

# *Illusory vowels in perceptual epenthesis: the role of phonological alternations\**

**Karthik Durvasula**

Michigan State University

**Jimin Kahng**

Northeastern Illinois University

---

Listeners often 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 incorporate surface phonotactic information and the acoustics of the speech tokens. In this article, inspired by Bayesian models of speech perception, we claim that the listener attempts to identify target phonemic representations during perception. 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) with native Korean listeners, and native English listeners as a control group, showing that Korean listeners perceive different sets of illusory vowels in different phonological contexts, in accordance with the phonological processes of vowel deletion and palatalisation in the language.

---

## **1 Introduction**

The phenomenon of illusory vowels has received a great deal of attention in recent literature (e.g. Dupoux *et al.* 1999, Dehaene-Lambertz *et al.* 2000, Berent *et al.* 2007, Kabak & Idsardi 2007, Berent *et al.* 2009, Monahan *et al.* 2009, Dupoux *et al.* 2011). The general finding of these

\* E-mail: DURVASUL@MSU.EDU, JKAHNG@NEIU.EDU.

This article was made possible by the help and support of many individuals. We would like to thank: first and foremost, the associate editor, three anonymous reviewers and the editors for valuable criticism that helped to improve the article greatly; second, Bill Idsardi, Alan Beretta, Yen-Hwei Lin and the members of the phonology-phonetics group at Michigan State University for many helpful discussions; third, Hongjun Seo and Boram Koo for helping us with experiment design; fourth, Alan Munn, Cristina Schmitt and Suzanne Wagner for help with experimental equipment; and finally, the audiences of *NELS* 43 and the 22nd Japanese/Korean Linguistics Conference for probing questions and helpful discussion.

studies is that listeners sometimes perceive illusory vowels in stimuli containing 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 between the consonants, 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 [ebuzo], given that [bz] is an illicit consonant sequence in Japanese, as shown 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 in most current psycholinguistic models of speech recognition, in which the primary units are segments and phonological/phonetic features (McClelland & Elman 1986, Kuhl 1993, Best 1994, Lahiri & Reetz 2002, 2010, Norris & McQueen 2008). They suggest that this can be remedied by having phonotactic constraints which refer to surface sequences of segments interact with categorisation in a single processing step. We argue in this article that the phenomenon of illusory vowels shows 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 (e.g. Feldman & Griffiths 2007, Bever & Poeppel 2010, Sonderegger & Yu 2010, Poeppel & Monahan 2011, Yu 2011, Wilson & Davidson *in press*), we claim that the task of the listener in speech perception is primarily one of reverse inference: it is to identify the best estimate of the intended underlying categories of the utterance for the incoming acoustic token.<sup>1</sup> In this case, the underlying category information we make reference to is the phonemic representation. The knowledge about which underlying categories map to which surface categories must include information about both phonological alternations and phonotactic constraints, which are therefore both 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, related work has argued for the need for the speech-perception mechanism to be sensitive to phonological alternations (Huang 2001, Hume & Johnson 2003, Boomershine *et al.* 2008, Johnson & Babel 2010). For example, Huang (2001) shows that the tone-sandhi alternation involving the contextual neutralisation of two otherwise contrastive tones in Mandarin Chinese (the low-falling-rising tone and the mid-rising tone)

<sup>1</sup> 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 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.

causes the two tones to be perceptually closer, and therefore more easily confusable for Mandarin Chinese listeners. In this article we extend the 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. Furthermore, we also present a particular point of view 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 of speech perception more generally, has a direct bearing on the theoretical literature on loanword adaptations, where there has been an extensive debate on the factors involved (e.g. Peperkamp 2005, Davidson 2007). Whereas some have claimed that perceptual factors are perhaps the primary factor influencing loanword-adaptation patterns (Peperkamp & Dupoux 2003, Peperkamp 2005), others have argued that perception is at best a minor factor in such patterns (Paradis & LaCharité 1997, Jacobs & Gussenhoven 2000, LaCharité & Paradis 2005, Uffmann 2006). The account proposed here suggests, contrary to these claims, that the perceptual mechanism uses the phonological system for inference in some detail, and it is therefore perhaps impossible to separate the effects of speech perception on loanword patterns from those of the phonological system.

With respect to the locus of perceptual epenthesis, while earlier work in the domain of illusory vowels (Dupoux *et al.* 1999, Dehaene-Lambertz *et al.* 2000) assumed that the relevant constraints driving the perceptual illusions were sequential phonotactic constraints, Kabak & Idsardi (2007) argue that the relevant phonotactic constraints driving such perceptual illusions are the syllable-structure constraints of the language.<sup>2</sup> They ran an AX discrimination task using 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 in Korean, and the corresponding consonant sequence,  $C_1C_2$ , was also illicit. In the other,  $C_1$  was a licit coda consonant, but the corresponding consonant sequence was illicit. They show that the perception of illusory vowels was consistently triggered by the first type of consonant sequence, but not by the second. They therefore argue that the illusory vowel phenomenon is better accounted for by syllable-structure constraints than by the surface consonant-sequence constraints of the language.

It has also been argued that the perception of illusory vowels is affected by the listener's knowledge of language universals related to the Sonority Sequencing Principle and syllable structure. In a series of experiments on Korean and English speakers, Berent and her colleagues show that universally dispreferred initial consonant sequences trigger a stronger perception of illusory vowels than universally preferred initial consonant sequences, even when neither sequence occurs in the subject's native language

<sup>2</sup> 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 or surface.

(Berent *et al.* 2007, Berent *et al.* 2008, Berent *et al.* 2009). For example, both [lb] and [bl] are illicit initial consonant sequences in Korean; however, only the former is a universally dispreferred sequence across the world's languages (Sievers 1881, Jespersen 1904, Hooper 1976, Steriade 1982, Selkirk 1984). Berent and her colleagues show that Korean speakers more readily misperceive the former than the latter.

Related work has shown that perceptual distortions are also driven by more abstract consonant-sequence constraints. Moreton (2002) demonstrates that subjects make use of abstract featural co-occurrence constraints, showing that English speakers misperceive words beginning with [dl] much more often than those with [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.<sup>3</sup>

It has also been shown that illusory vowels are only one of the many possible perceptual repairs for phonotactically illegal consonant sequences (Hallé *et al.* 1998, Davidson 2007, Davidson & Shaw 2012). 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, including consonant deletion, metathesis, prothesis, consonant change and perception of illusory vowels.<sup>4</sup> They further show that the likelihood of a particular repair was affected by the type of illicit consonant sequence presented to the subject.

As can be seen from the above review, the bulk of previous research assumes that perceptual epenthesis of illusory vowels is driven purely by surface phonotactics and the phonetic characteristics of acoustic tokens. However, this is not to say that there is no evidence of abstract knowledge being used.<sup>5</sup> As discussed above, Moreton (2002), Berent *et al.* (2007), Berent *et al.* (2008) and Berent *et al.* (2009) have indeed shown that listeners access relatively abstract knowledge. However, the knowledge that listeners seem to be using could be making reference to surface representations in a phonological sense (rather than to the acoustic/auditory signal), since Berent and colleagues' Sonority Sequencing Principle and Moreton's constraint on alveolar co-occurrence can both be thought of as surface phonotactic constraints, as is standard in the Optimality Theory tradition. Therefore, on the basis of the results, there is no evidence that a more abstract phonological level of representation, the phonemic level, is accessed during perception.

<sup>3</sup> This suggests that phonotactics is not simply a 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 from the behaviour of Korean listeners (in Kabak & Idsardi 2007), since the Korean listeners were at ceiling with some non-attested clusters.

<sup>4</sup> Similar repairs have been observed in loanword adaptations.

<sup>5</sup> Thanks to an anonymous reviewer for pointing out the importance of this fact.

With respect to the quality of the illusory vowel, Dupoux *et al.* (2011: 200) argue that it is ‘the phonetically minimal element of the language’, and therefore ‘the shortest vowel’ in the language (e.g. [u] in Japanese or [i] in Brazilian Portuguese). Their claim predicts that there can be at most one illusory vowel in a language.<sup>6</sup> 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 phonological alternations in the language. And in some contexts, it is even possible to trigger more than one illusory vowel, as long as the phonology of the language supports it.

Acoustic studies of Korean have shown that [i] is the shortest vowel in the language (Han 1964, Kim 1974, Chung *et al.* 1999).<sup>7</sup> The typical duration of [i] in phrase-initial contexts is around 144 ms; the duration of [i] and [u] in similar positions is around 160 ms and 165 ms 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 [i] to be the illusory vowel in all contexts.

We propose in what follows that, while it is certainly 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 (Feldman & Griffiths 2007, Bever & Poeppel 2010, Sonderegger & Yu 2010, Poeppel & Monahan 2011, Yu 2011, Wilson & Davidson *in press*) in claiming that the task of the listener in speech perception is primarily one of reverse inference – it is to identify the best estimate of the intended underlying categories (phonemic representations) of the utterance, given the acoustic token.<sup>8</sup> Knowledge of both phonological alternations and phonotactic constraints is required to reverse infer the phonemic 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.

<sup>6</sup> 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 coarticulatory traces in the spliced stimuli led to this particular result. This should be kept separate from cases where the consonant clusters were naturally produced, and therefore had no coarticulatory information, due to a spliced-out vowel. This was the case in their stimuli that were produced naturally with the consonant-sequence violation (e.g. [abda]). In such items, consistent with the claim of participants perceiving ‘the shortest vowel’, the Japanese speakers primarily perceived an /u/.

