

More generally, there is related work that has argued for the need for the speech perception mechanism to be sensitive to phonological alternations (Boomershine, Hall, Hume, & Johnson, 2008; Huang, 2001; Hume & Johnson, 2003; Johnson & Babel, 2010). For example, Huang (2001) showed that the tone-sandhi alternation involving the contextual neutralization of two otherwise contrastive tones in Mandarin Chinese (the low-falling-rising tone (214) and the mid-rising tone (35)) causes the two tones to be perceptual closer, and therefore, more confusable for Mandarin Chinese listeners. In the current article, we extend this previous line of work that argues for the importance of phonological alternations in speech perception by showing that the concept is crucial in understanding the phenomenon of illusory vowels

[1] It is important to note that we are *not* presenting a Bayesian model. However, the aspect of Bayesian models that is particularly relevant to the current article is that of reverse inference to hypotheses that account for the data, which in our case is reverse inference to the underlying/phonemic representation level. Therefore, what we show in this article is actually consistent with any view of speech perception that makes crucial reference to that concept.

(especially the facts presented herein). Furthermore, we also present a particular view of speech perception that can naturally account for such phonological sensitivity in speech perception.

As has been pointed out previously, a proper understanding of the phenomenon of illusory vowels, and speech perception more generally, has a direct bearing on the theoretical literature of loanword adaptations, where there is a debate on the factors affecting loanword adaptations (Davidson, 2007; Peperkamp, 2005; amongst others). Whereas some claim that perceptual factors are perhaps the primary factor influencing loanword adaptation patterns (Peperkamp, 2005; Peperkamp & Dupoux, 2003), others claim that perception is at best a minor factor in such patterns (Jacobs & Gussenhoven, 2000; LaCharite & Paradis, 2005; Paradis & LaCharite, 1997; Uffman, 2006). The proposed account in the current article suggests, contrary to these claims, the perceptual mechanism uses the phonological system for inference in quite some detail, and therefore, it is perhaps impossible to separate the effects of speech perception and those of the phonological system, on loanword patterns.

With respect to the locus of perceptual epenthesis, while earlier work in the domain of illusory vowels had assumed that the relevant constraints driving the perceptual illusions were sequential phonotactic constraints (Dehaene-Lambertz et al., 2000; Dupoux et al., 1999), Kabak & Idsardi (2007) argue that the relevant phonotactic constraints that drive such perceptual illusions are the syllable structure constraints of the language[2]. They ran an AX discrimination task on Korean speakers (with English speakers as controls) with two types of illicit consonant sequences. In one, the first consonant C_1 was an illicit coda consonant, and the corresponding consonant sequence C_1C_2 was also illicit in Korean. In the other, the consonant C_1 was a licit coda consonant, but the corresponding consonant sequence C_1C_2 was illicit in Korean. They

[2] While Kabak & Idsardi (2007) argue that listeners are trying to infer the most probable sequence of syllables, they are somewhat agnostic about whether the representations are underlying vs. surface representations.

showed that the perception of illusory vowels was consistently driven by the first type of consonant sequences but not by the second. Therefore, they argued that the illusory vowel phenomenon was better accounted for by the syllable structure constraints than the surface consonant sequence constraints in the language.

The perception of illusory vowels has also been argued to be modulated by the listener's knowledge of language universals related to the Sonority Sequencing Principle and syllable structure (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008; Berent et al., 2009, 2007). In a series of experiments on Korean and English speakers, Berent and colleagues show that universally dispreferred initial consonant sequences trigger a stronger perception of illusory vowels than universally preferred initial consonant sequences even when the subject's native language does not allow them experience of either sequence. For example, both [lb] and [bl] are illicit initial consonant sequences in Korean; however, the former is a universally dispreferred sequence relative to the latter across the world's languages (Hooper, 1976; Jespersen, 1904; Selkirk, 1984; Sievers, 1881; Steriade, 1982). Berent and colleagues show that Korean speakers more readily misperceive the former than the latter.

Related work on perceptual distortions has shown that such distortions are also driven by more abstract consonantal sequential constraints. Moreton (2002) shows that subjects make use of abstract featural co-occurrence constraints. He shows that English speakers misperceive words beginning with [dl] sequences much more than [bw] though both are nearly zero probability sequences in English. He argues that the asymmetry results from a specific featural co-occurrence constraint in English, a ban on two adjacent coronal consonants, which does not apply to a sequence of two adjacent labial consonants[3].

---

[3] This suggests that phonotactics is not a simple matter of keeping track of attested frequencies; it is equally important to recognise the type of representations over which the frequencies are tracked. A similar inference results

It has also been shown that illusory vowels are only one of the many possible perceptual repairs for phonotactically illegal consonant sequences (Davidson, 2007; Davidson & Shaw, 2012; Hallé, Segui, Frauenfelder, & Meunier, 1998). Davidson & Shaw (2012) show that when English subjects are auditorily presented with phonotactically illicit initial consonant sequences, they 'repair' the sequences in a variety of ways that include consonant deletion, metathesis, prothesis, consonant change, and perception of illusory vowels[4]. They further showed that the likelihood of a particular repair was modulated by the type of illicit consonant sequence presented to the subject.

As can be seen from the above review, the bulk of the research assumes that perceptual epenthesis of illusory vowels is driven purely by surface phonotactics and phonetic characteristics of acoustic tokens. This is however not to say that there is no evidence of abstract knowledge being used[5]. As discussed above, Berent et al (2007; et seq) and Moreton (2002) have indeed shown that listeners access relatively abstract knowledge. However, the knowledge that listeners seem to be using can be employed on surface representations in a phonological sense (not acoustic/auditory representations), since the Sonority Sequencing Principle that Berent and colleagues discuss and the constraint on alveolar co-occurrence that Moreton (2002) discusses can both be thought of as surface phonotactic constraints, as is standard in the tradition of Optimality Theory. Therefore, there is no need, based on those results, for an even more abstract phonological representation level, namely the underlying representation level.

With respect to the quality of the illusory vowel, Dupoux et al. (2011) argue that it is 'the phonetically minimal element of the language,' and therefore 'the shortest vowel' in the

---

from the behavior of Korean listeners (in Kabak & Idsardi, 2007) since the Korean listeners were at ceiling with some non-attested clusters.

[4] Similar repairs have been observed in loan-word adaptations.

[5] Thanks to an anonymous reviewer for highlighting this fact to us.

language ([ɯ] in Japanese, and [i] in Brazilian Portuguese). Their claim predicts that there can be at most one illusory vowel in a language[6]. We show that this claim is at best only partially consistent with what listeners actually do when encountering illicit sequences. We show that the quality of the illusory vowel is also modulated by the knowledge of the phonological alternations in the language. And in some contexts, it is even possible to induce more than one illusory vowel as long as the phonology of the language supports it.

Acoustic studies of Korean vowels have shown that the vowel [ɨ][7] is the shortest vowel in the language (Chung, Kim, & Huckvale, 1999; Han, 1964; Kim, 1974). The typical duration of the vowel [ɨ] 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 [ɨ] to be the illusory vowel in all contexts.

We propose in what follows immediately that, while it is surely true that surface phonotactics and the phonetic characteristics of acoustic tokens have an effect on perceptual epenthesis, the quality of the illusory vowel also depends on the phonological alternations in the language. As briefly discussed above, we take inspiration from Bayesian models of speech perception (Bever & Poeppel, 2010; Feldman & Griffiths, 2007; Poeppel & Monahan, 2011; Sonderegger & Yu, 2010; Wilson & Davidson, in press; Yu, 2011) in claiming that the task of the listener in speech perception is primarily a task of reverse inference - it is to identify the best

---

[6] In tokens where the illicit consonantal sequence was created by splicing out the medial vowel (for e.g., [abda] from [abida]), Dupoux et al (2011) showed that Japanese speakers primarily perceived an /i/. However, they suggest that remnant co-articulatory traces in the spliced stimuli led to this particular result. This should be kept separate from cases where there is no coarticulated information due to a spliced-out vowel to aid the listener. This was the case in their stimuli that were produced naturally with the consonant sequence violation (for e.g., [abda]). And in such items, consistent with the claim of participants perceiving 'the shortest vowel', the Japanese speakers primarily perceived an [u].

[7] There is some debate in the phonological literature on the use of the *unrounded high back vowel* [ɯ] for the Korean letter 으. 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 [ɨ] throughout.

estimate of the intended underlying categories (phonemic/underlying representations) of the utterance given the acoustic token[8]. Knowledge about both phonological alternations and phonotactic constraints is required to reverse infer the phonemic/underlying representations from the acoustic tokens. Therefore, both phonological alternations and phonotactic constraints are expected to play a role in speech perception, along with the phonetic characteristics of the language.

More specifically, in regard to the quality of the illusory vowel, we see the perceiver's task as attempting to repair the illicit phonotactic sequence with a vowel phoneme that best maps to the phonetic characteristics of the acoustic token. When no relevant phonological alternations bias listeners towards a certain vowel in the particular segmental context, the best vowel guess that repairs the particular phonotactic violation is indeed the phonetically minimal/shortest vowel in the inventory, a la Dupoux et al. (2011). This is because the phonetically shortest vowel is, in terms of duration, the closest amongst all the vowels in the inventory to the absence of a vowel. The illicit consonant sequences tested by Dupoux et al.'s (2011) were of the form $V_1C_1C_2V_2$. In Japanese and Brazilian Portuguese, the particular consonantal sequences, such as [..bd..], [..bg..], [..gn..], do not appear to be influenced by any phonological alternations that are relevant to the process of perceptual epenthesis (i.e., processes that bias listeners towards a certain vowel), so the best vowel guesses to make for the perceiver are the phonetically minimal vowels in the respective languages. However, when relevant phonological alternations do bias listeners towards particular vowel percepts in specific segmental context, the best guess depends on both the phonetics of the acoustic token and also the phonological alternations themselves. The types of phonological processes that are likely to play a role are those that bias the listener's

[8] A full Bayesian analysis will require corpus statistics in order to make precise quantitative predictions about the quality of the illusory vowel, and is well-beyond the scope of the current paper.

expectations about the quality of the illusory vowel. One such process is a consistent/regular vowel deletion process that targets a particular vowel. The presence of a regular process of vowel deletion that targets a particular vowel ($/V_1/ \rightarrow [Ø]$) in the phonology of the language supports the reverse inference of the same vowel in the phonemic representation when the surface representation has nothing (reverse inference: $[Ø] \rightarrow /V_1/$)[9]. For these reasons, in a phonotactically illicit consonantal context, where the condition can be perceptually repaired by a vowel, the best vowel to repair the situation is the phoneme $/V_1/$ that maps to [Ø] in the surface/acoustic representations. A second type of process that is likely to bias a listener's expectations about the vowel quality of the illusory vowel is one that involves allophonic mappings before a specific vowel ($/C_1/ \rightarrow [C_2]$ / _ V_2). In a phonotactically illicit consonantal context, where the condition can be perceptually repaired by a vowel, when the phonotactically illicit consonant is the allophone $[C_2]$, the phonemic consonant inferred is the corresponding phoneme $/C_1/$. In such situations, the best vowel to perceptually repair the context is the vowel $/V_2/$ next to which the phoneme $/C_1/$ surfaces as $[C_2]$, as this would also account for the acoustic properties of the illicit consonant.

