A Theory of Consonant Cluster Perception and Vowel Epenthesis

by

Suyeon Yun

A.B., Seoul National University (2007)
A.M., Seoul National University (2010)

Submitted to the Department of Linguistics and Philosophy in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016

Signature redacted

Author ....
Department of Linguistics and Philosophy
September 9, 2016

Signature redacted

Certified by... ....
Donca Steriade
Professor of Linguistics
Thesis Supervisor

Signature redacted

Certified by.... ....
Edward Flemming
Associate Professor of Linguistics
Thesis Supervisor

Signature redacted

Accepted by . ...
David Pesetsky
Head, Department of Linguistics and Philosophy

A Theory of
Consonant Cluster Perception and Vowel Epenthesis

by
Suyeon Yun

Submitted to the Department of Linguistics and Philosophy
on September 9, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This dissertation concerns cluster-dependent asymmetries in vowel epenthesis in loanword adaptation and in non-native cluster perception. The central argument is that auditory factors affect the relative perceptual similarity between consonant clusters and the corresponding epenthesis forms, which in turn plays an important role in determining the site of epenthesis in loanword adaptation. This dissertation provides an extended typology of vowel epenthesis sites in consonant cluster adaptation, considering a variety of clusters both in word-initial and in word-final positions. It will be argued that the cluster-dependent asymmetries in epenthesis sites are best explained by the auditory properties of consonant clusters, such as intensity rise. Specifically, if a cluster involves an intensity rise inside the cluster, epenthesis occurs inside the cluster; if a cluster involves an intensity rise outside the cluster, epenthesis occurs outside the cluster; and if a cluster involves two intensity rises, either internal or external epenthesis can occur. I argue that this is because the epenthetic vowel insertion where there is an intensity rise makes a perceptually less salient change from the original cluster than epenthesis where there is no intensity rise, based on the P-map hypothesis (Steriade, 2008) that an output involving a perceptually smaller change is more optimal. The results of several perception experiments support the hypothesis by showing that not only intensity rise but also C_1 voicing have a significant effect on the perceptual similarity between the consonant clusters and the corresponding epenthesis forms. Crucially, it will be shown that the novel generalization about vowel epenthesis sites and the results of perception experiments employing phonetically diverse stimuli can be best explained by the auditory properties, and not by the sonority profile, which has traditionally been used to explain these data.

Thesis Supervisor: Donca Steriade
Title: Professor of Linguistics

Thesis Supervisor: Edward Flemming
Title: Associate Professor of Linguistics

In memory of my best friend, Hongsun Im (1981-2010)
and his relentless passion for linguistics

Acknowledgments

A number of individuals have contributions to this dissertation in a variety of ways. First and foremost, I am extremely grateful to my committee members: Donca Steriade, Edward Flemming, Michael Kenstowicz, and Adam Albright. It has been a great honor working with these wonderful phonologists all this time, and I cannot thank them enough for all that they have done for me.

This dissertation has also been benefited indirectly by the other teachers I worked with at MIT, including David Pesetsky, Norvin Richards, Shigeru Miyagawa, Martin Hackl, Maziar Toosarvandani, and Noah Constant, who incented me to widen my research on linguistics beyond phonetics and phonology.

I would like to thank Jongho Jun, my advisor at Seoul National University. He introduced phonology and linguistics to me and has always been my strongest supporter. I would also like to thank Yoonjung Kang, who warmly encouraged me to keep working on this topic and motivated me to complete this dissertation.

My sincere thanks also go to Stefanie Shattuck-Hufnagel and Elizabeth Choi in the Research Laboratory of Electronics at MIT, who provided me an opportunity to join their team. Working with them has been a great joy for me, and the insights I got from the project with them also benefited this dissertation.

This dissertation would not have been possible without the assistance of the Office of the Dean for Graduation Education at MIT and the MIT Medical. In particular, I would like to thank Dean Blanche Staton, who gave me both material and emotional support when help was really needed, and Lisa Bosley, who emphasized with me whenever I was in trouble, as well as provided me with medical support.

It is a particular pleasure to acknowledge my debt to speakers of many different languages who were willing to share the knowledge of their native language. In particular, I am grateful to Maria Khtomsky, Sasha Podobraev, Igor Yanovich, and Sam Alxatib for recording my experimental stimuli several times. I also thank all of my language consultants and experimental participants for their time and energy.

I would also like to thank my fellow graduate students at MIT linguistics, who shared this journey with me. I was so lucky to have such a great cohort (Michelle Fullwood, Ayaka Sugawara, Isaac Gould, Ryo Masuda, Ted Levin, Gretchen Kern, Coppe van Urk, Sam Steddy, and Wataru Uegaki) and officemates (Claire Halpert, Hrayr Khanjian, Yusuke Imanishi, Michelle Fullwood, and Amanda Swenson). Thanks as well to Despina Ikonomou, Juliet Stanton, Sam Zukoff, Lilla Magyar, Ishani Guha, Benjamin Storme, Paul Marty, Peter Graff, Sam Alxatib, mitcho Erlewine, Michelle Yuan, Erin Olson, Ezer Rasin, Greg Padowski, Masako and Yuika Imanishi, and Doggy for their direct or indirect support for this dissertation.

I would also like to thank my dearest friends outside MIT, Youmi Jung and Suzy Ahn, who were always there for me and encouraged me not to give up this journey.

Finally, I would like to extend my deepest thanks to my parents, Seungho Yoon and Yunhee Park, and my brother Kanghyun Yoon, without whose love, support, and understanding I could never have completed this doctoral degree. And a very special thanks to my beloved dog Sally, who kept me smiling in every moment.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Consonant clusters CC are often a target of phonological modification, such as vowel epenthesis (CəC), consonant deletion (C), feature change, metathesis, etc. This dissertation examines vowel epenthesis as a modification to underlying consonant clusters from the joint perspectives of phonological typology and speech perception, focusing on cluster-dependent asymmetries observed in vowel epenthesis. The core data is the typology of vowel epenthesis in loanword phonology. Vowel epenthesis often takes place when a consonant cluster of a source language is not allowed in the borrowing language. In many cases, an illegal cluster could in principle be repaired by inserting a vowel between the two consonants (internal epenthesis; e.g., #CC → #CəC) or outside of the cluster (external epenthesis; e.g., #CC → #əCC). What interests us is that for most clusters only one pattern of epenthesis is observed across languages. For example, if [#kl] is repaired by epenthesis, the vowel is always epenthesized between the consonants, i.e., [#kəl], and if [#mb] is repaired by epenthesis, the vowel is always epenthesized outside the cluster, i.e., [#əmb]. The principle thesis is that phonological vowel epenthesis tends to occur where illusory vowel epenthesis occurs in speech (mis)perception, and the difference between clusters with respect to their ability to undergo epenthesis is determined by the degree of perceptual similarity between the original cluster and the epenthesized output. It has been known that listeners may perceive an illusory vowel that does not exist when they hear a non-native consonant cluster (Dupoux et al., 1999, a.o.); so a non-native cluster #CC may be perceived as #CəC or as #əCC. The working hypothesis is that clusters that undergo internal epenthesis, e.g., [#kl], are perceptually similar to the corresponding form with internal epenthesis, e.g., [#kəl], while clusters that undergo external epenthesis, e.g., [#mb], are perceptually similar to the corresponding form with external epenthesis, e.g., [#əmb], and this knowledge of perceptual similarity plays a role in forming the typology of vowel epenthesis.

The current perceptual similarity hypothesis is closely related to that of Fleischhacker (2001, 2005), in the sense that perceptual similarity in the acoustic signal is the determining factor of vowel epenthesis. In this respect, the current hypothesis is different from other similarity-based theories such as Steriade (2006) based on similarity in sonority profile, as well as from markedness-based theories such as Broselow (1992) and Gouskova (2001). Crucially, unlike previous similarity-based theories, the current theory identifies as the critical similarity factor that licenses epenthesis in consonant clusters the presence of an acoustic event in the phonetically realized input structure, intensity rise and C_1 voicing, in particular.

1.1 What is known about vowel epenthesis in consonant clusters

In the phonological literature, great attention has been given to cluster-dependent asymmetries in vowel epenthesis sites in loanword adaptation. Epenthesis may occur when a consonant cluster of a source language is phonotactically illegal in the borrowing language. For instance, English 'please' is adapted with an epenthetic vowel between the two consonants as [pɪlɪz] in Hindi (Singh, 1985). The most well-known finding about loan epenthesis in consonant clusters is that stop-sonorant clusters, such as [pl] in 'please', have an epenthetic vowel inserted between the two consonants in the cluster, whereas sibilant-stop clusters such as [sk] have an epenthetic vowel outside the cluster (Broselow, 1983, 1992; Singh, 1985; Fleischhacker, 2001). In Hindi, for example, 'school' beginning with the sibilant-stop [sk] is borrowed with an epenthetic vowel before the cluster as [ɪskul], in contrast with the cluster-internal epenthesis in 'please' (Singh, 1985).

The sonority profile of consonant clusters has been employed to predict the position of the epenthetic vowel in consonant cluster adaptation. For instance, Gouskova (2001) has argued, based on Russian loanwords in Kirghiz, that clusters involving a sonority rise, such as [#pl], are more likely to allow vowel epenthesis between the two consonants in the cluster than clusters involving a sonority plateau or fall, such as [#lb], which are likely to undergo external epenthesis outside the cluster. According to Gouskova (2001), this asymmetry originates from markedness over a syllable boundary. That is, a sonority rise across a syllable boundary is cross-linguistically marked (Syllable Contact Law; Murray and Vennemann, 1983; Vennemann, 1988), and internal epenthesis occurs between the two consonants, e.g., /#pl.../ → [#pə.l...], to avoid a marked structure, e.g., [#əp.l...], which would have resulted from external epenthesis. In contrast, external epenthesis occurs in clusters with level or falling sonority, e.g., /#lb.../ → [#əl.b...], because it does not create a sonority rise across the syllable boundary. In other words, the default site of epenthesis is cluster-external, and internal epenthesis is possible when the resulting structure violates the Syllable Contact Law.

On the other hand, Fleischhacker (2001, 2005) explains the asymmetry between stop-sonorant and sibilant-stop clusters based on perceptual similarity between the consonant cluster and the resulting epenthesized form. In particular, stop-sonorant clusters such as [#pl], are perceptually similar to the form with epenthesis, e.g., [#pəl], and this is why stop-sonorant clusters are adapted with internal epenthesis in many languages. In contrast, sibilant-stop clusters are often borrowed with external epenthesis because epenthesis in sibilant-stop clusters, e.g., /#sk/ → [#sək], yields a perceptually salient change, according to Fleischhacker. Fleischhacker (2001) also suggests an implicational hierarchy of vowel epenthesis in word-initial obstruent-initial clusters and argues that it also reflects the hierarchy of perceptual similarity between original clusters and repaired epenthesis forms, as presented in (1). That is, stop-sonorant (TR) clusters are the most likely to be split by an epenthetic vowel, followed by sibilant-liquid (SL), sibilant-nasal (SN), and sibilant-stop (ST) clusters in order, because #TR and #TəR are perceptually most similar to each other, and #ST and #SəT are the least.

(1) Partial implicational hierarchy in Fleischhacker (2001, 2005)
(S: sibilant, T: stop, N: nasal, L: liquid, R: sonorant)

	#ST		#SN		#SL		#TR
internal epenthesis CC → CəC		$<$		$<$		$<$	
perceptual similarity between CC-CəC		$<$		$<$		$<$	

Fleischhacker's (2001, 2005) generalization about epenthesis in obstruent-initial clusters is implemented employing the idea of sonority profile in Steriade (2006).Roughly speaking, /#pl.../ involves a sonority rise, and its corresponding form with internal epenthesis [#pəl...] also involves a sonority rise, albeit a bit steeper. So the change from the shallow sonority rise in /#pl/ to the steep sonority rise in [#pəl] is perceptually small, resulting in internal epenthesis. In /#sk.../, however, the sonority fall in /sk/ is perceptually very different from the sonority rise in [#sək] with internal epenthesis, and thus speakers are reluctant to apply internal epenthesis in ST clusters.

Sonority profile has also been argued to play an important role in consonant cluster perception. It is well known that speakers often misperceive phonological structures that are not attested in their language, and the misperception of a non-native consonant cluster CC as an epenthesized sequence CəC has attracted a considerable amount of attention. For example, *ebzo*, produced by a French speaker, is often misperceived as *ebuzo* with an illusory vowel by Japanese speakers, because such a consonant sequence is not legal in Japanese (Dupoux et al., 1999). Of special interest is that the likelihood of perceptual epenthesis may differ depending on the type of consonant cluster. Berent et al. (2007, 2008, 2009, *et seq.)* have argued that the likelihood of perceptual epenthesis is determined by the universal markedness of onset clusters, particularly by the sonority profile of the cluster. This is based on their experimental evidence that universally less marked clusters with rising sonority, e.g., *blif*, are more accurately perceived than universally more marked clusters with level sonority, e.g., *bdif*, which in turn are more accurately perceived than even more marked clusters with falling sonority, e.g., *lbif*.

On the other hand, there is a growing body of evidence that the phonetic details of a cluster have an influence on consonant cluster perception (Dupoux et al., 1999; Fleischhacker, 2001, 2005; Wilson et al., 2014; Durvasula and Kahng, 2016, a.o.). Wilson et al. (2014), for example, show that the longer the release duration is and the stronger the release amplitude is, the more likely a vowel is inserted after a cluster-initial stop in English speakers' non-native cluster production. Also, Durvasula and Kahng (2016) show that the longer the duration of closure voicing is, the more likely Korean listeners hear an illusory vowel after a cluster-initial stop.

1.2 What is not known about vowel epenthesis in consonant clusters

Although vowel epenthesis in consonant cluster adaptation and in speech perception has been studied from different points of view, there are still questions that need to be answered. First, previous studies have focused on vowel epenthesis in word-initial position, and little is known about vowel epenthesis in non-initial positions. What is the typological pattern of vowel epenthesis in non-initial consonant clusters? Is the same cluster-dependent asymmetry in epenthesis sites observed in non-initial positions as in word-initial position? What is the pattern of perceptual epenthesis in non-native clusters in non-initial positions? Do the loan vowel epenthesis and perceptual illusory vowel epenthesis pattern in the same fashion?

Once we figure out the pattern of vowel epenthesis in non-initial clusters, we need to investigate what the decisive factors that govern epenthesis are both in word-initial and in non-initial positions. Is it markedness or faithfulness that determines the epenthesis site? If it is markedness, is it related to sonority contour at a syllable boundary (Gouskova, 2001) or universal markedness within a syllable (Berent et al., 2007)? Or if it is faithfulness, is it sonority profile (Steriade, 2006) or a purely auditory property (Fleischhacker, 2001, 2005) to which the faithfulness is activated? Does the factor that determines cross-linguistic epenthesis patterns also govern perceptual epenthesis? This dissertation aims to answer these questions.

1.3 Overview of the dissertation

The first goal of this dissertation is to provide an extended typology of epenthesis sites in loanword adaptation, by considering various types of clusters, i.e., not only obstruent-initial but sonorant-initial clusters, both word-initially and word-finally. The present, larger typology will provide new generalizations about loan epenthesis sites, and uncover several interesting asymmetries that cannot be explained by the hypotheses relying on sonority contour in the previous studies. This dissertation will argue that the current typology of cluster-dependent asymmetries in epenthesis sites is best explained by auditory properties of consonant clusters, such as intensity rise. That is, vowel epenthesis occurs where there is a rise in intensity, which is schematically represented in (2). Specifically, if a consonant cluster involves an intensity rise, indicated by the arrow "↗", between the two consonants in the cluster, as in (2a), internal epenthesis occurs; if a cluster involves an intensity rise outside the cluster but not inside the cluster, as in (2b), external epenthesis occurs; if a cluster involves two intensity rises, one inside and the other outside the cluster, as in (2c), either internal or external epenthesis can occur.

(2) Consonant cluster types and epenthesis sites in loanword adaptation

	cluster	→	*epenthesis*
a.	#C↗C	→	#CəC
	C↗C#	→	CəC#
b.	#↗CC	→	#əCC
	CC↗#	→	CCə#
c.	#↗C↗C	→	#CəC~#əCC
	C↗C ↗#	→	CəC#~CCə#

I argue these cross-linguistic generalizations about epenthesis sites are based on the relative perceptual similarity between the consonant clusters and their corresponding vowel epenthesis forms. In other words, the epenthetic vowel inserted where there is an intensity rise makes a perceptually less salient change from the original cluster than epenthesis where there is no intensity rise, and this results in epenthesis at an intensity rise, based on the P-map hypothesis (Steriade, 2008) that an output involving a perceptually smaller change is more optimal. This hypothesis is supported by the results of several discrimination and identification experiments. This finding is intriguing because it shows that speakers' knowledge of perceptual similarity is actively applied to the adaptation of novel forms, and that this knowledge is applied in comparable ways across different languages.

This dissertation owes a great deal to Fleischhacker (2001, 2005), in that it explains cluster-dependent asymmetries in loan epenthesis based on perceptual similarity between the cluster and its matching epenthesis form. This dissertation, however, covers a broader range of typological data than Fleischhacker (2001, 2005), who focused on word-initial obstruent-initial clusters, and provides a more general account of the larger typology employing auditory features that can universally be applied to any kind of cluster.

The auditory property, intensity rise, will also be shown to affect the (mis)perception of an illusory vowel in consonant clusters in several perception experiments reported in this dissertation. It will also be shown that voicing of the first consonant in a cluster, as well as intensity rise, has a significant effect on the discrimination between consonant clusters and epenthesis forms. This dissertation also investigates the relative contribution of phonological factors like the sonority profile and auditory factors, i.e., intensity rise and C_1 voicing, in the perception of consonant clusters, by conducting several perception experiments employing diverse stimuli in terms of acoustic properties as well as sonority profiles. Crucially, it will be shown that it is auditory factors that significantly affect consonant cluster perception and vowel epenthesis, not the sonority profile.

1.4 Structure of the dissertation

The rest of this dissertation is organized as follows. In Chapter 2, I first introduce perceptual and acoustic bases of the auditory factors, intensity rise and C_1 voicing, that affect the perceptual similarity to the vowel-epenthesized forms, and provide experimental results that support the hypothesis that consonant clusters CC that involve an intensity rise or C_1 voicing are perceptually more similar to the vowel-epenthesized form CəC, compared to clusters that lack these properties. Chapter 3 presents the results of a typological survey of epenthesis sites in loan adaptation and suggests that the epenthesis typology can be explained by the presence or absence of intensity rise in the cluster. The results of similarity judgment and transcription experiments will provide further evidence that the relative perceptual similarity between consonant clusters and vowel epenthesis forms plays a crucial role in determining the site of epenthesis. Based on the P-map hypothesis, the epenthesis typology will be analyzed by the interaction between a perceptually-driven constraint ranking and other general markedness and faithfulness constraints. Chapter 4 discusses alternative analyses of cluster-dependent asymmetries in vowel epenthesis in loanword adaptation and cluster perception based on sonority sequencing and points out how they fail to generalize to the full set of data. Chapter 5 summarizes and concludes the dissertation.

Chapter 2

Auditory Factors Favoring Vowel Similarity

2.1 Perceptual and Acoustic Bases

This chapter concerns key auditory features that influence consonant cluster perception. I will argue that perceptual vowel epenthesis is more likely to occur in a consonant cluster CC when the cluster is sufficiently similar in auditory perception to the vowel-epenthesized counterpart CəC. Among the many acoustic properties that can appear in a consonant cluster, this dissertation hypothesizes that the two critical factors helping to determine the perceptual similarity between CC and CəC are intensity rise and C_1 voicing, each of which will be discussed in the following subsections.

2.1.1 Intensity rise

First, I hypothesize that a rise in intensity plays a key role in perceptual epenthesis in consonant clusters. In other words, if there is a rise in intensity in the transition between the two consonants in a cluster, the cluster is perceptually similar to the corresponding sequence with an epenthetic vowel, CəC, and is thus more readily split with an epenthetic vowel in loanword adaptation. The first question we can ask about this hypothesis is why the presence of an intensity rise makes a consonant cluster CC perceptually more similar to CəC than a cluster without an intensity rise. This is because vowels in general have greater intensity than consonants, resulting in a rise in intensity from the initial consonant to the vowel in CəC. Therefore, the intensity contour, rising from the first to the second segment, is similar in a consonant cluster involving an intensity rise and in its corresponding epenthesis form. This can be schematically presented as in Figure 2-1. The consonant cluster in Figure 2-1a involves a rise in intensity between C_1 and C_2, and so does its epenthesized counterpart in Figure 2-1b between C_1 and the vowel [ə]. Although the magnitude of the intensity rise is greater in Figure 2-1b since the intensity of the vowel is greater than that of C_2, whatever the kind of C_2 is, the patterns of intensity rise in word-initial sequences of two segments that listeners hear from the very beginning of the utterance are remarkably alike. Therefore, it would be plausible to assume that a consonant cluster involving a rise in intensity is perceived as similar to the matching vowel-epenthesized form, which also involves a rise in intensity.

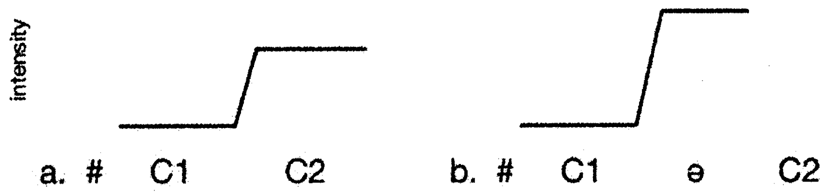


Figure 2-1: Schematic representations of the intensity contours in CC with an intensity rise and in CəC

Note that we assume here that what is relevant is the intensity rise present between C_1 and the epenthetic vowel, not the intensity fall between the vowel and C_2. It is true that vowel epenthesis introduces an intensity fall that is not present in a consonant cluster with rising intensity, as in Figure 2-1a. It is hypothesized here that the intensity rise between C_1 and the vowel is more important than the intensity fall between the vowel and C_2, because it is intuitive to assume that what listeners hear first, i.e., the intensity rise, plays a more important role in perception than what they hear next, i.e., the intensity fall. This would be especially true in this case since the window for the intensity rise is short and perhaps the auditory system cannot wait to get more information or cannot backtrack to override a prior decision. There is also literature claiming that an onset, i.e., rise in intensity, is more salient in the auditory system than an offset, i.e., fall in intensity. Phillips et al. (2002) show that sound onsets have a more elaborate neurophysiological representation and receive a heavier perceptual weighting than sound offsets. Also, Wright (2004) states that in CV sequences, the formant transitions of a vowel out of a consonant closure, as well as consonantal cues out of silence, receive a boost in auditory nerve fibres, while there is no equivalent boost in the formant transitions from a vowel to the following consonant in VC sequences.

