

**Speech perception is not isomorphic to phonology:
The case of perceptual epenthesis**

Bariş Kabak
(University of Konstanz)

and

William J. Idsardi
(University of Delaware)

Address for Correspondence:

Bariş Kabak
Department of Linguistics
University of Konstanz
Fach D 174
78457 Konstanz, Germany
Phone: +49-7531-884758
Fax: +49-7531-884160
E-mail: Baris.Kabak@uni-konstanz.de

Acknowledgements

This article is an extended version of an article that has appeared in the proceedings of the Berkeley Linguistics Society (Kabak & Idsardi, 2003), which was based on Kabak (2003), the doctoral dissertation of the first author. We would like to thank the audiences at the 29th Annual Meeting of the Berkeley Linguistics Society for comments and suggestions. We are also indebted to John Kingston, Anne Cutler, Donca Steriade, Aditi Lahiri, Jaye Padget, Benjamin Bruening, Irene Vogel, Juliette Blevins, and Colin Phillips for important criticism and suggestions.

ABSTRACT

We present the results from an experiment that tests the perception of English consonantal sequences by Korean speakers and demonstrate that perceptual epenthesis in a second language (L2) arises from syllable structure restrictions of the first language (L1), rather than linear co-occurrence restrictions. Our study replicates and extends Dupoux et al.'s (1999) results that suggested that listeners perceive epenthetic vowels within consonantal sequences that violate the phonotactics of their L1. Korean employs at least both kinds of phonotactic restrictions: (i) syllable structure restrictions that prohibit the occurrence of certain consonants in coda position (e.g., *[c.], *[g]), while allowing others (e.g., [k.], [l.]), and (ii) consonantal contact restrictions that ban the co-occurrence of certain heterosyllabic consonants (e.g., *[k.m]; *[l.n]) due to various phonological processes that alter such sequences on the surface (i.e., /k.m/ → [ŋ.m]; /l.n/ → [l.l]). The results suggest that Korean syllable structure restrictions, rather than consonantal contact restrictions, result in the perception of epenthetic vowels. Furthermore, the frequency of co-occurrence fails to explain the epenthesis effects in the percept of consonant clusters employed in the present study. We address questions regarding the interaction between speech perception and phonology and test the validity of Steriade's (2001 a, b) Perceptual-Mapping (P-Map) hypothesis for the Korean sonorant assimilation processes. Our results indicate that Steriade's hypothesis makes incorrect predictions about Korean phonology and that speech perception is not isomorphic to speech production.

Speech perception is not isomorphic to phonology: The case of perceptual epenthesis

INTRODUCTION

This article has two purposes: to demonstrate that perceptual epenthesis arises from syllable structure violations, rather than linear co-occurrence restrictions, and to show that speech perception is not always isomorphic to phonology. We first address the role of first language (L1) syllable structure restrictions in the perception of non-native second language (L2) speech. In particular, we report results from a speech perception study that focuses on the perception of illusory epenthetic vowels in sequences of speech strings that contain illicit consonant clusters. Our study replicates and extends previous studies on the perception of consonant clusters (especially Dupoux et al., 1999) and unconfounds the contribution of potential syllable structure constraints from purely sequential consonantal contact restrictions. Based on our findings, we suggest that L1 syllable structure restrictions result in the perception of epenthetic vowels and consonantal contact violations themselves do not explain perceptual epenthesis. We further show that explanations based on frequency fail to shed light on epenthesis effects in the percept of consonantal sequences employed in this study. The second aim of this article is to address questions pertaining to the interaction between speech perception and phonology. Can speech perception explain phonological processes? Does speech perception mirror speech production? In particular, we evaluate the findings from the present study in the context of Steriade's (2001 a, b) Perceptual-Mapping (P-Map) hypothesis and argue that Steriade's hypothesis makes incorrect predictions about Korean phonology. We suggest that speech perception is not always isomorphic to speech production. Our findings are

also consistent with the L2 literature in finding both commonalities and discrepancies between L1 and L2 phonology (e.g., Dulay, Burt, & Krashen, 1982).

PHONOTACTICS AND SPEECH PERCEPTION

There is an abundant literature on the role of phonotactics in adult speech perception. Most research in this field has focused on response biases in favor of perceiving ambiguous sound combinations as legal sequences, a pattern which is assumed to reflect the role of phonotactic knowledge in perception (e.g., Massaro & Cohen, 1983; Hallé et al., 1998; Pitt, 1998; Moreton, 2002). Other studies have looked at the origin of phonotactic knowledge, such as frequency and lexical effects (e.g., Luce & Pisoni, 1998; McClelland & Elman, 1996; Norris, 1994; Pitt & McQueen, 1998; Vitevitch and Luce, 1998). Over all, these studies demonstrate that listeners' information about the legality and the probability of phonotactic patterns influence the processing of spoken stimuli. Phonotactic knowledge in infant language processing has also been extensively investigated. The studies suggest that the knowledge about L1 phonotactic regularities emerge very early on in the course of acquisition (e.g., Jusczyk et al., 1994) and seem to bias infants' perception to prefer listening to words that accord with the phonotactics of their L1 (Jusczyk et al., 1993).

Since L1 phonotactic regularities have been shown to influence infants' perceptual abilities, will they also play a role in L2? There are several instances in which we can observe the impact of L1 phonotactic regularities on the realization of non-native speech. For instance, speakers of Korean and Japanese are known to insert epenthetic vowels

when they pronounce loan words involving sequences of segments that do not fit the syllable structure of their native languages:¹

- | | | | |
|-----|-----------------------------------|--------------|------------|
| (1) | a. [a.i.su.k ^h u.rim] | ‘ice cream’ | (Korean) |
| | b. [k ^h u.ri.su.ma.su] | ‘Christmas’ | (Korean) |
| | c. [ma.ku.do.na.ru.do] | ‘Mac Donald’ | (Japanese) |
| | d. [su.to.ra.i.ku] | ‘strike’ | (Japanese) |

Are the epenthetic vowels inserted in production, or in perception? In a series of behavioral experiments, Dupoux and colleagues (Dupoux et al., 1999; Dupoux, et al., 2001; Dehaene-Lambertz et al., 2000) compared Japanese listeners with French listeners in their perception of consonant clusters. For instance, in Dupoux et al. (1999), in an off-line phoneme detection task (Experiment 1) used a series of six items created from naturally produced nonce words (e.g., [abuno], [akumo], [ebuzo], [egudo], etc.) by gradually reducing the duration of the vowel [u] down to zero milliseconds. The participants were asked to respond whether each item they heard contained the sound [u]. Japanese listeners, unlike French listeners, overwhelmingly judged that the vowel was present at all levels of vowel length. Strikingly, this was the case seventy percent of the time even when the vowel had been completely removed (i.e. the zero ms condition). The French participants, on the other hand, judged that the vowel was absent in the no-vowel condition about 90% of the time and that a vowel was present only in 50% of the intermediate cases (Figure 1).

----INSERT FIGURE 1 ABOUT HERE----

¹ These transcriptions are given in the IPA except where noted; [u] is a high back unrounded vowel, [r] as an alveolar tap.

These results were confirmed in other experiments, which have led Dupoux and colleagues to conclude that the influence of native language phonotactics can be so robust that listeners "invent" illusory vowels to accommodate illicit sequences of segments in their L1. Follow-up studies, again on Japanese but employing different experimental paradigms such as a lexical decision task (Dupoux et al., 2001) and event-related potentials (Dehaene-Lambertz et al., 2000) have further confirmed perceptual epenthesis with Japanese participants.

Dupoux and colleagues' findings have yielded crucial implications for the potential role of perception in explaining production errors such as epenthesis and consequently also for our understanding of the nature of phonological representations. The researchers, however, have left the question open as to whether perceptual epenthesis stems from the fact that Japanese speakers never hear consonant clusters in their language or whether Japanese syllable structure is predominantly CV. That is, it is not clear from Japanese whether perceptual epenthesis arises due to restrictions on which consonants can cooccur in a sequence or due to syllable structure restrictions. Japanese syllables are predominantly CV, as Japanese licenses very few coda consonants, and then only under greatly restricted circumstances, and therefore displays a paucity of coda-onset clusters generally. In particular, a consonant can only be licensed in the coda position if it is the first member of a geminate (2a, b), or if it is a nasal homorganic to the following consonant (2c, d) (Itô, 1986; 1989). Coda consonants (other than the mora nasal) not carrying either of these two properties cannot occur in Japanese, as can be seen in the hypothetical examples given in (3) (from Itô (1989)).

- (2) a. kap.pa 'a legendary being'
 b. gak.koo 'school'
 c. tom.bo 'dragonfly'
 d. kaŋ.gae 'thought'
- (3) a. * kap.ta
 b. * tog.ba
 c. * pa.kat

Closer inspection on the stimulus items used in Dupoux et al. (1999) shows that they all contained illicit coda consonants (e.g., [eb.zo], [i].to], [ek.to], etc.). The study, thus, confounds whether the perceptual epenthesis induced in the percept of words such as [ebzo] is due to: (1) consonantal contact restrictions (i.e., the sequence [b.z] is impossible), or (2) syllable structure (coda) restrictions (i.e., [b] cannot be an independent coda). Since Dupoux and colleagues' primary aim was to document the perceptual vowel epenthesis per se, the restrictedness of Japanese phonology does not interfere with their findings and conclusions. It is important, however, to tease apart consonantal contact restrictions from coda restrictions since this has important consequences for phonological theories explaining phonotactic patterns. In particular, while most theories refer to a prosodic domain, namely the syllable, to explain consonantal phonotactics, there are also those that employ syllable-independent, string-based and linear statements (e.g., Dziubalska-Kolaczyk, 1994; Steriade, 1999; 2001a; Blevins, 2002). The theories that do not refer to the syllable or any other prosodic structure would explain the findings as a product of illegal consonant contact and would reject explanations in terms of syllable

structure violations. The present study successfully teases apart the two competing explanations by employing Korean, which offers a much more interesting array of consonant clusters.