In what follows, we briefly describe some regular phonological processes in Korean that are relevant for the phonological contexts tested in this paper. These processes exhibit exactly the above-mentioned characteristics needed to bias the perception of the illusory vowels. Korean has a phonological process of vowel deletion that targets the high central unrounded vowel /ɨ/ in certain environments during morphological concatenation (Ahn, 1985; Sohn, 1999). When the vowel /ɨ/ is in a vowel hiatus situation with another vowel due to morpheme concatenation, the

---

[9] Note, the presence of a structural change of vowel deletion that specifically targets a particular vowel, even if constrained to specific phonological environments, will increase the global probability of reverse inference to that particular vowel when there is no vowel correspondent in the acoustic token. Therefore, the presence of such a process will also increase the probability of reverse inference to that particular vowel in phonological environments that are different from the ones where the process typically occurs.

/ɨ/ always deletes (Table 1). Furthermore, [ɨ] is often deleted in Korean, especially in weak non-initial open syllables (Kang, 2003; Kim-Renaud, 1987). Therefore, following the logic of reverse inference discussed in detail above, [ɨ] is a good vowel to infer for a Korean listener in acoustic input where a vowel is not present but is expected based on the phonological patterns of the language. Finally, as mentioned earlier, /ɨ/ also has the shortest phonetic duration of all the vowels in the language. These facts allow /ɨ/ to be a good vowel for perceptual repairs in most contexts because it already varies with Ø (nothing) in the phonetic representations. We call this ***illusory vowel 1***.

Table 1

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

Process	Underlying/Phonemic Representations	Phonetic/Surface Representations	
Vowel Deletion			
/ɨ/ → Ø / __ + V	/k^hɨ + əto/	[k^hədo][10]	‘although (it is) big’
(or)			
/ɨ/ → Ø / V + __	/k^ha + ɨni/	[k^hani]	‘because we go’
Palatalization			
/$C_{alveolar}$/ → [palatal] / _ i	/pat^h + i/	[pac^{h}i]	‘dry field (Nom)’
	/os + i/	[oʃi]	‘clothes (Nom)’

[10] The phoneme /t/ maps to the allophone [d] intervocalically.

Furthermore, Korean has another phonological process of palatalization of alveolar consonants before /i/; for example, the phonemes /t^h/ and /c^h/ neutralise to the [c^h] and the phoneme /s/ surfaces as [ʃ] before /i/[11] (Table 1) (Ahn, 1985; Iverson, 1993; Sohn, 1999). For a Korean listener, when a palatal stop segment [c^h] is encountered in the acoustic token, there are two possible phonemic parses - it can either be from an alveolar stop phoneme /t^h/, or from a palatal stop phoneme /c^h/ (Table 2). For example, when a Korean listener hears a nonsense word such as [ec^hima], the surface consonant [c^h] is consistent with the reverse inference of either the phoneme /t^h/ or the phoneme /c^h/; thus, the inferred phonemic parses for the nonsense word could either be /ec^hima/ or /et^hima/. As proposed above, inferences about the phonemic/underlying representations of the presented nonsense words modulate the quality of the illusory vowel in illicit phonotactic contexts. More specifically, when a Korean listener encounters a nonsense word with a palatal sound [c^h] as the first consonant of an illicit syllable context (for example, [ec^hma]), the quality of the illusory vowel is modulated by the reverse inference about the phoneme that corresponds to the surface pronunciation [c^h] in the nonsense word; if the perceptual system infers the phoneme to be a palatal stop segment /c^h/, the /ɨ/ vowel (*illusory vowel 1)* is induced for reasons mentioned above; however, if the perceptual system infers the phoneme to be an alveolar stop segment /t^h/, then an /i/ vowel (we call this vowel ***illusory vowel 2***) is induced in the illicit syllable context, because the only way to get a phonetic [c^h] from the phoneme /t^h/ is to have a following phoneme /i/. Given this, we expect that the *same* illicit palatal coda can induce *both* an illusory /i/ and an illusory /ɨ/. Next, when an alveolar segment is encountered in the acoustic token, namely, [t^h] or [s], there is only one possible phonemic parse, the same alveolar phoneme, /t^h/ or /s/, (Table 2). In an illicit syllable context, the vowel /ɨ/

[11] Note, /t^h/ palatalization is blocked in tautomorphemic contexts, i.e., if both the /t^h/ and the /i/ are within the same morpheme, the palatalization rule is blocked. The /s/-palatalization process, however, happens in all contexts (Hong, 1997; Iverson, 1993, 2004).

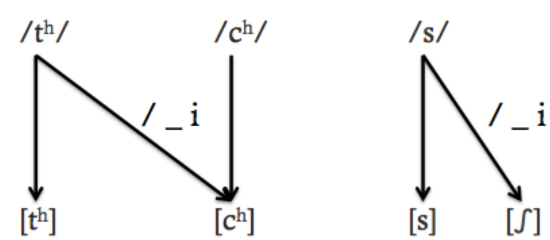
(*illusory vowel 1*) is induced for reasons mentioned above. Finally, when a palatal fricative, [ʃ], is encountered in the acoustic token, there is only one possible phonemic parse, the alveolar fricative /s/ (Table 2). However, if an alveolar fricative (/s/) is the inferred phoneme, then the /i/ vowel (*illusory vowel 2*) is induced in the illicit syllable context, because the only way to get a phonetic [ʃ] from a phonemic /s/ is to have a following phoneme /i/.

Table 2

Mappings and Neutralizations Resulting from Palatalization[12]

Phonemic Representation:

Phonetic Representation:



From the above discussion, it should be clear that unlike Dupoux et al. (2011), we predict different sets of illusory vowels in different illicit phonotactic contexts for Korean listeners. In illicit phonotactic contexts following alveolar contexts [t^h, s], we predict the illusory vowel to be [ɨ]; in those following the palatal stops [c^h], we predict the possibility of *both* [i] and [ɨ]; and in those following the palatal fricative [ʃ], we predict only the vowel [i].

[12] The table provides representative alveolar and palatal stop consonants. The processes described are true of all alveolar and palatal stop consonants. Furthermore, /s/ is the only fricative in Korean, and it has two surface variants [s] and [ʃ].

As a clarificatory note of our position, we would like to note that though we predict the possibility of *both* [i] and [ɨ] as illusory vowels for Korean listeners in the relevant palatal context [c^h], we do not think that both the illusory vowels are *simultaneously* perceived in the same nonceword phonemic percept by a Korean listener. It is possible that for any single presentation of an auditory input, two *separate* (nonceword) phonemic percepts are inferred simultaneously since both are consistent with the acoustic input, where each parse is assigned a certain probability conditioned by other aspects such as lexical frequencies of the relevant phonemes[13]. It is also possible that for any single presentation of an auditory input only a single (nonceword) percept is inferred in a probabilistic way.

In the following sections, we present the results of three experiments of identification and discrimination tasks on Korean subjects, with English subjects as controls to ensure that the differences in the acoustic tokens are not what are driving the perceptual epenthesis effects observed in the Korean subjects. Three different paradigms—AX task (Experiment 1), ABX task (Experiment 2), and identification task (Experiment 3)—were used to ensure that the effects are not artifacts of a certain experimental paradigm.

2. Experiment 1

Experiment 1 investigated perceptual epenthesis effects using an AX task, in which listeners heard two stimuli and decided whether the two stimuli were the same or different. In

[13] In fact, more generally, from a Bayesian perspective, it is possible to imagine that what is being inferred by a listener during speech perception is not a single percept but is a posterior probability distribution over different underlying/phonemic representational candidates. Thinking about it along these lines also allows one to better understand why the illusory vowel rates are never at ceiling in such experiments. Thanks to an anonymous reviewer for raising this possibility.

this paradigm, for instance, if listeners perceive an illusory vowel [ɨ][14] between consonants in a cluster [sm], they will demonstrate poor discrimination between [esma] and [esɨma]. Crucially, as per the claims in the previous section, we expect to see that Korean listeners should have much more difficulty than American English listeners in distinguishing the following two sets of stimulus pairs: (a) [etʰɨma-etʰma], [esɨma-esma], [ecʰɨma-ecʰma], (b) [ecʰima-ecʰma], [eʃima-eʃma]. In set (a), the Korean listeners are likely to perceive an illusory [ɨ] vowel in the second stimulus in each pair [etʰma, esma, ecʰma]; therefore, for the Korean listeners the pairs in (a) should be more confusable than for American English listeners as they are likely to sound more similar to each other. Similarly, in set (b), the Korean listeners are likely to perceive an illusory [i] vowel in the second stimulus of each pair [ecʰma, eʃma]; therefore, for the Korean speakers the pairs in (b) should be more confusable than for American English listeners as they are likely to sound more similar to each other.

2.1 Method

2.1.1 Participants

Twenty native Korean speakers (age 20 – 38, 10 men and 10 women) and 19 native American English speakers (age 19 – 23, 8 men and 11 women) participated in the experiment voluntarily. All the subjects were recruited at Michigan State University, East Lansing, USA and reported to have normal hearing. None of the Korean speakers learned English before the age of eleven, nor had they lived in English speaking countries more than four years except for one participant who started to learn English at the age of eight in Korea and lived in the US for ten years.

[14] The closest equivalent to the Korean [ɨ] in English is the vowel [ʊ]. We follow Kabak and Idsardi (2007) in expecting that the English speakers will confuse [ɨ] with [ʊ], and therefore will not have a problem in distinguishing stimuli containing it from other crucial stimuli.

2.1.2 Stimuli

The experimental stimuli consisted of those that were relevant for the current article and those that were relevant to another independent hypothesis (See Table A. **1** in Appendix). Thirty nonce words in the form $\mathbf{eC_1V_1C_2a}$ were used, in which C_1 was an alveolar/palatal/labial consonant [t^h, d, s, c^h, ʃ, b, m]; V_1 was [i, ɨ, Ø (Null)]; and C_2 was a labial stop/nasal consonant [p^h, m]. None of the stimuli were words in either Korean or in English. They had stress on the first vowel and were natural recordings by the first author, a trained male phonetician, who is a native speaker of Indian English and Telugu, and a near-native speaker of standard Hindi. The use of this particular speaker for recording stimuli was based on two reasons: (a) the speaker can naturally produce all the stimuli as they are phonotactically licit in his dialects of both Hindi and Telugu. On the other hand, the use of a native Korean speaker to record the stimuli would have only been possible if the speaker had neutralized their own linguistic biases, as many of the sequences are not licit in the languages. We strongly suspect that the use of Korean speakers to record stimuli would have introduced biases into the stimuli (in the form of very short excrescent vowels), especially for those sequences that are not licit in the relevant language, thereby making the interpretation of the results much more challenging, (b) the use of an American English speaker to record the stimuli was also avoided because a few that we tried had difficulty producing unstressed medial vowels that were unreduced (so, they couldn't block the vowel reduction process in their dialect). Furthermore, we did not want to introduce a bias that helped the control group, as the phonetic patterns would have been more natural for the American English listeners than the Korean listeners. Therefore, the interpretation of the results would have been confounded by this fact. For these reasons, we used the first author's voice for recording stimuli. Furthermore, the Korean-speaking co-author confirmed that the segmental and suprasegmental quality of the stimuli were controlled and were naturalistic.