<sup>7</sup> There is some debate in the phonological literature on the use of the unrounded high back vowel [u] for the Korean letter *우*. It has been 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 [i] throughout.

<sup>8</sup> A full Bayesian analysis would 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.

More specifically, in regard to the quality of the illusory vowel, we see the perceiver's task as being an attempt to repair the illicit phonotactic sequence with a vowel phoneme that best maps onto the phonetic characteristics of the acoustic token. When there are no relevant phonological alternations to bias listeners towards a certain vowel in a particular segmental context, the best guess to repair the particular phonotactic violation is indeed the phonetically minimal/shortest vowel in the inventory, *à la* Dupoux *et al.* (2011). This is because the shortest vowel is, in terms of duration, the closest in the inventory to the absence of a vowel. The illicit consonant sequences tested by Dupoux *et al.* were of the form  $V_1C_1C_2V_2$ . In Japanese and Brazilian Portuguese, consonantal sequences triggering the perception of illusory vowels, such as [bd bg gn], do not appear to be influenced by any phonological alternations relevant to the process of perceptual epenthesis (i.e. alternations that bias listeners towards a certain vowel), so the best vowel guesses 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 a specific segmental context, the best guess depends on both the phonetics of the acoustic token and the phonological alternations themselves. The types of phonological processes that are likely to play a role are those that bias the listener's expectations about the quality of the illusory vowel. One such process is a regular vowel-deletion process targeting a particular vowel ( $/V_1/ \rightarrow [\emptyset]$ ). The presence of such a process in the phonology of the language supports the reverse inference of the same vowel in the phonemic representation when the phonetic signal has nothing (reverse inference:  $[\emptyset] \rightarrow /V_1/$ ).<sup>9</sup> 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  $/V_1/$ , which maps to  $[\emptyset]$  in the surface/acoustic signal. 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 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 abovementioned characteristics

<sup>9</sup> The presence of a vowel-deletion process specifically targeting 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.



For example, when a Korean listener hears a nonsense word such as [ec<sup>h</sup>ima], the surface consonant [c<sup>h</sup>] is consistent with the reverse inference of either /t<sup>h</sup>/ or /c<sup>h</sup>/; thus the inferred phonemic parses for the nonsense word could be either /et<sup>h</sup>ima/ or /ec<sup>h</sup>ima/. As proposed above, inferences about the phonemic representations of the presented nonsense words affect 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<sup>h</sup>] as the first consonant of an illicit syllable context (e.g. [ec<sup>h</sup>ma]), the quality of the illusory vowel is modulated by the reverse inference about the phoneme that corresponds to the surface pronunciation [c<sup>h</sup>] in the nonsense word; if the perceptual system infers the phoneme to be a palatal stop /c<sup>h</sup>/, the /i/ vowel (which we refer to as illusory vowel 1) is induced, for the reasons mentioned above; however, if the perceptual system infers the phoneme to be an alveolar stop /t<sup>h</sup>/, then /i/ (illusory vowel 2) is induced in the illicit syllable context, because the only way that phonetic [c<sup>h</sup>] can result from the phoneme /t<sup>h</sup>/ is if it is followed by /i/. Given this, we expect that the *same* illicit palatal coda can induce *both* an illusory /i/ and an illusory /i/. When an alveolar segment, [t<sup>h</sup>] or [s], is encountered in the acoustic token, there is only one possible phonemic parse, the same alveolar phoneme, /t<sup>h</sup>/ or /s/ respectively, as in (2a). In an illicit syllable context, /i/ (illusory vowel 1) is perceived, as shown above. Finally, when a palatal fricative, [ʃ], is encountered in the acoustic token, there is only one possible phonemic parse, the alveolar fricative /s/, as in (2b). However, if an alveolar fricative (/s/) is the inferred phoneme, then /i/ (illusory vowel 2) is perceived 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/.

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 the alveolar consonants [t<sup>h</sup> s], we predict the illusory vowel to be /i/, in those following the palatal stop [c<sup>h</sup>], we predict the possibility of *both* /i/ and /i/, and in those following the palatal fricative [ʃ], we predict only the vowel /i/.

In clarification of our position, we would like to note that though we predict the possibility of *both* /i/ and /i/ as illusory vowels for Korean listeners in the relevant palatal context [c<sup>h</sup>], we do not think that both the illusory vowels are *simultaneously* perceived in the same nonce word by a Korean listener. It is possible that for a single auditory input, two *separate* nonce-word 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 the lexical frequencies of the relevant phonemes.<sup>13</sup> It is also possible that for any single

<sup>13</sup> 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 phonemic representational candidates. An analysis along these lines also allows one to better understand

presentation of an auditory input only a single percept is inferred in a probabilistic way.

In the following sections, we present the results of three experiments involving identification and discrimination tasks with Korean subjects, with American English subjects as controls to ensure that the differences in the acoustic tokens are not what are driving the perceptual epenthesis effects observed with the Korean subjects. Three different paradigms – an AX task (Experiment 1), an ABX task (Experiment 2) and an identification task (Experiment 3) – were used to ensure that the effects are not artefacts of a particular 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 this paradigm, if listeners perceive an illusory vowel /i/ between consonants in a cluster [sm], for instance, they will find it difficult to distinguish between [esma] and [es<sub>i</sub>ma].<sup>14</sup> Crucially, as claimed in the previous section, we expect that Korean listeners will have much more difficulty than English listeners in distinguishing the following two sets of stimulus pairs: (a) [et<sup>h</sup>ima–et<sup>h</sup>ma], [es<sub>i</sub>ma–esma], [e<sup>h</sup>ima–e<sup>h</sup>ma], (b) [e<sup>h</sup>ima–e<sup>h</sup>ma], [ef<sub>i</sub>ma–efma]. In set (a), the Korean listeners are likely to perceive an illusory /i/ in the second stimulus in each pair ([et<sup>h</sup>ma, esma, e<sup>h</sup>ma]); therefore, for the Korean listeners the pairs in (a) should be more confusable than for 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/ in the second stimulus of each pair ([e<sup>h</sup>ma, efma]); for the same reason, the pairs should again be more confusable for the Korean listeners.

### 2.1 Method

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

---

why the illusory vowel rates are never at ceiling in such experiments. Thanks to an anonymous reviewer for raising this possibility.

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

2.1.2 *Stimuli.* The experimental stimuli consisted of those that were relevant for this article and those that were relevant for another independent hypothesis (see Appendix: Table II for the full list). Thirty nonce words of the form [eC<sub>1</sub>V<sub>1</sub>C<sub>2</sub>] were used, in which C<sub>1</sub> was alveolar, palatal or labial [t<sup>h</sup> d s c<sup>h</sup> ʃ b m], V<sub>1</sub> was [ɪ i Ø] and C<sub>2</sub> was a labial stop or nasal [p<sup>h</sup> m]. None of the stimuli were words in either Korean or in English. They had stress on the first vowel, and were naturalistic 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. There were two reasons for the use of this particular speaker. Firstly, he could naturally produce all the stimuli, as they are phonotactically licit in his dialects of both Hindi and Telugu. The use of a native Korean speaker to record the stimuli would have only been possible if the speaker had neutralised their own linguistic biases, as many of the sequences are not licit in the language. 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. Secondly, the use of an English speaker to record the stimuli was also avoided, because those that we tried had difficulty in producing unstressed medial vowels that were unreduced (i.e. they couldn't block the vowel-reduction process in their dialect). Furthermore, we did not want to introduce a bias that would help the control group, as the phonetic patterns would have been more natural for the English listeners than for the Korean listeners. The interpretation of the results would therefore have been confounded by this. 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 was naturalistic and consistent between stimuli.

Each item was recorded using Praat (Boersma & Weenink 2012), with a Logitech USB desktop microphone (frequency response 100 Hz–16 KHz) at a 44 KHz 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 smooth ramping.

Table I shows all the clusters and the test items relevant to the current paper. All of the test items without intervening vowels, [et<sup>h</sup>ma, esma, ec<sup>h</sup>ma, efma], had an illicit coda in Korean, and the clusters were also all illicit linear sequencss, so that issues regarding the distinction between syllable-structure violation and surface phonotactic violation did not arise (Kabak & Idsardi 2007). As all the clusters violated both types of phonotactic constraints, they were expected to trigger perceptual epenthesis.