On the other hand, Figure 2-2 represents the intensity contour of a consonant cluster not involving an intensity rise and that of its epenthesized counterpart CəC. Intensity decreases from C_1 to C_2 in the consonant cluster in Figure 2-2a. If a vowel is inserted between the two consonants as in Figure 2-2b, however, intensity increases from C_1 to the following epenthetic vowel because a vowel has higher intensity than a consonant, as mentioned above. Therefore, the intensity contours in general in Figure 2-2, one with a decrease and the other with an increase, are quite different from each other, compared to the intensity patterns in Figure 2-1, indicating that a consonant cluster with no intensity rise and the matching epenthesis form would be perceptually more distinct from each other than a consonant cluster with an intensity rise and its epenthesis counterpart.

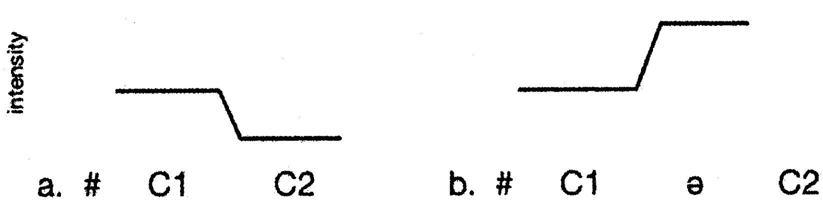


Figure 2-2: Schematic representations of the intensity contours in CC with no intensity rise and in CəC

Although only word-initial clusters are illustrated here for convenience sake, the same principle applies to word-medial and word-final clusters.

Let us see actual examples of consonant clusters with and without intensity rises, illustrated from Russian. In the following spectrograms, displayed with version 6.0 of Praat (Boersma and Weenink, 2015), time is plotted along the horizontal axis, and frequency (left) and intensity (right) are plotted along the vertical axis. The single hairline on the spectrograms indicates intensity as a function of time with a window length of 30 ms, while intensity in a particular frequency band is also represented by the darkness of the display.

First, stop-liquid clusters, such as [pl], illustrate well a rise in intensity at the transition between the two consonants. The spectrogram in Figure 2-3 shows that the intensity accelerates from the silence of [p] throughout the release burst of [p], as indicated by the arrow, and it continues to rise in the following [l] and the vowel. Vowels have an intensity maximum in the low and mid-frequency regions, and a peak in low frequency amplitude has been used for automatic vowel (or syllable) detection (e.g., Howitt, 2000). So in a vowel-epenthesized form CəC, there would be a clear rise in intensity from the initial consonant to the vowel [ə] in the low frequency region; thus it will be important to see whether an intensity rise in the low frequency region helps to determine the perceptual similarity between CC and CəC. That is, a consonant cluster CC involving an intensity rise in the low frequency region should be perceptually more similar to the matching CəC sequence than a cluster involving no intensity rise at low frequencies. Following Howitt (2000), we consider here intensity contours in the low frequency band between 300 Hz and 900 Hz. Figure 2-4 displays the low frequency band of [pli] in Figure 2-3, in which we see a sharper rise of intensity. The spectrograms presented from now on are from male Russian speakers' speech, unless otherwise noted. This pattern of rising intensity is analogous to that of the consonant-vowel sequence in CəC. Figure 2-5 displays the word-initial [pl] with an intervening vowel [ə], which shows a similar rise in intensity. Both in the cluster [pl] and in the epenthesized form [pəl], there is a sharp rise in intensity from the first segment [p] to the following segment [l] or [ə], so it is reasonable to assume that the two sequences are perceptually similar to each other.

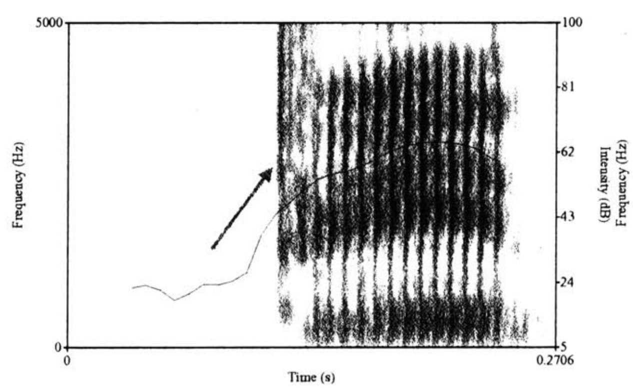


Figure 2-3: Spectrogram of [pli] from [plita] 'plait'

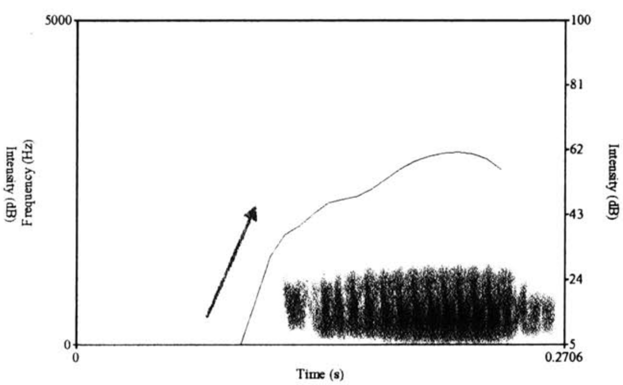


Figure 2-4: Spectrogram of [pli] in Figure 2-3, band-pass filtered

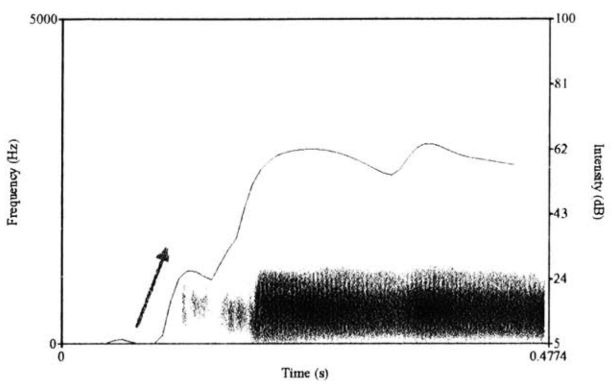


Figure 2-5: Spectrogram of [pəli], band-pass filtered

In contrast, most consonant clusters that do not involve an intensity rise involve a clear fall in intensity. The most common cluster showing no intensity rise consists of fricative-stop. For example, in the word-initial cluster [sp] in Figure 2-6, the sibilant fricative has comparatively strong energy, and there is a rapid drop in intensity as the stop closure starts, as marked by the arrow. In the band-pass filtered spectrogram in Figure 2-7, a similar intensity drop is observed.

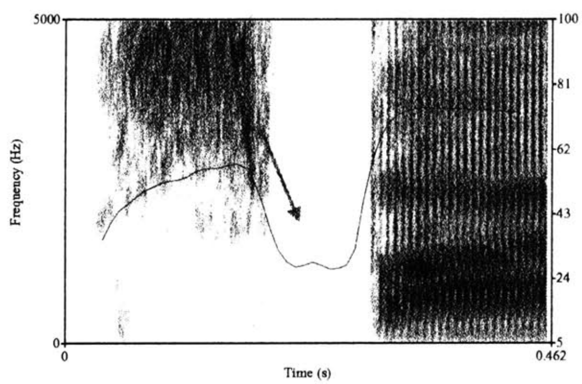


Figure 2-6: Spectrogram of [spa] from [spat^j] 'to sleep'

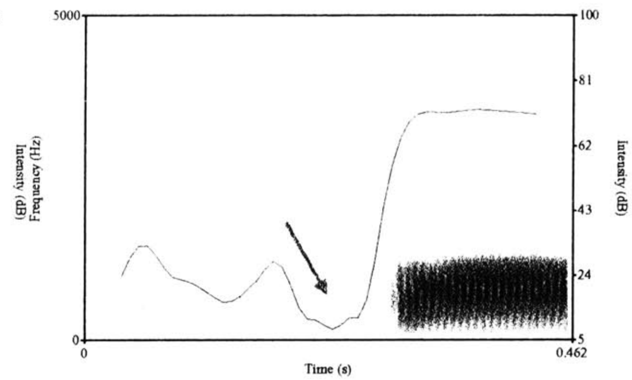


Figure 2-7: Spectrogram of [spa] in Figure 2-6, band-pass filtered

When a vowel is epenthesized between the word-initial fricative and the following stop, however, we see a very different intensity contour. As shown in Figure 2-8, there is a rise in intensity between the [s] and the following vowel, because vowels involve higher amplitude than consonants including fricatives. In comparison to the intensity fall in [sp] in Figure 2-7, the rising intensity contour in the [sə] sequence makes a large difference compared to the falling contour in the [sp] cluster, consistent with the schematic representations in Figure 2-2. Therefore, we can infer that consonant clusters like [sp] that do not involve a rise in intensity should be perceived quite differently from their epenthesized counterparts.

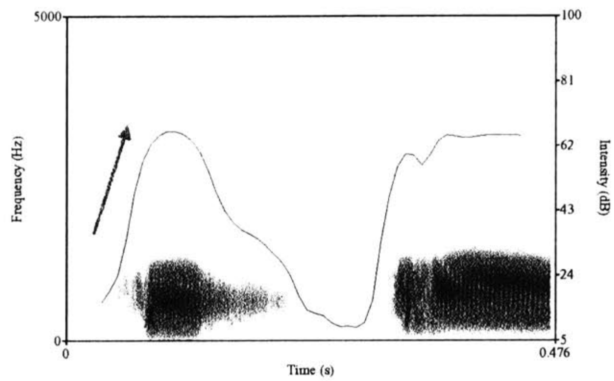


Figure 2-8: Spectrogram of [səpa], band-pass filtered

As intensity is the most reliable phonetic correlate of sonority (Parker, 2002), one may suggest that the intensity-based hypothesis is not essentially different from the sonority-based ones. However, there is a significant difference between sonority contours and intensity contours, and the two are not always correlated with each other. Sonority refers to relative loudness of sounds (Sievers, 1893; Ladefoged, 1975), and natural classes of sounds are ordered in a hierarchy with respect to sonority. While several different sonority hierarchies have been suggested in literature, the scale in (3) seems to be the most commonly used. This hierarchy does not distinguish stops, fricatives, and their voicing in the obstruent category, which may differ in sonority depending on the sonority hierarchy assumed and on the language.

(3) Sonority hierarchy (e.g., Clements, 1990)

	obstruents	nasals	liquids	glides	vowels
	1	2	3	4	5
⟵ *less sonorous*					*more sonorous* ⟶

Here the natural classes of sounds are arranged in order of sonority, from the most sonorous to the least sonorous, and a numeric index is given for each class. The sonority distance of a cluster C_1C_2 is calculated by subtracting the sonority level of C_1 from the sonority level of C_2. For example, the sonority distance of [bl], a cluster with rising sonority, is 2, subtracting 1 from 3, and that of [ns], a cluster with falling sonority, is −1, subtracting 2 from 1. So clusters with rising sonority have positive values of sonority distance, clusters with level sonority have a sonority distance of 0, and clusters with falling sonority have negative values of sonority distance.

Phonetic studies of sonority have provided actual intensity values for each sonority class, either integrated RMS (Root Mean Square) values averaged across entire phonemes (e.g., Ladefoged, 1975; Fry, 1979; Lavoie, 2000) or sound level extrema (maximum sound level in vowels and minimum sound level in consonants) (Parker, 2002, 2008; Jany et al., 2007). Those intensity values can replace the arbitrary numbers in (3) for each sonority class. Nevertheless, the sonority contours in consonant clusters should not change. Based on the intensity level of each sonority catgory, the sonority contour in C_1C_2 is calculated by subtracting the intensity value of C_1 from the intensity value of C_2, and a sonority class located in a higher position in the hierarchy in (3) involves a greater intensity than a class in a lower position. Based on Parker's (2008) intensity measurements of English consonants, for example, the sound level extreme of voiced stops relative to the vowel [a] is −21.1 and that of [l] is −11.8, and so the sonority distance of [bl] is thus 9.3, instead of 2 based on (3). For [ns], the sonority distance is −7.4, subtracting the sound level extreme of nasals −13.7 from that of fricatives −21.1, which replaces −1 based on (3). So positive sonority distance values based on (3) correspond to positive intensity differences and negative sonority distance values correspond to negative intensity differences. The point is that whether the sonority of sounds is quantified as abstract numbers or as phonetically measured intensity values, the relative differences between the sonority classes are not changed.[1]

Whereas sonority defines a property of the segment as a whole, intensity can change through the time course of a segment and can rise or fall in the transition between segments. An intensity contour refers to the actual intensity contour between the two consonants in a cluster, focusing on the transition between them. First, consider consonant clusters involving a sonority rise, which include obstruent-sonorant clusters and nasal-liquid clusters.[2] In these clusters, intensity increases at the transition between the two consonants. For example, the stop-liquid cluster [pl] illustrated in Figures 2-3 and 2-4 involves a steep rise at the release of the stop [p]. Since in a consonant cluster C_1C_2 with a sonority rise, intensity is always lower in C_1 than in C_2, consonant clusters involving an intensity rise also involve a sonority rise.

On the other hand, consonant clusters with a sonority plateau do not tend to involve a similar intensity plateau. A case of sonority plateau that also involves a plateau in intensity may solely be observed in a nasal-nasal cluster, as non-plateau intensity contours in sonority plateau clusters, i.e., stop-stop and fricative-fricative clusters, will be shown later. In the word-initial nasal-nasal cluster [nm] in Figures 2-9 and 2-10, spoken by a Russian speaker, the line indicating intensity is almost flat throughout the cluster. This is because the first nasal [n] is not audibly released.

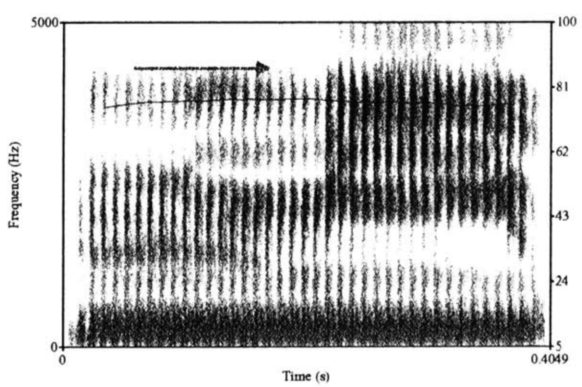


Figure 2-9: Spectrogram of [nmi] from [nmite] (nonce word)

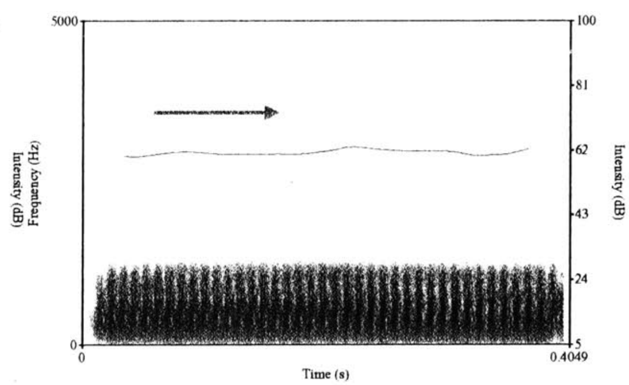


Figure 2-10: Spectrogram of [nmi] in Figure 2-9, band-pass filtered

In most cases, however, this kind of flat intensity is not observed in consonant clusters even when the sonority level of the consonants is identical. In particular, consider the pattern of

[1]Parker (2002) suggests a finer-grained sonority hierarchy of 16 steps based on the intensity measurement. Parker's hierarchy will have an empirical advantage compared to the sonority hierachy in (3), as well as have concrete phonetic evidence for the hierarchy. For the purpose of this study, however, both of the sonority hierarchies do not make different predictions in crucial cases which will be presented in what follows. For example, Parker's hierarchy, like the hierarchy in (3) does not differentiate sibilants and non-sibilants among fricatives, and thus predicts that any fricative-fricative clusters are of level sonority. So I will stick to the sonority hierarchy in (3) when making a comparison with the sonority-based hypothesis, for convenience sake.

[2]Consonant clusters of which second member is a glide are not considered here because in many languages it is unclear if the glide is part of the consonant cluster or part of syllable nucleus as a diphthong.

intensity in Figure 2-11. C_1 and C_2 belong to the same sonority category, such as stops, resulting in a sonority plateau. Considering the actual intensity of each consonant, the sonority contour is still level in the cluster because both integrated RMS values and lowest intensity values of C_1 and C_2 would be very similar to each other. If we concentrate on the intensity pattern at the transition between the two consonants, however, there is a clear rise in intensity, if C_1 is released. Therefore, this kind of consonant cluster involves an intensity rise in the sense defined earlier, even though it involves a sonority plateau.

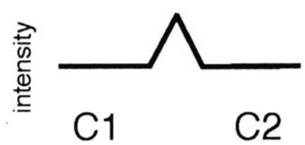


Figure 2-11: Possible intensity contour in sonority-plateau clusters

The intensity pattern schematized in Figure 2-11 is found in word-initial stop-stop clusters, as shown in Figures 2-12 and 2-13. Since both [k] and [t] are stops, the cluster forms a sonority plateau. Unlike the sonority contour, the intensity contour of the cluster involves a clear rise at the point where the first stop [k] is released with a burst, as indicated by the arrow.

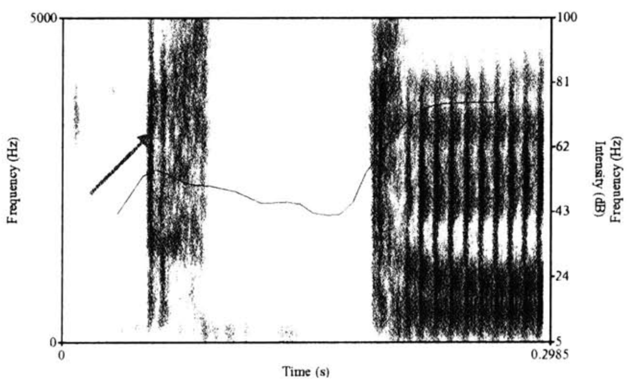


Figure 2-12: Spectrogram of [kto] 'who'

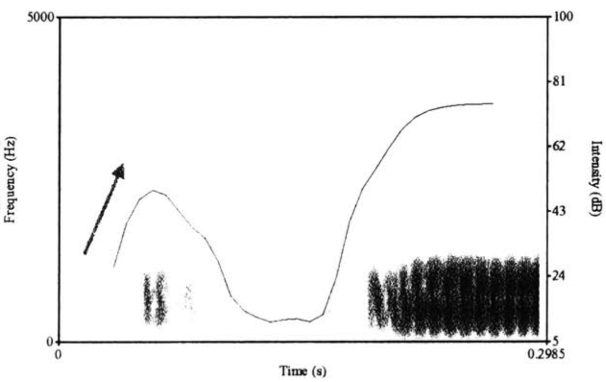


Figure 2-13: Spectrogram of [kto] in Figure 2-12, band-pass filtered

However, it is not always the case that stop-stop clusters involve an intensity rise. If the first stop of the cluster is unreleased, the cluster does not show any intensity rise in it. In English, for instance, the first stop in word-final stop-stop clusters may be unreleased, as shown in Figure 2-14. In this case, there is no rise in intensity within the cluster.

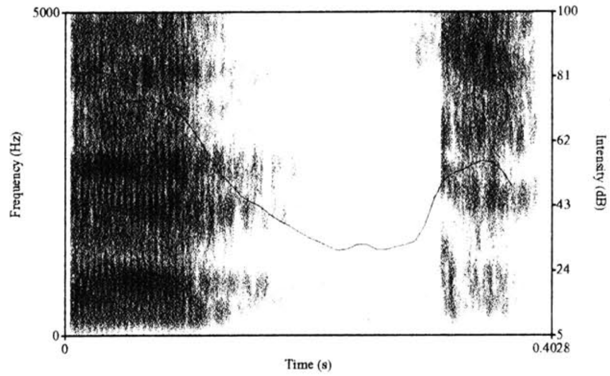


Figure 2-14: Spectrogram of [ɛkt] from English [əfɛkt], spoken by a female native speaker of English

Another sonority-plateau cluster, a nasal-nasal cluster, may involve an intensity rise or not depending on its phonetic realization, like stop-stop clusters. We have seen in Figures 2-9 and 2-10 that the first nasal in the nasal-nasal cluster [nm] is not audibly released and there is no intensity rise between the two nasals. However, it is also possible that the first nasal involves an audible vocalic release when preceding another nasal. The following nasal-nasal cluster [mn] in Russian serves as an example. The spectrogram in Figure 2-15 shows that the first nasal [m] involves a vocalic release at the transition into the following nasal [n], accompanied with a relatively slight rise in intensity, compared with the one we have seen in the [pl] cluster earlier. The magnitude of the intensity rise is sharper in the low frequency region as shown in Figure 2-16.

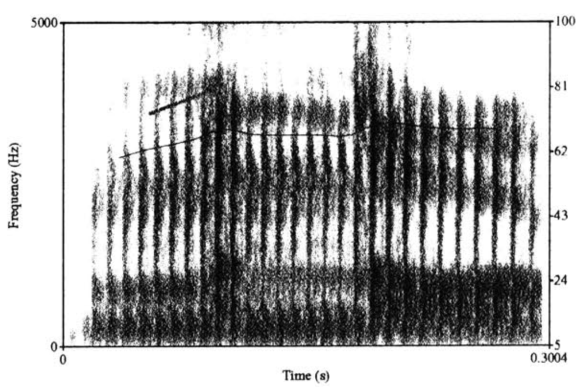


Figure 2-15: Spectrogram of [mno] from [mnogə] 'many'

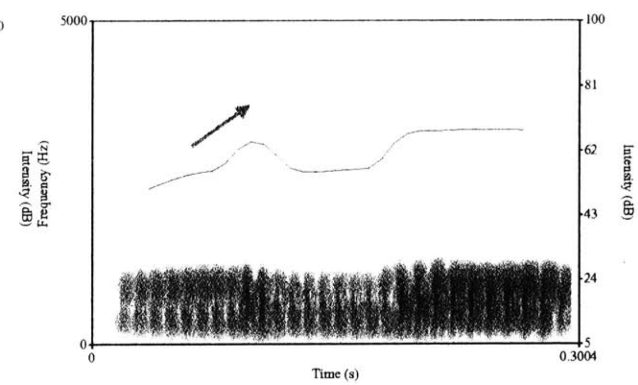


Figure 2-16: Spectrogram of [mno] in Figure 2-15, band-pass filtered

So far we have seen that depending on the release of the first nasal, nasal-nasal clusters may involve an intensity rise or not. Let us now consider one more example of the nasal-nasal cluster [mn] in Figures 2-17 and 2-18. In this case, the initial nasal [m] is not audibly released, and there is no abrupt rise in intensity between the two nasals resulting from a release. However, intensity is gradually increasing across the two nasals, although the magnitude of rise is far smaller than that of the released nasal-nasal cluster in Figures 2-15 and 2-16.