Each item was recorded 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). Two tokens were used for each item in the experiment. The stimuli were all normalised in Praat to have a mean intensity of 60 dB, and were then multiplied by a Hanning window, applied to the whole stimulus, to induce a smooth ramping.

Table 3 shows all the clusters and the test items relevant to the current paper. All of the test items without intervening vowels, [etʰma, esma, ecʰma, eʃma], had an illicit coda in Korean, and the clusters were also all illicit linear sequences in Korean, thereby eliminating issues regarding the distinction between syllable structure violation and surface phonotactic violation (Kabak & Idsardi, 2007). As all the clusters violated both types of phonotactic constraints, they were expected to induce perceptual epenthesis.

Table 3

Test Tokens in the Experiment

Cluster type	Cluster	Vowels		
		None	[ɨ]	[i]
Alveolar	tʰm	etʰma	etʰɨma	etʰima
	sm	esma	esɨma	esima
Palatal	cʰm	ecʰma	ecʰɨma	ecʰima
	ʃm	eʃma	eʃɨma	eʃima

2.1.3 Procedure

Following Kabak & Idsardi (2007) and Monahan et al. (2009), an AX discrimination (i.e., same/different) task was used to investigate a perceptual epenthesis effect. We tested all combinations of vowels [i, ɨ, Ø]. Therefore for the cluster [sm], the word-pairs were [esɨma-esima], [esɨma-esma], [esima-esma], [esɨma-esɨma], [esima-esima], and [esma-esma]. Word-pairs with different intervening vowels, such as [esɨma-esima] served as controls and were expected to be distinguished by all participants successfully.

There were two recordings used for each item. The order of tokens in a word-pair was counterbalanced. For instance, in the case of [esima-esma], there were four 'different' word-pairs [esima$_1$-esma$_1$], [esima$_1$-esma$_2$], [esima$_2$-esma$_1$], [esima$_2$-esma$_2$], and an additional four 'different' word-pairs in reversed order. All combinations of 'same' word pairs were also presented. For instance, in the case of [esima], there were four 'same' word-pairs [esima$_1$-esima$_1$], [esima$_1$-esima$_2$], [esima$_2$-esima$_1$], [esima$_2$-esima$_2$]. Each of the above word-pairs was presented twice. This amounted to a total of 720 test trials in the experiment.

The experiment was conducted individually in a quiet room using a laptop computer. The stimuli were presented to each participant through an AX discrimination task scripted in Praat with a low-noise headset (Koss R80 headphones). The participants were asked to listen to word-pairs of stimuli and determine whether the two stimuli were the 'same' or 'different' and click on the corresponding box 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 9 trials with another set of nonce words, [emɨma], [emima], and [emma][15], and they were not used in the

[15] Although English does not have singleton/geminate contrasts, the English participants were not expected to have trouble with [emma], as they were only asked to discriminate them from [emɨma] and [emima], but never from the singleton sequence [ema]. Therefore, even if they had perceived [emma] as [ema], they should have reliably discriminated it from the other practice items, and not found the practice task confusing. Furthermore, in the post-

actual experiment. The inter-stimulus interval and the inter-trial interval were both 1000ms. All the trials were randomised for each participant. The subjects were allowed to take a break after every 240 trials (roughly every 15 minutes), thus there were a total of two breaks during the experiment. Each subject took approximately 45 minutes to complete the experiment.

2.2 Results

As in Kabak & Idsardi (2007) and Monahan et al.'s (2009) paper, we took poorer discriminability between word-pairs with and without vowels, indicated by lower A-prime (A'), to suggest the induction of an illusory vowel (A' ≈ 0.5 reflects no discriminability; A' ≈ 1 reflects little to no confusion between word-pairs). A' is a nonparametric measure of discriminability that takes into account response bias (Macmillan & Creelman, 2004; Pollack & Norman, 1964). A' is presented instead of its parametric counterpart, d-prime (d') because with AX tasks it is actually not possible to assess if the d' parametric assumptions are upheld, and at least in some AX tasks the assumptions are not tenable (Stanislaw & Todorov, 1999). When the parametric assumptions are violated, d' is subject to vary with response bias (Stanislaw & Todorov, 1999).

Figure 1 shows average A' scores for English and Korean listeners on all the relevant word-pairs (see Table A. **2** in Appendix for the values). The A' scores for the control [ɨ-i] word-pairs ranged between 0.942 and 0.976, suggesting that both groups were successfully able to distinguish the control word-pairs which had two items with a different vowel.

test debriefing session, they consistently mentioned that both the practice task and the actual experiment were very straightforward.

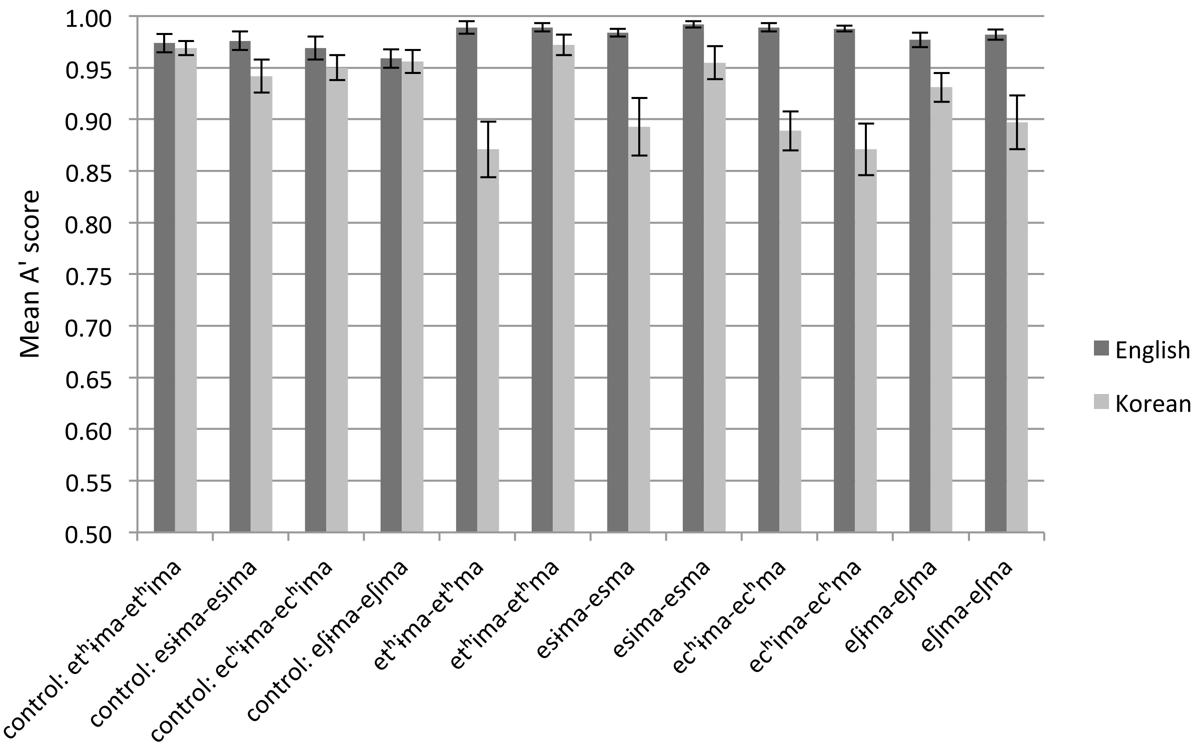


Figure 1. Average A' values for English and Korean listeners. Error bars represent standard errors.

Statistical analyses were conducted using SPSS Statistics 17.0 (SPSS Inc, 2008). As Mauchly's test showed that the assumption of sphericity was violated for the main effects of word-pair, $\chi^2(65) = 326.528$, $p < .001$, degrees of freedom were corrected using Greenhouse-Geisser ($\varepsilon = .330$). A mixed ANOVA of A' scores revealed a main effect of language, $F(1, 37) = 16.042$, $p <.001$, $\eta_p^2 = .302$, a main effect of word-pair, $F(3.634, 134.460) = 5.020$, $p = .001$, $\eta_p^2 = .119$, and an interaction of word-pair by language, $F(3.634, 134.460) = 7.809$, $p <.001$, $\eta_p^2 = .174$. Therefore, the Korean listeners achieved significantly lower A' scores than the English listeners for some word-pairs but not others.

In order to investigate on which word-pairs the two language groups performed in a statistically different way from the control [ɨ-i] word-pairs, we ran repeated measures of ANOVAs to compare A' values of Korean and English listeners against average control A'. We used the average of all the control A' scores since it is a more accurate estimate than the A' of an individual control word-pair (see Table A.3 in Appendix for the table of ANOVAs results). Therefore, in the post-hoc ANOVAs presented below, the factor word-pair had two levels (i.e., average control A', and the relevant test pair).

When the two language groups' A' scores of [et^{h}ɨma-et^{h}ma] were compared against the control A' scores, there was a main effect of word-pair, $F(1, 37) = 6.992$, $p = .012$, $\eta_p^2 = .159$, a main effect of language, $F(1, 37) = 14.212$, $p = .001$, $\eta_p^2 = .278$, and a significant interaction between word-pair and language, $F(1, 37) = 15.594$, $p < .001$, $\eta_p^2 = .297$. On the other hand, in the comparison between [et^{h}ima-et^{h}ma] and the controls for the two language groups, there was a main effect of word-pair, $F(1, 37) = 10.169$, $p = .003$, $\eta_p^2 = .216$, no main effect of language, $F(1, 37) = 2.144$, $p = .152$, $\eta_p^2 = .055$, and no interaction between word-pair and language, $F(1, 37) = 0.101$, $p = .752$, $\eta_p^2 = .003$. This suggests that the Korean listeners performed significantly worse only on [et^{h}ɨma-et^{h}ma] than on the control pairs compared to the English listeners but not on [et^{h}ima-et^{h}ma].