2.1.3 *Procedure.* Following Kabak & Idsardi (2007) and Monahan *et al.* (2009), an AX discrimination (i.e. same/different) task was used to

		vowels		
		[i]	[i]	∅
alveolar	et <sup>h</sup> ima	et <sup>h</sup> ima	et <sup>h</sup> ma	
	esima	esima	esma	
palatal	ec <sup>h</sup> ima	ec <sup>h</sup> ima	ec <sup>h</sup> ma	
	efima	efima	efma	

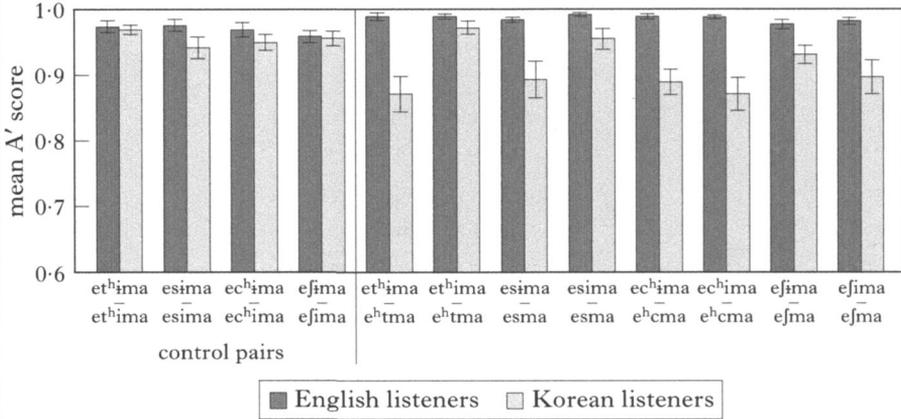
*Table 1*  
Test tokens in Experiment 1.

investigate the perceptual epenthesis effect. We tested all combinations of the vowels [i i ∅]. Thus, for the cluster [sm], the word-pairs were [esima–esima], [esima–esma], [esima–esma], [esima–esima], [esima–esima] and [esma–esma]. The word-pairs with different intervening vowels, such as [esima–esima], served as controls and were expected to be successfully distinguished by all participants.

Two recordings were 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<sub>1</sub>–esma<sub>1</sub>], [esima<sub>1</sub>–esma<sub>2</sub>], [esima<sub>2</sub>–esma<sub>1</sub>], [esima<sub>2</sub>–esma<sub>2</sub>], and an additional four ‘different’ word-pairs in reverse order. All combinations of ‘same’ word-pairs were also presented. For instance, in the case of [esima], there were four ‘same’ word-pairs: [esima<sub>1</sub>–esima<sub>1</sub>], [esima<sub>1</sub>–esima<sub>2</sub>], [esima<sub>2</sub>–esima<sub>1</sub>], [esima<sub>2</sub>–esima<sub>2</sub>]. Each of the above word-pairs was presented twice, giving 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 stimuli word-pairs to 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 nine trials with another set of nonce words, [emima], [emima] and [emma], which were not used in the actual experiment.<sup>15</sup> Both the interstimulus interval and the intertrial interval were 1000 ms. All the trials were randomised for each participant. The subjects were allowed to take a break after

<sup>15</sup> 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 it from [emima] 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-test debriefing session, they consistently stated that both the practice task and the actual experiment were very straightforward.



*Figure 1*  
 Mean A' (discriminability) values for English and Korean listeners in Experiment 1. Error bars represent standard errors.

every 240 trials (roughly every 15 minutes); thus there were two breaks. Each subject took approximately 45 minutes to complete the experiment.

**2.2 Results**

As in Kabak & Idsardi (2007) and Monahan *et al.* (2009), we took poorer discriminability between word-pairs with and without vowels, indicated by lower 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 non-parametric measure of discriminability that takes into account response bias (Pollack & Norman 1964, Macmillan & Creelman 2005). A' is presented instead of its parametric counterpart, 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 liable 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 Appendix: Table III for the values). The A' scores for the control [i-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.

Statistical analyses were conducted using SPSS. As Mauchly's test showed that the assumption of sphericity was violated for the main effects of Word-pair ( $\chi^2(65) = 326.528, p < 0.001$ ), degrees of freedom were corrected using Greenhouse-Geisser ( $\epsilon = 0.330$ ). A mixed ANOVA of A' scores revealed a main effect of Language ( $F(1,37) = 16.042, p < 0.001, \eta_p^2 = 0.302$ ), a main effect of Word-pair ( $F(3.634, 134.460) = 5.020,$

$p=0.001$ ,  $\eta_p^2=0.119$ ) and an interaction of Word-pair by Language ( $F(3.634,134.460)=7.809$ ,  $p<0.001$ ,  $\eta_p^2=0.174$ ). The Korean listeners therefore achieved significantly lower  $A'$  scores than the English listeners for some word-pairs, but not others.

In order to investigate for which word-pairs the performance of the two groups of listeners differed statistically from the control [i-i] pairs, we ran repeated measures ANOVAs to compare  $A'$  values for 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 Appendix: Table IV for the ANOVA 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 for [et<sup>h</sup>ima-et<sup>h</sup>ma] were compared against the control  $A'$  scores, there was a main effect of Word-pair ( $F(1,37)=6.992$ ,  $p=0.012$ ,  $\eta_p^2=0.159$ ), a main effect of Language ( $F(1,37)=14.212$ ,  $p=0.001$ ,  $\eta_p^2=0.278$ ) and a significant interaction between Word-pair and Language ( $F(1,37)=15.594$ ,  $p<0.001$ ,  $\eta_p^2=0.297$ ). On the other hand, in the comparison between [et<sup>h</sup>ima-et<sup>h</sup>ma] and the controls for the two language groups, there was a main effect of Word-pair ( $F(1,37)=10.169$ ,  $p=0.003$ ,  $\eta_p^2=0.216$ ), no main effect of Language ( $F(1,37)=2.144$ ,  $p=0.152$ ,  $\eta_p^2=0.055$ ) and no interaction between Word-pair and Language ( $F(1,37)=0.101$ ,  $p=0.752$ ,  $\eta_p^2=0.003$ ). This suggests that, compared to the English listeners, the Korean listeners performed significantly worse on [et<sup>h</sup>ima-et<sup>h</sup>ma] than on the control pairs, but not on [et<sup>h</sup>ima-et<sup>h</sup>ma].

A similar pattern was observed when the  $A'$  scores of [esima-esma] and [esima-esma] were compared against the control  $A'$  scores. When the two listener groups'  $A'$  scores of [esima-esma] were compared against the control  $A'$  scores, there was no main effect of Word-pair ( $F(1,37)=3.211$ ,  $p=0.081$ ,  $\eta_p^2=0.08$ ), but there was a main effect of Language ( $F(1,37)=8.566$ ,  $p=0.006$ ,  $\eta_p^2=0.188$ ) and an interaction between Word-pair and Language ( $F(1,37)=7.131$ ,  $p=0.011$ ,  $\eta_p^2=0.162$ ). In contrast, when [esima-esma] were compared to controls for the two language groups, there was a main effect of Word-pair ( $F(1,37)=5.581$ ,  $p=0.024$ ,  $\eta_p^2=0.131$ ), but not of Language ( $F(1,37)=3.794$ ,  $p=0.059$ ,  $\eta_p^2=0.093$ ), and no interaction between Word-pair and Language ( $F(1,37)=3.484$ ,  $p=0.07$ ,  $\eta_p^2=0.086$ ). In summary, for word-pairs with an alveolar cluster type, the Korean listeners were significantly worse than the English listeners for [et<sup>h</sup>ima-et<sup>h</sup>ma] and [esima-esma] compared to the control pairs, but not for [et<sup>h</sup>ima-et<sup>h</sup>ma] and [esima-esma].

When the two groups'  $A'$  scores for [ech<sup>h</sup>ima-ech<sup>h</sup>ma] were compared against the control  $A'$ , there was a main effect of Word-pair ( $F(1,37)=10.031$ ,  $p=0.003$ ,  $\eta_p^2=0.213$ ), a main effect of Language ( $F(1,37)=15.977$ ,  $p<0.001$ ,  $\eta_p^2=0.302$ ) and an interaction between Word-pair and Language ( $F(1,37)=27.428$ ,  $p<0.001$ ,  $\eta_p^2=0.426$ ). Furthermore, the same pattern was found when the  $A'$  scores for [ech<sup>h</sup>ima-ech<sup>h</sup>ma] were compared against the controls. There was a main effect of Word-pair

( $F(1,37) = 8.221$ ,  $p = 0.007$ ,  $\eta_p^2 = 0.182$ ), a main effect of Language ( $F(1,37) = 17.668$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.323$ ) and a significant interaction between Word-pair and Language ( $F(1,37) = 15.563$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.296$ ). Therefore, in comparison with the control pairs, the Korean listeners had significantly lower  $A'$  scores than the English listeners for both [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma] and [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma].<sup>16</sup>

For the comparison of  $A'$  scores for [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma] and the controls, there was a main effect of Language ( $F(1,37) = 7.301$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.165$ ); however, there was neither a main effect of Word-pair ( $F(1,37) = 0.900$ ,  $p = 0.349$ ,  $\eta_p^2 = 0.024$ ) nor an interaction between Word-pair and Language ( $F(1,37) = 3.188$ ,  $p = 0.082$ ,  $\eta_p^2 = 0.079$ ). In contrast, for the comparison between [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma] and the controls, there was no main effect of Word-pair ( $F(1,37) = 3.929$ ,  $p = 0.055$ ,  $\eta_p^2 = 0.096$ ); however, there was a main effect of Language ( $F(1,37) = 8.619$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.189$ ) and a significant interaction between Word-pair and Language ( $F(1,37) = 8.371$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.184$ ). Thus, for word-pairs with a [j], the Korean listeners performed significantly worse than the English listeners on [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma] compared to the controls, but not on [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma].