KOREAN AND ENGLISH PHONOTACTIC PATTERNS

Phonotactics, in its most commonly used sense, refers to the restrictions that govern sound sequences in a language. The coda or onset position of a syllable can be used to anchor particular phonological contexts where restrictions on consonants and their combinations may be observed. For instance, three-member consonantal onset clusters in English must start with /s/ (e.g., *spread*, *street*, *square*, *splash*, etc.), but codas do not obey the same requirement, nor the reverse requirement (e.g., *burst*). Likewise, English has some consonants that only appear in onset position (e.g. /h/) as well as consonants that only appear in coda position (e.g. /ŋ/). In Korean pronunciations, the alveo-palatal affricates (/c, c', c^{h/2}) can only occur in syllable onsets.

Korean and English have different restrictions on heterosyllabic consonantal contacts (i.e., [C₁.C₂]). Ewen and Van der Hulst (2001) offer a clear discussion of English medial clusters, demonstrating the utility of the syllable as an important domain over which to explain phonotactic regularities. Ewen and Van der Hulst propose that part of the phonological knowledge of a native speaker involves the specification of which consonant clusters are ill-formed in English, giving the examples in (4) (p. 123).

(4) *initial* *medial* *final*

² Following the Korean linguistics tradition, we will use /c/ to refer to Korean voiceless coronal affricate consonant. Accordingly, no diacritic marker on a consonant indicates a plain consonant, as opposed to

*km-	*-pkm-	*-pk
*mr-	*-kmr-	*-km
*mw-	*-tnw-	*-tn

They note that a fairly obvious redundancy arises in the above clusters: the constraints on medial clusters in English are not independent of those on initial and final clusters. Medial sequences must consist of a valid coda followed by a valid onset. Any other medial cluster is illicit. Other than this, English is quite profligate and seems to place no constraint on coda-onset contacts (see Lamontagne (1993) for a detailed discussion of the restrictions on English consonant clusters). It should be noted, however, that not all consonantal combinations are observed with equal frequency in English. For instance, while there are monomorphemic words with -kt- (e.g., *sector*, *doctor*, *October*, etc.), the reverse sequence -tk- is very difficult to find in the language (except for *catkin* (a type of flower), or proper names such as *Atkins* or *Ratko*). Despite its rarity, English speakers readily accept and produce pronunciations of loan words containing this cluster (e.g., *Kamchatka*).

In Korean, inflected or derived words provide potential contexts for heterosyllabic consonantal contacts, as consonants belonging to different morphemes come into contact at morpheme junctures. Korean employs a number of phonological processes that affect such heterosyllabic consonant clusters. For instance, due to the process of nasalization, a sequence of an oral consonant followed by a nasal consonant (C+N) never occurs (e.g.,

tense (e.g., /c'/) and aspirated (e.g., /c^h/). The symbol [j] represents the voiced allophone of the plain alveo-palatal affricate consonant (i.e., /c/). We use [y] to represent the palatal glide.

/puək^h/+/mun/ → [pu.əŋ.mun] 'kitchen door'; /os/+/noŋ/ → [on.noŋ] 'clothes chest')³

except where there is an intervening Intonational Phrase (IP) boundary.⁴ It should be noted that no IP boundary occurs inside inflected or derived words. In English, however, although some word-internal consonant clusters such as /gt/ are nearly impossible to find, compounds with such clusters do exist without undergoing any phonological alterations (e.g., *pigtail*, *ragtime*). Moreover, many unusual heterosyllabic combinations of consonants in loan words, novel words, and acronyms can easily be incorporated into English without modification and are pronounced without any difficulty by English speakers (e.g., [g.d]: *Magdeburg*; [ŋ.y]: *Pyongyang*; [t.k]: *dot com*, etc.). Since English seems to place virtually no additional restrictions on the type of *coda-onset* clusters, in the following we will focus only on Korean, which has some crucial phonotactic restrictions to separate consonantal contact restrictions from syllable structure restrictions.

Korean Phonology and Phonotactics

Korean has nineteen consonants and ten vowels (Tables 1 and 2, adapted from Sohn, 1999: 153, 156).

----INSERT TABLE 1 ABOUT HERE----

³ Korean examples in this article are adapted from Sohn (1994; 1999) unless otherwise noted.

⁴ The rules that apply across morphemes (see the next section) such as nasalization and lateralization have also applied historically to what are now unanalyzable combinations of morphemes (see for example Martin 1992: 52, who cites a fifteenth century form /punnon/ < */put^h-non/ 'igniting'). Since it is simply possible to lexically represent the now morpheme-internal sequence with two nasals, we cannot therefore find direct evidence for morpheme-internal applications of nasalization, as morphemes that involve CN sequences have been diachronically leveled out in favor of NN sequences. Similar considerations hold of borrowed words that have orthographic treatments in Hangul with 'nl' sequences, such as the Korean rendition of foreign names, for example Marilyn Monroe, orthographically 'monlo' but phonologically /monno/ (Lee and Ramsey, 2000:71). Since there is no evidence that nasalization does or does not apply

----INSERT TABLE 2 ABOUT HERE----

Korean has a three-way contrast for oral stops and affricates: plain (C), aspirated (C^h) and tense (C'). The tense consonants are phonetically both long and pronounced with a constricted glottis. Korean syllables are maximally CGVC. As in Japanese, the onset and coda cannot branch in Korean. The glide (G) in CGV is part of the nucleus. Korean restricts the occurrence of consonants in certain syllabic positions. For instance, only seven among nineteen consonantal phonemes (i.e., [p, t, k, m, n, ŋ, l]) can be pronounced in codas. The other consonants undergo various processes of neutralization when they end up in coda position. When coda consonants are in contact with other consonants at morpheme junctures, they undergo various assimilatory processes. Even consonants that cannot surface in coda position interact with adjacent consonants (e.g., /h/+t/→[t^h]; /coh/+ta/→[cot^ha] 'is good'). While assimilatory processes may alter the segmental realization of a given heterosyllabic consonant cluster on the surface, the consonants that form the contact can independently surface if they occur elsewhere. For instance, *[k.m] and *[l.n] are illicit but [k.t] and [l.t] are licit. Such illicit codas and consonantal contacts are repaired by various phonological rules. In the following, we will provide only those processes that are relevant for the present study. Thus Korean provides enough variety to distinguish clusters which are illicit because of the contact between the consonants versus those that are illicit because of the coda consonant alone.

morpheme-internally in Korean, in the absence of non-derived environment blocking effects, the simplest conclusion is that the rule applies both within morphemes as well as across morpheme boundaries.

First, strident consonants such as [c], [c^h], and [s] neutralize in coda position to the unreleased stop [t]. For instance, morphophonemic forms such as /nac/ ‘daytime’, /nac^h/ ‘face’ and /nas/ ‘sickle’ become homophonous when they are pronounced in isolation (i.e., [nat]). Second, Korean has a nasalization rule that turns stops into nasals before nasal segments (e.g., /k.m/ → [ŋ.m]: /hak+mun/ → [haŋ.mun] ‘learning’; /p.m/ → [m.m]: /cip+mun/ → [cim.mun] ‘house gate’). Third, a process known as lateralization affects nasal sounds after lateral sounds (/l.n/ → [ll]: /tal+nala/ → [tal.la.ra] ‘moon country’). Finally, voicing in Korean is predictable. Plain consonants become voiced between sonorants (e.g., /pa+po/ → [pa.bo] ‘idiot’). It should be noted that these processes are very general processes of Korean *without* exceptions. They not only apply within words but also across word boundaries, although not across Intonational Phrase boundaries (see Martin (1992: 30) and Kabak (2003) for an exhaustive list of possible [C₁.C₂] contacts and their surface realization). Finally, it should be noted that the Korean epenthetic vowels are [i] after palatal consonants and [u] in other contexts (e.g., [a.i.su.ku.rim] ‘ice cream’; [su.p^hən.ji] ‘sponge’).

THE PRESENT STUDY

Research Questions and Hypotheses

A closer inspection of Korean phonotactic patterns reveals that there are at least three reasons why a word in the form of [VC₁.C₂V] can be illicit. First, [C₁.C₂] could induce a contact violation. For instance, as explained above, words containing *[k.m] or *[l.n] sequences would be required to undergo nasalization and lateralization by the Korean production grammar. Second, [C₁] could be an illicit coda (e.g., *[c.], *[k^h.], *[h.], *[r.]

etc.). Third, [C₂] could be an illicit onset (e.g., *[.l], *[.ŋ], etc.). The first and the third reasons differ from the second in being sensitive to the nature of C₂. The present study is concerned only with the first two factors (see, however, Kabak (2003) for additional discussion of consonants disallowed in onsets).