A similar pattern was observed when the A' scores of [esɨma-esma] and [esima-esma] were compared against the A' of control. When the two language groups' A' scores of [esɨma-esma] were compared against control A' scores, there was no main effect of word-pair, $F(1, 37) = 3.211$, $p = .081$, $\eta_p^2 = .08$; but there was a main effect of language, $F(1, 37) = 8.566$, $p = .006$, $\eta_p^2 = .188$, and there was an interaction between word-pair and language, $F(1, 37) = 7.131$, $p = .011$, $\eta_p^2 = .162$. In contrast, when [esima-esma] were compared to controls for the two language groups, there was a main effect of word-pair, $F(1, 37) = 5.581$, $p = .024$, $\eta_p^2 = .131$, there was no main effect of language, $F(1, 37) = 3.794$, $p = .059$, $\eta_p^2 = .093$, and there was no interaction between word-pair and language, $F(1, 37) = 3.484$, $p = .07$, $\eta_p^2 = .086$. In summary, for word-pairs with an alveolar cluster type, the Korean listeners were significantly worse than the English listeners on [etʰɨma-etʰma] and [esɨma-esma] compared to the control pairs, but not on [etʰima-etʰma] and [esima-esma].

When the two groups' A' scores of [ecʰɨma-ecʰma] were compared against the control A', there was a main effect of word-pair, $F(1, 37) = 10.031$, $p = .003$, $\eta_p^2 = .213$, a main effect of language, $F(1, 37) = 15.977$, $p < .001$, $\eta_p^2 = .302$, and there was an interaction between word-pair and language, $F(1, 37) = 27.428$, $p < .001$, $\eta_p^2 = .426$. Furthermore, the same pattern was found when the A' scores of [ecʰima-ecʰma] were compared against the controls. There was a main effect of word-pair, $F(1, 37) = 8.221$, $p = .007$, $\eta_p^2 = .182$, there was a main effect of language, $F(1, 37) = 17.668$, $p < .001$, $\eta_p^2 = .323$, and there was a significant interaction between word-pair and language, $F(1, 37) = 15.563$, $p < .001$, $\eta_p^2 = .296$. Therefore, the Korean listeners had significantly lower A' scores than the English listeners on both [ecʰɨma-ecʰma] and [ecʰima-ecʰma][16] compared to the control pairs.

For the comparison of A' scores of [eʃɨma-eʃma] and the controls, there was a main effect of language, $F(1, 37) = 7.301$, $p = .01$, $\eta_p^2 = .165$; however, there was neither a main effect of word-pair, $F(1, 37) = 0.900$, $p = .349$, $\eta_p^2 = .024$, nor an interaction between word-pair and language, $F(1, 37) = 3.188$, $p = .082$, $\eta_p^2 = .079$. In contrast, for the comparison between [eʃima-eʃma] and the controls, there was no main effect of word-pair, $F(1, 37) = 3.929$, $p = .055$, $\eta_p^2 = .096$, however, there was a main effect of language, $F(1, 37) = 8.619$, $p = .006$, $\eta_p^2 = .189$, and a significant interaction between word-pair and language, $F(1, 37) = 8.371$, $p = .006$, $\eta_p^2 = .184$. Thus, for word-pairs with a [ʃ], the Korean listeners performed significantly worse than the English listeners on [eʃima-eʃma] compared to the controls, but not on [eʃɨma-eʃma].

[16] Visually inspecting the data showed that both the illusory vowels were perceived with both tokens [ecʰma$_1$] and [ecʰma$_2$].

As can be observed from the above statistical analysis, the Korean listeners were in fact significantly worse than the English listeners at discriminating the predicted word-pairs—[etʰɨma-etʰma], [esɨma-esma], [ecʰɨma-ecʰma], [ecʰima-ecʰma], and [eʃima-eʃma]—compared to the control [ɨ-i] word-pairs.

3. Experiment 2

The results of the AX task in Experiment 1 showed that Korean listeners perceived different sets of illusory vowels in different phonological contexts, according to the phonological processes of *Vowel Deletion* and *Palatalization* in Korean. However, given the somewhat high A-prime values for all pairs in Experiment 1, it is possible that the experimental results are actually the results of a more phonetic listening mode[17]. But, it is unclear what set of hypotheses of phonetic perception would result in this particular pattern of differences between the American English and Korean speakers. A more reasonable explanation, according to us, is that the observed differences are *smaller* as a result of the ease of an AX task; i.e., the differences are *smaller* because the task allows for a more phonetic perception. Nevertheless, given that such phonetic factors are commonly assumed to be strongly present in an AX task[18], in Experiment 2, we ran an ABX task, in which listeners were presented with three stimuli and compared whether the first or the second stimulus was more similar to the third stimulus. The ABX task is much more memory intensive and therefore is typically viewed as motivating more higher-level or phonological listening (Gerrits & Schouten, 2004). As discussed in relation to Experiment 1, we expect to see that Korean listeners should have much more difficulty than American English

[17] Thanks to the anonymous reviewers for pointing out this possibility and suggesting the use of an ABX task.
[18] Actually, the evidence for this view is in our opinion rather weak. We refer the reader to (Boomershine, Hall, Hume, & Johnson, 2008) for more discussion.

listeners in distinguishing the following two sets of stimulus pairs: (a) [etʰɨma-etʰma], [esɨma-esma], [ecʰɨma-ecʰma], (b) [ecʰima-ecʰma], [eʃima-eʃma].

3.1 Method

3.1.1 Participants

Seventeen native Korean speakers (age 20 – 31, 9 men, 8 women) and 17 native American English speakers (age 19 – 22, 2 men, 15 women) participated in the experiment. All the subjects were recruited at Michigan State University, East Lansing, USA and reported to have normal hearing. None of the Korean speakers came to the US or visited other English speaking countries before the age of 13, nor had they lived in English speaking countries more than four years.

3.1.2 Stimuli

The stimuli for Experiment 2 were the same 12 test items used in Experiment 1 as described in Table 3.

3.1.3 Procedure

In Experiment 2, we used an ABX task to investigate a perceptual epenthesis effect. As in Experiment 1, we tested all combinations of vowels [i, ɨ, Ø]. For example, for the cluster [sm], the AB word-pairs were [esɨma-esma], [esima-esma], and [esɨma-esima]. There were two recordings used for each item as in Experiment 1 and the order of tokens in an AB word-pair was counterbalanced. For instance, in the case of [esima-esma], there were four AB word-pairs [$esima_1$-$esma_1$], [$esima_1$-$esma_2$], [$esima_2$-$esma_1$], [$esima_2$-$esma_2$], and an additional four word-pairs in reversed order. To each of these AB word-pairs, either A or B was added as an X. When adding X's, the same token was never repeated in a single trial. Therefore, in the case of [esima-esma], there were four ABA word-triplets [$esima_1$-$esma_1$-$esima_2$], [$esima_1$-$esma_2$-$esima_2$], [$esima_2$-$esma_1$-$esima_1$], [$esima_2$-$esma_2$-$esima_1$], and an additional four ABB word-triplets [$esima_1$-$esma_1$-$esma_2$], [$esima_1$-$esma_2$-$esma_1$], [$esima_2$-$esma_1$-$esma_2$], [$esima_2$-$esma_2$-$esma_1$]. The same combinatorics was used for all the other clusters ([t^hm], [c^hm], and [ʃm]). This amounted to a total of 192 trials in the experiment.

The experiment was conducted in a quiet room with a group of at most 4 participants per session. The stimuli were presented to each participant through an ABX task scripted in Praat with a low-noise headset (Plantronics SupraPlus HW261). The participants were asked to listen to word-triplets of stimuli and determine whether the last sound was more similar to the first or the second and click on the corresponding box (1 or 2) on the screen with a mouse. All the instructions were in English for the English speakers ("Decide whether the last sound is more similar to the first or the second) and in Korean for the Korean speakers ("세번째 소리가 첫번째 소리와 비슷한지 두번째 소리와 비슷한지 고르세요"). 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. The inter-stimulus interval was 500ms and the inter-trial interval was 1500ms. All 192 trials were randomised for each participant. The subjects were allowed to take a break after half of the trials and the experiment took about 17 minutes.

3.2 Results

As in Experiment 1, we calculated A' as a measure of perceptual epenthesis. Figures 2 & 3 show average A' scores for English and Korean listeners on all the word-pairs for ABA and

ABB orders respectively (see Table A. **4** and Table A. **5** in Appendix for the values). Overall, the figures illustrate that English listeners have higher A' scores than Korean listeners. Interestingly, both English and Korean listeners seem to have higher A' scores for the ABB than the ABA order.

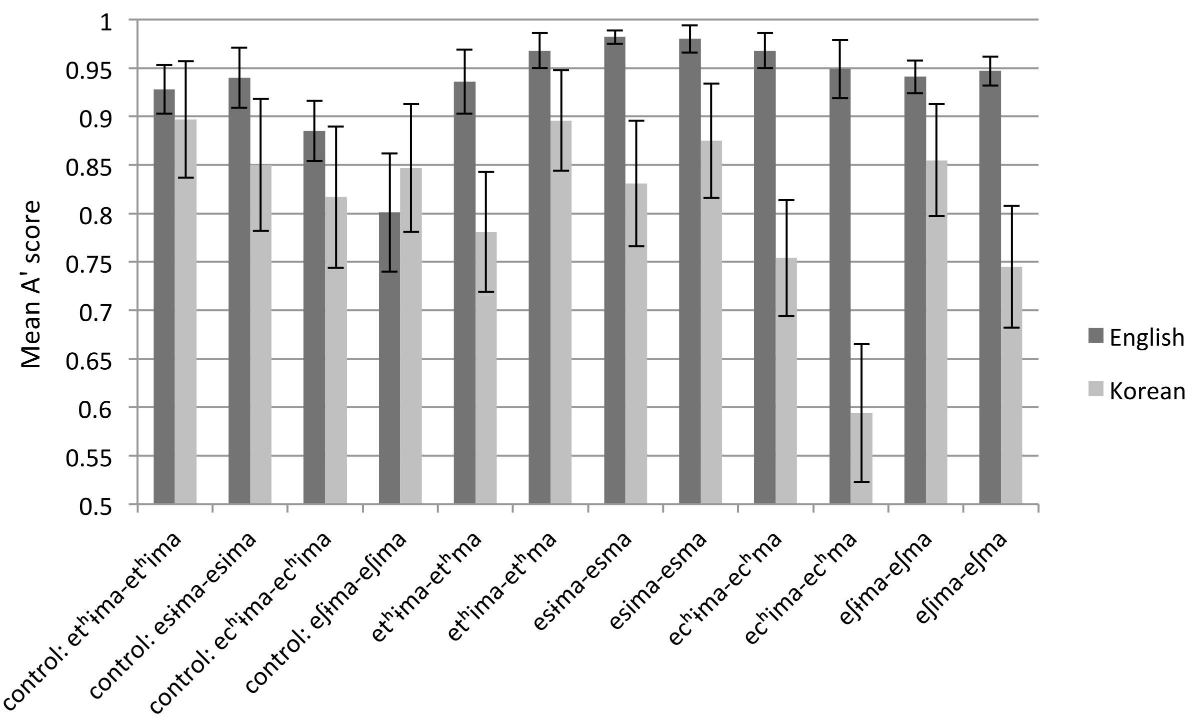


Figure 2. Average A' values for English and Korean listeners, for ABA order. Error bars represent standard errors.