As the above statistical analysis shows, the Korean listeners were, as predicted, significantly worse than the English listeners at discriminating the word-pairs [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma], [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma], [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma], [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma] and [e<sup>h</sup>i<sup>h</sup>ma–e<sup>h</sup>ma], as compared to the control [i–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, as would be expected on the basis of the phonological processes of vowel deletion and palatalisation in Korean. However, given the somewhat high  $A'$  values for all pairs in Experiment 1, it is possible that the experimental results are actually the result of a more phonetic listening mode.<sup>17</sup> But it is unclear what set of hypotheses of phonetic perception would result in this particular pattern of differences between the English and Korean speakers. A more reasonable explanation, we think, is that the observed differences are *smaller* as a result of the ease of the 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,<sup>18</sup> in Experiment 2 we ran an ABX task in which listeners were presented with three stimuli, and asked to

<sup>16</sup> Visual inspection of the data showed that both the illusory vowels were perceived with [e<sup>h</sup>ma<sub>1</sub>] and with [e<sup>h</sup>ma<sub>2</sub>].

<sup>17</sup> Thanks to two anonymous reviewers for pointing out this possibility and suggesting the use of an ABX task.

<sup>18</sup> Actually, the evidence for this view is in our opinion rather weak. We refer the reader to Boomershine *et al.* (2008) for more discussion.

decide whether the first or the second stimulus was more similar to the third stimulus. The ABX task is much more memory-intensive and is therefore typically viewed as encouraging higher-level or phonological listening (Gerrits & Schouten 2004). As discussed in relation to Experiment 1, we expect that Korean listeners should have much more difficulty than English listeners in distinguishing the following two sets of stimulus pairs: (a) [et<sup>h</sup>ima-et<sup>h</sup>ma], [esima-esma], [eç<sup>h</sup>ima-eç<sup>h</sup>ma], (b) [eç<sup>h</sup>ima-eç<sup>h</sup>ma], [ejima-efma].

### 3.1 Method

3.1.1 *Participants.* Seventeen native Korean speakers (9 men, 8 women; age 20–31) and 17 native English speakers (2 men, 15 women; age 19–22) participated in the experiment. All the subjects were recruited at Michigan State University, and reported that they had normal hearing. None of the Korean speakers had come to the U.S.A. or visited other English-speaking countries before the age of 13, nor had they lived in English-speaking countries for more than four years.

3.1.2 *Stimuli.* The stimuli for Experiment 2 were the same twelve test items used in Experiment 1, as described in Table I above.

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 the vowels [i i Ø]. For example, for the cluster [sm], the AB word-pairs were [esima-esma], [esima-esma] and [esima-esima]. Two recordings were used for each item, as in Experiment 1, and the order of tokens in each AB word-pair was counterbalanced. For instance, in the case of [esima-esma], there were four AB word-pairs, [esima<sub>1</sub>-esma<sub>1</sub>], [esima<sub>1</sub>-esma<sub>2</sub>], [esima<sub>2</sub>-esma<sub>1</sub>] and [esima<sub>2</sub>-esma<sub>2</sub>], and an additional four word-pairs in reverse order. Either A or B was added as an X to each of these AB word-pairs. When adding Xs, the same token was never repeated in a single trial. Therefore, in the case of [esima-esma], there were four ABA word-triplets [esima<sub>1</sub>-esma<sub>1</sub>-esima<sub>2</sub>], [esima<sub>1</sub>-esma<sub>2</sub>-esima<sub>2</sub>], [esima<sub>2</sub>-esma<sub>1</sub>-esima<sub>1</sub>], [esima<sub>2</sub>-esma<sub>2</sub>-esima<sub>1</sub>], and an additional four ABB word-triplets [esima<sub>1</sub>-esma<sub>1</sub>-esma<sub>2</sub>], [esima<sub>1</sub>-esma<sub>2</sub>-esma<sub>1</sub>], [esima<sub>2</sub>-esma<sub>1</sub>-esma<sub>2</sub>], [esima<sub>2</sub>-esma<sub>2</sub>-esma<sub>1</sub>]. The same permutations were used for the other clusters ([t<sup>h</sup>m ç<sup>h</sup>m f<sup>m</sup>]), giving a total of 192 trials in the experiment.

The experiment was conducted in a quiet room, with a group of at most four participants per session. The stimuli were presented to each participant as an ABX task scripted in Praat with a low-noise headset (Plantronics SupraPlus HW261). The participants were asked to listen to word-triplets, and to 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. 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 ('세 번째

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 twelve trials with another set of nonce words. The interstimulus interval was 500 ms and the intertrial interval was 1500 ms. All 192 trials were randomised for each participant. The subjects were allowed to take a break after half of the trials; the experiment took about 17 minutes.

### 3.2 Results

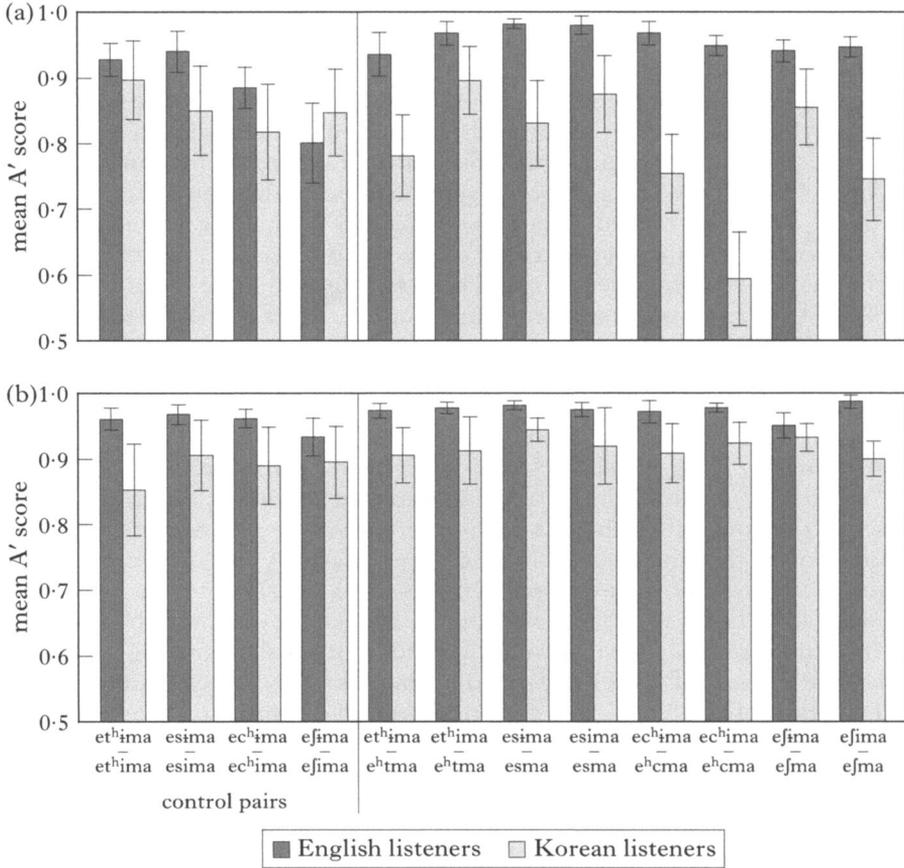
As in Experiment 1, we calculated  $A'$  as a measure of perceptual epenthesis. Figures 2a and b show the mean  $A'$  values for English and Korean listeners for all the word-pairs for ABA and ABB orders respectively (see Appendix: Tables V and VI for the individual values). Overall, the figures illustrate that the English listeners have higher  $A'$  values than Korean listeners. Interestingly, both the English and Korean listeners seem to have higher  $A'$  scores for the ABB order than for the ABA order.

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-subjects variable. The three-way mixed ANOVA for  $A'$  scores revealed that there was an effect of Language ( $F(1,32) = 4.377, p = 0.044, \eta_p^2 = 0.120$ ). There was a main effect of Word-pair ( $F(5.335,170.713) = 2.764,^{19} p = 0.018, \eta_p^2 = 0.079$ ) and an interaction of Word-pair by Language ( $F(5.335,170.713) = 2.992, p = 0.011, \eta_p^2 = 0.086$ ). There was also a main effect of Order with a very large effect size ( $F(1,32) = 24.476, p < 0.001, \eta_p^2 = 0.433$ ), and an interaction of Order by Language ( $F(1,32) = 5.74, p = 0.022, \eta_p^2 = 0.153$ ). There was an interaction of Word-pair by Order ( $F(5.619,179.798) = 3.217, p = 0.006, \eta_p^2 = 0.091$ ) and a three-way interaction between Word-pair, Order and Language ( $F(5.619,179.798) = 3.725, p = 0.002, \eta_p^2 = 0.104$ ).

As Order had a main effect with a very large effect size, participants' responses for ABA and ABB orders were analysed separately, using two two-way mixed ANOVAs, with Word-pair as a within-subject variable and Language as a between-subjects variable. A two-way mixed ANOVA for the ABA order revealed a main effect of Language ( $F(1,32) = 5.410, p = 0.027, \eta_p^2 = 0.145$ ), a main effect of Word-pair ( $F(5.350,171.214) = 4.056, p = 0.001, \eta_p^2 = 0.112$ ) and an interaction between Word-pair and Language ( $F(5.350,171.214) = 4.783, p < 0.001, \eta_p^2 = 0.130$ ). However, a two-way mixed ANOVA for the ABB order did not find a significant main effect of Language ( $F(1,32) = 2.643, p = 0.114$ ) or Word-pair ( $F(3.558,113.847) = 0.852, p = 0.485$ ), or an interaction between Word-pair and Language ( $F(3.558,113.847) = 0.479, p = 0.729$ ).