Given the two possible factors that induce a contact violation in Korean, we wish to investigate whether coda or contact violations cause perceptual epenthesis (e.g. whether words with [k.m] sequences are confusable with words with [kum] instead). A second question arises as to whether contact violations (e.g., *[k.m], *[l.n]) can be perceptually altered to fit their likely surface forms according to the rules and processes of the Korean production grammar. In particular, we want to test whether words with [k.m] and [l.n] can be confused with those with [ŋ.m] and [l.l] sequences, respectively. In fact, there is evidence from Korean loan adaptations that speakers employ native phonological rules, rather than epenthesis, to alter English words that contain such sequences. For example, loans such as *walnut*, *Telnet*, *Big Mac*, and *Packman* are commonly produced as having undergone the lateralization and nasalization rules where applicable by Korean speakers (e.g., *wa[l.l]ut*, *te[l.l]et*, *bi[ŋ.m]ac*, *Pa[ŋ.m]an* etc.). Thus, there are reasons to suspect that native phonological rules may apply in perception to alter sequences of sounds that violate the sound patterns of Korean. A third question concerns the phonological status of the violation: Could perceptual epenthesis be induced only by phonologically contrastive features of Korean such as [strident], [spread glottis], [nasal], etc., or could it also be induced by allophonic features violating the Korean phonological patterns (e.g. [voice])?

We propose two main hypotheses to be tested given the ways in which *[C₁.C₂] sequences can be illicit in Korean:

(4) The Consonantal Contact (CC) Hypothesis:

Korean listeners will hear epenthetic vowels when a given sequence is illicit in Korean.

(5) The Coda/Onset Identity (COI) Hypothesis:

Perceptual epenthesis will arise when there is a syllable structure violation.

A number of statements regarding our main hypotheses are in order. First, the CC Hypothesis relates perceptual epenthesis to illicit sequences. It is compatible with a string-based approach towards explaining sound distributions. In string-based phonotactics, the structure of speech input consists of linear strings of discrete abstract linguistic units (feature bundles or segments) that are ultimately bound by word or morpheme boundaries (cf. Steriade, 1999; Blevins, 2002). Steriade (1999), for instance, classifies the positions of segments not in syllabic terms (e.g., as onset vs. coda) or in linear terms (e.g., “before a vowel” vs. “after a consonant”) but as positions where certain featural contrasts are more versus less perceptible. Accordingly, certain consonant clusters are illegal since the featural contrast on one or both of the segments in the cluster is difficult to perceive in that context. The COI Hypothesis, on the other hand, is based on the view that phonological structure is hierarchical. In particular, the hypothesis crucially refers to syllable structure conditions in the L1 system, which specifically state that certain consonants can never surface as coda or as onset. That is, a given consonant may

crucially lack a particular positional identity (e.g., coda identity or onset identity) in the inventory. Specifically, it predicts that Korean listeners hear epenthetic vowels when *[C₁.] is an illicit coda consonant. Therefore, following a vowel, when the listener encounters a consonantal element that does not carry a coda identity (e.g., /c/), (s)he will automatically interpret the consonant as an onset.

Second, each hypothesis involves a different conception of perceptual epenthesis. The CC hypothesis motivates perceptual epenthesis to break up illicit sequences of consonants. In simple linear terms, *a là Steriade*, the percept of an epenthetic vowel ensures a buffer sound that breaks up the unwanted consonantal contact. Creating well-formed syllables, however, is the primary goal of the COI hypothesis. It should be noted that both hypotheses are unit-independent. That is, the unit of perception can be a segment or a bundle of features.

With regard to our secondary question as to the nature of L2 representations, we propose two conflicting sub-hypotheses:

(7) The Phonetic Processing Hypothesis:

All phonetic features in the L1, including allophonic ones, such as voicing in Korean, are represented upon the perception of L2 forms.

(8) The Phonological Processing Hypothesis:

Only contrastive (non-redundant) features in the L1 are represented in the L2 whereby predictable phonological information is suspended.

The Phonetic Processing hypothesis assumes that perceptual representations are fully equipped with all available surface phonetic information. This view is in part shared by Silverman (1992). Silverman hypothesizes that speakers match non-native signals to native feature matrices that most closely approximate the phonetic properties of the signal. Similarly, the Phonetic Processing hypothesis predicts that the material detected in the speech stream will be compared to the feature matrices of phonetic forms in the L1. Specifically, if the acoustic stream contains [voice], which closely approximates to the phonetic properties of the voiced segments in Korean, [voice] will be represented at the perceptual level although voicing is predictable in Korean. Thus, misplaced voiced segments such as [g.t] or [g.m] will be faithfully represented by Korean listeners by inducing perceptual epenthesis and violate the two hypotheses set earlier. This is because (1) these consonantal combinations never surface in Korean and thus violate the CC Hypothesis, and (2) [g] can never surface in the coda position and thus violate the COI hypothesis.

The Phonological Processing Hypothesis, on the other hand, assumes that speech perception functions in an abstract and minimal fashion. Accordingly, listeners filter the acoustic stream by detecting only a minimal amount of articulatory and/or acoustic cues, suppressing redundant and predictable information. These cues correspond to distinctive features that are employed to construct phonological contrasts. In this respect, the hypothesis is compatible with Underspecification Theory (cf. Archangeli, 1984; 1988) as well as the fundamentals of the Featurally Underspecified Lexicon Model (FUL) (e.g., Lahiri and Marslen-Wilson, 1991, 1992; Lahiri and Reetz, 1999). Korean voicing, under this hypothesis, is not represented by Korean listeners in their phonological perception

since voicing is a predictable phenomenon in Korean. It is assumed that when [voice] is not represented, [g] must be represented as the less specified plain velar stop, namely as /k/. If the CC hypothesis holds, under the Phonological Processing Hypothesis, only [g.t], which is perceived as /k.t/, should not induce perceptual epenthesis because [k.t] is a permissible combination in Korean. The cluster [g.m], which is perceived to be /k.m/, on the other hand, should induce perceptual epenthesis because it is still an illicit combination in Korean, which needs to undergo nasalization. On the contrary, if the COI hypothesis holds under the same hypothesis, both [g.t] and [g.m] should uniformly be unproblematic since [k.] is a legitimate coda consonant in Korean.

Design

An AX discrimination paradigm was employed, comparing pairs of nonce words with and without consonant clusters (e.g., [p^hakma] vs. [p^hak^huma]). The assumption was that if Korean listeners hear epenthetic vowels in consonant clusters, they are likely to interpret the two words in the pair to be the same. We constructed nonsense words of the form of [p^háC₁(V)C₂a] to create test pairs.⁵ Following the Korean epenthesis patterns, the vowel was [i] after palatals and [u], the closest approximation of [ɯ] elsewhere. C₁ varied between a permissible coda, [k] and [l], and an impermissible one, [c]. The onset of the second syllable, that is C₂, was either a stop ([t]) or a nasal ([m] or [n]).

⁵ We employ the usual conventions for aspiration in English stops, marking onsets but not codas. English coda consonants are not generally aspirated, and thus coda stops were not pronounced with aspiration in this study. It should be noted that English voiceless stops are adapted into Korean onsets with aspiration, regardless of their original position (e.g., *camera* [k^ha.me.ra]; *guitar* [ki.t^ha], examples from Kang 2001), even when the stops are clearly *not* aspirated in the English source (e.g., *spray* [su.p^hu.re.i]). The speaker employed in this study had very little rounding in [u]. He pronounced [c] and [j] as alveopalatal affricate, and had voicing into closure in the pronunciations of [g] and [j].

Consequently, the combination of C₁+C₂ produced either a permissible contact ([p^hákt^ha], [p^hált^ha]) or an impermissible one (*[p^hakma], *[p^halna]).

To address our second question in relation to the nature of phonological representations, we also employed the voiced counterparts of [k] and [c], that is [g] and [j], respectively, in the same cluster combinations (i.e., [p^hagt^ha], [p^hagma]; [p^hajt^ha], [p^hajma]). Furthermore, likely surface interpretations of *[p^hakma] and *[p^halna], that is [p^hajma] and [p^halla]/[p^hanna], with nasalization and lateralization respectively, were also included in the experiment. This was specifically to test whether Korean listeners apply sonorant assimilation rules such as nasalization and lateralization that are a crucial part of their production grammar to perceptually repair the illicit contacts. None of the test words corresponded to any existing word in Korean or English.

Table 3 summarizes our test variables and their surface permissibility in the Korean production grammar and Table 4 lists all the test words used in the experiment with an example word in English that contains the contact.