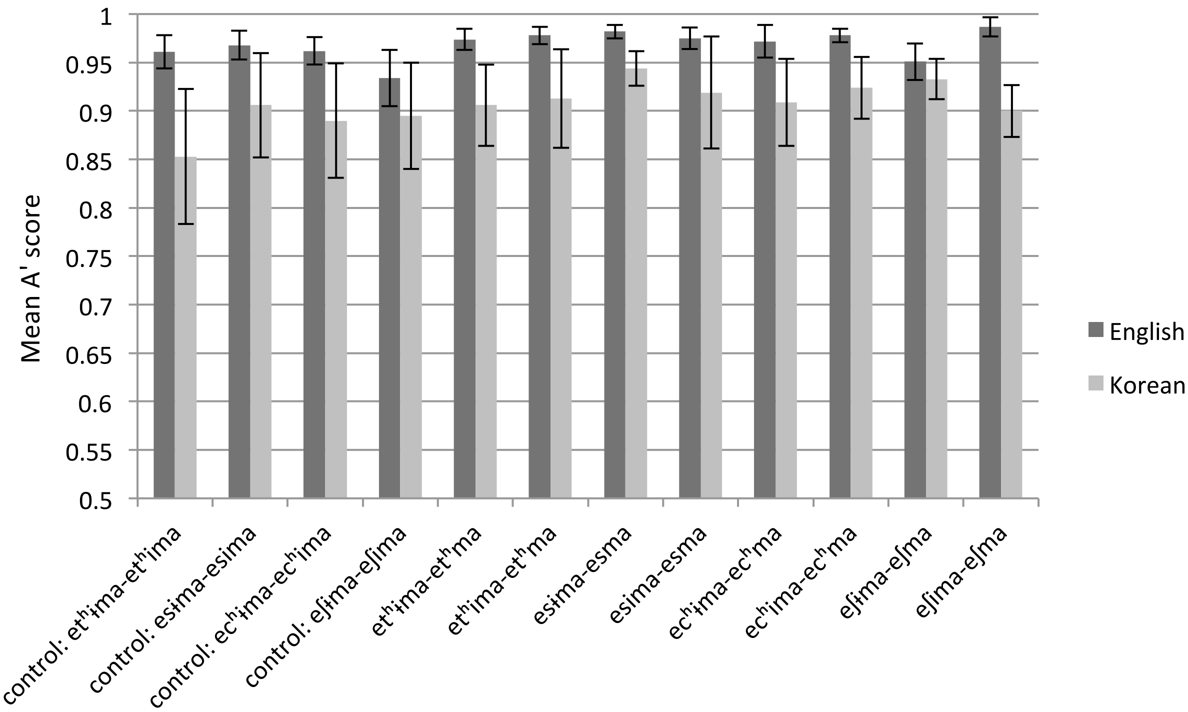


Figure 3. Average A' values for English and Korean listeners, for ABB order. Error bars represent standard errors.

In order to test statistical significance, a three-way mixed ANOVA was run with word-pair and order (i.e., ABA and ABB) as within subject variables and language (i.e., English and Korean) as a between subject variable. The three-way mixed ANOVA of A' scores revealed that there was an effect of language, $F(1, 32) = 4.377$, $p = .044$, $\eta_p^2 = .120$. There was a main effect of word-pair, $F(5.335, 170.713$[19]$) = 2.764$, $p = .018$, $\eta_p^2 = .079$, an interaction of word-pair by language, $F(5.335, 170.713) = 2.992$, $p = .011$, $\eta_p^2 = .086$. There was also a main effect of order with a very large effect size, $F(1, 32) = 24.476$, $p < .001$, $\eta_p^2 = .433$, and an interaction of order by language, $F(1, 32) = 5.774$, $p = .022$, $\eta_p^2 = .153$. There was an interaction of word-pair by

[19] When Mauchly's tests showed that the assumption of sphericity was violated, degrees of freedom were corrected using Greenhouse-Geisser.

order, $F(5.619, 179.798) = 3.217$, $p = .006$, $\eta_p^2 = .091$, and a three-way interaction between word-pair, order, and language, $F(5.619, 179.798) = 3.725$, $p = .002$, $\eta_p^2 = .104$.

As order had a main effect with a very large effect size, participants' responses for ABA and ABB orders were analyzed separately using two two-way mixed ANOVAs, with word-pair as a within-subject variable and language as a between-subject variable. A two-way mixed ANOVA for the ABA order revealed a main effect of language, $F(1, 32) = 5.410$, $p = .027$, $\eta_p^2 = .145$, a main effect of word-pair, $F(5.350, 171.214) = 4.056$, $p = .001$, $\eta_p^2 = .112$, also an interaction between word-pair and language, $F(5.350, 171.214) = 4.783$, $p < .001$, $\eta_p^2 = .130$. On the other hand, a two-way mixed ANOVA for the ABB order did not find a significant main effect of language, $F(1, 32) = 2.643$, $p = .114$, word-pair, $F(3.558, 113.847) = 0.852$, $p = .485$, or interaction between word-pair and language, $F(3.558, 113.847) = 0.479$, $p = .729$.

As only the ABA order showed a main effect of language and an interaction between word-pair and language, follow-up planned comparisons were conducted on the English and Korean listeners' responses for the ABA order (see Table A. **4** in Appendix for the table of t-tests results). The results showed that there was no significant difference between English and Korean listeners in the control word-pairs with different vowels, $t(32) = 0.475$, $p = .638$, for [et^{h}ɨma-et^{h}ima]; $t(32) = 1.199$, $p = .239$, for [esɨma-esima]; $t(21.580) = 0.852$, $p = .404$, for [ec^{h}ɨma-ec^{h}ima]; and $t(32) = -.504$, $p = .618$, for [eʃɨma-eʃima]. Among the test word-pairs, the English and Korean listeners were significantly different only for the predicted word-pairs, $t(32) = 2.217$, $p = .034$, for [et^{h}ɨma-et^{h}ma]; $t(16.379) = 2.292$, $p = .035$, for [esɨma-esma]; $t(19.003) = 3.444$, $p = .003$, for [ec^{h}ɨma-ec^{h}ma]; $t(21.664) = 4.577$, $p < .001$, for [ec^{h}ima-ec^{h}ma]; and $t(17.724) = 3.105$, $p = .006$, for [eʃima-eʃma]. The two language groups were not significantly different for the rest of the word-pairs, $t(32) = 1.310$, $p = .199$, for [etʰima-etʰma]; $t(17.854) = 1.708$, $p = .105$, for [esima-esma]; and $t(18.708) = 1.409$, $p = .175$, for [eʃɨma-eʃma].

To summarize the results of the ABX task in Experiment 2, stimuli order (i.e., ABA, ABB) had a main effect with a very large effect size, in which the Korean and English listeners had no significant difference in their responses to the ABB order, whereas they did show significant differences to the ABA order[20]. The effect of order could be explained by the fact that comparison to the second member of the triplet is going to be modulated by recency effects (Gerrits & Schouten, 2004). The listeners could have had lower memory load in the case of ABB trials as it is the second member of the triplet that is the same as the third. Given the lower memory load in the ABB trials, it is likely that the listeners used a more phonetic mode of perception.

The responses for the ABA order followed our predictions. Only Korean listeners but not English listeners perceived an [ɨ] between consonants in the clusters [tʰm] and [sm]. The Korean listeners also reported to have heard both [ɨ] and [i] for [cʰm] but they heard an [i] for [ʃm]. The results showed that there was no group difference in the control word-pairs with different vowels. However, it is interesting to see that the English listeners had relatively low A' scores for the control word-pairs with different vowels compared to the rest of the word-pairs with and without a medial vowel, which seems to reflect that they may have been influenced by English

[20] An anonymous reviewer asks why the results of the ABA order are more similar to that of the AX task than the results of the ABB order are, though it is possible to view the ABB order and the AX order as both involving local comparisons of identical stimuli. At this point, we can only speculate the possible reasons for this. First, while it is true that the ABB order does have the identical stimuli in adjacent positions, the participants in our experiment necessarily had to pay attention to both the stimulus adjacent to the crucial test item (X) and the non-adjacent one in a particular trial to arrive at their decision since they did not know which trial was likely to be an ABB trial in the experiment. So, it is not clear to us that the ABB trials are more like the AX task in our experiment. Furthermore, the ISIs in the experiments are substantially different for the two experiments (AX = 1000ms; ABX = 500ms); which means that adjacent stimuli in the ABX experiment might have been more affected by phonetic/auditory similarity than those in the AX task. In fact, the temporal proximity of the stimuli in the ABX task could potentially account for why the subjects were so good in the ABB trials. Perhaps, at such short ISIs participants still have access to fine-grained auditory representations in their short-term memory (Pisoni, 1973) that aids them with the task.

phonology, particularly the process of vowel reduction in unstressed syllables (Burzio, 1994). This issue definitely deserves a more thorough examination; however it is beyond the scope of the current article.

Furthermore, the fact that there was no observable effect of language in the ABB order also shows that the experimental results in both Experiment 1 and the ABA order of Experiment 2 were very unlikely to be due to a more phonetic perception mode or due to stimuli artifacts. If the results in Experiment 1 and the ABA order in Experiment 2 were either due to stimuli artifacts or a more phonetic mode of perception, then the same pattern of results should have been observable in the ABB results. This is not the case.

4. Experiment 3

Experiment 2 showed that the results of the ABX task also followed our predictions and demonstrated the same patterns as in the results of the AX task in Experiment 1. However, a potential problem with AX and ABX tasks is the locus of the difference perceived by the listener is unclear. For example, if the listener distinguishes the two stimuli [et^{h}ima-et^{h}ma], it is true that by hypothesis, the expected locus is indeed the medial vowel; however, it is not clear if the listener is distinguishing it based on the presence/absence of the medial vowel, or based on any other changes that they might have perceived in the consonants. More specifically, it was possible, in Experiment 1, that Korean listeners had a higher discriminability for the pair [et^{h}ima-et^{h}ma] than the pair [et^{h}ɨma-et^{h}ma] compared to English listeners because the first set involves a case of “perceptual palatalization”, wherein the [t^{h}] before [i] is perceived as a palatal consonant, i.e., [et^{h}ima] was perceived as [ec^{h}ima]. Therefore, the pair with both a consonantal and vowel difference in perception might have been discriminated better than that with just the presence versus absence of a vowel.