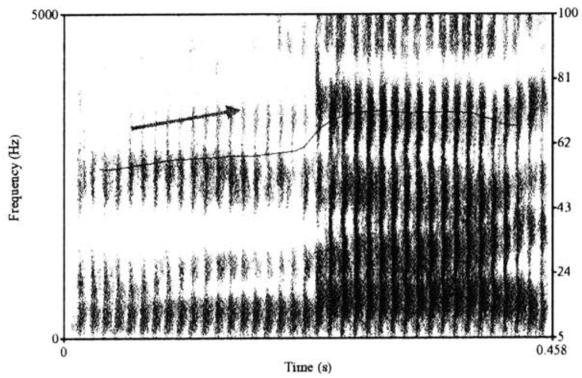


Figure 2-17: Spectrogram of [mna] from [mnate] (nonce word)

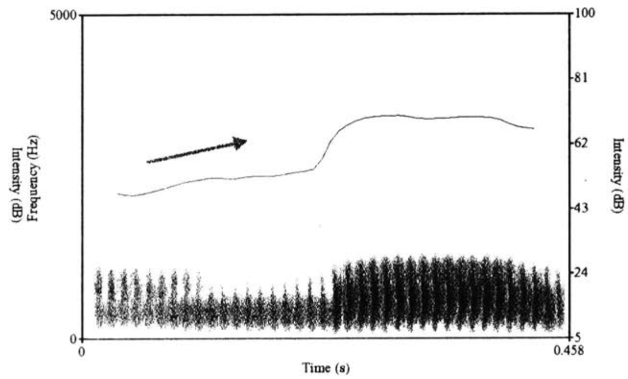


Figure 2-18: Spectrogram of [mna] in Figure 2-17, band-pass filtered

The next question is how much of rise in intensity can be regarded as a *perceptually salient* intensity rise. The magnitude of intensity difference between the two consonants in C_1C_2 would be huge when C_1 involves low intensity, such as in stops and non-sibilant fricatives, and C_2 involves high intensity, such as in nasals and liquids. Conversely, the magnitude of intensity rise would be slight when C_1 involves a high intensity and C2 involves a similar but slightly higher intensity. This is not the case for stop- or fricative-initial clusters because C_1's intensity is quite low; for stops, the intensity rise at a stop's release following the silence is abrupt, and the intensity level of non-sibilant fricatives is said to be equivalent to that of stops. For nasal-initial clusters, the intensity difference in nasal-nasal clusters would be smaller than that in nasal-liquid clusters since liquids have a higher intensity than nasals, and an intensity rise is observed in nasal-nasal clusters when the first nasal is audibly released and even when the first nasal does not involve an audible vocalic release, as in Figures 2-17 and 2-18. Liquids involve higher intensity compared to obstruents and nasals, and thus when a liquid precedes another consonant in a cluster, the intensity contour would not be rising. Hence, it is estimated that the threshold of counting as an intensity rise is at least greater than the rise in [mn] with the unreleased [m] in Figures 2-17 and 2-18 but less than in [mn] with the released [m] in Figures 2-15 and 2-16. Since intensity rises in other consonant clusters, such as [pl] and [kt], were much steeper than the one in the nasal-nasal clusters here, I assume that the threshold of intensity rise based on the measurement of nasal-nasal clusters works for all kinds of consonant clusters.

I measured the intensity values of several [mn] clusters spoken by seven native speakers of Russian, five males and two females. The clusters measured were extracted from: (i) two real Russian words, [mnemonitʃeskij] 'mnemonic' (R1 in Table 2.1) and [mnogə] 'many' (R2), which were recorded for reference, and (ii) nonce words, [mnate] (N1), [mnite] (N2), and [mnute] (N3), which were recorded for experimental stimuli that will be introduced in the following section. The recordings were made in a sound-attenuated booth in the MIT phonetics lab. Wearing a head-mounted condenser microphone, subjects read target words printed on paper. The utterances were recorded in mono at 44.1 kHz with the Amadeus software and saved to .aiff files. Measurement was conducted using version 6.0 of Praat. The words beginning with a nasal-nasal cluster that were subjected to measurement were first RMS-normalized using a Praat script and were band-pass filtered ranging from 300 Hz to 900 Hz with a window length of 30 ms. Intensity values were measured at several points in an interval of 60 ms (30 ms before and after the boundary between the two nasals). Based on the measured intensity values, I calculated rates of change for intensity relative to time (sec), by dividing the intensity difference by time, and chose the greatest value for each cluster. The results are presented in Table 2.1.

Table 2.1: Intensity rise in [mn] clusters (dB per second)

[m] released			[m] unreleased		
speaker	word	dB/sec	speaker	word	dB/sec
M1	N1	644.24	M1	N1	48.08
M2	N2	395.47	M2	N1	19.83
	N3	461.33			
F1	N1	253.91			
	N1	418.37			
M3	R1	216.67			
	R2	515.64			
M4	R1	217.88			
	R2	563.66			
F2	R1	321.33			
	R2	472.43			
M6	R1	250.83			
	R2	209.98			

It was found that two unreleased nasal-nasal clusters appeared in the pronunciation of a nonce word, wheareas both of the real words were spoken with a released first nasal in the cluster. The rates of change for intensity rise in unreleased nasal-initial clusters were lower than 50 dB per second. The rates of change in released nasal-initial clusters varied, from 210 dB per second to 644 dB per second, all of which were much higher than those in unreleased nasal-initial clusters. So we can set the threshold value for the intensity rise between 50 and 200 dB per second.

To sum up, stop-stop and nasal-nasal clusters that involve a sonority plateau may involve an intensity rise when it increases by at least 200 dB per second, and otherwise they involve no intensity rise.

Fricative-fricative clusters, which also involve a sonority plateau, show different intensity patterns. Although both non-sibilant fricatives and sibilant fricatives are categorized as fricatives, it is well known that sibilant fricatives have stronger energy than non-sibilants. Therefore, clusters consisting of a sibilant and a non-sibilant, either sibilant-non-sibilant or non-sibilant-sibilant combinations, involve change in intensity between the two fricatives. If the first fricative in a fricative-fricative cluster is a non-sibilant and the second one is a sibilant, such as [fs], we observe a clear intensity rise between the two fricatives as in Figures 2-19 and 2-20, which are spectrograms of the cluster spoken by a female native speaker of Russian.

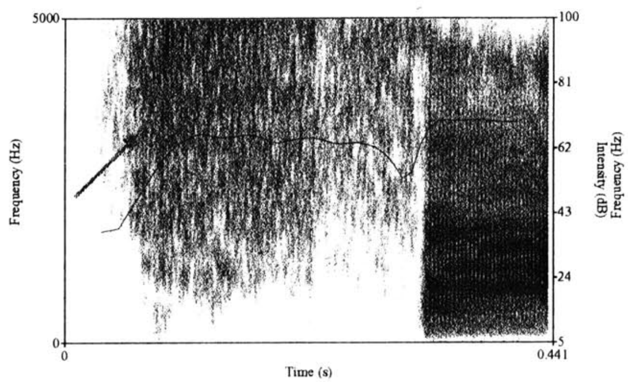


Figure 2-19: Spectrogram of [fs] from [fsate] (nonce word)

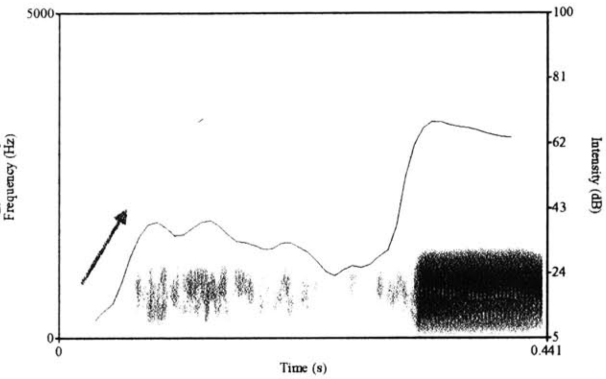


Figure 2-20: Spectrogram of [fs] in Figure 2-19, band-pass filtered

In contrast, if the first fricative is a sibilant and the second one is a non-sibilant in a fricative-fricative cluster, e.g., [sf], as in Figures 2-21 and 2-22, the intensity of the first fricative is higher than the second one and we see a drop in intensity. In this case, there is no AD. Hence, even though both non-sibilant-sibilant clusters and sibilant-non-sibilant clusters are of level sonority, the former involves an intensity rise but the latter does not.

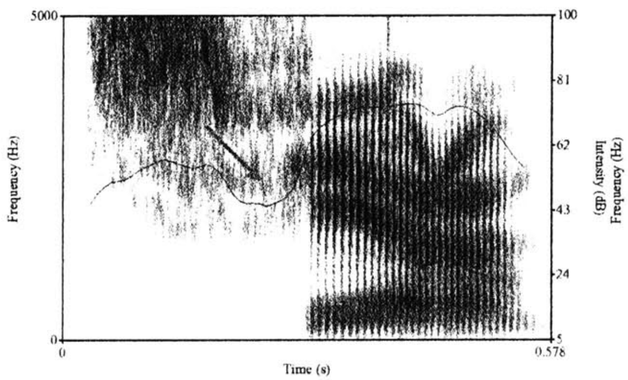


Figure 2-21: Spectrogram of [sfʲira] 'sphere'

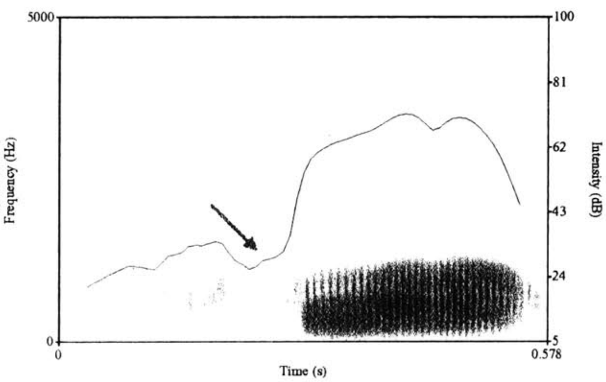


Figure 2-22: Spectrogram of [sfʲira] in Figure 2-21, band-pass filtered

In addition, fricative-stop clusters are classified as a sonority plateau cluster according to the sonority scale in (3). They also involve no intensity rise, regardless of whether the initial fricative is a sibilant or a non-sibilant, because the fricative has stronger intensity than the following stop closure, as shown in Figures 2-6 and 2-7 earlier.

Let us now consider the intensity countours of falling sonority clusters. Many of the consonant clusters with a sonority fall involve an intensity fall as well, and there is no intensity rise in them. Lateral-initial clusters are examples. In the cluster [lg], for instance, the liquid [l] involves a higher intensity than the following stop closure, represented with the silence with a voicebar, and the intensity contour is falling in the cluster, as shown in Figure 2-23. Tracking the intensity contour in just the low frequency region in Figure 2-24, the drop in intensity is by far clearer. The same holds for the other lateral-initial clusters because laterals involve higher intensity than the following consonant – stops, fricatives, or nasals.[3] So we can conclude that there is no intensity rise in lateral-initial clusters.

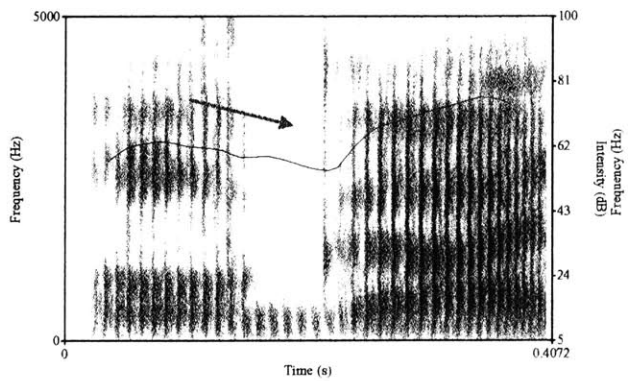


Figure 2-23: Spectrogram of [lga] from [lgat^j] 'to lie'

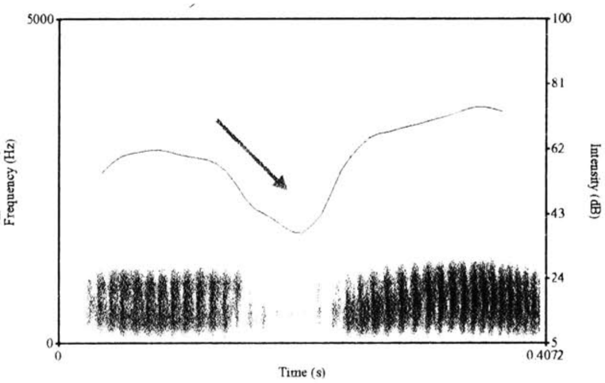


Figure 2-24: Spectrogram of [lga] in Figure 2-23, band-pass filtered

However, there are falling sonority clusters that involve an intensity rise. Considering a schematic representation of a consonant cluster C_1C_2 in Figure 2-25, the sonority contour between the two consonants is falling, because the intensity value of C_1, either the integrated RMS or the lowest sound level, is higher than that of C_2. Looking at the transition between the two consonants, however, intensity increases at the end of C_1 and decreases in C_2. This rise in intensity can be counted as an intensity rise, perceptually salient.

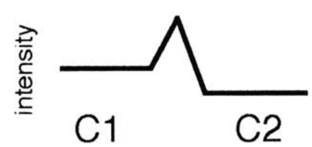


Figure 2-25: Possible intensity contour in sonority-falling clusters

Trill-initial clusters show the intensity pattern in Figure 2-25. Trills normally consist of two or three periods of vibration, and each vibration includes a closed phase with weaker energy and an open phase with stronger energy, creating a rise in intensity. Crucially, the last vibration may count as an intensity rise at the transition into the following C. In Figure 2-26, for example, the word-initial trill involves two periods of vibration, and we see two intensity peaks. The second rise in intensity, marked with an arrow, is regarded as an intensity rise that forms an AD.

For similar reasons, I assume that clusters that begin with a tap or a flap also involve an intensity rise. Taps and flaps have a single short closure, in contrast with several in trills, and there is an intensity rise following the closure.

As demonstrated above, lateral-initial clusters and trill-initial clusters are both liquid-initial and involve falling sonority, but they are different in terms of intensity rise. Trill-initial clusters involve an intensity rise but lateral-initial clusters do not.

[3]There is no liquid-liquid cluster attested word-initially in Russian. Word-finally, a cluster such as [rl] is possible, but it is not lateral-initial.

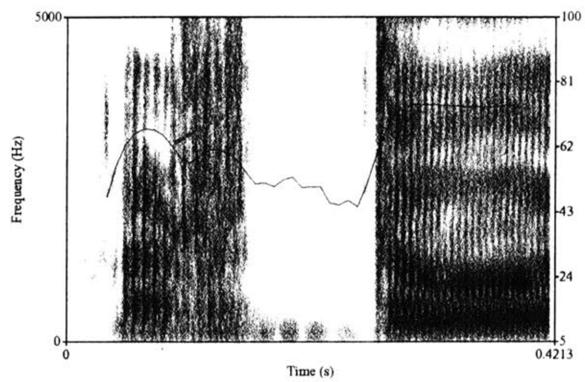


Figure 2-26: Spectrogram of [rtu] from [rtut^j] 'mercury'

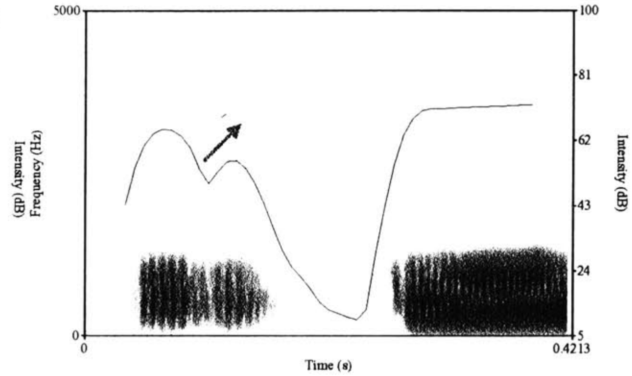


Figure 2-27: Spectrogram of [rtu] in Figure 2-26, band-pass filtered

To summarize, here is one way of categorizing consonant clusters with respect to separability by a vowel: clusters involving an intensity rise and clusters involving no intensity rise. Specifically, consonant clusters with an intensity rise involve a rise in intensity higher than 200 dB per second. Table 2.2 summarizes the consonant clusters in terms of the presence or absence of intensity rise as discussed in this section.

Table 2.2: Consonant clusters with or without an intensity rise

intensity rise	**no intensity rise**
stop (released)-initial clusters	stop (unreleased)-initial clsuters
non-sibilant-sibilant fricative clusters	sibilant-non-sibilant fricative clusters
fricative-sonorant clusters	fricative-stop clusters
nasal (released)-initial clusters	nasal (unreleased)-initial clusters
trill-initial clusters	lateral-initial clusters

Notice that the classification in Table 2.2 is based on the intensity rise *within* a consonant cluster, which we have focused on thus far. The presence of intensity rise within the cluster affects the perceptual similarity between the cluster CC and its matching sequence with an epenthetic vowel CəC. Although we have considered the presence or absence of intensity rise within the cluster so far, I will broaden the definition by including the difference between the ilence at the onset of the utterance and the first segment. An intensity rise may also exist from silence to the word-initial consonant that involves a great intensity. Sonorants involve a high intensity, and so a rise in intensity is associated with the start of the word-initial sonorant. For example, the word-initial [lg] cluster involves an intensity rise from silence to the initial [l], while there is no intensity rise between the two consonants in the cluster. This is shown in Figures 2-28 and 2-29, which are the same spectrograms as Figures 2-23 and 2-24 with an additional blue arrow indicating the initial intensity rise.

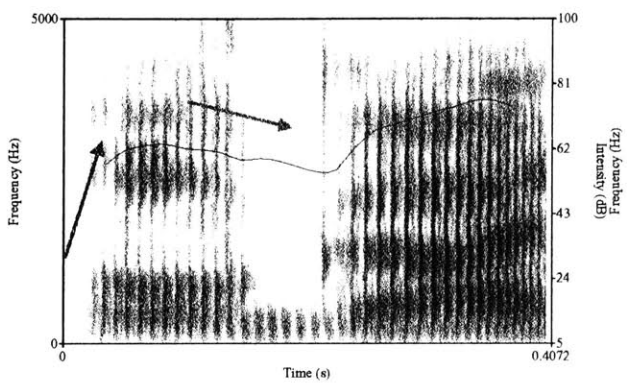


Figure 2-28: Spectrogram of [lga] from [lgat^j] 'to lie'

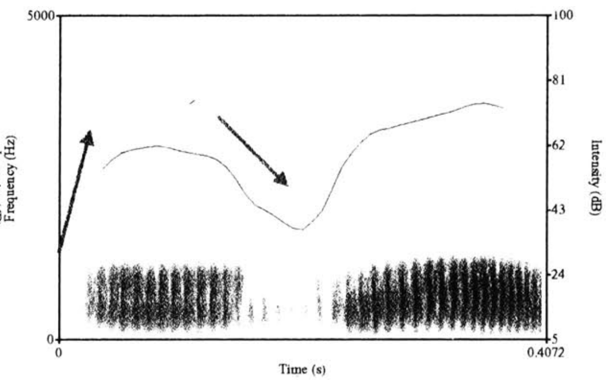


Figure 2-29: Spectrogram of [lga] in Figure 2-23, band-pass filtered

Among obstruents, sibilant fricatives involve a high amplitude, and we observe a rise in intensity from the initial silence to the sibilant. Figures 2-30 and 2-31 below repeat the spectrograms in Figures 2-6 and 2-7 with another arrow added to indicate initial intensity rise. The rise in intensity from the silence to the sibilant is quite steep and the rate of change is higher than 200 dB per second; in the case of [sp] in Figures 2-30 and 2-31, it is 674 dB per second. Therefore, although the word-initial [sp] cluster does not involve an intensity rise between the two consonants [s] and [p], it does involve an intensity rise before the cluster, from the silence to the [s].

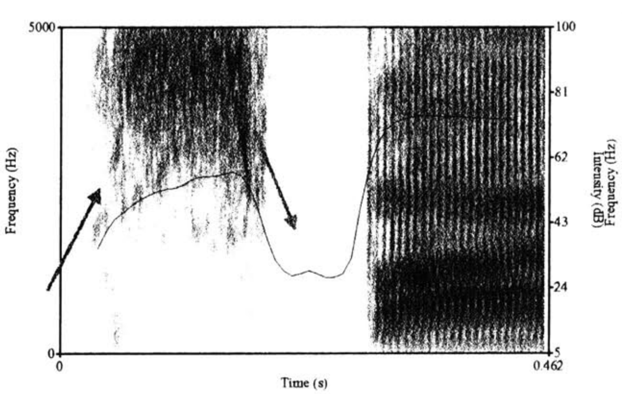


Figure 2-30: Spectrogram of [spa] from [spat^j] 'to sleep'

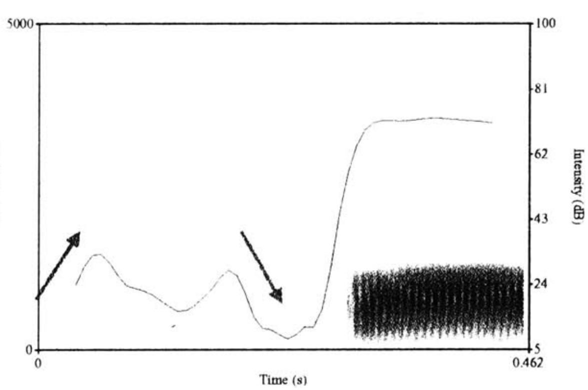


Figure 2-31: Spectrogram of [spa] in Figure 2-6, band-pass filtered

As summarized in Table 2.3, we hypothesize that word-initial clusters beginning with a sibilant fricative or a sonorant involve an intensity rise, regardless of whether they also have an intensity rise within the cluster. For stop- and non-sibilant-initial clusters, the transition from silence to C_1 is assumed to fall below the threshold that defines an intensity rise that in turn motivates epenthesis.

Table 2.3: Consonant clusters with or without an intensity rise outside the word-initial cluster

intensity rise	**no intensity rise**
sonorant-initial clusters	stop-initial clusters
sibilant-initial clusters	non-sibilant-initial clusters

There is no comparable case in word-final clusters because a vowel precedes the final cluster. Since this intensity rise is present outside the cluster, it is predicted that its presence or absence will not have a decisive influence on perceptual similarity between CC and CəC but will do so on the perceptual similarity between CC and əCC, with an epenthetic vowel outside the cluster. In this chapter, we focus on the perceptual similarity between CC and CəC, and the perceptual similarity between CC and əCC will be discussed in the next chapter.