----INSERT TABLE 3 ABOUT HERE----

----INSERT TABLE 4 ABOUT HERE----

Using a high impedance microphone (Audio-Technica, Omnidirectional, ATR35S), each nonce item was produced several times by a male native speaker of American English, who is a trained phonetician, in a sound attenuated room and recorded directly to compact disc with a JVC XL-R5020 compact disc recorder. The recorded tracks were then transferred onto a Mac OS 9.1 computer, where they were converted to

22kHz (16-bit resolution; 1-channel) and stored as ".aiff" files to be used in PsyScope 1.2.5 PPC (Cohen, MacWhinney, Flatt and Provost, 1993). AX test pairs were created by pairing different exemplars of test words containing consonant clusters with words where the test clusters were separated by a vowel (e.g., [p^hacma] vs. [p^hac^hima]; [p^hakt^ha] vs. [p^hak^hut^ha]), with a 1500 ms interval between each word and a 2000 ms interval between trials. Based on the principles of the Signal Detection Theory (Green and Swets, 1974), we measured d-prime (d'), the ability to discriminate between same and different pairs. Therefore, we also included different renditions of the same words (e.g., [p^hak^huma] vs. [p^hak^huma]; [p^hakma] vs. [p^hakma]). An experimental block was created with 2 random repetitions of each of the test and filler pairs. No two pairs contained identical recordings (e.g., [[p^hakma]₁ vs. [p^hakma]₂; [p^hakma]₃ vs. [p^hakma]₄; [p^hakt^ha]₁ vs. [p^hak^hut^ha]₂; [p^hak^hut^ha]₃ vs. [p^hakt^ha]₄). Since Korean listeners were potentially at a disadvantage as most of the stimulus pairs were expected to yield a “same” response from them, we created filler pairs which should pose no discrimination difficulty for Korean listeners (e.g., [p^hal.t^ha] vs. [panna]; [p^hakt^ha] vs. [p^hakma], etc.). We created an experimental block from these items, which contained 118 trials (39 test pairs plus 20 filler pairs presented in both possible orders). The experimental block was presented 5 times in a row, with 4 self-terminated rest periods in between, in order to get 10 total repetitions for each pair. The items in each block were automatically randomized by PsyScope each time a new block was presented. We employed a practice session for every subject before the actual experiment. Thus, another set of cluster-and epenthesis-words was created for the session (e.g., [p^habma], [p^halt^ha], [p^habuma], [p^halut^ha], etc.). The session contained 5

“same” and 5 “different” pairs, which were expected to cause no problem for both groups of subjects. None of the practice items were included in the actual experiment.

Twenty-five native speakers of Korean and 25 native speakers of English at the University of Delaware participated in the experiment. All the Korean speakers were residing in the USA for educational purposes and none of them had started learning English before the age of 12. None of the English subjects had any knowledge of Korean.

Each subject participated in the AX discrimination task in a single testing session. They were specifically told that they would hear an American man saying nonsense words of English in pairs, and their task was to determine whether the man repeated the same word the second time or said a different word, pressing <A> on the keyboard for same, and <L> for different. The subjects heard the stimuli over headphones at a comfortable intensity for the subject. Unlike in Dupoux and colleagues’ experiments, there was no training session. Likewise, no feedback was given regarding subjects’ responses during the practice session.

Predictions

The two main hypotheses and the two sub-hypotheses make different predictions with respect to the outcome of the experiment. The predictions of the four combinations of hypotheses considered above are summarized in Table 5, where X indicates a situation where the Korean participants are predicted to fail to discriminate the contrast between [paC₁C₂a] and [paC₁VC₂a].

----INSERT TABLE 5 ABOUT HERE----

Under the Consonantal Contact (CC) hypothesis, all of the cases that yield illicit consonantal contacts, namely [k.m], [l.n], [c.m] and [c.t^h], are expected to be misperceived by Korean participants. This is because they all contain contacts that are not permissible in Korean. If on the other hand, the Coda/Onset Identity (COI) hypothesis were true, all contacts that contain a permissible coda consonant such as [k.] and [l.] (i.e., [k.m], [k.t^h], [l.n], [l.t^h]) are *not* expected to be misperceived. With respect to voiced consonants, if Korean participants use a phonetic processing approach and thereby represent voicing then both [g] and [j] should behave like [c] because they can also never occur as codas in Korean. To the contrary, under a phonological processing hypothesis, if voicing is perceptually suppressed and represented as plain then the same consonants, namely [g] and [j], should behave like [k] and [c], respectively. It should be noted that under all possible combinations of the hypotheses, clusters with strident C₂'s are expected to be misperceived while [k.t^h] and [l.t^h] should not be misperceived since they include both a permissible contact and a permissible coda consonant. The combinations of the same hypotheses, however, make different predictions with respect to other contacts that will crucially differentiate between the hypotheses. Therefore, this experiment guarantees, regardless of theoretical choice, that there is both a good condition (i.e., [k.t^h], [l.t^h]), and a bad condition (i.e., [c.m], [c.t^h], etc.).

Signal Detection Theory Statistics

Before turning to the results of the experiment, we will briefly review the use of Signal Detection Theory (SDT) used in this study. As mentioned above, we have used the principles of SDT (Green & Swets 1974) to calculate discriminability scores (*d'*) for each subject on each “different” pair. To do this we need to obtain two values: (1) *p*(H), the

proportion of “different” responses when the words in the pair are different (also known as the “hit rate”), and (2) $p(F)$, the proportion of “different” responses when the pairs presented are the same (also known as the “false alarm rate”).⁶ The two scores are converted into z-scores, and the initial estimate for d' is obtained by subtracting $z(F)$ from $z(H)$. The initial estimate was corrected using the correction for “same-different” designs in Table (A.5.3.) of MacMillan and Creelman (1991). The d' values obtained for each test pair per subject ranged from 4.21 to 0. It should be noted that 4.21 was the practical maximum given the number of items presented to the listener, and that 0 was the theoretical as well as practical minimum in the present study. There were cases where some subjects had no misses. Thus, $p(H)$ was 1. Likewise, there were also cases where a subjects had no false alarms; therefore, $p(F)$ was 0. We amended such cases, following Kadlec (1999), by substituting a value of 0.5 for misses and/or false alarms for each subject that had zero for these parameters. On the other hand, there were cases where

⁶ It should be noted that the AX paradigm involved in the present experiment is neither a classic roving design nor a classic fixed design (MacMillan and Creelman, 1991: 147-159). In roving designs, stimuli are (equally) spaced in some physical units along a single continuum and presented within the same block. An analysis based on a decision procedure called the differencing strategy is used in roving designs, where the two observations on a trial are subtracted, and the result is compared to a criterion (see especially Figures 6.1 and 6.3 in MacMillan and Creelman (1991) for further explanation of this difference). In fixed experiments, a single stimulus pair is used and the analysis based on an Independent-Observation Model is employed, where the decision rule is to decide separately whether each item is S_1 or S_2 , then report whether these subdecisions are the same or different. Since the test doublets were drawn randomly from a pool of exemplars, they were different from each other in the present study and since they were presented in the same block, it cannot readily be classified as a classic fixed design. Furthermore, the words contained in the doublets were not intended to vary equi-distantly from one another in a continuum. An example of a roving design would be the one in which the length of the “epenthetic vowel” was manipulated (as in the case of Dupoux et al., 1999). Therefore, the present experiment is also not a classic roving design, either. However, the intended listening strategy that was imposed on the subjects is compatible with the decision rule given by the Independent-Observation Model. Specifically, the AX paradigm was designed to induce a phonological analysis of the stimuli, where the optimal decision rule for subjects was to construct a percept of each word in the doublet and then report whether these percepts were the same or different. The basic premise of the Differencing Model, however, leads to an acoustic comparison of the test words, where the acoustic images of the two words are subtracted. Given the 1500 ms of ISI used in the experiment, such a strategy is unlikely. Therefore, the pairs in this experiment are best analyzed in accordance with the Independent-Observation Model.

subjects had no hits or false alarms. This was typically the case for Korean subjects, who could not detect a difference upon hearing some “different” pairs, but were perfectly able to report that they heard the same words when they heard “same” pairs. In such a situation, both the $p(F)$ and $p(H)$ are 0. If $p(F)$ is equal to, or greater than $p(H)$, the d' is defined to be zero in SDT, indicating that the subject has no sensitivity to the signal. The smallest non-zero d' score that can be obtained in the experiment was 0.89, which occurs when the hit rate is 1 and the false alarm rate is 0 (which is corrected by adding 0.5). The best and worst possible performances and their calculated d' -prime scores are shown in Table 6.

Results

Descriptive Statistics

Figure 2 provides the average d' scores for both groups for the various clusters we employed. As expected, the English group successfully discriminated all test clusters (except for the anomalously poor performance on [g.m]), with an average d' of 3.38. An apparent three-way grouping, however, emerged from the Korean data: (1) [c.m], [j.m], [j.t^h], [c.t^h], where the mean d' values were below 1; (2) [g.m], [k.m], [g.t^h], which formed an intermediate category with d' scores roughly between 2 and 3; and (3) [l.n], [l.t^h], [k.t^h], where the d' scores were very close to the English group's mean d' scores. The first group with strident C_1 (which cannot occur in the Korean coda position), had an average d' of 0.53, indicating an inability to discriminate. When considering the *median* d' scores, the Korean group has a d' value of 0 on these clusters. For the other clusters, however, the median d' scores were similar to those for the English group. Both the Korean and the English groups' performances on [g.m] were somewhat degraded. Except for this cluster,

all the mean d' scores can be said to exhibit a flat performance, ranging between 3 and 4. While there is no immediate explanation for why both groups suffered some degree of difficulty with [g.m], it is instructive to note that this performance was not nearly as bad as the Korean group's performance on strident clusters.

---INSERT FIGURE 2 ABOUT HERE---

An important statistical question arises here with respect to which d' scores should be considered as “bad”, “moderate”, and “successful” performance. Given that the range of individual d' scores vary between 0 and 4.21, for instance, should a d' score of 2.57 indicate a “moderate” or “successful” discrimination ability? Likewise, should a d' value of 1.80, as in the Korean [g.m], be grouped with scores below 1, which show almost no discrimination ability? A peculiarity of the d' calculation has consequences for our interpretations of these scores. There is a gap in the possible d' scores for this size of experiment between 0 and 0.89. That is, the smallest non-zero d' score that can be obtained in this experiment was 0.89 (see Table 6). Therefore, the mean d' of 0.53 observed with the Korean subjects for the stridents can only be obtained when a significant number of subjects had d' scores of 0. In fact, the *median* d' score in this case was zero, showing that more than half of the d' scores are 0.