For this reason, in Experiment 3, we decided to run an identification task, in which listeners heard a stimulus and decided whether there was a vowel between the two consonants; and if there was a vowel, they decided which vowel it was. The identification task was different from the AX and ABX tasks in Experiments 1 and 2 in that Experiment 3 required participants to focus on the medial vowel. It was clearly a more metalinguistic task. Given that the identification task is more metalinguistic, and that it forces the participants to focus on just one part of the stimuli, it is possible that there could be slightly stronger task-related effects due response bias, selective attention focused on particular parts of the stimuli, and the effect these have on auditory coding (Caporello Bluvas & Gentner, 2013). Despite these concerns, it is useful to run an identification task as it gives us yet another perspective into what is happening during the perception of the relevant stimuli.

Following the view of perception laid out in the introduction, unlike the American English listeners, we expect the Korean listeners to hear illusory vowels in two sets of stimuli, (a) In the stimuli [et^{h}ma], [ec^{h}ma] and [esma], we expect the Korean listeners to hear the illusory vowel [ɨ], (b) In the stimuli [ec^{h}ma] and [eʃma], we expect the Korean listeners to hear the illusory vowel [i].

4.1 Method

4.1.1 Participants

The participants were the same as in Experiment 2.

4.1.2 Stimuli

The stimuli were the same 12 test items used in Experiments 1 and 2 as described in Table 3. There were two recordings used for each item as in Experiments 1 and 2, and they were each presented twice; therefore, there were 4 tokens of each item/nonce-word, and a total of 48 tokens in the experiment.

4.1.3 Procedure

In Experiment 3, we used an identification task to investigate a perceptual epenthesis effect. The experiment was conducted in a quiet room with a group of at most 4 participants per session. Experiment 3 drew participants' attention to the medial vowel in the stimuli. Therefore, Experiment 3 was conducted after Experiment 2 (after a short break) so as not to have the participants to focus on only the vowel in Experiment 2. The stimuli were presented to each participant through an identification task scripted in Praat with a low-noise headset (Plantronics SupraPlus HW261). The participants were asked to listen to a stimulus and determine whether the medial vowel was [ɨ], [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[21]. 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

[21] Thanks to an anonymous reviewer for raising an important point for any researcher working with English orthography in behavioural experiments. In Exp. 3, we used "u" as the letter to represent [ʊ], as it is used to signify the sound in words such as *pull* and *put*. We are of course aware that the letter "u" does not uniquely identify the phoneme [ʊ]. However, the spelling "oo", which is also used in English to represent the same sound appeared to us (impressionistically) to be more ambiguous. In fact, informal discussions with native English speakers prior to the experiment suggested to us that they prefer "u" to "oo" to represent the vowel [ʊ]. Finally, that the English listeners in Exp. 3 had no problem associating "u" with [ʊ] is further supported by the fact that the average identification rates of "u" in stimuli with the [ʊ] counterpart in the test items (i.e., et^{h}ɨma, esɨma, eʃɨma,ec^{h}ɨma) was about 96% (Appendix A.6).

모음을 고르세요”). 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. The inter-trial interval was 1000ms. All 48 trials were randomised for each participant.

4.2 Results

In Experiment 3, participants heard stimuli and determined whether the medial vowel was [ɨ], [i], or nothing. Responses to all the stimuli can be found in Table A. 6 in Appendix. Figures 4-5 illustrates percentages of vowel responses (i.e., [ɨ], [i], nothing) for eCma stimuli. The figure shows that the English listeners correctly identified the absence of the vowels in all cases, whereas the Korean listeners identified an [ɨ] for [$\mathrm{et^hma}$] and [esma], an [i] for [$\mathrm{ec^hma}$] and [eʃma]. Korean listeners also identified an [ɨ] for [$\mathrm{ec^hma}$].

To examine whether Korean and English listeners responded differently when they heard stimuli with no medial vowels, separate two-way mixed ANOVAs were run for eCma stimuli (i.e., $\mathrm{et^hma}$, esma, $\mathrm{ec^hma}$, eʃma), with response (i.e., [ɨ], [i], or nothing) as a within-subject variable and language (i.e., Korean, English) as a between-subject variable (see Table A. 7 in Appendix for the table of ANOVAs results). For [$\mathrm{et^hma}$], there was a main effect of response, $F(1.000, 32.000) = 40.403$, $p < .001$, $\eta_p^2 = .558$, and an interaction between response and language, $F(1.000, 32.000) = 67.814$, $p < .001$, $\eta_p^2 = .679$. For [esma], there was a main effect of response, $F(1.114, 35.650) = 42.398$, $p < .001$, $\eta_p^2 = .570$, and an interaction between response and language, $F(1.114, 35.650) = 82.694$, $p < .001$, $\eta_p^2 = .721$. For [$\mathrm{ec^hma}$], there was a main effect of language, $F(1, 32) = 10.667$, $p = .003$, $\eta_p^2 = .250$; response, $F(1.575, 50.410) = 22.884$, $p < .001$, $\eta_p^2 = .417$, and an interaction between response and language, $F(1.575, 50.410) = 41.937$, $p < .001$, $\eta_p^2 = .567$. For [eʃma], there was a main effect of response, $F(2, 64) = 32.667$,

$p < .001$, $\eta_p^2 = .505$, and an interaction between response and language, $F(2, 64) = 41.692$, $p < .001$, $\eta_p^2 = .566$. To summarize the results of the four mixed ANOVAs, for all four stimuli, English and Korean listeners responded differently.

In order to test our predictions, follow-up planned comparisons were conducted on the responses to the stimuli with no medial vowel. Planned comparisons showed that the English and Korean listeners' responses followed the predictions (see Table A. 8 in Appendix for the table of planned comparisons). For [et^{h}ma], Korean listeners identified [ɨ] more than English listeners, $t(16) = 8.235$, $p < .001$ but none of the English or Korean listeners identified an [i]. For [esma], there was a group difference in [ɨ] identification, $t(16) = 9.123$, $p < .001$ but not in [i] identification, $t(16) = 1.852$, $p = .083$. When presented with [ec^{h}ma], Korean listeners identified [i] more than English listeners, $t(16) = 5.886$, $p < .001$ and they also identified [ɨ] more than English listeners with approaching significance, $t(17.658) = 2.000$, $p = .061$. When presented with [eʃma], Korean listeners identified [i] more than English listeners, $t(19.938) = 6.500$, $p < .001$; however there was no statistical group difference in [ɨ] identification, $t(16) = 1.646$, $p = .119$.

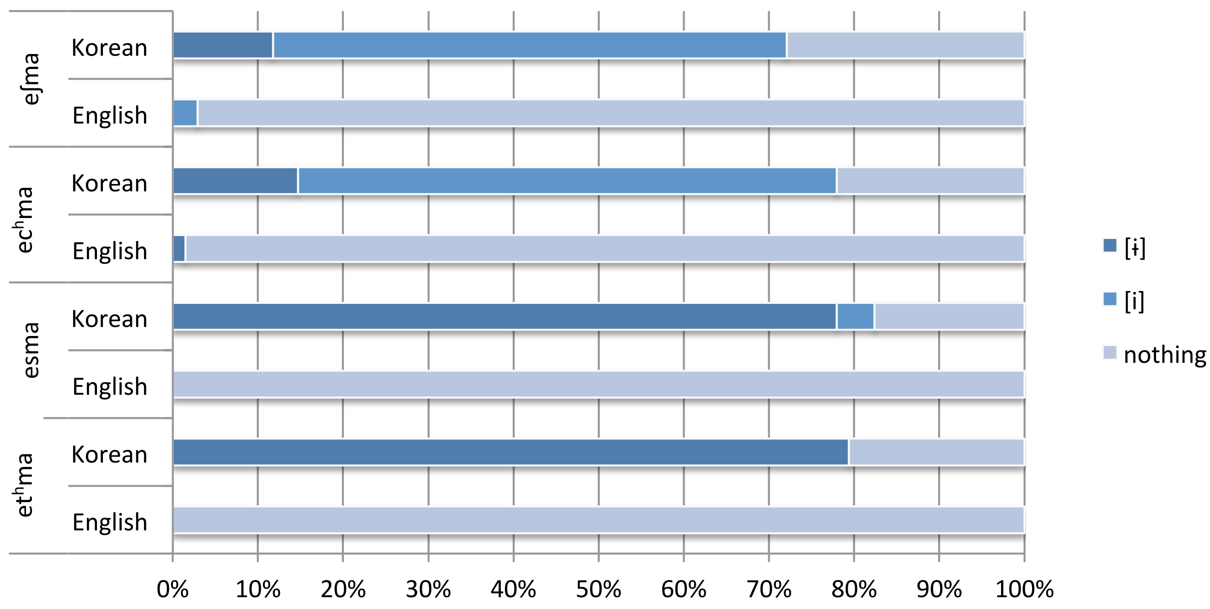


Figure 4. Percentages of vowel responses (i.e., [ɨ], [i], nothing) for eCma stimuli.

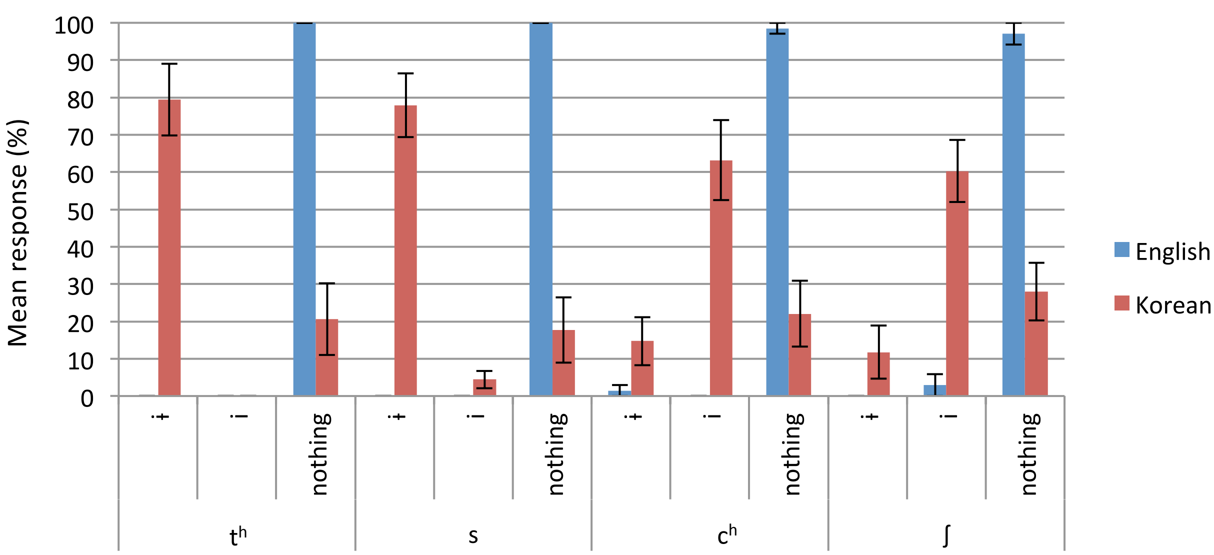


Figure 5. Percentages of vowel responses (i.e., [ɨ], [i], nothing) for eCma stimuli. Error bars represent standard errors. [Note: C = consonant]

In summary, the results of the identification task in Experiment 3 showed that for [etʰma] and [esma], the Korean listeners perceived an illusory [ɨ] more than the English listeners. For [ecʰma], the Korean listeners perceived both illusory [ɨ] and [i]. And for [eʃma], they perceived an illusory [i] more than the American English listeners; the Korean listeners also perceived a statistically non-significant number of illusory [ɨ] compared to the American English listeners.