2.1.2 C_1 voicing

Another feature that may affect the perceptual similarity between a consonant cluster and its matching sequence with an epenthetic vowel is the voicing of the first consonant in the cluster. Namely, a consonant cluster that begins with a voiced consonant, $C_{[+voice]}C$ is hypothesized to be perceptually more similar to the vowel-epenthesized form, $C_{[+voice]}$əC, compared to the perceptual similarity between consonant clusters that begin with a voiceless consonant, $C_{[-voice]}C$, and its vowel-epenthesized form, $C_{[-voice]}$əC. This is simply because the epenthetic vowel is voiced.[4] In the case of vowel epenthesis in a voiced consonant-initial cluster, $C_{[+voice]}$əC, periodic vibration of the vocal folds starts in the first consonant, and vowel insertion causes an intensity boost, with voicing countinuously present during the entire sequence. In contrast, when the first consonant is voiceless in $C_{[-voice]}$əC, vowel insertion not only results in an intensity boost but also initiates voicing. Since voicing is not present before the vowel begins here, vowel epenthesis would be perceptually more salient in $C_{[-voice]}C$ than in $C_{[+voice]}C$ clusters.

Let us take a stop-stop cluster, with both of the stops voiced, as an example. Figures 2-32 and 2-33 illustrate the word-initial cluster [gd] in Russian. Here we see voicing in the word-initial stop [g], in the form of dark regular striations of very low frequency, as opposed to the silence in the word-initial [k] in [kto] shown earlier in Figures 2-12 and 2-13. So if a vowel is inserted after the [g], it will induce a perceptually less salient change to the cluster than when a vowel is inserted after the [k], because the former causes only the rise in intensity at the stop release and the latter inaugurates voicing as well as an intensity rise.

[4]This might not always be true and an epenthetic vowel might be fully or partially devoiced. For example, devoicing of high vowels in Japanese (e.g., Tsuchida, 1997) may apply to an epenthetic vowel. In cases where the epenthetic vowel is not voiced, however, it tends not to be transcribed as a vowel in the loanword sources. The current assumption would not apply where voiceless epenthetic vowels appear, but such cases are rare.

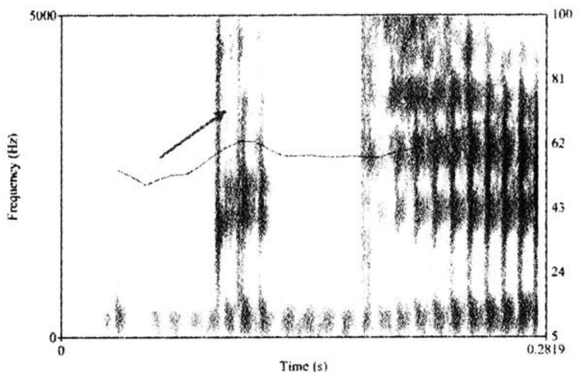


Figure 2-32: Spectrogram of [gd^{j}e] 'where'

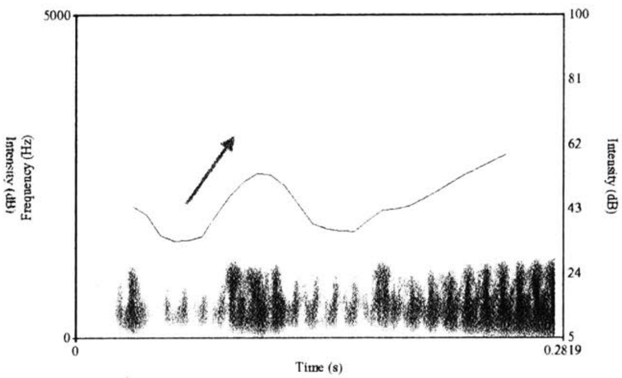


Figure 2-33: Spectrogram of [gd^{j}e] in Figure 2-32, band-pass filtered

2.1.3 Summary

We have explored two acoustic properties that may affect the perceptual similarity between a consonant cluster CC and its corresponding epenthesis form CəC: intensity rise and C_1 voicing. Since the epenthetic vowel involves an intensity rise from the preceding consonant and is voiced, it brings about a less noticeable perceptual alteration when the original cluster involves an intensity rise and/or a voiced initial consonant compared to when it involves no intensity rise and/or a voiceless initial consonant. More specific predictions are as follows: if a cluster involves both intensity rise and C_1 voicing, vowel epenthesis causes the least noticeable perceptual change; if a cluster involves either an intensity rise or C_1 voicing but not both, the perceptual change that vowel epenthesis brings about is greater compared to the one in a cluster with both an intensity rise and C_1 voicing but smaller than the one in a cluster with neither an intensity rise nor C_1 voicing. This is summarized in Table 2.4, where Δ(A-B) is read as "the perceptual distinctness between A and B".

Table 2.4: Degrees of vowel similarity

intensity rise	C_1 voicing		Δ(CC-CəC)
✓	✓	↑	more similar
✓	✗		
✗	✓		
✗	✗	↓	less similar

Tables 2.5 and 2.6 summarize the relevant acoustic features by cluster types, in which T stands for stops, F for non-sibilant fricatives, S for sibilant fricatives, N for nasals, R for sonorants, C for consonants. A checkmark indicates the presence of the factor and hence a more optimal epenthesis site in loanword adaptation.

Table 2.5: Auditory factors in word-initial clusters

cluster types	acoustic features	intensity rise		C_1
		#_C_1	C_1_C_2	[+vc]
#T$_{[-vc]}$**C** e.g., #kt, #kn, #pl, #tr	intensity rise with a release burst	✗	✓	✗
#T$_{[+vc]}$**C** e.g., #gd, #gn, #bl, #dr	intensity rise with a release burst voiced C_1	✗	✓	✓
#F$_{[-vc]}$**T** e.g., #fk, #xt	intensity fall from F to T	✗	✗	✗
#F$_{[+vc]}$**T** e.g., #vd, #ɣd	intensity fall from F to T voiced C_1	✗	✗	✓
#F$_{[-vc]}$**S, #F**$_{[-vc]}$**R** e.g., #fs, #fl, #θr	intensity rise from F to S/R	✗	✓	✗
#F$_{[+vc]}$**S, #F**$_{[+vc]}$**R** e.g., #vz, #vl, #ðr	intensity rise from F to S/R voiced C_1	✗	✓	✓
#S$_{[-vc]}$**T, #S**$_{[-vc]}$**F** e.g., #sk, #sp, #ʃt, #sf	intensity rise from silence to S intensity fall from S to T/F	✓	✗	✗
#S$_{[+vc]}$**T, #S**$_{[+vc]}$**F** e.g., #zd, #zv	intensity rise from silence to S intensity fall from S to T/F voiced C_1	✓	✗	✓
#N⌜C e.g., #m⌜d, #m⌜n, #m⌜l, #m⌜r	intensity rises from silence to N intensity rise with an audible release voiced C_1	✓	✓	✓
#N⌝C e.g., #m⌝b, #m⌝n	intensity rise from silence to N voiced C_1	✓	✗	✓
#lC e.g., #lg, #lv, #ln	intensity rise from silence to [l] voiced C_1	✓	✗	✓
#rC e.g., #rt, #rf, #rm	intensity rises before & after [r] voiced C_1	✓	✓	✓

Table 2.6: Intensity rise and C_1 voicing in word-final clusters

cluster type	acoustic features	intensity rise		C_1
		$C_1\_C_2$	$C_2\_\#$	[+vc]
T˹T˹# e.g., p˹k˹#, g˹d˹#	intensity rise with a release burst of T_1 intensity rise with a release burst of T_2	✓	✓	✓/✗
rT˹# e.g., rk˹#, rd˹#	intensity rise at the end of [r] intensity rise with a release burst of T_2	✓	✓	✓
other CT˹# e.g., p˺k˹#, g˺d˹#, fk˹#, st˹#, lk˹#	intensity rise with a release burst of T_2	✗	✓	✓/✗
T˹F#, TS# e.g., k˹f#, k˹s#, g˹z#	intensity rise with a release burst	✓	✗	✓/✗
FS# e.g., fs#, vz#	intensity rise from F to S	✓	✗	✓/✗
rF#, rS# e.g., rf#, rz#, rʃ#	intensity rise at the end of [r]	✓	✗	✓
N˺N#, lN˺# e.g., m˺n˺#, lm˺#	no intensity rise	✗	✗	✓
other CN˺# e.g., k˹m˺#, fm˺#, sn#, zm˺#, m˹n˺#	intensity rise from the obstruent to N *or* intensity rise with an audible release of C_1	✓	✗	✓/✗
Cl# e.g., g˹l#, fl#, sl#, zl#, ml#	intensity rise from C_1 to [l]	✓	✗	✓/✗
Cr# e.g., g˹r#, fr#, sr#, zr#, mr#	intensity rise from C_1 to [r] intensity rise at the end of [r]	✓	✓	✓/✗

In the next section, we will see whether these hypotheses can be experimentally supported; do subjects perceive consonant clusters with an intensity rise and a voiced C_1 as more similar to the epenthesis form than consonant clusters with no intensity rise and a voiceless C_1?

2.2 Experiments

This section provides experimental evidence for the hypothesis that consonant clusters involving an intensity rise and/or a voiced C_1 are perceived as more similar to the matching vowel epenthesis forms than clusters without them. The perceptual hypotheses, laid out in the previous section, were tested in three discrimination experiments, with speakers of three different languages: English, Korean, and Mandarin Chinese. In particular, the hypotheses tested in the experiments are stated in (4) below.

(4) a. **Hypothesis 1:** A consonant cluster involving an intensity rise is perceptually more similar to its epenthetic counterpart than a cluster involving no intensity rise.

b. **Hypothesis 2:** A consonant cluster that begins with a voiced consonant is perceptually more similar to its epenthetic counterpart than a cluster that begins with a voiceless consonant.

Much scholarly work has been done on the perception of consonant clusters and illusory vowels with various hypotheses from different perspectives. It has been shown that nativeness of consonant clusters affects their perception (e.g., Dupoux et al., 1999; Berent et al., 2007). That is, no matter what the consonant cluster is, listeners can perceive it more accurately if it exists in their native language. Dupoux et al. (1999) show that when Japanese listeners hear a sequence including a consonant cluster, e.g., [ebzo], not allowed in their native language, they tend to misperceive it with an illusory vowel, [ebuzo], while French listeners, whose native language permits such a sequence, correctly perceive it without the illusory vowel. In Berent et al.'s (2007) experiments, cross-linguistically marked clusters, such as [#lb], are rarely misperceived by Russian listeners whose native language has the clusters in question, whereas English listeners, whose native language lacks the clusters, have problems in perceiving such clusters.

Acoustic properties have also been said to play a role in non-native cluster perception. Relevant to intensity rise, notably, Fleischhacker (2001, 2005) investigated perceptual similarity between consonant clusters and their matching epenthesis forms. Fleischhacker (2001) conducted similarity rating experiments; English speakers listened to native English words that began with a consonant cluster, either stop-initial (stop-liquid) or fricative-initial (fricative-stop, fricative-nasal, fricative-liquid, and fricative-glide), and their matching epenthesis form, either prothesized #əCC or anaptycized #CəC, and rated how similar the pairs were to each other. The results showed that obstruent-sonorant clusters were perceived as more similar to the epenthesis forms CəC than sibilant-stop clusters. This is consistent with the current hypothesis about intensity rise because obstruent-sonorant clusters exhibit an intensity rise but sibilant-stop clusters do not. On the other hand, Fleischhacker (2005) examined the perceptual similarity between a cluster and its epenthesis counterpart with an AX discrimination experiment employing English listeners. For stimuli, [CCɑ]-[CəCɑ] pairs were used, created by a diphone synthesizer. The types of consonant clusters studied included more stop-initial clusters (stop-stop, stop-fricative, stop-nasal, and stop-glide) and a nasal-nasal cluster, in addition to the native English clusters used in Fleischhacker (2001). In Fleischhacker's (2005) stimuli, native and non-native clusters in English were mixed, and the only cluster involving no intensity rise for certain was [sk], a native cluster. The discrimination accuracy of [sk] was poorer than that of non-native clusters with an intensity rise, such as [kt] and [kn], but it is not clear whether this should be attributed to the nativeness or to the absence of an intensity rise in the cluster [sk]. The nasal-nasal [mn] cluster presumably was a cluster involving no intensity rise, and it exhibited higher accuracy than clusters like [kt] and [kn]. Hence, Fleischhacker's (2005) discrimination results may partially support the current hypothesis. Meanwhile, since Fleischhacker (2001, 2005) focused on stop- and sibilant-initial clusters, no voiced obstruents or sonorants were employed in the initial position of consonant clusters.

Davidson and Shaw (2012) conducted AX and ABX discrimination experiments using English listeners. They investigated discrimination between consonant clusters and their possible perceptual repairs: C_1 change, C_1 deletion, prothesis, and epenthesis. The stimulus clusters involved fricative-nasal, fricative-stop, stop-nasal, and stop-stop clusters, including [sm] and [sp] that exist in English, recorded by an English-Russian bilingual speaker. Focusing on their epenthesis results for our purpose, stop-stop clusters, involving an intensity rise (release), were significantly less accurately discriminated from their epenthesized counterparts compared to fricative-stop clusters, involving no intensity rise. Also, fricative-nasal and stop-nasal clusters were better discriminated than stop-stop clusters. This appears to go against the present hypothesis about intensity rise because it is expected that fricative-nasal and stop-nasal clusters involving an intensity rise are harder to discriminate from the epenthesis forms than stop-stop clusters. We cannot automatically conclude that this is against the current hypothesis, however, because the participants were speakers of English, a language that contains several word-initial consonant clusters. In other words, English speakers may have extended their knowledge of existing word-initial clusters, such as stop-liquid and fricative-liquid, to the novel clusters in the experiment, stop-nasals and fricative-nasals. (cf. Daland et al., 2011). Therefore, the experimental results from Davidson and Shaw (2012) can be regarded as at least a partial support for the current hypothesis. Both voiced and voiceless obstruents were used in the cluster-initial position, but they did not report whether there was a significant difference in the results depending on the C_1 voicing.

Durvasula and Kahng (2016) tested the effect of stop voicing in the perception of an illusory vowel in intervocalic stop-nasal clusters, e.g., [ekma] and [egma]. One identification experiment showed that Korean listeners perceived more illusory vowels after a voiced stop than after a voiceless stop. Moreover, when employing five different degrees of voicing in the stop closure, it turned out that the more the portion of the stop closure was voiced, the more likely the participant heard an illusory vowel after it. This serves as evidence for the current hypothesis about C_1 voicing.[5] Hwang (2011) also reports a similar voicing effect of stops observed in Korean listeners.

My experiments are intended to contribute to a better understanding of non-native consonant cluster perception. With regard to intensity rise, I employ a greater number of clusters than the previous research, including sonorant-initial clusters and test whether the intensity rise in

[5] Durvasula and Kahng's (2016) hypothesis was that people use their phonological knowledge in speech perception not only at the word level but also beyond the word level. In Korean, there is no voicing contrast in stops, and stops preceding a nasal undergo nasal assimilation, e.g., /...km.../ → [...ŋm...], indicating that both *[...km] and *[...gm...] are prohibited word-medially. At the level of the Intonational Phrase, however, stops can surface before a nasal, e.g., [...k $\#_{IP}$ m...]. So Durvasula and Kahng's interpretation of their experimental results is that Korean listeners use their knowledge of phrase-level phonotactics to better perceive a voiceless stop than a voiced stop in pre-nasal position. This is supported by results of English listeners who speak a native language that allows both [...km] and [...gm...] and perceive a voiced stop as well as a voiceless stop before a nasal. According to the current hypothesis about C_1 voicing, having a vowel after a voiced consonant is a perceptually less salient modification to the original sequence than having a vowel after a voiceless consonant, as stated previously. This hypothesis is based on auditory and phonetic grounds and does not refer specifically to knowledge of native language phonotactics. Consistent results across participants speaking different languages in what follows will provide support for the current hypothesis.

audibly released nasals and trills has a significant effect on consonant cluster perception. As for C_1 voicing, I test the voicing effect in word-initial and word-final positions, with more types of consonant clusters than the previous research. One advantage of my experiments is that the interaction between intensity rise and C_1 voicing can be tested; Experiments 2 and 3 use stimuli with all four possible combinations of intensity rise and C_1 voicing, as illustrated below.

Table 2.7: Types of stimuli

	intensity rise	no intensity rise
C_1 voiced	Type 1	Type 2
C_1 voiceless	Type 3	Type 4

In addition, my experiments employ participants from several different language backgrounds: Mandarin Chinese and Korean speakers bilingual in English, Korean monolingual speakers, and English native speakers. Comparisons between the participant groups may confirm that the knowledge of consonant cluster perception is universal or show how knowledge of native language phonotactics can affect consonant cluster perception.

In the next subsections, three discrimination experiments conducted for this study will be presented.

2.2.1 Experiment 1

2.2.1.1 Methods

2.2.1.1.1 Participants

Participants for Experiment 1 were 24 Korean speakers and 12 Mandarin Chinese speakers. All the Mandarin speakers and 4 out of 24 Korean speakers were undergraduate or graduate students at MIT, and 20 Korean speakers were undergraduate students at Seoul National University, Seoul, Korea, who had learned English in school for more than 10 years. None of the participants reported any speaking/hearing problems. Participants were paid $10 for their participation.

Both Korean and Mandarin Chinese lack consonant clusters at word edges. So I assume that phonological knowledge of their native language has no significant influence on the experimental results. The participants, however, were familiar with English consonant clusters, and it is possible that they may have used their knowledge of English phonotactics in the experiment, although there were no native English clusters included in experimental stimuli.

2.2.1.1.2 Stimuli

The stimuli were three types of nonce words beginning with a consonant cluster, [CCáte], [CCíte], and [CCúte], and their matching vowel epenthesis forms, [CəCáte], [CəCíte], and [CəCúte]. The stimuli were recorded by a male native speaker of Russian, who speaks a native language involving many different consonant clusters and can naturally produce the clusters in the stimuli. Recording was conducted in a sound-attenuated booth in the Phonetics Lab in the MIT linguistics department. The speaker wore a head-mounted microphone (Shure SM10A) and read the target material written on paper twice. The sound files were recorded in mono using the Amadeus software at a sampling rate of 44.1 kHz and saved to .aiff files. The speaker was directed to read the stimulus words with a stress on the vowel following the initial cluster or the epenthesized cluster, so that the epenthetic vowel was not stressed.

The recorded stimulus words were manipulated using Praat, so that parts in words other than the initial C(ə)C did not affect the discrimination. First, the final syllable [-te] in the words was cross-spliced in all stimuli. Using the PSOLA function, the pitch contours of the stimulus word and durations of the stressed vowel that follows the consonant cluster were manipulated to be indistinguishable in each pair of the cluster form and the epenthesis form. I used the natural pronunciation of the epenthetic vowels, and the duration of each vowel differed. Durations of the epenthetic vowels in the stimuli are given in Appendix A. Since previous research has established that the duration of epenthetic vowels affects consonant cluster perception (e.g., Dupoux et al., 1999), this durational difference between stimuli may be a potential confound of the results, but in this experiment the vowel duration turned out not to have a significant effect on the results.

The target consonant clusters are listed in Table 2.8, which classifies the clusters according to intensity rise, C_1 voicing, and sonority contour. $C^{\ulcorner}$ indicates that the consonant (stop or nasal) is audibly released, as opposed to unreleased $C^{\urcorner}$.

Table 2.8: Consonant clusters used in the stimuli (Experiment 1)

sonority	#	intensity rise	no intensity rise
rise	9	*C_1 voiceless* #t$^{\ulcorner}$m, #k$^{\ulcorner}$m, #k$^{\ulcorner}$n, #p$^{\ulcorner}$n *C_1 voiced* #d$^{\ulcorner}$m, #b$^{\ulcorner}$n, #g$^{\ulcorner}$m, #g$^{\ulcorner}$n, #m$^{\ulcorner}$l	
plateau	10	*C_1 voiceless* #k$^{\ulcorner}$p, #k$^{\ulcorner}$t, #p$^{\ulcorner}$k, #p$^{\ulcorner}$t *C_1 voiced* #g$^{\ulcorner}$b, # d$^{\ulcorner}$b, #d$^{\ulcorner}$g, #b$^{\ulcorner}$d, #m$^{\ulcorner}$n	*C_1 voiced* #n$^{\urcorner}$m
fall	10	*C_1 voiced* #m$^{\ulcorner}$d	*C_1 voiced* #m$^{\urcorner}$k, #m$^{\urcorner}$g, #m$^{\urcorner}$d, #n$^{\urcorner}$g, #n$^{\urcorner}$k #lk, #lb, #lg, #lp, #lm

Rising-intensity clusters were composed of stop-initial clusters and two nasal-initial clusters, [#m$^{\ulcorner}$n] and [#m$^{\ulcorner}$d]. First, word-initial clusters that began with a stop always involved an audible release, as previously seen in Figures 2-3, 2-4, 2-12, and 2-13, featuring a rise in intensity; in stop-stop clusters, particularly, a release burst was the only acoustic cue for the first stop.

On the other hand, nasal-initial clusters may be audibly released (Figures 2-15 and 2-16) or not (Figures 2-17 and 2-18), as previously stated. In the nasal-initial clusters recorded for the experiment, most of the nasals were not audibly released, and therefore involved no intensity rise, but all [#m$^{\ulcorner}$n] clusters and one [#m$^{\ulcorner}$d] cluster showed an audible release at the end of the initial [m], accompanied by an intensity rise. Specifically, all the occurrences of [#m$^{\ulcorner}$n] cluster involved an intensity rise as in Figures 2-15 and 2-16, in contrast with the other nasal-nasal cluster [#n$^{\urcorner}$m] as in Figures 2-9 and 2-10.[6] In the case of the nasal-stop clusters, only [#m$^{\ulcorner}$d]

[6]This is unexpected, given that front-to-back clusters have greater gestural overlap than back-to-front clusters (e.g., Byrd, 1996; Chitoran et al., 2002). One possible reason that [mn] clusters involved a vocalic release but [nm] did not could be that [mn] is a native cluster in Russian but [nm] is not. In other words, no audible release in [nm] clusters was possibly the result of careful pronunciation.

in [mdite] involved an intensity rise, as shown in Figures 2-34 and 2-35. The same [md] clusters pronounced in [mdate] and [mdute] did not involve an audible release of the [m] and intensity decreased from the [m] to the following [d], as shown in Figures 2-36 and 2-37 with the initial part of [mdute]. There was no nasal place assimilation in any nasal-stop cluster.