As only the ABA order showed a main effect of Language and an interaction between Word-pair and Language, follow-up planned comparisons

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



*Figure 2*  
 Mean A' (discriminability) values for English and Korean listeners in Experiment 2: (a) for the ABA order; (b) for the ABB order. Error bars represent standard errors.

were conducted on the English and Korean listeners' responses for the ABA order (see Appendix: Table V for the *t*-test results). The results showed that there was no significant difference between English and Korean listeners in the control Word-pairs with different vowels ( $t(32) = 0.475$ ,  $p = 0.638$  for [et<sup>h</sup>ima–et<sup>h</sup>ima];  $t(32) = 1.199$ ,  $p = 0.239$  for [esima–esima];  $t(21.580) = 0.852$ ,  $p = 0.404$  for [ech<sup>h</sup>ima–ech<sup>h</sup>ima];  $t(32) = -0.504$ ,  $p = 0.618$  for [efima–efima]). Among the test word-pairs, the English and Korean listeners were significantly different only for the predicted word-pairs ( $t(32) = 2.217$ ,  $p = 0.034$  for [et<sup>h</sup>ima–et<sup>h</sup>ma];  $t(16.379) = 2.292$ ,  $p = 0.035$  for [esima–esma];  $t(19.003) = 3.444$ ,  $p = 0.003$  for [ech<sup>h</sup>ima–ech<sup>h</sup>ma];  $t(21.664) = 4.577$ ,  $p < 0.001$  for [ech<sup>h</sup>ima–ech<sup>h</sup>ma];  $t(17.724) = 3.105$ ,  $p = 0.006$  for [efima–efma]). The two Language groups

were not significantly different for the rest of the word-pairs ( $t(32) = 1.310$ ,  $p = 0.199$  for [et<sup>h</sup>ima–et<sup>h</sup>ma];  $t(17.854) = 1.708$ ,  $p = 0.105$  for [esima–esma];  $t(18.708) = 1.409$ ,  $p = 0.175$  for [ejima–ejma]).

To summarise the results of the ABX task in Experiment 2, 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 for the ABA order.<sup>20</sup> The effect of Order could be explained by the fact that comparison to the second member of the triplet is affected 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 the Korean listeners perceived an /i/ between consonants in the [t<sup>h</sup>m] and [sm] clusters. They also reported that they heard both /i/ and /i/ in [c<sup>h</sup>m], but /i/ in [jm]. 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 in comparison to the rest of the word-pairs, both with and without a medial vowel, which seems to reflect that they may have been influenced by English 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 for both Experiment 1 and the ABA order of Experiment 2 were very unlikely to be due to a more phonetic perception mode or to stimuli artefacts. If this had been the case, then the same pattern of results should have been observable in the ABB results.

<sup>20</sup> 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, though it is possible to view both the ABB and the AX orders as involving local comparisons of identical stimuli. At this point, we can only speculate about the possible reasons for this. First, while it is true that the ABB order does have 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 any 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 interstimulus intervals in the experiments are substantially different for the two experiments (AX = 1000 ms; ABX = 500 ms), which means that adjacent stimuli in the ABX task might have been more affected by phonetic 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 interstimulus intervals, participants still have access to fine-grained auditory representations in their short-term memory (Pisoni 1973), which aids them in the task.

## 4 Experiment 3

Experiment 2 showed that the results of the ABX task also followed our predictions, demonstrating the same patterns as the results of the AX task in Experiment 1. However, a potential problem with AX and ABX tasks is that the locus of the difference perceived by the listener is unclear. For example, if the listener distinguishes the two stimuli [et<sup>h</sup>ima–et<sup>h</sup>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 on the basis of the presence/absence of the medial vowel, or on the basis of any other changes that they might have perceived in the consonants. More specifically, it is possible that the Korean listeners in Experiment 1 had a higher discriminability for the pair [et<sup>h</sup>ima–et<sup>h</sup>ma] than for [et<sup>h</sup>ima–et<sup>h</sup>ma] in comparison to English listeners because the first pair involves a case of ‘perceptual palatalisation’, where the [t<sup>h</sup>] before [i] was perceived as a palatal consonant, i.e. [et<sup>h</sup>ima] was perceived as /e<sup>h</sup>ima/. Therefore, the pair in which both the first consonantal and the medial vowel were perceived to be different might have been discriminated better than that with just the presence *vs.* absence of a vowel.

For this reason, in Experiment 3 we decided to run a task in which listeners heard a stimulus, and had to decide whether there was a vowel between the two consonants, and if there was a vowel, what 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 to response bias, selective attention focused on particular parts of the stimuli, and the effect these have on auditory coding (Caporello Bluvás & Gentner 2013). Despite these concerns, it is useful to run an identification task, as it can give us yet another perspective into what is happening during the perception of the relevant stimuli.

Following the view of perception laid out in §1, we expect the Korean listeners, unlike the English listeners, to hear illusory vowels in two sets of stimuli; (a) in the stimuli [et<sup>h</sup>ma], [e<sup>h</sup>ma] and [esma], we expect the Korean listeners to hear the illusory vowel /i/, (b) in the stimuli [e<sup>h</sup>ma] and [efma], we expect them 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 twelve test items used in Experiments 1 and 2 (see Table I above). There were two recordings used for each item as in Experiments 1 and 2, and they were each presented twice; there were therefore four tokens of each item, and a total of 48 tokens in the experiment.

4.1.3 *Procedure.* We used an identification task to investigate the perceptual epenthesis effect. The experiment was conducted in a quiet room, with a group of at most four participants per session. Experiment 3 drew participants' attention to the medial vowel in the stimuli. Therefore, it was conducted after Experiment 2 (after a short break), so as not to have the participants focus only on the vowel in Experiment 2. The stimuli were presented to each participant as 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  $\emptyset$ , and click on the corresponding box on the screen (the actual choices were 'u', 'i' and 'nothing' for the English listeners, and '으', '이' and '없음' for the Korean listeners).<sup>21</sup> 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'). Before the actual experiment, each participant completed a practice session, to ensure familiarity with the task. The practice session had nine trials with a different set of nonce words. The intertrial interval was 1000 ms. All 48 trials were randomised for each participant.

## 4.2 Results

Participants had to determine whether the medial vowel in a stimulus was [ɨ], [i] or  $\emptyset$ . The mean percentage of vowel responses to the stimuli can be found in Table VII in Appendix. Figure 3 gives percentages of vowel responses (i.e. [ɨ i  $\emptyset$ ]) for the [eCma] stimuli (where C = consonant). The figure shows that the English listeners in general correctly identified the absence of the vowels in all cases, whereas the Korean listeners generally identified an [ɨ] in [et<sup>h</sup>ma] and [esma], and an [i] in [ec<sup>h</sup>ma] and [efma]. Korean listeners also identified an [ɨ] for [ec<sup>h</sup>ma].

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. [et<sup>h</sup>ma, esma, ec<sup>h</sup>ma, efma]), with Response (i.e. [ɨ i  $\emptyset$ ]) as a within-subject variable and Language (i.e. Korean, English) as a between-subjects variable (see Appendix: Table VIII for the ANOVA results). For [et<sup>h</sup>ma], there was a main effect of Response ( $F(1,32) = 40.403$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.558$ ), and an interaction between Response and Language ( $F(1,32) = 67.814$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.679$ ).

<sup>21</sup> Thanks to an anonymous reviewer for raising an important point for any researcher working with English orthography in behavioural experiments. In Experiment 3, we used ⟨u⟩ as the letter to represent /u/, 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 /u/. 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 /u/. Finally, that the English listeners in Experiment 3 had no problem associating ⟨u⟩ with /u/ is further supported by the fact that the average identification rates of ⟨u⟩ in stimuli with the /u/ counterpart in the test items (i.e. [et<sup>h</sup>ɪma, esɪma, efɪma, ec<sup>h</sup>ɪma]) was about 96% (Appendix: Table VII).

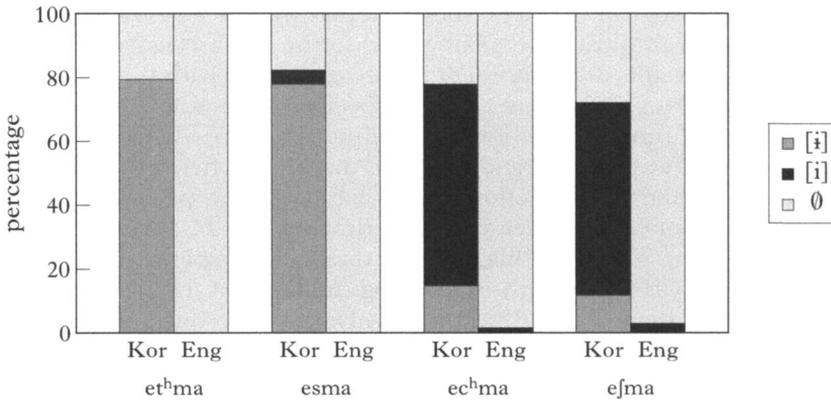


Figure 3

Percentages of vowel responses for [eCma] stimuli.