Overall, the results of Experiment 3 were also consistent with the expectations laid out earlier. However, there are two aspects of the results in Experiment 3 that need more discussion and future exploration: (a) Although, we predicted that Korean listeners will hear more of both [i] and [ɨ] in [ecʰma] compared to the American English listeners, we made no further predictions about which would be identified at a higher level. At least from Experiments 2 & 3, there is clearly a preference for [i]. Whether this is a bias due to the experimental task or a more general bias due to the phonological facts of the Korean needs further investigation. (b) There is also a small, but non-significant, level of the perception of illusory [ɨ] in [eʃma] in the Experiment 3. Again, it is unclear if this is due to some facts about the auditory coding of segments that are not separate phonemes. Perhaps, the auditory segment [ʃ] is more likely to be coded as the auditory segment [s] (i.e., the more general member of the phonemic pair)[22] because the focus on the medial vowel hampers with the coding of adjacent consonants. A second possibility is that the vowel [ɨ] is a more *default* illusory vowel in Korean given its participation in general vowel deletion processes and it being the shortest vowel in the language. A third interesting explanation suggested to us by an anonymous reviewer is that of the possibility of their being an illusory [j] after [ʃ] (since, palatalization in Korean is also triggered before the palatal glide [j]), and consequently an illusory [ɨ] after the [j], thereby sometimes resulting in the phonemic percept

[22] We are suggesting that it is possible that the allophone [ʃ] might be more confusable with [s], but not vice-versa, given that /s/ is the phonemic counterpart. If, in fact, the [ʃ] is more asymmetrically confused with [s], then we would expect some illusory [ɨ] vowels in [ʃ] contexts.

/sjɨm/ when presented with [ʃm]. This third account is still consistent with the overall picture presented in this article of reverse inference to the underlying representation. With respect to all three possibilities mentioned above, it is important to notice that the presence of illusory vowel [ɨ] with [ʃ] is the smallest (and somewhat inconsistent in all three experiments), thereby suggesting that the locus of the explanation for this particular effect might be different than the ones we have been discussing in this article. Again, none of these possibilities takes away from the predictions in the current paper, but they do suggest very interesting further inquiry.

5. Discussion

In this paper, we showed that the location and quality of the illusory vowels in illicit phonotactic sequences of consonants is modulated by the native phonology of the listeners, using an AX discrimination task, an ABX task, and an identification task on Korean speakers with English speakers as a control group. Contrary to Dupoux et al.'s (2011) claim that the illusory vowel is the phonetically minimal or shortest vowel in the language, we showed that it is possible to obtain more than one illusory vowel in the same language, and even in the same context as long as the phonology of the language and the acoustic tokens themselves motivate such a re-analysis of the illicit sequences. The phonological processes of *Vowel Deletion* and *Palatalization* in Korean provide specific expectations of illusory vowels in different consonantal contexts. In consonantal sequence contexts where the first (coda) consonant is an alveolar consonant (namely, [et^hma], [esma]), the phonological alternations lead to the expectation of the vowel [ɨ] (*illusory vowel 1)*; in consonant contexts where the first (coda) consonant is an aspirated palatal stop consonant ([ec^hma]), the phonological alternations in the language lead to the expectation of both the vowel [ɨ] (*illusory vowel 1)* and the vowel [i] (*illusory vowel 2)*; and finally, in consonant contexts where the first (coda) consonant is a palatal fricative consonant ([eʃma]), the phonological alternations in the language lead to the expectation of the vowel [i] (*illusory vowel 2)*. We showed that the observed cases of illusory vowel perception were exactly the ones expected.

Our results clearly indicate that listeners can hear different illusory vowels in different contexts modulated by language-specific factors. In contrast, the expectation with regards to illusory vowels as per (Dupoux et al., 2011) is that the illusory vowel is [ɨ], perhaps due to its phonetically minimal characteristics. However, this doesn't account for the specific patterns of illusory vowels observed in the data. If this were the hypothesis, it is unclear why [c^hm] and [ʃm] trigger an illusory [i] for Korean listeners. Though, this is not to say that the proposed account is not partially compatible with the claim that the illusory vowel in some contexts can be the shortest vowel in the language (Dupoux et al., 2011). In illicit phonotactic contexts where the phonology of the language does not bias the listener towards a particular vowel (or set of vowels), the illusory vowel could indeed be expected to be the phonetically minimal or shortest vowel.

Furthermore, the patterns of illusory vowel perception observed *cannot* be explained based purely on surface phonotactic patterns in the language. It is true that the illusory vowels were perceived by the Korean listeners in phonotactically illicit sequences [*t^hm, *c^hm, *sm, *ʃm]. However, again, the focus needs to be on the *quality* of the illusory vowel perceived. The perception of the illusory vowel [i] in the contexts [ʃm] could indeed be alternatively explained by surface phonotactic constraints that ban [ʃ] from being followed by any vowel except [i] in Korean (since, only [ʃi] are possible in Korean). Similarly, one could also account for the absence of the illusory vowel [i] in the context [sm] by appealing to a surface phonotactic constraint banning [*si]. However, attempting to account for all the illusory vowels observed in this article using purely surface-phonotactics is problematic for the following reasons: (a) The account proposed for the absence of the illusory vowel [i] in [sm] contexts by itself does not explain why some other illusory vowel other than [ɨ] is not inferred in the [sm] contexts[23]. (Note: [sam], [sem] and [som] are also possible sequences in Korean.); (b) On a similar note, the purely surface phonotactic account cannot explain why some other vowels other than [i] and [ɨ] are not possible illusory vowels in the [c^hm] context (Note: [c^ham], [c^hom], and [c^hem] are possible sequences in Korean); (c) Finally, and most importantly, the purely surface phonotactic account cannot explain why [i] is not a possible illusory vowel in the [t^hm] context though [t^him] is a possible sequence in Korean[24]. And in parallel to (a-b), it also does not easily account for why other vowels are also not possible illusory vowels. In contrast to the problems associated with a purely surface phonotactic account, the account based on phonological alternations laid out earlier in the paper is able to accurately predict the quality of the illusory vowel in different contexts.

The account of illusory vowels motivated by the current paper provides an explanation for the somewhat unexpected results presented by Monahan et al (2009). Monahan and colleagues attempted to obtain more than one illusory vowel for Japanese speakers. Based on loan-word patterns in Japanese such as [makɯdonarɯdo] ‘McDonald’s’, it is possible to hypothesise that the illusory vowel next to non-coronal consonants (e.g., [k], [g]…) be the vowel [ɯ], and that next to coronal consonants [e.g., [t], [d]…) be the vowel [o]. However, as they

[23] One could of course argue that Experiment 3 (that involved the direct identification of the illusory vowel) did not include any of the other vowels. However, this counter argument is weakened by the fact that the loanword data in Korean show exactly the same pattern in that they show *only* epenthetic [ɨ] in [sm] contexts.
[24] As noted in footnote 11, the stop palatalization process is blocked within morphemes.

show, while Japanese speakers confuse [egɯma] and [egma], they do not seem to confuse [etoma] and [etma]. From the perspective developed in the current article, there appear to be no native Japanese phonological processes that motivate other possible illusory vowels in the contexts they tested. So, by our account, the only illusory vowel expected for the contexts they tested is the [ɯ], as it is the shortest vowel in the language[25]. We further predict that there will be other illusory vowels in Japanese. Japanese has a similar process of palatalizing alveolar consonants before /i/ as in Korean; therefore the account proposed here predicts that the set of illusory vowels induced next to illicit palatal codas in Japanese should include the vowel /i/.

Finally, the article provides support for the view that speech perception involves the reverse inference to the underlying/phonemic representation level. Such a conception of speech perception, according to us, falls out quite naturally from a Bayesian perspective, and therefore, we see it as support more generally for the Bayesian view of speech perception (Bever & Poeppel, 2010; Feldman & Griffiths, 2007; Poeppel & Monahan, 2011; Sonderegger & Yu, 2010; Wilson & Davidson, in press; Yu, 2011). Having said this, it is important to re-iterate the point made earlier (fn. 1) that what we show in this article is consistent with any view of speech perception that makes crucial reference to the concept of reverse inference to the underlying/phonemic representation level.

Finally, in line with some previous research on the topic (Boomershine, Hall, Hume, & Johnson, 2008; Huang, 2001; Hume & Johnson, 2003; Johnson & Babel, 2010), the results of the current article show that speech perception is modulated by not only the acoustics of the speech

[25] Coronal stops in Japanese cannot be followed by [ɯ] (*tɯ, *dɯ). Therefore, it is reasonably clear that inferring the illusory vowel [ɯ] does not perfectly repair the illicit phonotactics in nonsense words such as [edzo]. However, this still leaves open the question of why loanwords with an illicit coronal coda consonant are adapted into Japanese with an [o] repair, and not some other vowel (apart from [ɯ]). If the account of illusory vowels presented in this paper is on the right track, this suggests a non-perceptual explanation for the [o]-insertion repair involved in loanwords with coronal coda stops in Japanese.

tokens and the surface phonotactics of a language, but also by the *phonological alternations*, and thereby by the phoneme to allophone mappings of a language.

References

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

Berent, I., Lennertz, T., Jun, J., Moreno, M., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Acadamy 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.

Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: A (re-)emerging program of research for language and vision. *Biolinguistics*, *4.2-3*, 174–200.

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/.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony vs. contrast on speech perception. In P. Avery, B. E. Dresher, & K. D. Rice (Eds.), *Phonological Contrast: Perception and Acquisition* (pp. 146–172). New York: Mouton de Gruyter.

Burzio, L. (1994). *Principles of English Stress*. Cambridge: Cambridge University Press.

Caporello Bluvas, E., & Gentner, T. Q. (2013). Attention to natural auditory signals. *Hearing Research*, *305*, 10–18. doi:10.1016/j.heares.2013.08.007

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

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. doi:10.1162/089892900562390

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.

Feldman, N. H., & Griffiths, T. L. (2007). A Rational Account of the Perceptual Magnet Effect (pp. 257–262). Austin, TX: Cognitive Science Society.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. doi:10.3758/BF03194885

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.