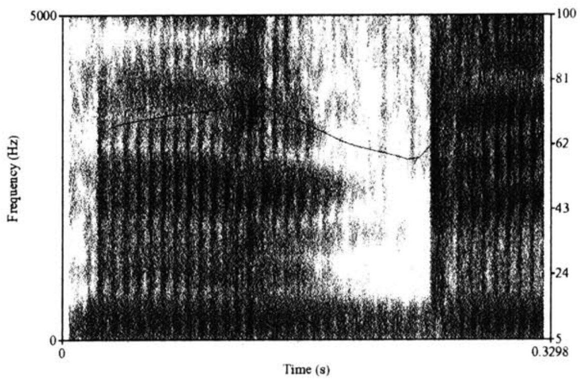


Figure 2-34: #m˹d

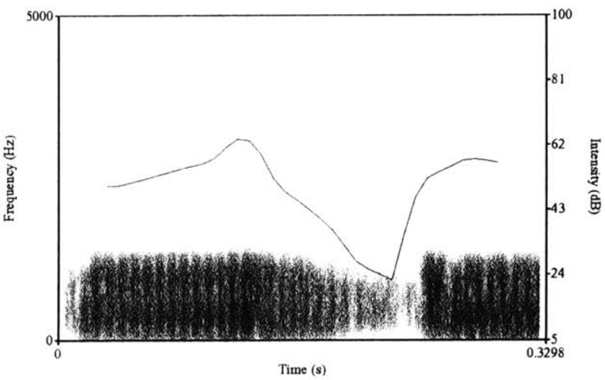


Figure 2-35: #m˹d (band-pass filtered)

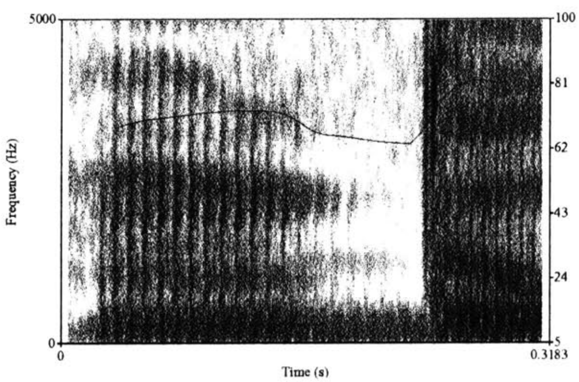


Figure 2-36: #m˺d

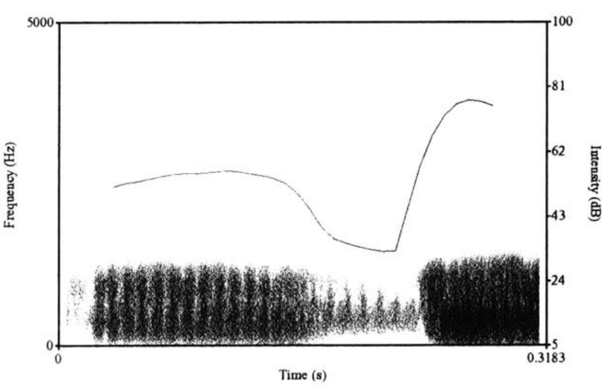


Figure 2-37: #m˺d (band-pass filtered)

Five lateral-initial clusters were also included as clusters with no intensity rise. In this experiment, clusters that begin with a voiceless consonant were not used for the no intensity rise material, and trill-initial clusters were not used for the intensity rise material.

2.2.1.1.3 Procedure

The experiment was carried out in the Experimental Syntax and Semantics Lab at MIT using an iMac and in the Phonetics Lab in the Linguistics Department at Seoul National University using a Macbook Air.

The task was AX discrimination; for each trial the participants heard a pair of audio stimuli and determined whether the two stimuli were the same or different. Two types of stimuli, one beginning with a consonant cluster and the other its matching epenthesis form, were used; the pairs were constructed to be the same (AA or XX) or different (AX or XA), as illustrated in Table 2.9.

Table 2.9: AX discrimination paradigm

	A	X
A	CCáte-CCáte	CCáte-CəCáte
X	CəCáte-CCate	CəCáte-CəCáte

Instructions were presented verbally in English and in writing on the screen. After understanding the procedure, the participants put on headphones (Audio-Technica ATH-M50 or Sennheiser HD280) and started the experiment by pressing a key on the keyboard when they were ready. There was first a practice session that comprised six trials so that the participants could become accustomed to the task. The practice trials were same or different pairs of real English words, e.g., *plight* vs. *polite*, which contained either an initial cluster or the same consonant sequence separated by a vowel. The real experiment following the practice trials consisted of 702 trials separated into 11 blocks, presented in random order. To be specific, the 29 stimulus clusters in Table 2.8 and their epenthetic counterparts in three vowel contexts ([–áte], [–íte], and [–úte]) were presented in four types of pairs in Table 2.9 twice, with six dummy trials. The inter-stimulus interval was 350 ms, and inter-trial interval was 500 ms. After completing each block, the participants could take a self-terminated rest period as long as they wanted, and the whole procedure lasted around 45-50 minutes.

2.2.1.2 Results

Discrimination results from Korean participants and from Chinese participants are presented in Sections 2.2.1.2.1 and 2.2.1.2.2, respectively. I report both percentage correct and d' and use both measures for statistical analysis. d' is calculated by subtracting the z-score of the proportion of false alarms from the z-score of the proportion of hits. In the AX discrimination task, hit refers to "different" responses for "different" pairs (AX or XA), and false alarm refers to "different" responses for "same" pairs (AA or XX). We can use d' for this AX discrimination experiment, based on an Independent-Observation Model (MacMillan and Creelman, 2005), assuming that subjects form a percept of each stimulus (A or X), and decide and answer whether the two are the same or different. The d' scores reported here were calculated using the *psyphy* package (Knoblauch, 2014) in R (R Core Team, 2016).

To evaluate the statistical significance of the differences between the independent variables, I fitted mixed effects logistic regression models to the percentage correct data and mixed effects probit regression models to the d' data, using the *glmer* function from the *lme4* package (Bates et al., 2015) in R. Fixed effects included intensity rise or C_1 voicing, or both. Random intercepts for clusters and participants were also included, unless otherwise noted. In most cases, random slope models did not converge. To indicate statistical significance, z scores from the Wald test will be reported along with p-values. The z value is the coefficient estimate divided by the standard error. If the z value is close to zero and fails to reject the null hypothesis that the coefficient is zero, the independent variable does not have a significant effect on the model, because the independent variable having a coefficient that is really small compared to its standard error is assumed to be of no help to predict dependent variable.

A mixed effects logistic regression model was also fit to the accuracy results with the predictors of the duration of the epenthetic vowels and the type of vowel in the templates with random intercepts of clusters and participants. It turned out that these factors did not influence the results, both for Korean and for Chinese participants.

2.2.1.2.1 Korean listeners

Let us first see the results for Korean listeners. Table 2.10 and Figure 2-38 show percentage correct of the clusters classified by both intensity rise and C_1 voicing. Recall that there were no clusters that involved no intensity rise and began with a voiceless consonant for this experiment. A mixed effects logistic regression model was fit to the percentage correct results with the predictors of intensity rise and C_1 voicing, including cluster and participant as random effects. The dependent variable was accuracy that was coded either as 1 (correct) or as 0 (incorrect). In the current model and hereafter, both intensity rise and C_1 voicing were treatment-coded; the baseline values were having an intensity rise and voiced C_1. It turned out that when C_1 was voiced, clusters with an intensity rise were significantly less accurately distinguished from the vowel epenthesis forms than clusters than clusters with no intensity rise (β=0.49, z=4.26, p<.001). Also, in intensity-rising clusters, the participants revealed lower accuracy when C_1 was voiced than when C_1 was voiceless (β=0.4, z=3.39, p<.001).

Table 2.10: Percentage correct by intensity rise and C_1 voicing (Korean listeners)

	intensity rise	no intensity rise
C_1 voiced	59.93%	66.54%
C_1 voiceless	66.64%	

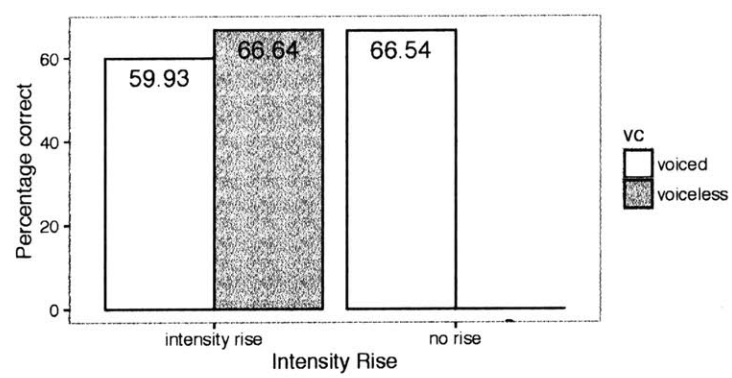


Figure 2-38: Percentage correct by intensity rise and C_1 voicing (Korean listeners)

Mean d' scores by intensity rise and C_1 voicing are illustrated in Table 2.11 and in Figure 2-39, which shows compatible results with those of percentage correct in Table 2.10. A mixed effects probit regression model was fit to response (same or different, sum-coded) with the predictors of intensity rise and C_1 voicing. It confirmed that intensity rise was a significant predictor of the discriminability between cluster forms and epenthesis forms when C_1 was voiced (β=0.86, z=4.92, p<.001), and so was C_1 voicing when the cluster involved an intensity rise (β=0.8, z=4.05, p<.001).

Table 2.11: d' scores by intensity rise and C_1 voicing (Korean listeners)

	intensity rise	no intensity rise
C_1 voiced	1.80	2.08
C_1 voiceless	2.26	

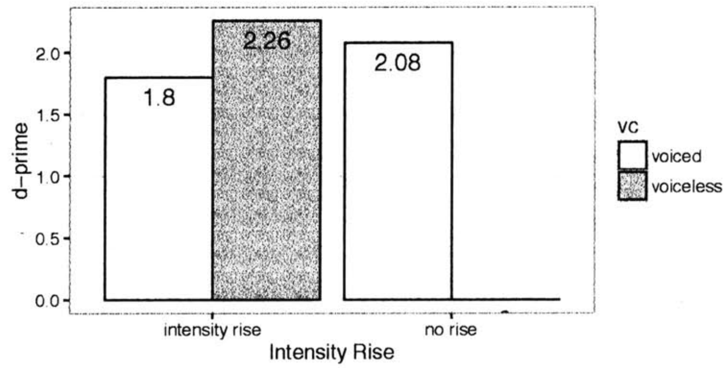


Figure 2-39: d' scores by intensity rise and C_1 voicing (Korean listeners)

2.2.1.2.2 Chinese listeners

Overall, Mandarin Chinese listeners show slightly lower percentage correct and d' scores than Korean listeners, but the pattern of effects of intensity rise and C_1 voicing they exhibit are consistent with those seen in Korean listeners. Table 2.12 and Figure 2-40 present mean percentage correct, and Table 2.13 and Figure 2-41 present d' scores by intensity rise and C_1 voicing. According to a mixed effects logistic regression model, the effect of intensity rise was significant when C_1 was voiced (β=0.26, z=3.47, p<.001), and the effect of C_1 voicing was also significant when the cluster displayed an intensity rise (β=0.32, z=3.84, p<.001). Also, a mixed effects probit regression model confirmed that the presence of intensity rise and that of C_1 voicing made d' for the distinction between a cluster and its epenthetic counterpart significantly lower when C_1 was voiced (β=0.37, z=2.92, p<.01), and when there was an intensity rise in the cluster (β=0.54, z=3.8, p<.001), respectively.

Table 2.12: Percentage correct by intensity rise and C_1 voicing (Chinese listeners)

	intensity rise	no intensity rise
C_1 voiced	56.74%	62.57%
C_1 voiceless	63.97%	

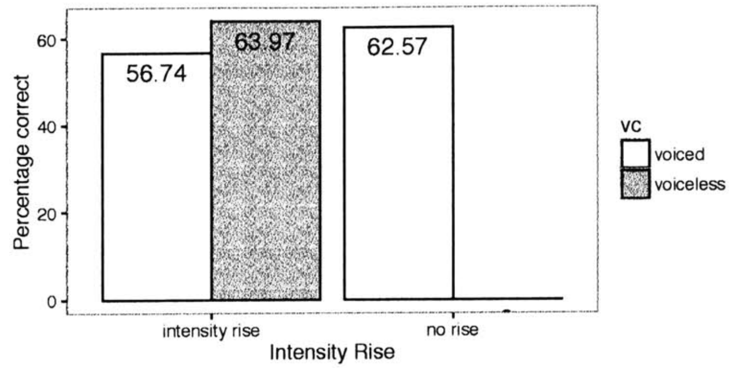


Figure 2-40: Percentage correct by intensity rise and C_1 voicing (Chinese listeners)

Table 2.13: d' scores by intensity rise and C_1 voicing (Chinese listeners)

	intensity rise	no intensity rise
C_1 voiced	1.31	1.62
C_1 voiceless	1.69	

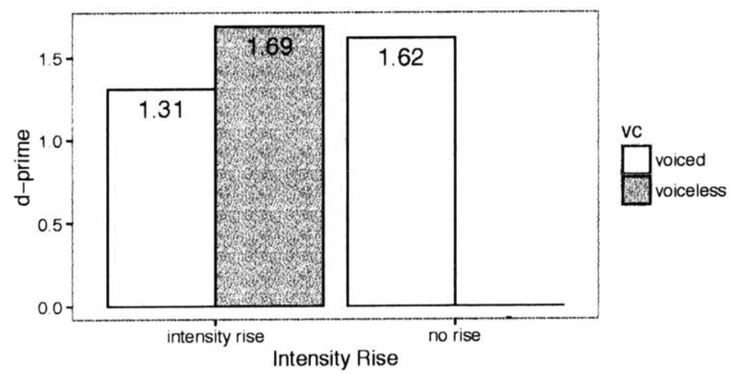


Figure 2-41: d' scores by intensity rise and C_1 voicing (Chinese listeners)

2.2.1.3 Discussion

It is shown that both intensity rise and C_1 voicing played a role in perceiving the difference between a consonant cluster and its corresponding epenthesis form, both in Korean and Chinese listeners' results. It should be underscored that even the same type of clusters or exactly the same clusters may show different discriminabilities from the epenthesis forms depending on their phonetic realizations, i.e., having an intensity rise or not. As previously mentioned, the initial [m] in [mn] was audibly released featuring an intensity rise, whereas the initial [n] in [nm] was not in the stimuli, although both [mn] and [nm] were composed of two nasals. As illustrated in Table 2.14, both Korean and Chinese participants showed lower accuracies and discriminabilities in the [mn] cluster with an intensity rise than in the [nm] cluster with no intensity rise.

Table 2.14: Percentage correct and d' scores in [m⌜n] and [n⌝m]

	[m⌜n]		[n⌝m]	
	Korean	Chinese	Korean	Chinese
percentage correct	58.93%	59.47%	67.86%	67.8%
d'	1.76	1.62	2.24	1.91

Also, the stimuli included an [m⌜d] cluster in [mdíte] where the first [m] involved an audible release and thus involved an intensity rise, as illustrated in Figures 2-34 and 2-35. In contrast, the other two [md] clusters in [mdâte] and [mdúte] involved no intensity rise, as presented in Figures 2-36 and 2-37 above. The results reflect this acoustic asymmetry, as shown in Table 2.15. That is, the [m⌜d] cluster featuring an intensity rise was harder to discriminate from the matching epenthesis form than the [m⌝d] cluster with no intensity rise.

Table 2.15: Percentage correct and d' scores in [m⌜d] and [m⌝d]

	[m⌜d]		[m⌝d]	
	Korean	Chinese	Korean	Chinese
percentage correct	56.55%	61.36%	84.52%	75.57%
d'	1.48	1.70	3.00	2.23

Note that there were no clusters that began with a voiceless consonant without having an intensity rise (Type 4 in Table 2.7) in Experiment 1. It should be noted that statistical significance was not always observed when only one of the factors, intensity rise or C_1 voicing, was included as the predictor, whereas the statistical significance was enhanced when both of the factors were used as the predictors, as shown above. This might be because intensity rise was confounded by C_1 voicing and vice versa. Since all clusters involving no intensity rise, expected to have a higher discriminability, began with a voiced consonant, expected to have a lower discriminability, the effect of intensity rise was not significant in percentage correct results of Korean listeners and both percentage correct and d' results of Chinese listeners because it was partially overridden by the effect of C_1 voicing. No significant effect of C_1 voicing in Korean listeners' percentage correct results may also come from a possible confound of having an intensity rise, because all voiceless-consonant-initial clusters involved an intensity rise. In addition, we were not able to see whether intensity rise makes a significant difference in discrimination when C_1 was voiceless and whether C_1 voicing does so when there is no intensity rise. Experiment 2 overcomes this weakness.

2.2.2 Experiment 2

Experiment 2 employed for stimuli consonant clusters containing a fricative and a trill, which were not used in Experiment 1, as well as types of clusters used in Experiment 1, testing the discriminability of a greater number of clusters. As will be shown later, this allows us to see whether there is a significant interaction between intensity rise and C_1 voicing, as well as to confirm the hypotheses with more types of consonant clusters.

2.2.2.1 Methods

2.2.2.1.1 Participants

The participants of Experiment 2 were 19 native speakers of Korean and 28 native speakers of American English. The Korean participants were all monolingual Korean speakers 30-50 years of age, unlike the Korean participants in Experiment 1. Notice, though, that it is hardly ever possible these days to find purely monolingual speakers of Korean due to English education in school. So the participants selected for this experiments were persons who have not studied English or other foreign languages for 10-30 years after they graduated from high school. They also did not usually watch subtitled movies or TV shows in English, so I considered them to have almost no experience with word-initial consonant clusters for a reasonably long time. So their results can be compared with the results from the Korean participants in Experiment 1 who were bilingual in Korean and English to see whether the knowledge of a second language has a significant effect on consonant cluster perception.

The English participants were college students or affiliates in the Boston area, recruited from the subject pool of the MIT Behavioral Research Lab. All of the participants self-reported that they had no experience of a foreign language involving many consonant clusters (e.g., Russian, Arabic, German, Greek, etc.) None of them reported a hearing problem, and the participants were compensated for their participation ($10). For English speakers, I analyzed only 23 participants, excluding five people who made more than one wrong answer out of the sixteen fillers, which were easily distinguishable as same or different pairs of CV(C) words without any consonant clusters (e.g., [da] vs. [bol]). There was no Korean participant who did not meet this criterion.

2.2.2.1.2 Stimuli

The stimuli used in Experiment 2 were nonce words that begin with an initial consonant cluster [#CC-] and their epenthetic counterparts [#CəC-], as in Experiment 1. One exception is the word-final cluster /kt/, which will be described later in detail. The stimuli were recorded by five native speakers of Russian in the same recording setting as in Experiment 1. They were instructed to read a list of the nonce words, written in Russian orthography including stress markers on the non-epenthetic vowel. Epenthetic vowels in the stimuli were unstressed in all cases. Since some clusters showed variations in their phonetic realizations, as will be discussed in detail later, the stimuli had to be chosen from different speakers, but each pair composed of a cluster and its epenthesis counterpart was picked from the same speaker's speech.

Each initial cluster was embedded in a template [CCát], which has its correspondent epenthetic form [CəCát], and the final /kt/ cluster was embedded in [napíCC], or [napíCəC] with epenthesis. The duration of the stressed vowel in each stimulus was in range of 150–200 ms, and was kept nearly identical in each pairing of a cluster form and its epenthetic counterpart, by manipulation with Praat. The durations of the epenthetic vowels ranged around 60-80 ms, which appear in Appendix 2B. The other parts of the stimulus, i.e., [-t] in word-initial and [nap-] in word-final stimuli, were cross-spliced. The pitch pattern was also manually manipulating using Praat to be intistinguishable between each cluster and its epenthesis form, which resulted in a natural falling tone in most cases.

For a classification of the stimulus clusters according to their sonority profile, intensity rise, and C_1 voicing, see Table 2.16. All obstruent-obstruent clusters agreed in voicing, since Russian does not allow obstruent clusters that differ in voicing.

Table 2.16: Consonant clusters used in the stimuli (Experiment 2)

sonority	#	intensity rise	no intensity rise
rise	24	*C_1 voiceless* #k$^{\ulcorner}$f, #k$^{\ulcorner}$s, #k$^{\ulcorner}$m, #k$^{\ulcorner}$l, #k$^{\ulcorner}$r #fm, #fl, #fr #sm, #sl, #sr *C_1 voiced* #g$^{\ulcorner}$v, #g$^{\ulcorner}$z, #g$^{\ulcorner}$m, #g$^{\ulcorner}$l, #g$^{\ulcorner}$r #vm, #vl, #vr, #zm, #zl, #zr, #m$^{\ulcorner}$l, #m$^{\ulcorner}$r	
plateau	14	*C_1 voiceless* #k$^{\ulcorner}$p, k$^{\ulcorner}$t$^{\ulcorner}$#, #fs *C_1 voiced* #gd, #m$^{\ulcorner}$n	*C_1 voiceless* k$^{\urcorner}$t$^{\ulcorner}$#, #sf, #fk, #sk *C_1 voiced* #vd, #zd, #vz, #zv, #m$^{\urcorner}$n
fall	20	*C_1 voiced* #m$^{\ulcorner}$k, #m$^{\ulcorner}$f, #m$^{\ulcorner}$s, #rk, #rg, #rf, #rv, #rs, #rz, #rm	*C_1 voiced* #m$^{\urcorner}$k, #m$^{\urcorner}$f, #m$^{\urcorner}$s, #lk, #lb, #lf, #lv, #ls, #lz, #lm

As in Experiment 1, the stop in stop-initial clusters is always released and thus exhibits an intensity rise. To see whether intensity rise plays a role in the perception of stop-initial clusters, the word-final stop-stop cluster [kt#] was also included, which is shown in Figure 2-42. Since the first stop in stop-stop clusters in Russian is always clearly released even word-finally (cf. Kochetov, 2008), as in Figure 2-42, I created another word-final [kt] cluster where the release burst of the first stop was removed, as shown in Figure 2-43.