For [esma], there was a main effect of Response ( $F(1.114, 35.650) = 42.398$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.570$ ), and an interaction between Response and Language ( $F(1.114, 35.650) = 82.694$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.721$ ). For [e<sup>h</sup>ma], there were main effects of Language ( $F(1, 32) = 10.667$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.250$ ) and Response ( $F(1.575, 50.410) = 22.884$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.417$ ), and an interaction between Response and Language ( $F(1.575, 50.410) = 41.937$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.567$ ). For [efma], there was a main effect of Response ( $F(2, 64) = 32.667$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.505$ ), and an interaction between Response and Language ( $F(2, 64) = 41.692$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.566$ ). To summarise the results of the four mixed ANOVAs, English and Korean listeners responded differently for all four stimuli.

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 conformed to the predictions (see Appendix: Table IX for the results). For [e<sup>t</sup><sup>h</sup>ma], Korean listeners identified [ɨ] more often than English listeners ( $t(16) = 8.235$ ,  $p < 0.001$ ), but none of the English or Korean listeners identified an [i]. For [esma], there was a group difference in [ɨ] identification ( $t(16) = 9.123$ ,  $p < 0.001$ ), but not in [i] identification ( $t(16) = 1.852$ ,  $p = 0.083$ ). When presented with [e<sup>h</sup>ma], Korean listeners identified [i] more than English listeners ( $t(16) = 5.886$ ,  $p < 0.001$ ), and they also showed a marginally significant increase in [ɨ] compared to English listeners ( $t(17.658) = 2.000$ ,  $p = 0.061$ ). When presented with [efma], Korean listeners identified [i] more often than English listeners ( $t(19.938) = 6.500$ ,  $p < 0.001$ ); however, there was no statistical group difference in [ɨ] identification ( $t(16) = 1.646$ ,  $p = 0.119$ ).

In summary, the results of the identification task in Experiment 3 showed that the Korean listeners perceived an illusory /ɨ/ more often in [e<sup>t</sup><sup>h</sup>ma] and [esma] than the English listeners. In [e<sup>h</sup>ma], the Korean listeners perceived both illusory /ɨ/ and /i/, and in [efma], they perceived an illusory /i/ more

often than the English listeners; they also perceived a statistically non-significant number of illusory /i/'s compared to the English listeners.

Overall, the results of Experiment 3 were consistent with the expectations laid out above. However, there are two aspects of the results that need more discussion and future exploration. First, although we predicted that Korean listeners would hear both /i/ and /i/ in [e<sup>h</sup>ma] more often than the English listeners, we made no predictions about which of the two would be perceived more frequently. At least in Experiments 2 and 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. Secondly, there is also a small, but non-significant, number of cases of the perception of illusory /i/ in [ɛfma] in Experiment 3. Again, it is unclear if this is due to facts about the auditory coding of segments that are not distinct phonemes. Perhaps the auditory segment [ɲ] is more likely to be coded as [s] (i.e. the more general member of the phonemic pair) because the focus on the medial vowel hampers the coding of adjacent consonants.<sup>22</sup> A second possibility is that the vowel /i/ is a more *default* illusory vowel in Korean, given its participation in general vowel-deletion processes and its being the shortest vowel in the language. A third interesting explanation suggested to us by an anonymous reviewer is the possibility of there being an illusory /j/ after the phoneme inferred for [ɲ] (since palatalisation in Korean is also triggered before the palatal glide /j/), and consequently an illusory /i/ after the /j/, thereby sometimes resulting in the phonemic percept /sjim/ for [ɲm]. This third account is 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 perceptual illusion of the vowel /i/ was least in the context of [ɲ] (and was somewhat inconsistent in all three experiments), suggesting that the locus of the explanation for this particular effect might be different from the ones we have been discussing in this article. Again, none of these possibilities detracts from the predictions in the current paper, but they do suggest very interesting directions for further inquiry.

## 5 Discussion

In this paper we have shown 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 have shown that it is possible to obtain more than one illusory vowel in the same language, and even in the same context, as

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

long as the phonology of the language and the acoustic tokens themselves motivate such a reanalysis of the illicit sequences. The phonological processes of vowel deletion and palatalisation 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 (i.e. [et<sup>h</sup>ma] and [esma]), the phonological alternations in the language lead us to expect /i/ as the illusory vowel; in consonant contexts where the first (coda) consonant is an aspirated palatal stop consonant ([e<sup>h</sup>ma]), we expect either /ɨ/ or /i/; and finally, in consonant contexts where the first (coda) consonant is a palatal fricative consonant ([e<sup>f</sup>ma]), we expect /i/. We have shown 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, Dupoux *et al.*'s (2011) approach predicts that the illusory vowel in Korean will be /ɨ/, perhaps due to its phonetically minimal characteristics. However, this does not account for the specific patterns of illusory vowels observed in the data. If this were the hypothesis, it is unclear why [c<sup>h</sup>m] and [ʃm] would trigger an illusory /i/ for Korean listeners. However, 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 purely on the basis of surface phonotactic patterns in the language. It is true that illusory vowels were perceived by the Korean listeners in phonotactically illicit sequences \*[t<sup>h</sup>m c<sup>h</sup>m 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 [ʃm] context could be alternatively explained by surface phonotactic constraints that ban [ʃ] from being followed by any vowel except [i] in Korean (since only [ʃi] is possible in Korean). Similarly, one could also account for the absence of the illusory vowel /i/ in the [sm] context by appealing to a surface phonotactic constraint banning \*[si]. However, attempting to account for all the illusory vowels observed in this paper using purely surface phonotactics is problematic for a number of reasons. Firstly, the account proposed for the absence of the illusory vowel /i/ in [sm] contexts does not by itself explain why some illusory vowel other than /ɨ/ is not inferred in the [sm] contexts (note that [sam sem som] are also possible sequences in Korean.<sup>23</sup> Secondly, on a similar note, the purely surface phonotactic account cannot explain why some vowel other than /i/ and /ɨ/ is not a possible illusory vowel in the [c<sup>h</sup>m]

<sup>23</sup> One could of course argue that Experiment 3 (involving the direct identification of the illusory vowel) did not include any of the other vowels. However, this 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.

context ([c<sup>h</sup>am c<sup>h</sup>em c<sup>h</sup>om] are possible sequences). Finally, and most importantly, the purely surface phonotactic account cannot explain why /i/ is not a possible illusory vowel in the [t<sup>h</sup>m] context, even though [t<sup>h</sup>im] is a possible sequence in Korean.<sup>24</sup> In parallel with the first two reasons, it also does not easily account for why there are no other possible illusory vowels. In contrast to the problems associated with a purely surface phonotactic account, the account based on phonological alternations laid out above 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), who attempted to obtain more than one illusory vowel for Japanese speakers. Based on the patterns of loanwords in Japanese such as [makuodonarudo] ‘McDonald’s’, it is possible to hypothesise that the illusory vowel adjacent to non-coronal consonants (e.g. [k g]) is /u/, and adjacent to coronal consonants (e.g. [t d]) /o/. However, as they show, while Japanese speakers confuse [eguma] 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 tested by Monahan *et al.* On our account, the only illusory vowel expected for the contexts they tested is /u/, as it is the shortest vowel in the language.<sup>25</sup> We further predict that there will be other illusory vowels in Japanese. Japanese has a process palatalising alveolar consonants before /i/ which is similar to Korean; the account proposed here therefore predicts that the set of illusory vowels triggered adjacent to illicit palatal codas in Japanese should include the vowel /i/.

Finally, this article has provided support for the view that speech perception involves the reverse inference to the phonemic representation level. Such a conception of speech perception, we believe, falls out quite naturally from a Bayesian perspective, and we therefore see it as offering support more generally for the Bayesian view of speech perception (Feldman & Griffiths 2007, Bever & Poeppel 2010, Sonderegger & Yu 2010, Poeppel & Monahan 2011, Yu 2011, Wilson & Davidson in press). Having said this, it is important to reiterate the point made earlier (note 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 phonemic representation level.

Finally, in line with some previous research on the topic (Huang 2001, Hume & Johnson 2003, Boomershine *et al.* 2008, Johnson & Babel 2010), the results of the current article show that speech perception is modulated

<sup>24</sup> As observed in note 11, the stop-palatalisation process is blocked within morphemes.

<sup>25</sup> Coronal stops in Japanese cannot be followed by [u] (\*[tu, du]). It is therefore clear that inferring the illusory vowel /u/ 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 in Japanese with [o], rather than any other vowel. 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.

not only by the acoustics of the speech 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.