Analysis of Variance

A 5x2x2 (i.e., C_1 (5) by C_2 (2) by Language Group (2)) repeated measures Analysis of Variance (ANOVA) of d' revealed statistically significant effects throughout the model. All main factors were significant: language, $F(1, 48)=184.627$, $p<.0001$, first consonant

(C₁), $F(4, 45)=96.131$, $p<.0001$ and second consonant (C₂), $F(1, 48)=55.555$, $p<.0001$. Moreover, all two-way interactions were also significant, though C₂ by language was marginal: C₁ by language, $F(4, 45)=57.871$, $p<.0001$, C₂ by language, $F(1, 48)=4.159$, $p<.047$, and C₁ by C₂, $F(4, 45)=18.492$, $p<.0001$. Ordinarily, these two-way interactions could be taken as the initial evidence for an effect of consonantal contact since both C₁ and C₂ seem to contribute to the difference between the Korean and English scores. However, the three-way interaction for C₁ by C₂ by language is also significant, $F(4, 45)=6.08$, $p<.001$, which necessitates a careful inspection of the post-hoc tests. Figure 3 graphically illustrates the upper and lower bounds of 95% confidence intervals for both groups.

---INSERT FIGURE 3 ABOUT HERE----

We see in Figure 3 that the English group has relatively small standard errors, except for [g.m], which only overlaps with [c.m]. A Post-hoc Scheffé test with an alpha level of .01 suggests that the English [c.m] is only significantly different from the English [k.t^h]. The difference between the two is unclear and uninteresting for our current purposes. However, [g.m], being significantly different from all other clusters (except [c.m]), does merit further consideration. A closer inspection reveals that there were vast individual differences in the English group for this item, suggesting that some of the participants experienced some trouble with the cluster. Given that English contains quite a number of simplex words with this cluster (e.g., *dogma*, *pigment*, *sigma*, *pragmatic*, etc.) compared to those with [c.m] or [c.t], which we cannot observe in simplex words, the difficulty experienced by the English group does not find an immediate explanation.

Other than this small problem, the English group successfully discriminates all other clusters.

When we consider the Korean groups' d' scores, Scheffé and Tukey-Kramer post-hoc tests consistently divide the consonant clusters into three groups: (1) [c.m], [j.t^h], [j.m], [c.t^h], (2) [g.m], [k.m], [g.t^h], (3) [l.n], [l.t^h], [k.t^h], ordered from the lowest to the highest d' . With the data divided into three distinct groups, however, the main hypotheses of the study cannot be evaluated in a simple and straightforward fashion. First, it cannot be said that the intermediate performance on [k.m], [g.m] and [g.t^h] provides sufficient evidence for the CC hypothesis because the cluster [l.n], which also induces a consonantal contact violation in Korean, is among the top three successfully discriminated clusters in the Korean group. Furthermore, the Korean performance on [l.n] overlaps with the English group's performance on the same cluster. Indeed, this finding, by itself, disputes the CC hypothesis. Second, the Korean group's performance on these clusters is not as bad as the performance on the strident clusters, which yielded near complete indiscriminability. Third, our hypotheses were constructed in such a way that they allow for only two distinct levels of performance on the test consonant clusters: (1) those that yield poor performance, and (2) those that yield good performance. Therefore, it is necessary to distinguish relative degrees of badness in order to better evaluate the hypotheses. How can the discriminability indices on the consonant clusters be grouped if only two different groups of performance were to be made?

To answer this question, we ran a multivariate hierarchical Cluster Analysis (CA) on the Korean data using SPSS. CA is a multivariate analysis technique that organizes information about variables and sorts cases into clusters in such a way that the degree of

association is strong between the members of the same cluster and weak between the members of different clusters. CA produces a dendrogram (Figure 4), which is a tree structure, and graphically clusters cases (in the case of the present study, the test consonant clusters) by calculating multivariate correlation between cases and then clusters them from strongest mutual correlation to weakest.

----INSERT FIGURE 4 ABOUT HERE---

The scale at the bottom of the dendrogram in Figure 4 provides an index of an arbitrary measure of dissimilarity. The clusters are joined at increasing levels of dissimilarity. That is, the farther the clustering occurs, the more dissimilarity there is amongst the members of the cluster. From left to right, the CA on the Korean data divides the consonant clusters into 8, 6, 3, and 2 groups, then combines them finally into one big group. The 8 and 6-way clustering cannot be interpreted in any phonologically meaningful sense for our purposes. The 3-way clustering, (1) [c.m], [j.t^h], [j.m], [c.t^h], (2) [g.m], [k.m], (3) [g.t^h], [l.n], [l.t^h], [k.t^h], on the other hand, does not provide sufficient evidence against the COI hypothesis and the Phonological processing hypothesis because there is still a difference between the middle group (i.e., [g.m] and [k.m]), and the lower group (the strident clusters). Furthermore, not all clusters with voiced segments occur in the middle group (i.e., [g.t^h] is grouped with the clusters that induced high d' scores), suggesting that the voicing information on C₁ does not matter for the Korean listeners, which rules out the Phonetic Processing hypothesis. As mentioned before, another cluster that induces a contact violation, namely [l.n], has been discriminated by the Korean listeners as well as the English group thus the middle group cannot be said to indicate the

validity of the CC Hypothesis. The two-way grouping groups the test clusters in a more homogenous way such that all the strident clusters are clustered on one side and the rest of the test clusters on the other. This grouping matches the predictions of the last column of the evaluation table provided in Table 3 above, where only strident codas are expected to be bad. This confirms only the Coda/Onset Identity (COI) hypothesis in combination with the Phonological processing hypothesis.

The [l]-clusters, namely, [l.n] and [l.t^h], form a replication of the same question that [k.m] and [k.t^h] aimed to test in the study. In the following, we will use these clusters as an independent test for the evaluation of the main hypotheses. According to the CC hypothesis, [ln] is expected to be problematic while both clusters should be successfully discriminated under the COI hypothesis. This is because [l] is a permissible coda consonant. An analysis of the lower and upper bounds for the clusters [l.n] and [l.t^h] within 95% confidence rate reveal that these clusters are not significantly different from each other. In addition, both the English and the Korean listeners do not differ from one another on these clusters. A power analysis reveals that any real difference in d' scores for these clusters must be less than 0.6 units of d' within 90% confidence (or 0.8 units of d' with 99% confidence) if such a difference existed. The statistics on the [l]-clusters, thus, provide no support for the CC hypothesis with such small probable differences, yet again support the COI hypothesis.

Furthermore, the voiced coda consonants do not cause perceptual epenthesis in the Korean group, suggesting that the voicing information is suppressed in speech perception. In particular, [j] was mapped to /c/, which induced perceptual epenthesis because it is still a strident. When [g] was mapped to /k/, however, it did not provoke

epenthesis since [k] is a permissible coda in Korean. This supports phonological processing of features based on their abstract underspecified representations in the L1 of the listener.

It should be remembered that to guard the conclusions on perceptual epenthesis against other types of phonological adjustments such as lateralization and nasalization, we also included a number of doublets which compared [k.m] and [l.n] with their likely output forms in the Korean production grammar, [ŋ.m] and [l.l] or [n.n], respectively. The mean d' scores for these pairs revealed that they are all relatively similar and very high in both the English (mean d' = 3.72) and the Korean (mean d' = 3.59) groups, indicating that the Korean listeners did not confuse the illicit consonant clusters with their likely output forms in Korean. Table 7 provides the mean d' scores for these items.

----INSERT TABLE 7 ABOUT HERE----

There are no statistically significant differences between the English and Korean performances on these clusters, suggesting that not all phonological processes are relevant for the Korean listeners' perception, an important finding that provides evidence against the P-map hypothesis (Steriade, 2001a). Our study suggested that the difference between [k.m] and [ŋ.m], and [l.n] and [l.l] was very noticeable to Korean listeners; nevertheless, the alternation from these input forms to output forms is made in Korean production, a point we will come back in the following sections.

DISCUSSION

In this study, we have motivated two fundamentally different hypotheses based on two different approaches to phonotactics, which make crucially different predictions about the way L1 phonotactic knowledge influences the perception of L2 consonant clusters. The results of our experiment showed that the Korean listeners can distinguish a given consonantal sequence C_1C_2 containing a permissible coda consonant (C_1) from its epenthetically adjusted counterpart (i.e., C_1VC_2) regardless of whether the following consonant (C_2) is a plosive or a nasal sound. Consequently, clusters (e.g. [k.m]) that never occur in Korean can nonetheless be successfully perceived by Korean listeners. The response patterns can only be explained if we assume that the L1 syllable structure constraints, rather than contact violations, influence the perception of consonant clusters. This is in support of the COI hypothesis, which is syllabically motivated, and in support of those psycholinguistic models that recognizes the role of syllables (e.g., Cutler & Norris, 1998; Cutler, Demuth & McQueen, 2002).