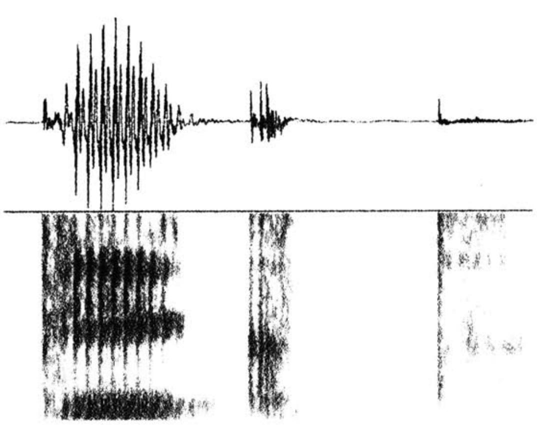


Figure 2-42: k$^{\ulcorner}$t$^{\ulcorner}$#

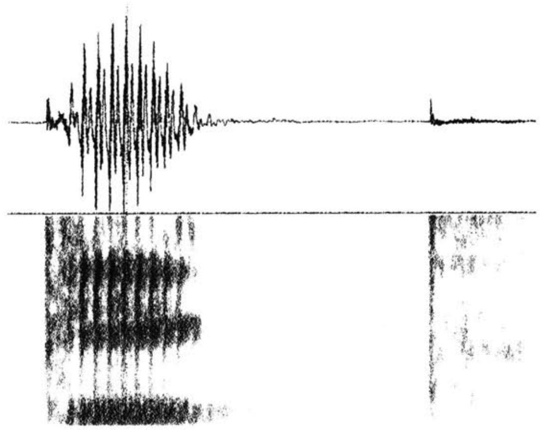


Figure 2-43: k$^{\urcorner}$t$^{\ulcorner}$#

Among the stimuli were fricatives, both sibilant and non-sibilants, in either position in the cluster. Fricative-sonorant clusters involve a rise in intensity, because the intensity level was lower in fricatives than in nasals or liquids. On the other hand, fricative-stop or fricative-fricative clusters involved no intensity rise, except for the [fs] cluster in which the initial non-sibilant [f] has lower intensity than the following sibilant [s], as discussed earlier in Figures 2-19 and 2-20. Unlike the voiceless non-sibilant-sibilant cluster [fs], its voiced counterpart [vz] was classified as a cluster with no intensity rise, even though the non-sibilant [v] preceded the sibilant [z]. Looking at Figure 2-44, intensity is increasing from [v] to [z], but in Figure 2-45, with a high-frequency region over 900 Hz filtered out, [v] displays the higher intensity than [z].[7] As discussed in the previous section, we focus on intensity in the low-frequency band, and classify [vz] as a cluster involving no intensity rise. Conversely, [v] following the word-initial [z] in [#zv] did not involve high intensity as in [vz], as illustrated in Figures 2-46 and 2-47. Therefore, [#zv] was not included in the intensity rise group of the clusters.

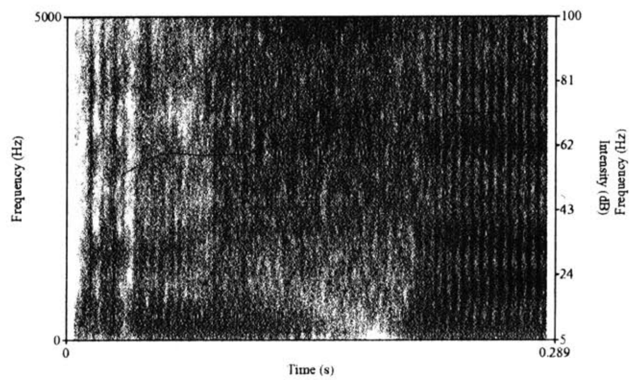


Figure 2-44: #vz

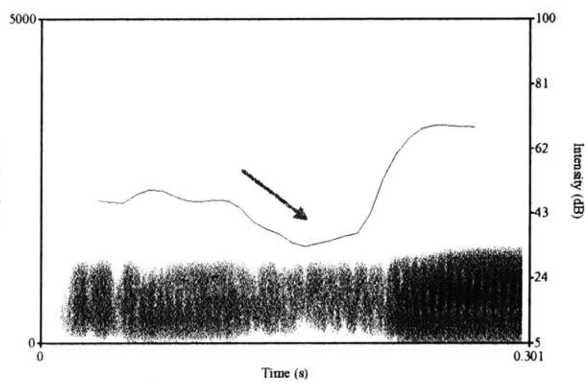


Figure 2-45: #vz (band-pass filtered)

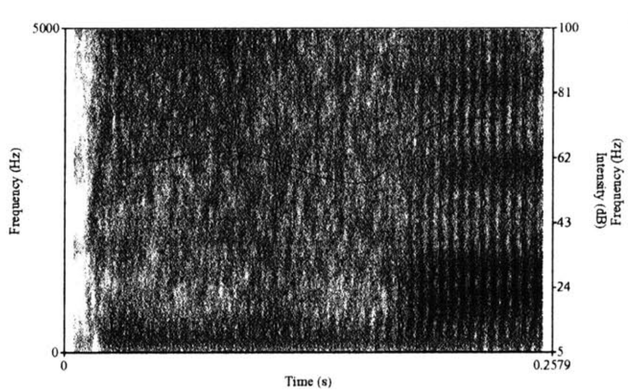


Figure 2-46: #zv

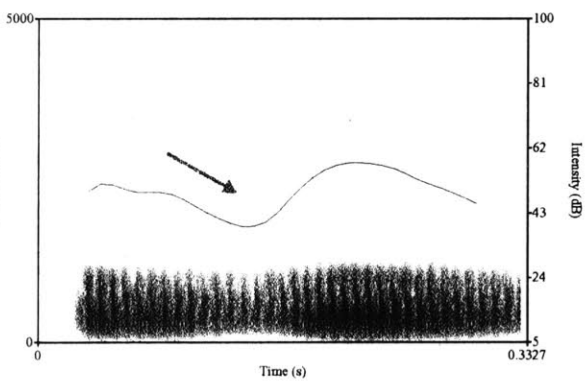


Figure 2-47: #zv (band-pass filtered)

All nasal-initial clusters except nasal-liquid clusters involved two kinds of tokens, one with a

[7]This may be Russian-specific because of the special status of [v] in Russian. Russian [v] may behave like a sonorant or like an obstruent, depending on the type of phonological process (Padgett, 2002, a.o.). In word-initial [vz], it is predicted that [v] maintains its voicing, as shown in Figure 2-44. It is not apparent, however, that the [v] preceding a voiced obstruent always features high amplitude as in Figure 2-44, because to my knowledge, previous studies of Russian [v] have focused on whether it is voiced or not, not focusing on its amplitude. The stimulus [vz] in this experiment may contribute to but does not make any generalizations about phonetic realizations of Russian [v].

released nasal and the other with an unreleased nasal, similar to the released and unreleased [md] pair in Figures 2-34/2-35 and 2-36/2-37. It should, however, be noted that there were only a few nasal-obstruent clusters featuring an audible release, whereas there were many such nasal-nasal clusters. For the nasal-liquid clusters [ml] and [mr], an unreleased nasal did not occur.

For liquid-initial clusters, trill-initial clusters were included in the stimuli as intensity-rising clusters, whereas lateral-initial clusters were included as non-intensity-rising clusters.

2.2.2.1.3 Procedure

As in Experiment 1, an AX dicrimination task was used. The stimuli were played in a randomized order. Each trial presented a pair of the same or different stimuli with 350 ms of inter-stimulus interval. The participants were instructed to press the key marked "kat^h*im*" (meaning 'same' in Korean; written in Korean) for the Korean speakers and "same" for the English speakers if they thought the two items sounded identical, and to press the key marked "*tarim*" (meaning 'different' in Korean; written in Korean) for the Korean speakers or "diff" for the English speakers if they thought the two items sounded different from each other. This instruction, in Korean for Korean speakers and in English for English speakers, was presented on the screen throughout the experiment, as well as being delivered by the experimenter and presented on the screen before the experiment started. The participants were also informed that the presented words were not Korean or English words.

The experiment was preceded by a practice with eight trials. A total of 480 trials were presented divided by 10 blocks, involving two repetitions of 58 types of consonant clusters, four types of pairs (AA, XX, AX, XA), and eight fillers for a sanity check. The sanity check items consisted of different pairs of mono- or bisyllabic Russian words that are so easy to discriminate (e.g., [bol] vs. [da]). The whole procedure took 25-30 minutes for the Korean speakers and 20-25 minutes for the English speakers.

The Korean participants did the experiment through Praat on a Macbook Air in a quiet room. The participants wore the Audio-Technica or Sennheiser headphones used in Experiment 1. For the English participants, an iMac and Sony headphones were used and the experiment was conducted in the MIT Behavioral Research Lab.

2.2.2.2 Results

Sections 2.2.2.2.1 and 2.2.2.2.2 present the results from the Korean participants and from the English participants, respectively. Regardless of the participant group, it was shown that the differences in the talker of the stimuli and the duration of the epenthetic vowels did not have a significant effect on the experimental results.

2.2.2.2.1 Korean listeners

Here are the results from the Korean participants. The percentage correct results are given in Table 2.17 and in Figure 2-48 by intensity rise and C_1 voicing. The interaction between intensity rise and C_1 voicing was significant ($\chi^2(1)$=7.75, p<.01), confirmed by a Likelihood Ratio Test. To be specific, a mixed effects logistic regression model showed that intensity rise was a significant predictor when C_1 was voiced (β=0.92, z=2.87, p<.01), but not when C_1 was voiceless. On the other hand, C_1 voicing was a significant predictor when there was a rise in intensity (z=6.47, p<.001), but not when there was no rise in intensity.

Table 2.17: Percentage correct by intensity rise and C_1 voicing (Korean listeners)

	intensity rise	no intensity rise
C_1 voiced	57.24%	72.11%
C_1 voiceless	78.10%	75.82%

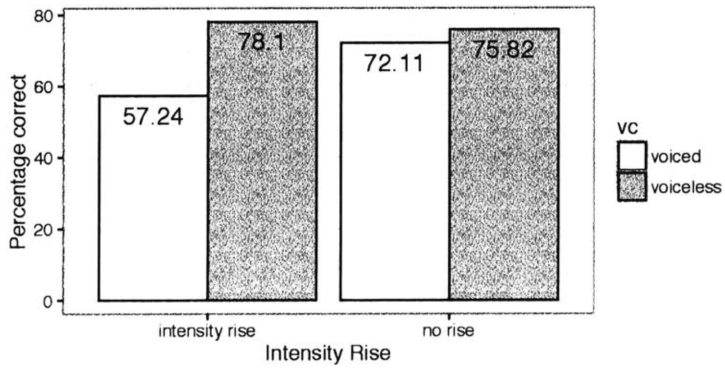


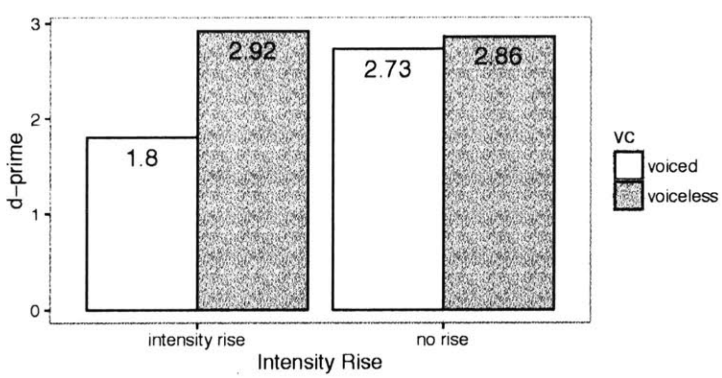
Figure 2-48: Percentage correct by intensity rise and C_1 voicing (Korean listeners)

d' results in Table 2.18 and in Figure 2-49 show the same pattern and statistical significance. A mixed effects probit regression model confirmed that intensity rise was a significant predictor of the discrimination between CC and CəC only when C_1 was voiced (β=1.09, z=8.1, p<.001), and so was C_1 voicing only when there was a rise in intensity (β=1.31, z=9.99, p<.001).

Table 2.18: d' scores by intensity rise and C_1 voicing (Korean listeners)

	intensity rise	no intensity rise
C_1 voiced	1.80	2.73
C_1 voiceless	2.92	2.86

Figure 2-49: d' scores by intensity rise and C_1 voicing (Korean listeners)



2.2.2.2.2 English listeners

Turning now to the results from the English listeners, Tables 2.19 and 2.20 illustrate percentage correct and d' results classified by intensity rise and C_1 voicing, and the results seem quite similar to what we have seen in Tables 2.17 and 2.18 for Korean.

Table 2.19: Percentage correct by intensity rise and C_1 voicing (English listeners)

	intensity rise	no intensity rise
C_1 voiced	70.26%	80.59%
C_1 voiceless	85.44%	83.77%

Table 2.20: d' scores by intensity rise and C_1 voicing (English listeners)

	intensity rise	no intensity rise
C_1 voiced	2.46	2.96
C_1 voiceless	3.22	3.01

However, these results were confounded by the nativeness of the stimulus clusters. That is, the stimuli included native English clusters, such as [sk] and [kl], which the participants were familiar with and could discriminate well. Indeed, the native clusters in English were significantly better discriminated than non-native ones (β=0.95, z=4.44, p<.001).

Considering results from clusters that do not exist in English only, the mean percentage correct and d' scores decrease, but they are still higher than those of Korean participants in general. More importantly, the results show a very similar pattern to the results from the Korean participants seen in section 2.2.2.2.1. Table 2.21 and Figure 2-50 illustrate the accuracy proportions by intensity rise and C_1 voicing on non-native trials. The two significant predictors interacted with each other at a statistically significant level ($\chi^2(1)$=5.34, p<.05). Similarly to the results of the Korean participants, a mixed effects logistic regression model showed that intensity rise was a significant predictor only when C_1 was voiced (β=1, z=2.38, p<.05), and C_1 voicing was significant only when there existed an intensity rise in the cluster (β=0.64, z=3.38, p<.001).

Table 2.21: Percentage correct on non-native trials (English listeners)

	intensity rise	no intensity rise
C_1 voiced	70.49%	80.59%
C_1 voiceless	81.47%	74.73%

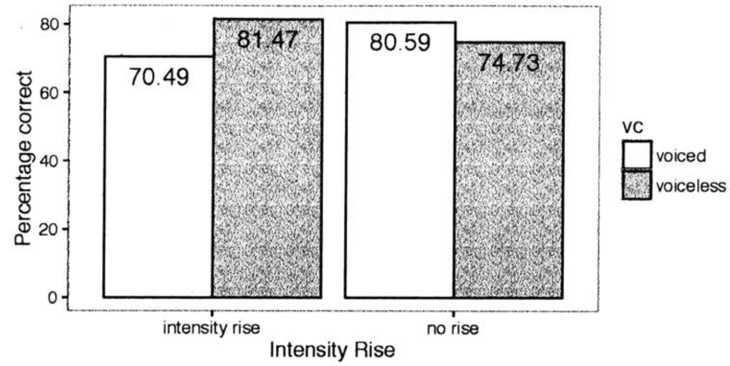


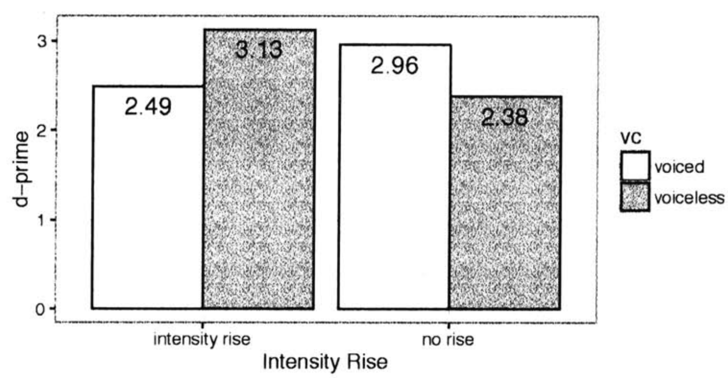
Figure 2-50: Percentage correct on non-native trials (English listeners)

On the other hand, d' results in Table 2.22 and Figure 2-51 exhibit statistical significance in every contrast. Mixed effects probit regression models including participant as a random effect showed that the effect of intensity rise was highly significant both when C_1 was voiced (β=0.65, z=6.07, p<.001) and when C_1 was voiceless (β=1.53, z=6.51, p<.001), and the effect of C_1 voicing was also significant regardless of whether there was an intensity rise in the cluster (β=0.82, z=6.4, p<.001) or not (β=0.71, z=3.58, p<.001). Unpredicted is that voiceless-consonant-initial clusters were less accurately discriminated from the epenthesis forms than voiced-consonant-initial clusters in consonant clusters involving no intensity rise.

Table 2.22: d' scores on non-native trials (English listeners)

	intensity rise	no intensity rise
C_1 voiced	2.49	2.96
C_1 voiceless	3.13	2.38

Figure 2-51: d' scores on non-native trials (English listeners)



2.2.2.3 Discussion

In Experiment 2, we established that intensity rise and C_1 voicing interact with each other, and intensity rise and C_1 voicing are significant predictors of the discriminability between clusters and their matching epenthesis forms conditionally. A question is why the predictors are significant only in certain conditions. First, the effect of intensity rise was significant only when the consonant cluster begins with a voiced consonant because clusters that involve an intensity rise and begin with a voiceless consonant showed quite high accuracy, contrary to the prediction that they would show lower accuracy than clusters involving no intensity rise that begin with a voiceless consonant. The d' results of English listeners show contradictory effects of an intensity rise depending on whether the cluster begins with a voiced or voiceless consonant. Let us consider the stimulus clusters that begin with a voiceless consonant in more detail. As illustrated in Table 2.16, the stop [k], and the fricatives [f] and [s] are the initial consonant in such clusters. In the Korean listeners' results, intensity rise significantly affects the discriminability between [k]-initial clusters and their epenthesis counterparts. [k˺t˹#] is the only [k]-initial cluster involving no intensity rise, and the accuracy proportion is 75%. The accuracy proportion of the other word-final [k]-initial cluster [k˹t˹#] that involves an intensity rise inside the cluster is 68.12%, lower than the one without an intensity rise, as predicted by the hypothesis. For English listeners, there is no non-native stop-initial cluster without an intensity rise. Hence, the results in clusters that begin with the voiceless stop display the expected pattern.

It is fricative-initial clusters that show an unexpected pattern among the clusters beginning with a voiceless consonant; fricative-initial clusters involving no intensity rise were not really better discriminated than those involving an intensity rise. Figures 2-52 and 2-53 present the accuracy proportions of the fricative-initial clusters for Korean and English listeners, respectively.

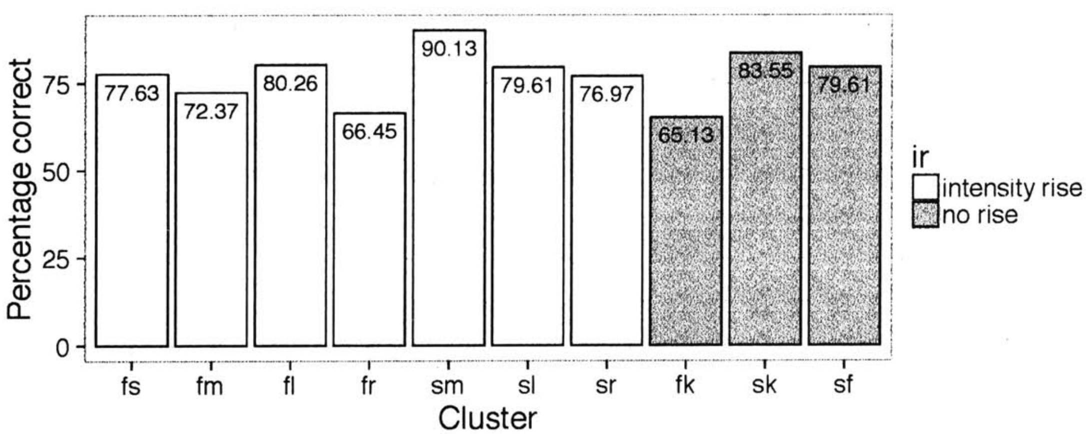


Figure 2-52: Percentage correct on fricative-initial clusters (Korean listeners)

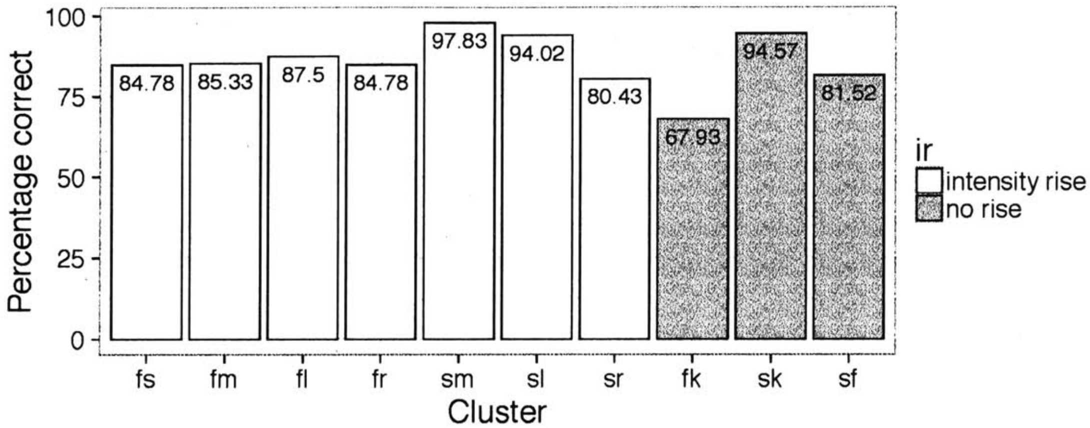


Figure 2-53: Percentage correct on fricative-initial clusters (English listeners)

Considering Korean listeners' results in Figure 2-52, we see that the percentage correct of [fk] is the lowest, even though it does not involve an intensity rise. On the other hand, all fricative-initial clusters involving an intensity rise exhibit a percentage correct higher than the mean percentage correct of the intensity-rising clusters, 64.73%. Similar patterns are observed in English listeners' results in Figure 2-53, where English native clusters are also included. The [fk] cluster shows the lowest percentage correct, as in the Korean listeners' results above.[8] Non-native clusters involving an intensity rise, [fs] and [fm], are well discriminated comparable to the native clusters, [fl] and [fr], even though non-native clusters of English were less accurately discriminated than native clusters in general, as mentioned previously.

This mysterious result of fricative-initial clusters is similar to Fleischhacker's (2005) results in the AX discrimination task mentioned earlier. All fricative-initial clusters in her experiment are [s]-initial, which exist in English. The percentages correct are ordered from the highest to the lowest: [sn] > [sl] > [sw] > [sk]. The lowest accuracy for the discrimination of [sk] is not expected by Fleischhacker's hypothesis, as well as by the current one. The results in Davidson

[8]Although the unexpectedly low percentage correct of the [fk] cluster is noticeable, it is hard to attribute the result to a stimulus artifact. First, the duration of the epenthetic vowel in the [fk] is 61.2 ms and is not particularly shorter than the vowels in the other clusters. The mean and maximum intensity of the epenthetic vowel are also not very different compared to the other clusters. As shown in (1) below, the intensity of the epenthetic vowel in [fk] is slightly lower than the other clusters, but is higher than the one in [fs], which displays a higher accuracy than [fk]. Crucially, cluster was included as a random effect in the regression models. There is also no data on epenthesis in [fk] in loan adaptation, and so this result does not go against the empirical data.

(1) Mean and maximum intensity of the epenthetic vowel in voiceless fricative-initial clusters (dB)

cluster	mean dB	maximum dB
fk	62.1	63.34
fs	61.41	62.6
fm	66.03	67.18
fl	64.1	66.08
fr	64.81	66.44
sk	65.28	66.89
sf	64.39	65.58
sm	66.66	67.42
sl	65.92	66.49
sr	66.03	68.34

and Shaw (2012) also show similar high accuracy for the discrimination of fricative-stop clusters, involving no intensity rise. The discriminability between fricative-stop clusters ([vb], [zb], [fp], [sp]) and their epenthesized counterparts is overall comparable to the discriminability between fricative-nasal clusters ([vm], [zm], [fm], [sm]) and their epenthesis forms; the former was slightly higher in the AX discrimination, and the latter was so in the ABX discrimination. Although it is not clear at the moment why this unexpected result is observed in voiceless fricative-initial clusters, it is consistent with the results of the previous studies. I leave this question for future research.