Hong, S. (1997). Palatalization and Umlaut in Korean. *University of Pennsylvania Working Papers in Linguistics*, *4*(3), 87–132.

Huang, T. (2001). The interplay of perception and phonology in tone 3 sandhi in Chinese Putonghua. In E. Hume & K. Johnson (Eds.), *Studies on the Interplay of Speech Perception and Phonology* (pp. 23–42).

Hume, E., & Johnson, K. (2003). The Impact of Partial Phonological Contrast on Speech Perception. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2385–2388). Barcelona.

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

Iverson, G. K. (2004). Deriving the Derived Environment Constraint in Non-Derivational Phonology. *Studies in Phonetics, Phonology and Morphology*, *11*, 1–23.

Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, *38*, 127–136.

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

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.

Kuhl, P. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht: Kluwer Academic Publishers.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In *Papers in laboratory phonology* (Vol. 7, pp. 637–676). Berlin: Mouton.

Lahiri, A., & Reetz, H. (2010). Distinctive Features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*, 44–59.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Taylor & Francis.

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

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.

Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*, 357–395.

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

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*, 253–260.

Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, *26*(7), 935–951.

Pollack, I., & Norman, D. A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science*, (1), 125–126.

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

Sonderegger, M., & Yu, A. C. L. (2010). A rational account of perceptual compensation for coarticulation. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 375–380). Austin, Tx: Cognitive Science Society.

SPSS Inc. (2008). SPSS Statistics for Windows (Version 17.0). Chicago: Spss Inc: SPSS Inc.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *31*(1), 137–149.

Wilson, C., & Davidson, L. (in press). Bayesian Analysis of Non-native Cluster Production. In *Proceedings of NELS 40*. MIT, Cambridge, MA.

Yu, A. C. L. (2011). On measuring phonetic precursor robustness: A response to Moreton 2008. *Phonology*, *28*(3), 491–518.

Appendix

Table A. 1
Test Tokens in Experiment 1

		Vowels			Items relevant to Experiment 1
		[ɨ]	[i]	None	
Illicit coda	Alveolar	etʰɨma	etʰima	etʰma	X
		esɨma	esima	esma	X
		esɨpʰa	esipʰa	espʰa	
	Palatal	ecʰɨma	ecʰima	ecʰma	X
		ecʰɨpʰa	ecʰipʰa	ecʰpʰa	
		eʃɨma	eʃima	eʃma	X
Licit coda	Alveolar	edɨpʰa	edipʰa	edpʰa	
		edɨma	edima	edma	
Filler	Labial	ebɨpʰa	epipʰa	ebpʰa	
		emɨpʰa	emipʰa	empʰa	

Table A. 2
Means and standard errors of A' values for English and Korean listeners in Experiment 1

Pairs	English listeners (N = 19)		Korean listeners (N = 20)	
	M	*SE*	*M*	*SE*
etʰɨma-etʰima	0.974	0.009	0.969	0.007
etʰɨma-etʰma	0.989	0.006	0.871	0.027
etʰima-etʰma	0.989	0.004	0.972	0.01
esɨma-esima	0.976	0.009	0.942	0.016
esɨma-esma	0.984	0.004	0.893	0.028
esima-esma	0.992	0.003	0.955	0.016
ecʰɨma-ecʰima	0.969	0.011	0.950	0.012
ecʰɨma-ecʰma	0.989	0.004	0.889	0.019
ecʰima-ecʰma	0.988	0.003	0.871	0.025
eʃɨma-eʃima	0.959	0.009	0.956	0.011
eʃɨma-eʃma	0.977	0.007	0.931	0.014
eʃima- eʃma	0.982	0.005	0.897	0.026

Table A. 3

Results of ANOVAs comparing A' scores of Korean and English listeners against average control A' in Experiment 1

	F	df	p	η_p^2
et^{h}ɨma-et^{h}ma				
Word-pair	6.992	1, 37	.012	.159
Language	14.212	1, 37	.001	.278
Word-pair x Language	15.594	1, 37	$< .001$	.297
et^{h}ima-et^{h}ma				
Word-pair	10.169	1, 37	.003	.216
Language	2.144	1, 37	.152	.055
Word-pair x Language	0.101	1, 37	.752	.003
esɨma-esma				
Word-pair	3.211	1, 37	.081	.080
Language	8.566	1, 37	.006	.188
Word-pair x Language	7.131	1, 37	.011	.162
esima-esma				
Word-pair	5.581	1, 37	.024	.131
Language	3.794	1, 37	.059	.093
Word-pair x Language	3.484	1, 37	.070	.086
ec^{h}ɨma-ec^{h}ma				
Word-pair	10.031	1, 37	.003	.213
Language	15.977	1, 37	$< .001$	.302
Word-pair x Language	27.428	1, 37	$< .001$	.426
ec^{h}ima-ec^{h}ma				
Word-pair	8.221	1, 37	.007	.182
Language	17.668	1, 37	$< .001$	.323
Word-pair x Language	15.563	1, 37	$< .001$	.296
eʃɨma-eʃma				
Word-pair	0.900	1, 37	.349	.024
Language	7.301	1, 37	.010	.165
Word-pair x Language	3.188	1, 37	.082	.079
eʃima-eʃma				
Word-pair	3.929	1, 37	.055	.096
Language	8.619	1, 37	.006	.189
Word-pair x Language	8.371	1, 37	.006	.184

Table A. 4
Means and standard errors of A' values and t-tests results for the ABA order in Experiment 2

	English Listeners (N = 17)		Korean Listeners (N = 17)				
	M	*SE*	*M*	*SE*	*t*	*df*	*p*
et^{h}ɨma-et^{h}ima	.928	.025	.897	.060	0.475	32	.638
esɨma-esima	.940	.031	.850	.068	1.199	32	.239
ec^{h}ɨma-ec^{h}ima	.885	.031	.817	.073	0.852	21.58	.404
eʃɨma-eʃima	.801	.061	.847	.066	-.504	32	.618
et^{h}ɨma-et^{h}ma	.936	.033	.781	.062	2.217	32	.034
et^{h}ima-et^{h}ma	.968	.018	.896	.052	1.310	32	.199
esɨma-esma	.982	.007	.831	.065	16.379	2.292	.035
esima-esma	.980	.014	.875	.059	1.708	17.854	.105
ec^{h}ɨma-ec^{h}ma	.968	.018	.754	.060	3.444	19.003	.003
ec^{h}ima-ec^{h}ma	.949	.030	.594	.071	4.577	21.664	< .001
eʃɨma-eʃma	.941	.017	.855	.058	1.409	18.708	.175
eʃima-eʃma	.947	.015	.745	.063	3.105	17.724	.006

Table A. 5
Means and standard errors of A' values for the ABB order in Experiment 2

Pairs	English listeners (N = 17)		Korean listeners (N = 17)	
	M	*SE*	*M*	*SE*
et^{h}ɨma-et^{h}ima	0.961	0.017	0.853	0.070
et^{h}ɨma-et^{h}ma	0.974	0.011	0.906	0.042
et^{h}ima-et^{h}ma	0.978	0.009	0.913	0.051
esɨma-esima	0.968	0.015	0.906	0.054
esɨma-esma	0.982	0.007	0.944	0.018
esima-esma	0.975	0.011	0.919	0.058
ec^{h}ɨma-ec^{h}ima	0.962	0.014	0.890	0.059
ec^{h}ɨma-ec^{h}ma	0.972	0.017	0.909	0.045
ec^{h}ima-ec^{h}ma	0.978	0.007	0.924	0.032
eʃɨma-eʃima	0.934	0.029	0.895	0.055
eʃɨma-eʃma	0.951	0.019	0.933	0.021
eʃima- eʃma	0.987	0.010	0.900	0.027

Table A. 6
Means and standard errors of percentages of vowel responses in Experiment 3

	English listeners			Korean listeners		
	ɨ	i	nothing	ɨ	i	nothing
etʰɨma	98.53 (1.47)	0 (0)	1.47 (1.47)	60.29 (10.51)	1.47 (1.47)	38.24 (10.74)
etʰima	0 (0)	100 (0)	0 (0)	0 (0)	94.12 (5.88)	5.88 (5.88)
etʰma	0 (0)	0 (0)	100 (0)	79.41 (9.64)	0 (0)	20.59 (9.64)
esɨma	98.53 (1.47)	0 (0)	1.47 (1.47)	73.53 (9.7)	0 (0)	26.47 (9.7)
esima	0 (0)	100 (0)	0 (0)	4.41 (3.21)	88.24 (6.11)	7.35 (3.56)
esma	0 (0)	0 (0)	100 (0)	77.94 (8.54)	4.41 (2.38)	17.65 (8.78)
ecʰɨma	95.59 (3.21)	0 (0)	4.41 (3.21)	63.24 (9.62)	4.41 (3.21)	32.35 (9.29)
ecʰima	2.94 (2.94)	95.59 (3.21)	1.47 (1.47)	0 (0)	97.06 (2.94)	2.94 (2.94)
ecʰma	1.47 (1.47)	0 (0)	98.53 (1.47)	14.71 (6.45)	63.24 (10.74)	22.06 (8.81)
eʃɨma	91.18 (5.23)	1.47 (1.47)	7.35 (4.68)	48.53 (10.16)	0 (0)	51.47 (10.16)
eʃima	1.47 (1.47)	97.06 (2.01)	1.47 (1.47)	4.41 (3.21)	91.18 (4.26)	4.41 (3.21)
eʃma	0 (0)	2.94 (2.94)	97.06 (2.94)	11.76 (7.15)	60.29 (8.32)	27.94 (7.69)

Table A. 7
Results of ANOVAs in Experiment 3

	F	*df*	*p*	η_p^2
[etʰma]				
Language	.000	1, 32	1.000	.000
Response	40.403	1.000, 32.000	< .001	.558
Response x Language	67.814	1.000, 32.000	< .001	.679
[esma]				
Language	.000	1, 32	1.000	.000
Response	42.398	1.114, 35.650	< .001	.570
Response x Language	82.694	1.114, 35.650	< .001	.721
[ecʰma]				
Language	10.667	1, 32	.003	.250
Response	22.884	1.575, 50.410	< .001	.417
Response x Language	41.937	1.575, 50.410	< .001	.567
[eʃma]				
Language	.000	1, 32	1.000	.000
Response	32.667	2, 64	< .001	.505
Response x Language	41.692	2, 64	< .001	.566

Table A. 8
Results of planned comparisons of the English and Korean listeners' responses in Experiment 3

Stimuli	Response	*t*	*df*	*p*
$et^h ma$	ɨ	8.235	16	< .001
esma	ɨ	9.123	16	< .001
	i	1.852	16	.083
$ec^h ma$	ɨ	2.000	17.658	.061
	i	5.886	16	< .001
eʃma	ɨ	1.646	16	.119
	i	6.500	19.938	< .001