Perhaps, the most crucial finding of the present study is that not all bad contact cases behave the same in Korean perception, which is summarized in Table 8, where line (==) indicates the separation into two clusters determined by the Cluster Analysis.

---INSERT TABLE 8 ABOUT HERE---

These results also constitute important evidence against views that attribute perceptual preference for certain consonant clusters to the frequency with which those clusters occur in the language thus to the listeners' differing experience with them (e.g., Vitevitch and Luce, 1998; 1999). If perceptual epenthesis were a means by which the perceptual system

biases processing of clusters that have zero frequency, then all the illicit consonant clusters in the present study would be more susceptible to epenthesis. Our findings, however, demonstrate that Korean listeners' exhibit poor performance on only a certain set of consonant clusters (i.e., the strident cases) although all the illicit consonantal sequences have zero frequency of occurrence in Korean production. It is instructive to quote Kim-Renaud (1995: 223) here: "a sequence of l and n in either order is absolutely not permissible on the surface in Korean, and a complete assimilation of n to l occurs whenever such a sequence arises". The same also goes for k and m. Consequently, a phonological influence of L1 phonotactic knowledge, rather than an effect of frequency, best explain the Korean groups' performance. Our findings are consistent with those of Moreton (2002), who argues for structural (featural) differences, rather than frequency differences, influencing English listeners' perceptual biases against certain illicit consonant clusters such as [dl] compared to [bw], both of which have zero frequency in English.

Finally, we have demonstrated that the processes of nasalization and lateralization do not play the same role as epenthesis in perception such that the illicit consonant clusters are not misperceived as their likely output forms, as we have seen in Table 4 above. In the case of comparably impossible clusters that contained [strident] onsets, our results suggested that Korean listeners did perceptually change the input to fit into the L1 sound patterns. It is instructive here to note that the perceptual strategy that was employed by Korean listeners with [strident] codas is presumably not the same as in their production system. This is because all [strident] codas are neutralized in Korean. It should be noted that although we did not include any test pair to compare [p^hact^ha] or

[p^hacma] with their neutralized counterparts: [p^hat't^ha] and [p^hatma]/[panma], respectively, we can confidently say that the Korean listeners should be able to distinguish them given that the discriminability scores between pairs containing illicit sequences of consonants and their likely surface realizations (e.g., [pakma] vs. [paŋma]) were very high. Indeed, one reasonable interpretation of these results is that stridency is so salient that Korean listeners could not suppress it. Hearing stridency can be said to make Korean listeners place the strident segment in an onset position, which automatically evokes perceptual epenthesis. The different mechanisms employed by the perceptual and production systems are graphically illustrated in Figure 5.⁷

---INSERT FIGURE 5 ABOUT HERE---

Having demonstrated that perceptual phenomena are not simple inversions of phonological phenomena in the production system, the findings of the present study are difficult to interpret in models that incorporate listeners' knowledge of perceptibility of sound contrasts to predict phonological alternation, the centerpiece of Steriade's model (e.g., Steriade 1999, 2001 a, b), which we will discuss in the following section.

THE P-MAP HYPOTHESIS

The P-map Hypothesis (Steriade 2001 a, b) has been proposed to account for directional asymmetries in phonological processes based on an assumption that less perceptible

⁷ Naturally, the assumed intermediate representation is different in production and perception. While in production, the feature [strident] is associated with a coda position and then undergoes change, the same feature is never in a coda position in perception. Rather, the original stimulus, that is, the English production, has the feature [strident] linked to a coda.

changes are cross-linguistically preferred since they involve repair strategies that mediate between a maximally similar underlying representation and a surface form. According to Steriade, speakers have the knowledge of the listener's ability to perceive differences and such knowledge of perceptibility contrasts are encoded in the phonology as similarity rates. That is, speakers choose less noticeable repairs since they know they are less noticeable by listeners. The P-map hypothesis has its roots in Steriade's (1999: 4) Licensing by Cue hypothesis: "the likelihood that distinctive values of the F-contrast will occur in a given context is a function of the *relative perceptibility* of the F-contrast in that context" (emphasis added). D' scores give a direct measurement of relative perceptibility, allowing a direct test of the Licensing by Cue hypothesis. Accordingly, the probability that a consonant undergoes a phonological process (such as deletion, triggering epenthesis, or blocking vowel deletion) correlates with the quality and quantity of the auditory/phonetic cues associated with the contrast in a given context. For instance, voiceless and voiced obstruents are neutralized before obstruents and word-finally as opposed to intervocalic contexts because the contrast between the two is more perceptible in intervocalic contexts as more of the acoustic cues to voicing such as formant values of adjacent vowels, closure duration, VOT value, are available. For Steriade, voicing is likely to be maintained or lost depending on the speakers' precise knowledge of the perceptibility of voicing contrasts in particular phonological contexts that make the contrast more or less salient for the listener. This knowledge is the P-map, "the repository of speakers' knowledge, rooted in observation and inference, that certain contrasts are more discriminable than others, and that the same contrast is more salient in some positions than others" (Steriade 2001b: 236).

Testing the P-Map

According to Steriade's (2001b: 222) "Perceptual Similarity to Input" idea, "the likelihood that a lexical representation R will be realized as R' is a function of the perceived similarity between R and R'". Since Steriade explicitly claims that perceptual similarity is psychologically real and rooted in observation and inference, we can maintain the position that it is empirically testable. Indeed, the degree of similarity between any two features, by definition, should be inversely related to their discriminability. The methodology employed in the present study can, therefore, constitute an empirical test for the validity of Steriade's statement. Essentially, the discriminability indices that were obtained from the present study are the inverse of 'perceived similarity'. The SDT analysis computes a discriminability value by comparing hit rates with false alarm rates. Thus, "perceived similarity" can be calculated with reference to the index of discriminability, which must be the inverse of similarity. The mathematical model in (9) illustrates the conversion of discriminability scores on [k.m] vs. [ŋ.m] into "perceived similarity" scores:

$$(9) \quad P([\dots\eta.m\dots] | /...k.m.../) \sim \frac{1}{10^{d'([\dots\eta.m\dots], [\dots k.m\dots])}}$$

Accordingly, the probability of producing [ŋ.m] given /k.m/ is proportional to the reciprocal of the discriminability score obtained on [ŋ.m] vs. [k.m]. It should be noted

that since division by 0 is undefined, 10 is used as the base of the exponent for correction because if d' is 0, the probability must be 1. Since discriminability is the inverse of similarity according to the model we have employed, the probability approaches 1 when a given d' score approaches 0. Our findings showed that Korean group's d' scores on [p^hakma] vs. [p^haŋma] (or vice versa) is 3.73. According to (9), if d' ([p^hakma] vs. [p^haŋma]) $\gg 0$ then $p([p^haŋma] | /p^hakma/) ≈ 0 . Thus, the /k.m/ \rightarrow [ŋ.m] process in Korean cannot be explained by Steriade's notion of *Perceptual Similarity to Input*.$

The same equation can also be applied to [l.n] vs. [l.l]. Our results show that the Korean group's d' scores on this pair is also very high (3.56). Such a high score results in probability scores that are very close to 0, which again contradicts Steriade's hypothesis. In summary, if assimilatory processes were predicted by an index of perceived similarity between the assimilated variant and the underlying representation, Korean listeners should confuse clusters such as [k.m] with its likely output form [ŋ.m] in Korean. The way loan words containing [k.m] and [l.n] is adapted in Korean gives us every reason to expect that native phonological rules may affect perception. It turns out however that neither epenthesis nor native phonological rules affect the perceptual processing of such strings. Thus, the best working hypothesis is that violations involving syllables structure instead of consonantal contact affect perception, and neither nasalization nor lateralization have any basis in perception.

Finally, we observed that strident codas are perceptually altered via epenthesis by Korean listeners, which suggests that they are indistinguishable from onset stridents. However, no such phonological alteration is attested in the synchronic phonology of Korean. It should be noted that the phonetic properties of stridency and perceptual

epenthesis are confounded in our study since the only cases that invoked perceptual epenthesis involved a strident segment and the segments we employed in the experiment were strident. Thus, one could suggest that the distinction between [c] and [ci] is minimal, thus increasing the likelihood of perceiving an epenthetic vowel according to the P-map hypothesis. However, we know of no cross-linguistic tendency that particularly requires stridents to be prevocalic. More importantly, Korean phonology alone does not employ epenthesis to salvage stridency in the coda position although it is phonetically very salient but rather it removes it from the surface representation through neutralization. However minimal the perceptual similarity between [c] and [ci] may be, our perceptual results do not suggest that the neutralization of strident segments in the coda position is based in perception. Nor do they explain lateralization and nasalization processes in Korean. It is also true that stridency is particularly salient acoustically; we leave for future research tests of other languages with coda restrictions on other features.

SUMMARY AND CONCLUSIONS

In this article, based on a perceptual experiment, we have demonstrated that language-specific co-occurrence restrictions do not explain the perceptual epenthesis phenomenon in words containing sequentially illicit consonantal sequences. The results obtained from the study suggest that perceptual epenthesis is evoked when the members of the illicit sequences incur a syllable structure violation. The interpretation of our results, therefore, constitutes evidence against theories and analyses that use syllable-independent and linear statements to explain the consonantal distributions (e.g., Dziubalska-Kolaczyk, 1994; Steriade 1999; 2001a; Blevins, 2002). Another important finding regards the

underspecified nature of L2 percepts. We have claimed that L2 representations suspend featural information in the stimuli if the detected features correspond to those that are underspecified in the L1. Accordingly, the present study demonstrated that [voice] was suspended in the representations of Korean listeners.