Recall that no voicing effect was observed in clusters with no intensity rise. The non-intensity-rising clusters that begin with a voiceless consonant are [k˺t˺#], [sf], [fk], and [sk], and those that begin with a voiced consonant are [vd], [vz], [zd], [zv], and [m˺n]. Figures 2-54 and 2-55 illustrate the accuracy proportions of these clusters for Korean and English listeners, respectively.

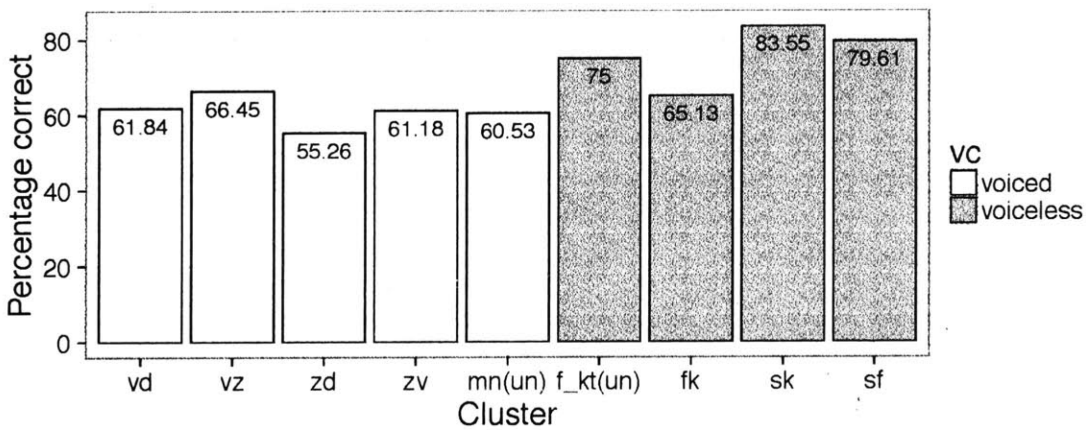


Figure 2-54: Percentage correct on non-intensity rise clusters (Korean listeners)

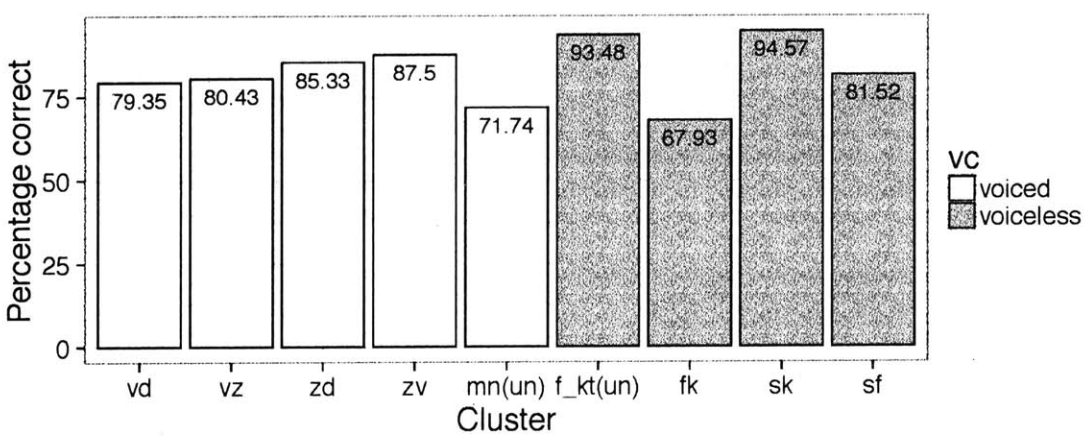


Figure 2-55: Percentage correct on non-intensity rise clusters (English listeners)

Looking at the Korean listeners' results in Figure 2-54, clusters with no intensity rise that begin with a voiced consonant exhibit higher accuracy than all clusters that begin with a voiceless consonant except for the [fk] cluster, which shows a comparable percentage correct to the clusters that begin with a voiced consonant. So it seems likely that the low accuracy in the [fk] cluster cancels the voicing difference in clusters with no intensity rise.

On the other hand, three of the clusters involving no intensity rise that begin with a voiceless consonant, i.e., [k˺t˹#], [sf], and [sk], are native clusters in English, and therefore in Figure 2-55 the actual comparison is made between all the non-intensity-rise clusters beginning with a voiced consonant and the [fk] cluster. As in Korean listeners' results, the [fk] cluster shows an unexpectedly low percentage correct. Thus, we can conclude that the no significant difference in C_1 voicing in clusters involving no intensity rise is again attributed to the unpredicted pattern of the [fk] cluster and thus does not reject the current hypothesis.

In short, the current hypotheses about intensity rise and C_1 voicing holds true with the exception of fricative-initial clusters. This ungeneralizable pattern of fricative-initial clusters, which is similar to ones in the previous research, is in need of further investigation.

2.2.3 Experiment 3

Although the role of intensity rise and C_1 voicing in consonant cluster perception is investigated only with word-initial clusters in the previous section, it is expected that both intensity rise and C_1 voicing play a role in consonant clusters at any position. In other words, it is hypothesized that word-medial and word-final clusters are also perceptually more similar to their matching epenthesis forms when involving an intensity rise and/or C_1 voicing than when not involving these factors. Experiment 3 was conducted to test this hypothesis.

2.2.3.1 Methods

2.2.3.1.1 Participants

Participants in Experiment 3 were 37 native speakers of Korean who lived in Korea, recruited online. All of them self-reported that they learned English in school but had hardly used it since they graduated from high school. Ages varied from 20s to 50s. No participant reported speaking/hearing problems. $5 were paid for their participation. I analyzed 35 participants, excluding two who made more than two errors out of eight sanity check trials, which were also used in Experiment 2.

2.2.3.1.2 Stimuli

Stimuli for Experiment 3 involve nonce words ending in a consonant cluster and their epenthetic counterparts. The final clusters were embedded in one of the three templates, [netáCC], [betáCC], or [mitáCC], and in the matching epenthesis forms, [netáCəC], [betáCəC], or [mitáCəC], in which stress always fell on the second vowel [a]. The stimuli were chosen from the recorded speech of a native speaker of Ukrainian from Kiev, Ukraine, who was also bilingual in Russian, and the speech of three native speakers of Russian. Since Russian speakers devoice the final voiced obstruents, the Ukrainian speaker's recordings were used for word-final clusters ending in a voiced obstruent. An example of word-final obstruent clusters, [gd#], is presented in Figure 2-56. The recording was conducted in the same setting as in Experiments 1 and 2.

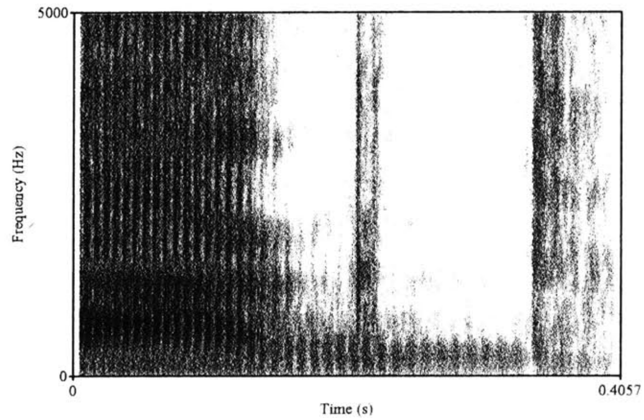


Figure 2-56: Spectrogram of [agd] from the stimulus [betagd]

The consonant clusters used in the stimuli are listed in Table 2.23. Most of these word-final clusters have word-initial counterparts employed in either Experiment 1 or Experiment 2. Notice that there were not two types of nasal-obstruent clusters, released and unreleased, unlike the stimuli in Experiment 2; in word-final position, pre-obstruent nasals were not audibly released in any cluster.

Table 2.23: Consonant clusters used in the stimuli of Experiment 3

sonority	#	intensity rise	no intensity rise
rise	11	*C_1 voiceless* k⌜m#, k⌝l#, fn#, fl# *C_1 voiced* g⌜z#, g⌜m#, g⌝l#, vm#, vl# m⌝l#, m⌜r#	
plateau	10	*C_1 voiceless* k⌜t⌜#, k⌜f#, fs# *C_1 voiced* g⌜d⌜#, rl#	*C_1 voiceless* fk⌜#, sk⌜#, sf# *C_1 voiced* zg#, m⌝n#
fall	11	*C_1 voiced* rk⌜#, rs#, rm#	*C_1 voiced* m⌝k⌜#, m⌝f#, m⌝g⌜#, m⌝z# lk⌜#, lg⌜#, lf#, lm#

Durations of the epenthetic vowels ranged from 50 ms to 80 ms, which can be found in Appendix C. As in the previous experiments, the other factors, i.e., pitch pattern, duration of the stressed vowel, and the rest of the stimulus, were controlled so as not to affect the results.

2.2.3.1.3 Procedure

An AX discrimination task was employed as in Experiments 1 and 2. Unlike Experiments 1 and 2, however, Experiment 3 was conducted online using Experigen (Becker and Levine, 2013). First, the participants read the instructions, written in Korean, on the first page of the experiment webpage, and were asked to do the experiment in a quiet place putting on headphones, not using loudspeakers. Then they were asked to click a play button at the bottom of the page, which played a song that was RMS normalized with the stimuli, and to adjust volume to a comfortable listening level. After a short demographic survey, there was a practice session using five trials, and the actual experiment followed. For each trial, the participants were solicited to click a play button with a mouse, which played a pair of stimuli, with 500 ms of initial silence and 300 ms of inter-stimulus interval. Once the playback was done, two buttons appeared: one marked "*kat^hɨm*" (meaning 'same' in Korean; written in Korean) and the other marked "*tarɨm*" (meaning 'different' in Korean; written in Korean). And the participants clicked the "*kat^hɨm*" button if they thought the two words they heard sounded the same or the "*tarɨm*" button if they thought the two words sounded different from each other.

The 32 clusters listed in Table 2.23 were presented in four types of pairs (AA, XX, AX, XA) with eight sanity-checking fillers. Therefore, there were 134 trials in total, and the trials were randomized for each participant. There was no repetition because I aimed to keep the experiment short enough, supposing that participants would be more easily distracted in the online experiment setting than in the lab. Also, there was no scheduled break, but the participants were allowed to have a break, if they wanted, before clicking the play button for each trial. The procedure took around 10-15 minutes.

2.2.3.2 Results

As in Experiments 1 and 2 employing word-initial clusters, intensity rise was a significant predictor in word-final clusters too. Unexpectedly, however, there was no significant effect of C_1 voicing. Table 2.24 and Figure 2-57 shows percentage correct results by intensity rise and by C_1 voicing. A mixed effects logistic regression model confirmed that intensity rise was a significant predictor when C_1 was voiced (β=1.72, z=6.05, p<.001) and when C_1 is voiceless (β=1.07, z=2.03, p<.05). In contrast, C_1 voicing is not a significant predictor regardless of whether there is an intensity rise or not in the cluster. Also, a Likelihood Ratio Test confirmed that the interaction between intensity rise and C_1 voicing was statistically significant ($\chi^2(1)$=3.91, p<.05).

Table 2.24: Percentage correct by intensity rise and C_1 voicing (word-final clusters)

	intensity rise	no intensity rise
C_1 voiced	66.75%	91.09%
C_1 voiceless	72.21%	84.20%

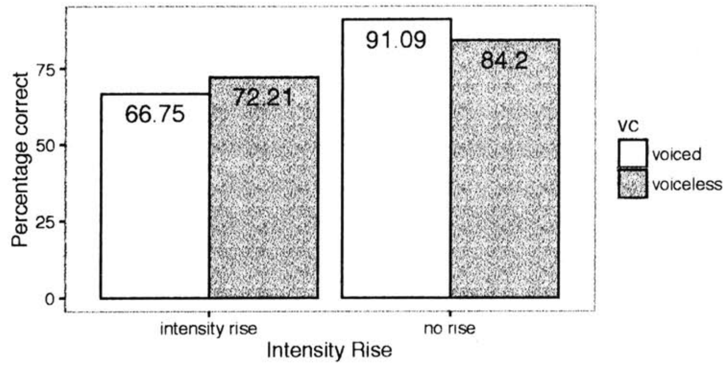


Figure 2-57: Percentage correct by intensity rise and C_1 voicing (word-final clusters)

Mean d' scores in Table 2.25 and in Figure 2-58 show similar tendencies to the percentage correct in Table 2.24. Namely, the effect of intensity rise was statistically significant when C_1 was voiced (β=0.74, z=8.89, p<.001). In d', unlike in percentage correct, the significance of the intensity rise effect was marginal when C_1 was voiceless (β=0.23, z=1.95, p=.515). Again, C_1 voicing did not yield a significant difference.

Table 2.25: d' scores by intensity rise and C_1 voicing (word-final clusters)

	intensity rise	no intensity rise
C_1 voiced	2.35	3.58
C_1 voiceless	2.60	2.94

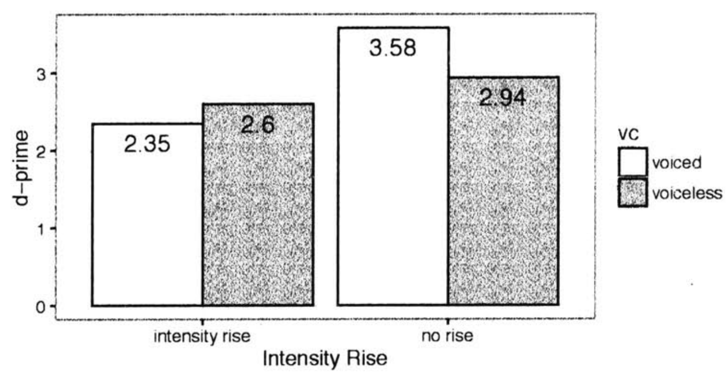


Figure 2-58: d' scores by intensity rise and C_1 voicing (word-final clusters)

2.2.3.3 Discussion

In Experiment 3, intensity rise was a significant predictor of discrimination between clusters and epenthesis forms, but C_1 voicing was not, unlike in Experiments 1 and 2. This is due to the higher accuracy and discriminability in clusters that involve no intensity rise and begin with a voiced consonant. Table 2.26 presents percentages correct in clusters involving no intensity rise. We see here that most of the clusters beginning with a voiced consonant, most of which were [m]- or [l]-initial, show a percentage correct higher than 90%.

Table 2.26: Percentage correct in clusters with no intensity rise

C_1 voicing	cluster	percentage correct
C_1 voiceless	fk	83.21
	sk	82.84
	sf	86.57
C_1 voiced	zg	88.72
	mn	90.30
	mk	77.61
	mf	95.56
	mg	94.74
	mz	92.48
	lk	93.28
	lg	92.48
	lf	94.78
	lm	90.91

One possible interpretation of this result is that the Korean participants were able to discriminate [m]-initial and [l]-initial clusters well under the influence of native phonotactics. Korean allows [mC] and [lC] clusters in word-medial intervocalic position. Also, vowel epenthesis tends to occur in Korean loanword adaptation after a word-final singleton consonant (e.g., 'gag' → [kæki), as well as after a cluster-final consonant (e.g., 'land' → [ɾænti]). So it is possible that the Korean participants heard an illusory vowel after the final consonant and perceived the word-final clusters as word-medial, which are familiar to them. For example, the stimulus [betamg] may have been perceived as [betamgi] by the Korean listeners, and the [mg] cluster is licit in Korean phonotactics and was easy to discriminate. This explanation, however, is not applicable to the [mn] and [lm] clusters that end in a nasal because the word-final nasal was not audibly released and thus was not likely to be perceived with an illusory vowel after it. For the [mn] and [lm] clusters, there is a possibility of misperception as a singleton because the two consonants are relatively similar to each other, compared to the clusters that consist of a sonorant and an obstruent. This hypothesis needs to be substantiated by additional research.

2.2.4 Summary

This section has provided experimental evidence for the hypothesis that the two auditory factors, intensity rise and C_1 voicing would have a crucial influence on the perceptual similarity between consonant clusters and their corresponding epenthesis forms. The discrimination results show that consonant clusters involving an intensity rise are harder to discriminate from the epenthesis forms than clusters involving no intensity rise, and that consonant clusters beginning with a voiced consonant are harder to discriminate from the epenthesis forms than clusters beginning with a voiceless consonant. It should be stressed that even the same cluster showed different discriminabilities from the epenthesis form depending on its phonetic realization; for example, participants were significantly better at discriminating the [md] cluster with an unreleased [m] from its epenthesis counterpart than the [md] with a released [m]. However, there are still unexplained results which need further investigation. In particular, the current hypothesis does not hold for the discrimination in word-initial sibilant-initial clusters. Also, we need to find out additional factors that can explain no C_1 voicing effect in word-final clusters without an intensity rise, which I leave for future research.

It should also be noted that the experimental results were largely consistent across participant groups. Regardless of the native language of the participants, intensity rise and C_1 voicing were mostly the significant predictor of the discrimination between the consonant clusters and the vowel epenthesis forms. This indicates that the current results are not derived from language-specific knowledge but from universal knowledge of perceptual similarity/distinctness.

The next chapter will show how this knowledge of perceptual similarity plays a role in the typology of vowel epenthesis in loanword adaptation.

Appendix 2A
Duration of the epenthetic vowels in Experiment 1 (msec)

cluster	following vowel		
	[a]	[i]	[u]
bd	57.9	53.6	56.8
bn	53.7	63.5	61.1
db	61.3	62.8	65.3
dg	62.3	67.7	59.3
dm	44	67.5	62.3
gb	56.8	64.5	68.4
gm	50.3	62.4	64.4
gn	41.7	74.8	81.1
km	60.7	40.8	53.9
kn	45.7	60.6	56.3
kp	45.3	35.5	52.1
kt	34.9	30.5	46.2
lb	37.8	38.1	36.3
lg	43.4	43.9	39.5
lk	61.6	44.8	32.3
lm	71.1	55.5	30.7
lp	34	34	43.2
md	71.5	71.6	60.1
mg	57.7	72.3	65.1
mk	60.9	75.1	71.4
ml	44.9	38.6	58.3
mn	75.1	80.5	57
ng	74.5	62.8	64.1
nk	78.8	74.1	77.6
nm	44.4	61.1	75.4
pk	58.7	40.8	50.2
pn	61.6	86	76.9
pt	41.9	51.7	46.2
tm	56.5	70.7	61.1

Appendix 2B
Duration of the epenthetic vowels in Experiment 2 (msec)

cluster	ə duration	cluster	ə duration	cluster	ə duration	cluster	ə duration
final kt	51.1	fk	61.2	fl	92.9	fm	70.7
fr	84.9	fs	65.7	gd	54.9	gl	79.8
gm	55.3	gr	73.3	gv	64.2	gz	72.2
kf	52	kl	101	km	69.9	kp	51.4
kr	44.1	ks	52.8	lb	62.2	lf	71.8
lk	69.8	lm	90.8	lv	76.2	lz	75.7
mf	81.8	mk	72.8	ml	83.2	mn	94.5
mr	78.5	ms	70.8	rf	69.3	rg	90.2
rk	58.4	rm	70.8	rs	72.8	rv	78.7
rz	88.7	sf	67.5	sk	67	sl	90.4
sm	80.5	sr	74.4	vd	73.7	vl	71.8
vm	60.2	vr	70	vz	82.4	zd	75
zl	73.4	zm	78.2	zr	99.3	zv	88

Appendix 2C
Duration of the epenthetic vowels in Experiment 3 (msec)

cluster	ə duration	cluster	ə duration	cluster	ə duration	cluster	ə duration
fk	65.9	fl	51	fn	53.7	fs	62.1
gd	61.7	gl	72.8	gm	58.3	gz	66.7
kf	48.4	kl	61.4	km	53.7	kt	49.4
lf	68.2	lg	66.8	lk	62.1	lm	57.6
mf	56.4	mg	53.8	mk	63.5	ml	47.2
mn	68.9	mr	75.8	mz	75.9	rk	60.1
rl	75.1	rm	73.3	rs	59.8	sf	46.4
sk	47.5	vl	74.9	vm	61.9	zg	72.9

Chapter 3

A typology of epenthesis sites in loanword adaptation

This section explores cross-linguistic patterns of epenthesis positions. In loanword adaptation, when a consonant cluster in a source language is phonotactically illegal in the borrowing language, vowel epenthesis may take place to repair the illegal cluster. For example, English 'cloth' is adapted as [kɪlɔθ] in Hindi (Singh, 1985), with an inserted vowel between the two consonants. When vowel epenthesis takes place to repair a consonant cluster not allowed in a certain position, there are two possible epenthesis sites in principle: (i) between the two consonants (internal epenthesis, e.g., [#kla] → [#kəla]), and (ii) outside of the cluster (external epenthesis, e.g., [#kla] → [#əkla]). In Hindi, unlike the example above, 'school' is borrowed as [ɪskul], with external epenthesis. Of special interest is that the sites of vowel epenthesis differ depending on the cluster, and for most clusters only one pattern of epenthesis is observed across languages. For example, if [#kl] is repaired by epenthesis, the vowel is always epenthesized between the consonants, and if [#mb] is repaired by epenthesis, the vowel is always epenthesized outside the cluster.

This cluster-dependent asymmetry in epenthesis sites has been paid much attention by phonologists (e.g., Singh, 1985; Broselow, 1992; Fleischhacker, 2001, 2005; Gouskova, 2001; Steriade, 2006, a.o.). Previous analyses of loanword adaptation have argued that the position of vowel epenthesis depends on the sonority profile of the cluster. Clusters with rising sonority, e.g., the stop-sonorant cluster in 'cloth', are more likely to undergo internal epenthesis than clusters with level/falling sonority, e.g., the sibilant-stop cluster in 'school'. Most of the previous literature has only focused on epenthesis in stop-sonorant and sibilant-stop clusters in word-initial position (e.g., Singh, 1985; Broselow, 1992; Fleischhacker, 2001, 2005). It appears that this is because words beginning with those clusters in English are borrowed in many languages and are well-studied. On the other hand, Gouskova (2001) expands the sonority-based generalization to clusters with falling sonority other than sibilant-stop clusters, based on Russian loanwords in Kirghiz, as Russian allows far more various types of consonant clusters than English at word edges. For example, the Russian word *trupka* 'pipe' is borrowed into Kirghiz as [turupke] with internal epenthesis, whereas *lbovskij* 'nonce last name' is borrowed as [ylbovskij] with external epenthesis. All of these studies, however, have centered around the patterns of vowel epenthesis that occur in word-initial position only, and few have attempted to address the patterns of vowel epenthesis in non-initial position.