**Appendix**

		vowels		
		[ɨ]	[i]	∅
illicit coda	alveolar	et <sup>h</sup> i <sup>h</sup> ma	et <sup>h</sup> ima	et <sup>h</sup> ma
		esima	esima	esma
		esip <sup>h</sup> a	esip <sup>h</sup> a	esp <sup>h</sup> a
	palatal	ec <sup>h</sup> i <sup>h</sup> ma	ec <sup>h</sup> ima	ec <sup>h</sup> ma
		ec <sup>h</sup> ip <sup>h</sup> a	ec <sup>h</sup> ip <sup>h</sup> a	ec <sup>h</sup> p <sup>h</sup> a
		ef <sup>h</sup> ima	ef <sup>h</sup> ima	ef <sup>h</sup> ma
licit coda	alveolar	edip <sup>h</sup> a	edip <sup>h</sup> a	edp <sup>h</sup> a
filler	labial	ebip <sup>h</sup> a	epip <sup>h</sup> a	ebp <sup>h</sup> a
		emip <sup>h</sup> a	emip <sup>h</sup> a	emp <sup>h</sup> a

*Table II*

Complete list of test tokens in Experiment 1. Items relevant to the experiment are given in Table I above.

pairs	English listeners		Korean listeners	
	<i>mean</i>	SE	<i>mean</i>	SE
et <sup>h</sup> i <sup>h</sup> ma–et <sup>h</sup> ima	0.974	0.009	0.969	0.007
et <sup>h</sup> i <sup>h</sup> ma–et <sup>h</sup> ma	0.989	0.006	0.871	0.027
et <sup>h</sup> ima–et <sup>h</sup> ma	0.989	0.004	0.972	0.010
esima–esima	0.976	0.009	0.942	0.016
esima–esma	0.984	0.004	0.893	0.028
esima–esma	0.992	0.003	0.955	0.016
ec <sup>h</sup> i <sup>h</sup> ma–ec <sup>h</sup> ima	0.969	0.011	0.950	0.012
ec <sup>h</sup> i <sup>h</sup> ma–ec <sup>h</sup> ma	0.989	0.004	0.889	0.019
ec <sup>h</sup> ima–ec <sup>h</sup> ma	0.988	0.003	0.871	0.025
ef <sup>h</sup> ima–ef <sup>h</sup> ima	0.959	0.009	0.956	0.011
ef <sup>h</sup> ima–ef <sup>h</sup> ma	0.977	0.007	0.931	0.014
ef <sup>h</sup> ima–ef <sup>h</sup> ma	0.982	0.005	0.897	0.026

*Table III*

Means and standard errors of A' values for English listeners (*n* = 19) and Korean listeners (*n* = 20) in Experiment 1.

		<i>F</i>	<i>df</i>	<i>p</i>	$\eta_p^2$
eth <sup>h</sup> ima–et <sup>h</sup> ma	Word-pair	6.992	1,37	0.012	0.159
	Language	14.212	1,37	0.001	0.278
	Word-pair × Language	15.594	1,37	<0.001	0.297
eth <sup>h</sup> ima–et <sup>h</sup> ma	Word-pair	10.169	1,37	0.003	0.216
	Language	2.144	1,37	0.152	0.055
	Word-pair × Language	0.101	1,37	0.752	0.003
esima–esma	Word-pair	3.211	1,37	0.081	0.080
	Language	8.566	1,37	0.006	0.188
	Word-pair × Language	7.131	1,37	0.011	0.162
esima–esma	Word-pair	5.581	1,37	0.024	0.131
	Language	3.794	1,37	0.059	0.093
	Word-pair × Language	3.484	1,37	0.070	0.086
ec <sup>h</sup> ima–ec <sup>h</sup> ma	Word-pair	10.031	1,37	0.003	0.213
	Language	15.977	1,37	<0.001	0.302
	Word-pair × Language	27.428	1,37	<0.001	0.426
ec <sup>h</sup> ima–ec <sup>h</sup> ma	Word-pair	8.221	1,37	0.007	0.182
	Language	17.668	1,37	<0.001	0.323
	Word-pair × Language	15.563	1,37	<0.001	0.296
ef <sup>h</sup> ima–ef <sup>h</sup> ma	Word-pair	0.900	1,37	0.349	0.024
	Language	7.301	1,37	0.010	0.165
	Word-pair × Language	3.188	1,37	0.082	0.079
ef <sup>h</sup> ima–ef <sup>h</sup> ma	Word-pair	3.929	1,37	0.055	0.096
	Language	8.619	1,37	0.006	0.189
	Word-pair × Language	8.371	1,37	0.006	0.184

*Table IV*

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

pairs	English listeners		Korean listeners		<i>t</i>	<i>df</i>	<i>p</i>
	<i>mean</i>	SE	<i>mean</i>	SE			
et <sup>h</sup> ima-et <sup>h</sup> ima	0.928	0.025	0.897	0.060	0.475	32	0.638
et <sup>h</sup> ima-et <sup>h</sup> ma	0.936	0.033	0.781	0.062	2.217	32	0.034
et <sup>h</sup> ima-et <sup>h</sup> ma	0.968	0.018	0.896	0.052	1.310	32	0.199
esima-esima	0.940	0.031	0.850	0.068	1.199	32	0.239
esima-esma	0.982	0.007	0.831	0.065	16.379	2.292	0.035
esima-esma	0.980	0.014	0.875	0.059	1.708	17.854	0.105
ech <sup>h</sup> ima-ech <sup>h</sup> ima	0.885	0.031	0.817	0.073	0.852	21.580	0.404
ech <sup>h</sup> ima-ech <sup>h</sup> ma	0.968	0.018	0.754	0.060	3.444	19.003	0.003
ech <sup>h</sup> ima-ech <sup>h</sup> ma	0.949	0.030	0.594	0.071	4.577	21.664	<0.001
efjima-efjima	0.801	0.061	0.847	0.066	-0.504	32	0.618
efjima-efjma	0.941	0.017	0.855	0.058	1.409	18.708	0.175
efjima-efjma	0.947	0.015	0.745	0.063	3.105	17.724	0.006

*Table V*

Means and standard errors of A' values and *t*-test results for English listeners (*n* = 17) and Korean listeners (*n* = 27) for the ABA order in Experiment 2.

pairs	English listeners		Korean listeners	
	<i>mean</i>	SE	<i>mean</i>	SE
et <sup>h</sup> ima-et <sup>h</sup> ima	0.961	0.017	0.853	0.070
et <sup>h</sup> ima-et <sup>h</sup> ma	0.974	0.011	0.906	0.042
et <sup>h</sup> ima-et <sup>h</sup> ma	0.978	0.009	0.913	0.051
esima-esima	0.968	0.015	0.906	0.054
esima-esma	0.982	0.007	0.944	0.018
esima-esma	0.975	0.011	0.919	0.058
ech <sup>h</sup> ima-ech <sup>h</sup> ima	0.962	0.014	0.890	0.059
ech <sup>h</sup> ima-ech <sup>h</sup> ma	0.972	0.017	0.909	0.045
ech <sup>h</sup> ima-ech <sup>h</sup> ma	0.978	0.007	0.924	0.032
efjima-efjima	0.934	0.029	0.895	0.055
efjima-efjma	0.951	0.019	0.933	0.021
efjima-efjma	0.987	0.010	0.900	0.027

*Table VI*

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

	English listeners			Korean listeners		
	i	i	∅	i	i	∅
et <sup>h</sup> ima	98.53 (1.47)	0 (0)	1.47 (1.47)	60.29 (10.51)	1.47 (1.47)	38.24 (10.74)
et <sup>h</sup> ima	0.00 (0)	100 (0)	0 (0)	0 (0)	94.12 (5.88)	5.88 (5.88)
et <sup>h</sup> ma	0 (0)	0 (0)	100 (0)	79.41 (9.64)	0 (0)	20.59 (9.64)
esima	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)
ech <sup>h</sup> ima	95.59 (3.21)	0 (0)	4.41 (3.21)	63.24 (9.62)	4.41 (3.21)	32.35 (9.29)
ech <sup>h</sup> ima	2.94 (2.94)	95.59 (3.21)	1.47 (1.47)	0 (0)	97.06 (2.94)	2.94 (2.94)
ech <sup>h</sup> ma	1.47 (1.47)	0 (0)	98.53 (1.47)	14.71 (6.45)	63.24 (10.74)	22.06 (8.81)
efjima	91.18 (5.23)	1.47 (1.47)	7.35 (4.68)	48.53 (10.16)	0 (0)	51.47 (10.16)
efjima	1.47 (1.47)	97.06 (2.01)	1.47 (1.47)	4.41 (3.21)	91.18 (4.26)	4.41 (3.21)
efjma	0 (0)	2.94 (2.94)	97.06 (2.94)	11.76 (7.15)	60.29 (8.32)	27.94 (7.69)

*Table VII*

Means and standard errors (in parentheses) of percentages of vowel responses in Experiment 3.

		<i>F</i>	<i>df</i>	<i>p</i>	$\eta_p^2$
et <sup>h</sup> ma	Language	0.000	1,32	1.000	0.000
	Response	40.403	1,32	<0.001	0.558
	Response × Language	67.814	1,32	<0.001	0.679
esma	Language	0.000	1,32	1.000	0.000
	Response	42.398	1.114,35.650	<0.001	0.570
	Response × Language	82.694	1.114,35.650	<0.001	0.721
ec <sup>h</sup> ma	Language	10.667	1,32	0.003	0.250
	Response	22.884	1.575,50.410	<0.001	0.417
	Response × Language	41.937	1.575,50.410	<0.001	0.567
efma	Language	0.000	1,32	1.000	0.000
	Response	32.667	2,64	<0.001	0.505
	Response × Language	41.692	2,64	<0.001	0.566

*Table VIII*  
Results of ANOVAs in Experiment 3.

	response	<i>t</i>	<i>df</i>	<i>p</i>
et <sup>h</sup> ma	ɪ	8.235	16	<0.001
esma	ɪ	9.123	16	<0.001
	i	1.852	16	0.083
ec <sup>h</sup> ma	ɪ	2.000	17.658	0.061
	i	5.886	16	<0.001
efma	ɪ	1.646	16	0.119
	i	6.500	19.938	<0.001

*Table IX*  
Results of planned comparisons of English and Korean listeners' responses in Experiment 3.