In addition, we have demonstrated that not all phonological phenomena have relevance to speech perception, as processes such as neutralization, lateralization, and nasalization were not employed by Korean listeners to repair misplaced sequences of consonants which would be altered by the application of these processes according to Korean phonology. This finding provides evidence against models that prioritize the role of speech perception in explaining synchronic phonological processes (Steriade, 1999; 2001a, b). Equally importantly, we have shown that the effects we found with respect to Korean speakers' perception of epenthetic vowels cannot be explained by the frequency of the sequences that are involved since certain consonant clusters ([k.m] and [l.n]) were successfully discriminated by Korean listeners although they have zero probability of occurrence in Korean.

REFERENCES

- Archangeli, D. (1984). *Underspecification in Yawelmani phonology and morphology*. Ph.D. Dissertation (published in Garland in 1988). Massachusetts Institute of Technology.
- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183-207 (1988).
- Blevins, J. (2002). The independent nature of phonotactic constraints: an alternative to syllable-based approaches. In C. Féry and G. van de Weijer (Eds.), *The syllable in Optimality Theory* (pp. 375-403). Cambridge: Cambridge University Press.

- Cohen J.D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25, 257- 271.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.
- Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science*, 13, 258-262.
- Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience*, 12, 635-647.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. New York: Cambridge University Press.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1568-1578.
- Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, 16, 491-505.
- Dziubalska-Kolaczyk, K. (1994). Acquisitional evidence against the phonological syllable. *Papers and Studies in Contrastive Linguistics*, 29, 17-30.
- Ewen, C., & van der Hulst, H. (2001). *The Phonological structure of words: an introduction*. Cambridge: Cambridge University Press.
- Green, D. M., & Swets, J. A. (1974). *Signal Detection Theory and Psychophysics*. Huntington, New York: Krieger.
- Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation?. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 592-608.
- Itô, J. (1986). *Syllable theory in prosodic phonology*. Unpublished Ph.D. Dissertation. University of Massachusetts-Amherst.
- Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language and Linguistic Theory*, 7, 217-259.
- Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V., & Jusczyk, A. (1993). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language*, 32, 402-420.

- Jusczyk, P., Luce, P., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, 33, 630-645.
- Kabak, B. (2003). *The perceptual processing of second language consonant clusters*. Unpublished Ph.D. Dissertation. University of Delaware.
- Kabak, B., & Idsardi, W. (2003). Syllabically conditioned perceptual epenthesis. *Proceedings of the Berkeley Linguistic Society*, 29.
- Kadlec, H. (1999). Statistical properties of d' and [beta] estimates of signal detection theory. *Psychological Methods*, 4, 22-43.
- Kang, Y. (2001). Adaptation of English post-vocalic word-final stops to Korean loanwords. Paper presented at NELS 32. City University of New York/ New York University.
- Kim-Renaud, Y-K. (1995). *Korean consonantal phonology*. Seoul: Hanshin.
- Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: a phonological approach to recognition of lexicon. *Cognition*, 38, 245-294.
- Lahiri, A., & Marslen-Wilson, W. D. (1992). Lexical processing and phonological representation. In G.J. Docherty and D.R. Ladd (Eds.), *Papers in laboratory phonology II* (pp. 229-254). Cambridge: Cambridge University Press.
- Lahiri, A., & Reetz, H. (1999). The FUL speech recognition system. *Journal of the Acoustical Society of America*, 105: 1091.
- Lamontagne, G. (1993). *Syllabification and consonant cooccurrence restrictions*. Unpublished Ph.D. Dissertation, University of Massachusetts-Amherst.
- Lee, I. & Ramsey, S. R. (2000). *The Korean language*. Albany: State University of New York Press.
- Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, 19, 1-36.
- Martin, S. (1992). *A reference grammar of Korean*. Rutland: Charles E. Tuttle Company.
- Massaro, D. W. & Cohen, M. M. (1983). Phonological constraints in speech perception. *Perception & Psychophysics*, 34, 338-348.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.
- Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, 84, 55-71.

- Pitt, M. A. (1998). Phonological processes and the perception of phonetically illegal consonant clusters. *Perception and Psychophysics*, 60, 941-951.
- Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated between monosyllabic spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 378-387.
- Silverman, D. (1992). Multiple scansions in loanword phonology: Evidence from Cantonese. *Phonology*, 9, 289-328.
- Sohn, H-M. (1994). *Korean*. New York: Routledge.
- Sohn, H-M. (1999). *The Korean language*. Cambridge: Cambridge University Press.
- Steriade, D. (1999). Alternatives to syllabic analysis of consonantal phonotactics. In O. Fujimura and B. Palek (Eds.), *Proceedings of the 2nd Linguistics and Phonetics Conference* (pp. 205-245). Prague: Charles University Press.
- Steriade, D. (2001a). The Phonology of perceptibility effects: The P-map and its consequences for constraint organization. Unpublished Manuscript. University of California, Los Angeles.
- Steriade, D. (2001b). Directional asymmetries in place assimilation: a perceptual account; In E. Hume and K. Johnson, *The role of speech perception in phonology* (pp. 219-250). New York: Academic Press.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: levels of processing in perception of spoken words. *Psychological Science*, 9, 325-329.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374-408.

TABLES

Table 1: Consonantal inventory of Korean

		Labial	Alveolar	Alveo-palatal	Velar	Glottal
Plosives	Plain	p	t	c	k	
	Tense	p'	t'	c'	k'	
	Aspirated	p ^h	t ^h	c ^h	k ^h	
Fricative	Plain		s			h
	Tense		s'			
Nasal		m	n		ŋ	
Liquid			l			

Table 2: Vowel inventory of Korean

	-Back		+Back	
	-Round	+Round	-Round	+Round
High	i	y	ɯ	u
Mid	e	ø	ə	o
Low	ɛ		a	

Table 3: Status of test clusters in Korean

		C_2		
		<i>Oral Stop (i.e., [t^h])</i>	<i>Nasal (i.e., [n]or[m])</i>	
C_1	Licit	[k]	Licit	Illicit
		[l]	Licit	Illicit
	Illicit	[c]	Illicit	Illicit
		[j]	Illicit	Illicit
		[g]	Illicit	Illicit

Table 4: Test words used in the experiment

<i>Cluster</i>	<i>Cluster-words</i> ([CVC ₁ C ₂ V])	<i>Epenthesis-words</i> ([CVC ₁ VC ₂ V])	<i>Example English word</i> <i>with the cluster</i>
[kt ^h]	[p ^h ákt ^h a]	[p ^h ák ^h ut ^h a]	<i>doctor, factory</i>
[km]	[p ^h ákma]	[p ^h ák ^h uma]	<i>acme</i>
[gt ^h]	[p ^h ágt ^h a]	[p ^h águt ^h a]	<i>pigtail, ragtime</i>
[gm]	[p ^h ágma]	[p ^h águma]	<i>dogma, segment</i>
[ct ^h]	[p ^h áct ^h a]	[p ^h ác ^h it ^h a]	<i>pitch-tracker</i>
[cm]	[p ^h ácma]	[p ^h ác ^h ima]	<i>Richmond, attachment</i>
[jt ^h]	[p ^h ájt ^h a]	[p ^h ájit ^h a]	<i>vegetable (dialectal)</i>
[jm]	[p ^h ájma]	[p ^h ájima]	<i>arrangement</i>
[lt ^h]	[p ^h ált ^h a]	[p ^h álut ^h a]	<i>saltines, shelter</i>
[ln]	[p ^h álna]	[p ^h álna]	<i>walnut, vulnerable</i>
[ŋm]	[p ^h áŋma]	-	<i>Ingmar, kingmaker</i>
[nn]	[p ^h ánna]	-	<i>pinenut⁸</i>
[ll]	[p ^h állla]	-	<i>mail-list</i>

⁸ Variation in geminate reduction can be observed in the pronunciation of these words in English.