REFERENCES

Ahn, Sang-Cheol (1985). *The interplay of phonology and morphology in Korean*. PhD dissertation, University of Illinois at Urbana-Champaign.

Berent, Iris, Tracy Lennertz, Jongho Jun, Miguel A. Moreno & Paul Smolensky (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences* **105**. 5321–5325.

Berent, Iris, Tracy Lennertz, Paul Smolensky & Vered Vaknin-Nusbaum (2009). Listeners' knowledge of phonological universals: evidence from nasal clusters. *Phonology* **26**. 75–108.

- Berent, Iris, Donca Steriade, Tracy Lennertz & Vered Vaknin (2007). What we know about what we have never heard: evidence from perceptual illusions. *Cognition* **104**. 591–630.
- Best, Catherine T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In Judith C. Goodman & Howard C. Nusbaum (eds.) *The development of speech perception: the transition from speech sounds to spoken words*. Cambridge, Mass: MIT Press. 167–224.
- Bever, Thomas G. & David Poeppel (2010). Analysis by synthesis: a (re-)emerging program of research for language and vision. *Biolinguistics* **4**. 174–200.
- Boersma, Paul & David Weenink (2012). *Praat: doing phonetics by computer* (version 5.3.20). <http://www.praat.org>.
- Boomershine, Amanda, Kathleen Currie Hall, Elizabeth Hume & Keith Johnson (2008). The impact of allophony versus contrast on speech perception. In Peter Avery, B. Elan Dresher & Keren Rice (eds.) *Contrast in phonology: theory, perception, acquisition*. Berlin & New York: Mouton de Gruyter. 145–171.
- Burzio, Luigi (1994). *Principles of English stress*. Cambridge: Cambridge University Press.
- Caporello Bluvas, Emily & Timothy Q. Gentner (2013). Attention to natural auditory signals. *Hearing Research* **305**. 10–18.
- Chung, Hyunsong, Kyongsok Kim & Mark Huckvale (1999). Consonantal and prosodic influences on Korean vowel duration. In *Proceedings of EuroSpeech99*. Vol. 2. Budapest, Hungary. 707–710.
- Davidson, Lisa (2007). The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* **24**. 261–286.
- Davidson, Lisa & Jason A. Shaw (2012). Sources of illusion in consonant cluster perception. *JPh* **40**. 234–248.
- Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux & Ariel Gout (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience* **12**. 635–647.
- Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* **25**. 1568–1578.
- Dupoux, Emmanuel, Erika Parlato, Sónia Frota, Yuki Hirose & Sharon Peperkamp (2011). Where do illusory vowels come from? *Journal of Memory and Language* **64**. 199–210.
- Feldman, Naomi H. & Thomas L. Griffiths (2007). A rational account of the perceptual magnet effect. In Danielle S. McNamara & J. Gregory Trafton (eds.) *Proceedings of the 29th Annual Cognitive Science Society*. Austin: Cognitive Science Society. 257–262.
- Gerrits, E. & M. E. H. Schouten (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics* **66**. 363–376.
- Hallé, Pierre A., Juan Segui, Uli Frauenfelder & Christine Meunier (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* **24**. 592–608.
- Han, Miekko S. (1964). *Duration of Korean vowels*. Los Angeles: Acoustics Phonetics Research Laboratory, University of Southern California.
- Hong, Soonhyun (1997). Palatalization and umlaut in Korean. *University of Pennsylvania Working Papers in Linguistics* **4:3**. 87–132.
- Hooper, Joan B. (1976). *An introduction to natural generative phonology*. New York: Academic Press.
- Huang, Tsan (2001). The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. *OSU Working Papers in Linguistics* **55**. 23–42.
- Hume, Elizabeth & Keith Johnson (2003). The impact of partial phonological contrast on speech perception. In Solé *et al.* (2003). 2385–2388.

- Iverson, Gregory K. (1993). (Post) lexical rule application. In Sharon Hargus & Ellen M. Kaisse (eds.) *Studies in lexical phonology*. San Diego: Academic Press. 255–275.
- Iverson, Gregory K. (2004). Deriving the Derived Environment Constraint in non-derivational phonology. *Studies in Phonetics, Phonology and Morphology* 11. 1–23.
- Jacobs, Haike & Carlos Gussenhoven (2000). Loan phonology: perception, salience, the lexicon and OT. In Joost Dekkers, Frank van der Leeuw & Jeroen van de Weijer (eds.) *Optimality Theory: phonology, syntax, and acquisition*. Oxford: Oxford University Press. 193–210.
- Jespersen, Otto (1904). *Lehrbuch der Phonetik*. Leipzig & Berlin: Teubner.
- Johnson, Keith & Molly Babel (2010). On the perceptual basis of distinctive features: evidence from the perception of fricatives by Dutch and English speakers. *JPh* 38. 127–136.
- Kabak, Barış & William J. Idsardi (2007). Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints? *Language and Speech* 50. 23–52.
- Kang, Yoonjung (2003). Perceptual similarity in loanword adaptation: English post-vocalic word-final stops in Korean. *Phonology* 20. 219–273.
- Kim, Kong-On (1974). *Temporal structure of Spoken Korean: an acoustic phonetic study*. PhD dissertation, University of Southern California.
- Kim-Renaud, Young-Key (1987). Fast speech, casual speech and restructuring. *Harvard Studies in Korean Linguistics* 2. 341–359.
- Kuhl, Patricia K. (1993). Innate predispositions and the effects of experience in speech perception: the native language magnet theory. In Bénédicte de Boysson-Bardies, Scania de Schonen, Peter W. Jusczyk, Peter MacNeilage & John Morton (eds.) *Developmental neurocognition: speech and face processing in the first year of life*. Dordrecht: Kluwer. 259–274.
- LaCharité, Darlene & Carole Paradis (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *LI* 36. 223–258.
- Lahiri, Aditi & Henning Reetz (2002). Underspecified recognition. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory Phonology* 7. Berlin & New York: Mouton de Gruyter. 637–675.
- Lahiri, Aditi & Henning Reetz (2010). Distinctive features: phonological underspecification in representation and processing. *JPh* 38. 44–59.
- McClelland, James L. & Jeffrey L. Elman (1986). The TRACE model of speech perception. *Cognitive Psychology* 18. 1–86.
- Macmillan, Neil A. & C. Douglas Creelman (2005). *Detection theory: a user's guide*. 2nd edn. Hillsdale: Erlbaum.
- Monahan, Philip J., Eri Takahashi, Chizuru Nakao & William J. Idsardi (2009). Not all epenthetic contexts are equal: differential effects in Japanese illusory vowel perception. In Shoichi Iwasakai, Hajime Hoji, Patricia M. Clancy & Sung-Ock Sohn (eds.) *Japanese/Korean linguistics*. Vol. 17. Stanford: CSLI. 391–405.
- Moreton, Elliott (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84. 55–71.
- Norris, Dennis & James M. McQueen (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review* 115. 357–395.
- Paradis, Carole & Darlene LaCharité (1997). Preservation and minimality in loanword adaptation. *JL* 33. 379–430.
- Peperkamp, Sharon (2005). A psycholinguistic theory of loanword adaptations. *BLS* 30. 341–352.
- Peperkamp, Sharon & Emmanuel Dupoux (2003). Reinterpreting loanword adaptations: the role of perception. In Solé *et al.* (2003). 367–370.
- Pisoni, David B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics* 13. 253–260.

416 *Karthik Durvasula and Jimin Kahng*

- Poeppel, David & Phillip J. Monahan (2011). Feedforward and feedback in speech perception: revisiting analysis by synthesis. *Language and Cognitive Processes* **26**. 935–951.
- Pollack, Irwin & Donald A. Norman (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science* **1**. 125–126.
- Selkirk, Elisabeth (1984). On the major class features and syllable theory. In Mark Aronoff & Richard T. Oerhle (eds.) *Language sound structure*. Cambridge, Mass.: MIT Press. 107–136.
- Sievers, Eduard (1881). *Grundzüge der Phonetik, zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf & Härtel.
- Sohn, Ho-Min (1999). *The Korean language*. Cambridge: Cambridge University Press.
- Solé, M. J., D. Recasens & J. Romero (eds.) (2003). *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona: Causal Productions.
- Sonderegger, Morgan & Alan C. L. Yu (2010). A rational account of perceptual compensation for coarticulation. In Stellan Ohlsson & Richard Catrambone (eds.) *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*. Austin: Cognitive Science Society. 375–380.
- Stanislaw, Harold & Natasha Todorov (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers* **31**. 137–149.
- Steriade, Donca (1982). *Greek prosodies and the nature of syllabification*. PhD dissertation, MIT.
- Uffmann, Christian (2006). Epenthetic vowel quality in loanwords: empirical and formal issues. *Lingua* **116**. 1079–1111.
- Wilson, Colin & Lisa Davidson (in press). Bayesian analysis of non-native cluster production. *NELS* **40**.
- Yu, Alan C. L. (2011). On measuring phonetic precursor robustness: a response to Moreton. *Phonology* **28**. 491–518.