Table 5: Discrimination predictions

	CC Hypothesis		COI Hypothesis	
	Phonetic	Phonological	Phonetic	Phonological
k.t^h				
g.t^h	X		X	
k.m	X	X		
g.m	X	X	X	
l.n	X	X		
l.t^h				
c.t^h	X	X	X	X
c.m	X	X	X	X
j.t^h	X	X	X	X
j.m	X	X	X	X

Table 6: D' performance limits

	Ceiling	Floor
Hits (H)	10	0.5
Misses (M)	0.5	10
False Alarms (FA)	0.5	0.5
Correct Rejections (CR)	20	20
p(H)	0.952	0.048
p(FA)	0.024	0.024
z(H)	1.668	-1.668
z(FA)	-1.970	-1.970
z(H)-z(FA)	3.639	0.302
p(c)	0.966	0.560
d'	4.216	0.898
Beta	1.733	1.733

Table 7: Korean mean d' scores for the discrimination of test clusters with their likely output forms

Test Doublet	mean d' score	
	Korean	English
p ^h akma vs. p ^h aŋma	3.73	3.69
p ^h alna vs. p ^h alla	3.56	3.74
p ^h alna vs. p ^h anna	3.48	3.75

Table 8: Summary of Korean results

Sequence	Type	mean d'
kt ^h	OK	3.58
ln	Bad contact	3.42
lt ^h	OK	3.28
gt ^h	Bad contact	2.72
km	Bad contact	2.28
gm	Bad contact	1.80
<hr/>		
ct ^h	Bad coda	0.89
jm	Bad coda	0.48
jt ^h	Bad coda	0.47
cm	Bad coda	0.30

FIGURES

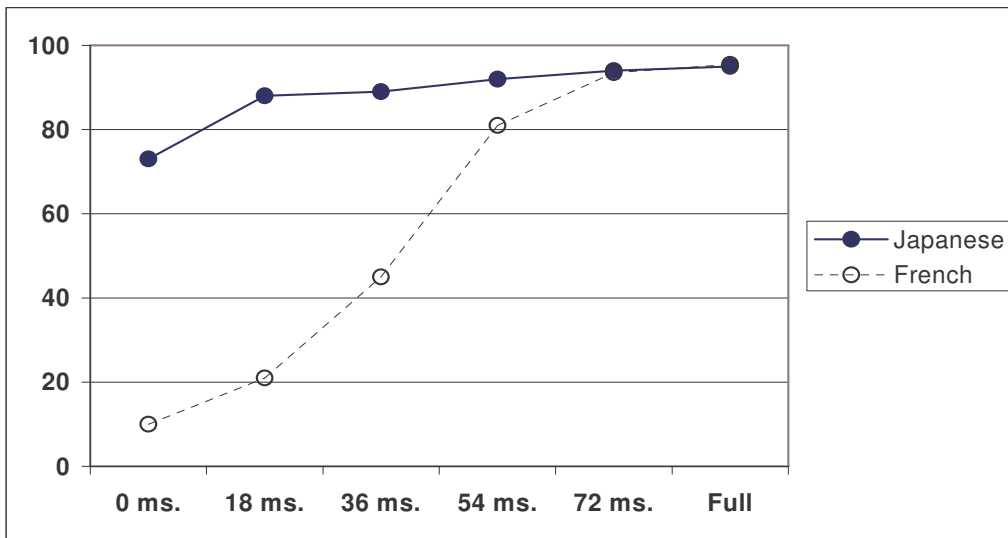


Figure 1: Percept [u] vowel judgments as a function of vowel duration (adapted from Dupoux et al., 1999).

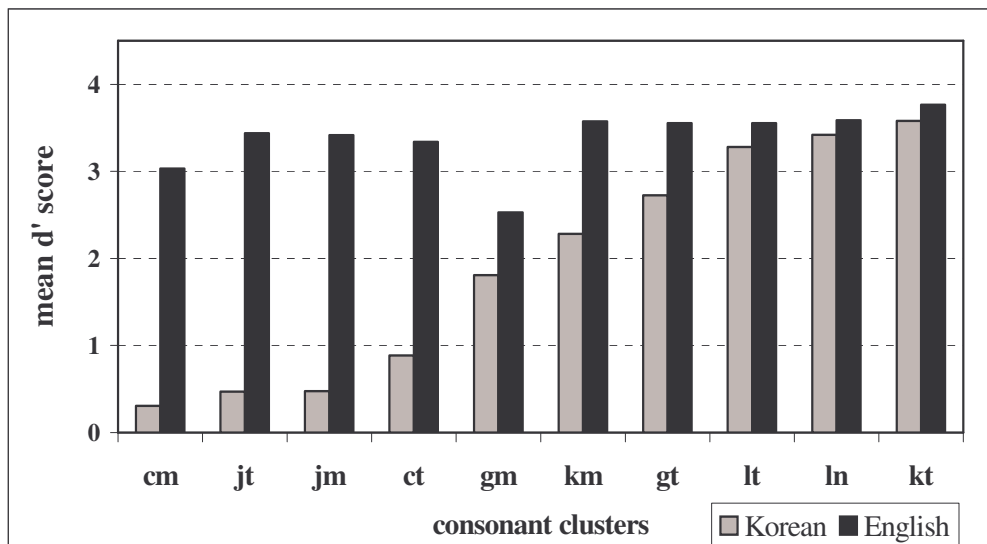


Figure 2: The mean d' scores for the discrimination of [CVCCV] vs. [CVCVCV].

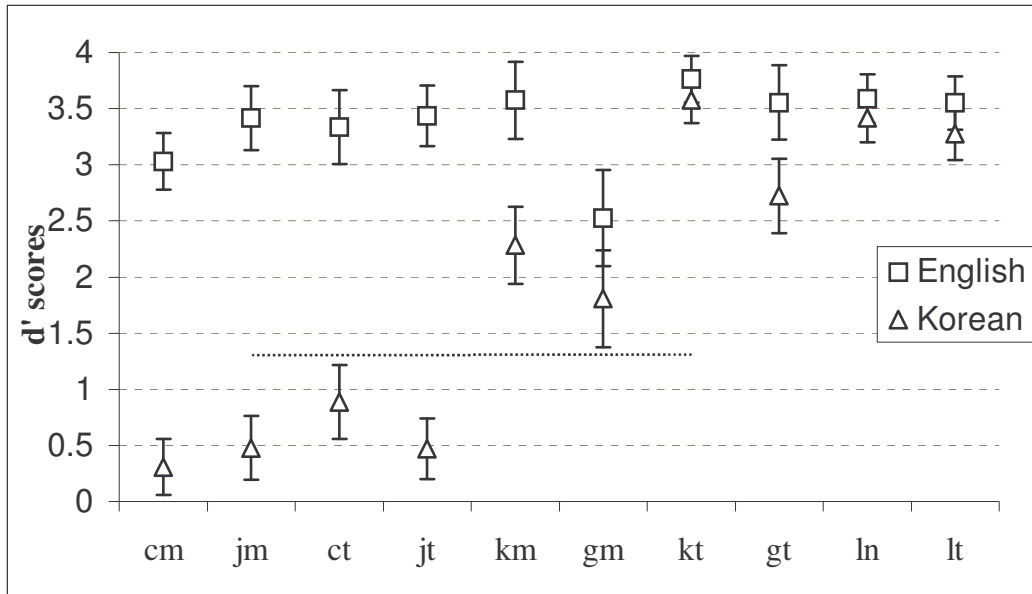


Figure 3: Korean and English mean d' scores for the discrimination of [CVCCV] vs. [CVCVCV] words (ordering by increasing d' values for Korean) with 95% confidence intervals. The line indicates that the confidence intervals for [ct] and [gm] in the Korean group do not overlap.

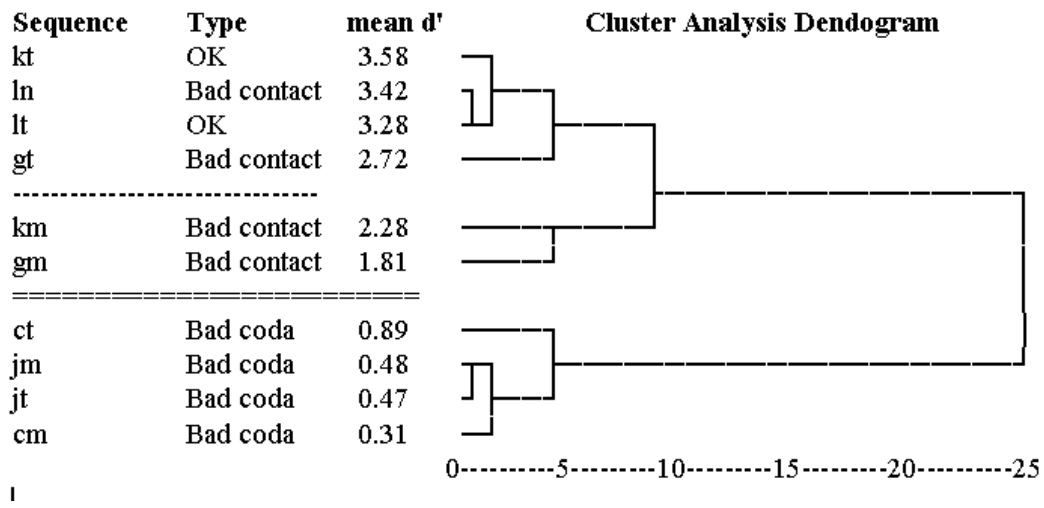


Figure 4: Korean d' cluster analysis dendogram

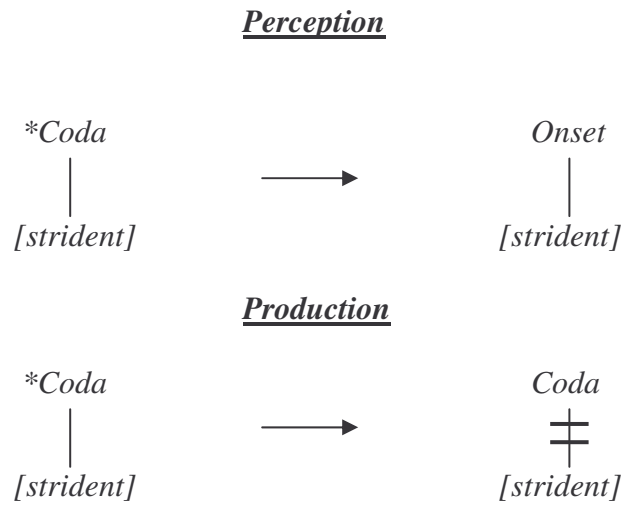


Figure 5: Korean perception vs. production grammar