This study aims to provide a more comprehensive coverage of the sites of vowel epenthesis in loanword adaptation, by considering a variety of biconsonantal clusters, both in word-initial and word-final positions. So this study extends the typology of consonant clusters beyond sibilant-stop and obstruent-sonorant clusters. Also, epenthesis sites in word-final position are considered: internal epenthesis (e.g., [akl#] → [akəl#] vs. external epenthesis (e.g., [akl#] → [aklə#]). Epenthesis in word-medial clusters is not considered since there is only one possible epenthesis site in word-medial biconsonantal clusters (e.g., $VC_1C_2V \rightarrow VC_1əC_2V$), and thus no cluster-dependent asymmetries are possible.

This study focuses on the cases where language-specific phonotactics do not obligatorily determine the epenthesis site. Although this confound is ignored in most previous studies, in many cases, only one pattern of epenthesis, either internal or external, is possible because of a ban on certain sequences in the native phonology of the borrowing language. An example of a markedness-driven cluster-dependent asymmetry is observed in Yakut (Sakha). Obstruent-liquid clusters are adapted with internal epenthesis in Yakut, as shown in (5a), whereas sibilant-stop clusters are adapted with external epenthesis, as in (5b). However, these examples are not included in the current database because the obstruent-liquid clusters are not allowed even in intervocalic position and thus external epenthesis is not a possible repair strategy. The examples in (6) show that the obstruent-liquid clusters are also illegal word-medially and repaired by an epenthetic vowel between the two consonants. So for Russian *klad*, for example, *[ɯklaat] with external epenthesis, which includes a phonotactically illegal sequence in Yakut, is not a phonotactically legal structure, and [kɯlaat] with internal epenthesis is thus the only possible epenthesis repair. In this case, the choice in epenthesis sites between obstruent-liquid and sibilant-stop clusters is not in play, and thus this type of case is not considered in the current study.

(5) Russian loanwords in Yakut (Vasilyeva, 2010)

a. Internal epenthesis in obstruent-sonorant clusters

Russian	*Yakut*	*gloss*
klad	kɯlaat	'treasure'
platʲə	bɯlaaçça	'dress'
sluʐba	suluuspa	'service'
korablʲ	xaraabɨl	'ship'

b. External epenthesis in fricative-stop clusters

Russian	*Yakut*	*gloss*
stol	ostuol	'table'
ʃkaf	ɯskaap	'closet'
spitsɨ	ispiisse	'knitting needles'

(6) Epenthesis in word-medial clusters (Pakendorf and Novgorodov, 2009; Vasilyeva, 2010)

Russian	*Yakut*	*gloss*
kukla	kuukula	'doll'
kreslo	kiriehile	'armchair'
kabluk	χobuluk	'heel'

Therefore, considered here are the cases in which both internal epenthesis and external epenthesis result in a sequence permitted in the borrowing language and hence can both be possible epenthesis repairs for the consonant cluster at the word edge. Loanwords in Bezhta, for example, show such an asymmetry in epenthesis sites free from markedness restrictions. As exemplified in (7), in Bezhta, word-initial stop-liquid clusters are borrowed with internal epenthesis

(7a, b), whereas sibilant-stop clusters are borrowed with external epenthesis (7c). In this language, external epenthesis before stop-liquid clusters does not cause a problem in the native phonology, as we see them occurring intervocalically and adapted intact in (8a, b). For sibilant-stop clusters, internal epenthesis, e.g., *[ʃikola] for *ʃkola* in (7c), should also possible in principle, because a sibilant can initiate a word, as shown in (8c).

(7) Epenthesis in word-initial clusters in Bezhta (Comrie and Khalilov, 2009)

	Russian	*Bezhta*	*gloss*
a.	klej	keley	'glue'
b.	krolik	korolik	'rabbit'
c.	ʃkola	iʃkola	'school'

(8) No epenthesis in word-medial clusters in Bezhta (Comrie and Khalilov, 2009)

	Source	*Bezhta*	*gloss*
a.	aklemi (Georgian)	aklamo	'camel'
b.	pikru (Avar)	pikro	'idea'
c.	ʃakɬi (Avar)	ʃakɬi	'doubt'

The goal of this chapter is twofold: (i) to find generalizations about epenthesis sites in loanword adaptation; and (ii) to investigate the factor(s) that determine epenthesis sites when the native phonology of the borrowing language does not compel one site. By finding universal generalizations about epenthesis sites that are not predictable from language-specific phonology, we can figure out universal factors that derive the typology of epenthesis sites. Section 3.1 presents the results of a survey of epenthesis sites and suggests that the presence or absence of intensity rise in the consonant cluster plays a crucial role in determining the site of vowel epenthesis. Section 3.2 lays out the perceptual hypothesis about the typology of epenthesis sites based on perceptual similarity between clusters and their matching forms with internal or external epenthesis. Section 3.3 provides experimental evidence supporting the hypothesis, and a formal analysis of the typology is given in Section 3.4.

3.1 Crosslinguistic survey

This section presents the results of a cross-linguistic survey of epenthesis sites in loanword adaptation. I conducted a survey of languages in which consonant clusters are not allowed to appear either in word-initial or in word-final position, and vowel epenthesis takes place as a repair. Loanwords from 50 languages were collected both from the literature and from my own fieldwork. The entire list of languages, examples, and references appears in Appendix 3. Unless otherwise noted, examples are from my own elicitations. The data from the fieldwork include only loanwords that are borrowed from a donor language and used in the everyday language of the consultants and do not include the consultants' on-line adaptations, while some examples from the literature include on-line adaptation and second language errors. Notice that clusters involving a glide are not included in the data, because the syllabic status of the glide, i.e., whether it is part of an onset cluster or complex nucleus, may not be clear in each language.

At the outset, it is imperative to state the scope of the present survey. It only concerns cases of epenthesis of *a single vowel* in clusters consisting of *two consonants.* So epenthesis of two vowels, after each consonant, e.g., 'milk' → [miruɯkuɯ] (Japanese; Hirano, 1965), and epenthesis after a single consonant, e.g., 'pad' → [pʰæti] (Korean; Kang, 2003) are not taken into consideration.

As stated earlier, this thesis will be limited to consideration of cases of epenthesis when both internal and external epenthesis are possible. Although this makes the number of available examples rather small, it allows us to see where epenthesis is preferred in an environment that is phonotactically undetermined.

The following subsections present the survey results according to the position of the epenthetic vowel: (i) internal epenthesis (Section 3.1.1), (ii) external epenthesis (Section 3.1.2), and (iii) variation between internal and external epenthesis (Section 3.1.3).

3.1.1 Internal epenthesis

The types of consonant clusters that are always adapted with internal epenthesis are summarized in (9). Examples by contexts and clusters are provided in the following tables, in which the numbers in the second column refer to the number of languages that show the epenthesis pattern in question.

(9) Internal epenthesis in consonant clusters
(T: stop, F: non-sibilant fricative, S: sibilant fricative, N: nasal, L: liquid, l: lateral, C: consonant)

a. Word-initial

cluster		#	example
#TC	#TT	6	gdansk (Polish; city name) → [gədænsk] (English)
	#TF	1	kvadrat (Russian; 'types') → [kuwa:dra:t] (Farsi; Bashiri, 1994)
	#TN	6	kniga (Russian; 'book') → [kiniga] (Even) (Choi, 2010)
	#TL	22	klub (Russian; 'club') → [kɨlub] (Tatar)
#FL		10	'Florida' → [feloɾida] (Farsi; Shademan, 2002), [pɪlorida] (Uyghur)
#NL		3	mladic (Serbian; last name) → [məladitʃ] (English)

b. Word-final

cluster		#	example
TS#		3	qudsˤ (Arabic; 'holy') → [qutus] (Uyghur)
CN#	TN#	6	hukm (Arabic; 'law') → [hokum] (Indonesian; Tadmor, 2009b)
	FN#	2	rahn (Arabic; 'hostage') → [refem] (Portuguese)
	SN#	7	ʒism (Arabic; 'body') → [ʒisim] (Indonesian; Tadmor, 2009b)
	NN#	1	gimn (Russian; 'hymn') → [gimun] (Kirghiz)
	LN#	5	'film' → filim (Uyghur)
Cl#	Tl#	12	korablʲ (Russian; 'ship') → [karabɨl] (Tatar)
	Fl#	3	sahl (Arabic; 'easy') → [sahal] (Nubian; Versteegh et al., 2006)
	Nl#	4	kremlʲ (Russian; 'the Kremlin') → [kɯremɯl] (Kirghiz)

First, any word-initial cluster beginning with a stop undergoes internal epenthesis between the stop and the following consonant. For example, Polish *gdansk* is borrowed into English as [gədænsk], even though *[əgdænsk] is phonotactically also possible. The number of languages that undergo internal epenthesis in stop-fricative clusters is particularly small because the clusters in question are cross-linguistically rare and in some cases other kinds of repair, such as deletion, are used to deal with illicit stop-fricative clusters. Also, internal epenthesis takes place in clusters that consist of a non-sibilant fricative and a liquid. Due to the rarity of source words that begin with a cluster consisting of a non-sibilant fricative and a following stop, fricative, or nasal, the examples are mostly limited to epenthesis in [f]-liquid clusters. In the present survey, all loanwords borrowed from words with initial [f]-liquid clusters have an epenthetic vowel after the initial [f]. The last word-initial cluster that undergoes internal epenthesis is the nasal-liquid cluster; the nasal-liquid cluster [ml] often appears in the Slavic languages and is borrowed with internal epenthesis.

Word-finally, stop-sibilant clusters show internal epenthesis, as illustrated in (9b). Also, internal epenthesis takes place in any cluster that ends in a nasal or a lateral liquid. No matter what the type of preceding consonant, the epenthetic vowel is placed before the final nasal or lateral.[1]

It should be stressed that most of the consonant clusters in (9) that undergo internal epenthesis involve an intensity rise within the cluster. Namely, epenthesis occurs at an intensity rise. Word-initially, all the clusters in (9a) involve an intensity rise in the cluster; in stop-initial clusters, cluster-initial stops involve a rise in intensity accompanied by a release burst, and in fricative-liquid and nasal-liquid clusters, intensity increases from the fricative and the nasal to the following liquid. In word-final position, first, stop-sibilant clusters involve an intensity rise when the stop is released, which is the case in the Arabic loanword in (9b). Nasal-final and lateral-final clusters involve an intensity rise too when the preceding consonant is either a stop or a fricative. Word-final nasal-nasal and liquid-nasal clusters are cases that show internal epenthesis but no intensity rise. It is possible that the first nasal in nasal-nasal clusters involves an audible release and thus involves an intensity rise, as seen in the example of word-initial nasal-nasal clusters previously. If this is true, there is an intensity rise between the two nasals, and epenthesis takes place at the rise. If this is not the case, however, no intensity rise is present in word-final nasal-nasal clusters as well as in lateral-nasal clusters, indicating that epenthesis takes place at a non-rise position. Notice, however, that there is no intensity rise in either position in those clusters; since the word-final nasal in Russian and English does not involve an audible release, unlike what is observed in French word-final nasals, no intensity rise is created at the end of the nasal. Hence, internal epenthesis in word-final nasal-nasal and lateral-nasal clusters does not at least form a counterexample to the generalization that epenthesis occurs at an intensity rise. In other words, if there is one intensity rise in a cluster, that is where epenthesis occurs.

3.1.2 External epenthesis

Let us now consider conosnant clusters that are always repaired by external epenthesis, which are presented in (10). Unlike the clusters that undergo internal epenthesis and involve an intensity rise within the cluster, most of the clusters that undergo external epenthesis involve an intensity rise outside the cluster.

[1] There is no case of epenthesis repair in word-final liquid-lateral clusters that is not determined phonotactically. The only case of internal epenthesis in liquid-lateral clusters is found in Hebrew, in which English 'earl' is adapted as [eʁel] (Cohen, 2009). This example, however, is not counted in the current study, because Cohen (2009) argues that it is likely that the loanword has an orthographic input when the English pronunciation is not acoustically similar to the Hebrew pronunciation, which is the case in [eʁel]. In the other cases of adaptation of English *rl* [rl] clusters, the [ɹ] deletes, often accompanied with compensatory lengthening of the preceding vowel (e.g., 'earl' → [əl] (Korean)).

(10) External epenthesis in consonant clusters
(T: stop, F: non-sibilant fricative, S: sibilant fricative, N: nasal, L: liquid, l: lateral, C: consonant, O: obstruent)

a. Word-initial[2]

cluster		#	example
#FT		1	vdova (Russian; 'widow') → [ogdo:bo] (Yakut; Pakendorf and Novgorodov, 2009)
#NO	#NT	3	mbira (Shona; 'mbira') → [əmbirə] (English)
	#NF	1	nʒamena (French; *N'Djamena*, capital of Chad) → [ɨncamena] (Korean)

b. Word-final

cluster		#	example
CT#	TT#	4	'Egypt' → [icipt^{h}i] (Korean)
	ST#	7	most (Russian; 'bridge') → [mosta] (Ewen; Choi, 2010)
	NT#	7	'land' → [lɛntɨ] (Korean)
	lT#	7	ʃolk (Russian; 'silk') → [ʃelki] (Tatar)
NS#		3	'France' → [fransi] (Tatar)

First, in word-initial position, it is observed that external epenthesis occurs in consonant clusters consisting of a non-sibilant fricative and a stop, when epenthesis is employed as a repair of illegal consonant clusters. The Yakut example in (10a) is the only example that adapts a word-initial fricative-stop cluster. As mentioned earlier, however, Russian [v] may behave as a glide phonetically and phonologically, and so may involve higher intensity compared to the other fricatives. If this is the case in the [vd] cluster in *vdova*, the cluster is not a fricative-stop cluster phonetically, and there will be a rise in intensity from silence to the initial [v]. Also, word-initial nasal-obstruent clusters undergo external epenthesis. Many nasal-obstruent clusters appear as homorganic, as observed in [mb] in *mbira* in (10aa), but even when the place of articulation is not exactly identical external epenthesis is preferred, as shown in *N'Djamena* → [ɨncamena] (Korean). Nasals also have high amplitude, and we assume an intensity rise from silence to the initial nasal.

Let us now consider cases of external epenthesis in word-final position. When the word-final consonant is a stop, the epenthetic vowel goes after the stop, as long as the illegal cluster is repaired by epenthesis, not by the other repairs, such as consonant deletion or feature change. In particular, there are a number of languages that undergo deletion of the final stop, e.g., 'pump' → [pam] (Marshallese; Shinohara, 2006), which are not included in the current database because this study concentrates on epenthesis patterns. The type of repair of stop-final clusters, epenthesis or deletion, might be based on the phonetic realization of the stop, i.e., whether it is released or not. If the final stop tends to be unreleased, it is likely to be repaired by deletion, and if the stop is released, it is likely to be repaired by epenthesis. The fact that word-initial stops rarely undergo deletion in loanword adaptation supports this hypothesis. Also, Kang (2003), for example, shows that vowel epenthesis after word-final stops in English loanwords in Korean

[2]The example of *lbovskij* in Kirghiz (Gouskova, 2001) mentioned earlier is not included here because according to Kara (2003), words beginning with [r] or [l] are usually pronounced with a vowel before the liquid, even when the liquid is a singleton onset (e.g., /lajyk/ → [ylajyk] 'sufficient'). Therefore, it is not obvious whether external epenthesis in [ylbovskij] is due to the markedness restriction on a word-initial cluster with a sonority fall or due to the ban on word-initial [l] in general.

is closely correlated with the stop release. In the examples in (10b), I assume that the stop in question is fully released and creates a rise in intensity. The examples in (10b) show external epenthesis after the word-final stop, even when internal epenthesis before the stop should also be allowed according to native language phonotactics. In addition, nasal-sibilant clusters have external epenthesis word-finally; I do not have relevant examples of epenthesis that occurs in lateral-fricative clusters. There is no intensity rise within the word-final nasal-sibilant cluster as well as outside the cluster.

It should be emphasized that the consonant clusters showing external epenthesis in (10) do not involve an intensity rise, unlike the clusters showing internal epenthesis in (9) but involve an intensity rise outside the cluster, i.e., from silence to the initial consonant in word-initial position and from the final stop closure to its release in word-final position. One possible counterexample is word-final stop-stop clusters; depending on the phonetic realization of the first stop, there could be an intensity rise (when the stop is released) or not (when the stop is not released). In English, the first stop in word-final stop-stop clusters tends to be unreleased, and so the 'Egypt' example above does not involve an intensity rise in the cluster. However, another example from Kazakh, for instance, involves Russian stop-stop clusters as a source, e.g., *abyekt* (Russian; 'object') → [äbyekti] (Krippes, 1994). As mentioned earlier, stops are clearly released in Russian even in pre-stop position, and so there must be an intensity rise between the two stops in the final cluster. However, there is also an intensity rise after the final stop, which is also audibly released in Russian (Kochetov, 2008). There are thus two rises in intensity, one between the two stops and the other after the final stop. This is reflected in loanword adaptation. In Korean, for example, English word-final TT clusters are adapted with single external epenthesis (e.g., 'compact' → [k^həmp^hɛkt^hɨ]), but Russian TT clusters also optionally allow internal epenthesis along with external epenthesis (e.g., *relikt* (Russian; proper name) → [lellikt^hɨ]~[lellik^hɨt^hɨ]). Thus, epenthesis still does not occur in a non-rise position, and so external epenthesis in word-final stop-stop clusters is not a counterexample.

A counterexample to the generalization is external epenthesis in word-final nasal-sibilant clusters because they do not display an intensity rise outside the cluster. This will be discussed in detail in the following section.

3.1.3 Variation between internal and external epenthesis

This subsection presents cases that show epenthesis patterns that vary between internal and external epenthesis. The variation may depend on the source languages, borrowing languages and dialects, and lexical items, or free variation between the two epenthesis forms.

3.1.3.1 Word-initial sibilant-initial clusters

Sibilant-initial clusters in word-initial position do not show an exceptionless epenthesis pattern cross-linguistically. As summarized in (11), the predominant pattern is external epenthesis in sibilant-stop clusters, while there exist two languages that employ internal epenthesis only and three languages that vary between internal epenthesis and external epenthesis.

(11) Epenthesis in sibilant-stop clusters

epenthesis	#	example
internal	3	stakan (Russian; 'glass cup') → [sitakkaanaq] (Aleut Eskimo; Koo, 1980)
external	23	stakan (Russian; 'glass cup') → [ɯstakan] (Kirghiz; Gouskova, 2001)
variation	6	'school' → [sikuulaq] vs. ʃkola (Russian; 'school') → [ɨskuuluq] (Yupik Eskimo; Koo, 1982)

For the cases of variation, Yupik Eskimo shows a variable epenthesis pattern depending on the source language.[3] Similarly, Indonesian borrowed a sibilant-stop cluster from Sanskrit with external epenthesis (e.g., *stri* → [istri] 'woman') and the cluster from English with internal epenthesis (e.g., *stamp* → [sətæmp]) (Yuliati, 2014). On the other hand, Farsi shows dialectal variation between internal and external epenthesis; Russian *ʃkaf* 'cupboard' is normally borrowed with external epenthesis ([ʔeʃkap]) but in some dialects it is borrowed with internal epenthesis ([ʃekaf]) (Kambuziya and Hashemi, 2011). Hindi, Uyghur and Tatar show lexical variation; for instance, in Hindi, English 'school' is borrowed with external epenthesis [ɪskul] (Singh, 1985), and 'score' is borrowed with internal epenthesis [sɪkor] (M. Ohala, p.c.). Also, according to my Hindi consultant, even the same word may show different epenthesis patterns depending on the dialect (e.g., 'spelling' → [səpeliŋ]~[ispeliŋ]).[4] Korean (Fleischhacker, 2001) and Japanese (Zuraw, 2007) are also reported to show internal epenthesis in sibilant-stop clusters but are excluded from the current database because neither language allows the intervocalic sibilant-stop clusters in the native phonology that would have appeared in the case of external epenthesis. There is no rise in intensity between the initial sibilant and the next stop, while there is an intensity rise from silence to the sibilant. So the predominant pattern of external epenthesis in word-initial sibilant-stop clusters is consistent with the generalization that epenthesis occurs at the position of intensity rise, although there exist a few cases of internal epenthesis and variation.

Next, some clusters that are composed of a sibilant fricative and a non-sibilant fricative undergo internal epenthesis while the others undergo external epenthesis. The number of languages showing each epenthesis pattern and examples are given in (12).

(12) Epenthesis in sibilant-fricative clusters

epenthesis	#	example
internal	6	sv^ječka (Russian; 'candle') → [siveska] (Evenki; Choi, 2010)
external	2	'sphere' → [ɪsfiʌr] (Hindi; Singh, 1985)
variation	1	sv^jinka (Russian; 'pig') → [ɕitiinkaaq] vs. sv^jit^jir (Russian; 'sweater') → [iswataq]/[isvataq] (Yupik Eskimo; Koo, 1982)

The cluster that is repaired by internal epenthesis is the [sv] cluster of Russian, not only in the Evenki example in (12) but also in the other languages. See Appendix 3 for the examples. It can be again said that epenthesis occurs at a rise in intensity. In [sv], the preceding consonant [s]

[3]Koo (1982) states that [sikuulaq] is from English and [ɨskuuluq] is from Russian, but no evidence for this is provided.

[4]According to the consultant, no dialectal variation in epenthesis sites is observed in stop-initial and non-sibilant-initial clusters, which are always repaired by internal epenthesis.

is a sibilant but is voiceless, while the following [v] is a non-sibilant but is voiced. This indicates that in the low frequency region, the level of intensity must be greater in the [v] than in the [s], resulting in a rise in intensity between the two. In contrast, clusters repaired by external epenthesis do not involve an intensity rise within the cluster. For the [sf] cluster in the Hindi example in (12), the sibilant [s] involves a greater intensity than the non-sibilant [f], resulting in an intensity fall. The other example of external epenthesis is observed in Kirghiz; Russian *zveno* is adapted in Kirghiz as [uzvana] 'chain link' (Gouskova, 2001). Likewise, the sibilant [z] involves a greater intensity than the following [v], and there is no intensity rise between them. There is an intensity rise from silence to the initial sibilant in [sf] and [zv], so external epenthesis takes place where intensity is rising.

Sibilant-nasal clusters also show variable epenthesis patterns, which are summarized in (13). Like sibilant-stop clusters, three patterns of epenthesis, i.e., internal epenthesis only, external epenthesis only, and variation between the two, occur in sibilant-nasal clusters, although the total number of languages is much smaller. The type of variation between internal epenthesis and external epenthesis may be a free variation as in Tatar in (13), and may also depend on the particular lexical items, as in Kazakh, which will be shown in detail later.

(13) Epenthesis in sibilant-nasal clusters

epenthesis	#	example
internal	1	'Smith' → [sɨmis] (Uyghur)
external	4	Smolny (Russian; proper name) → [ismolnin] (Uzbek; Akiner, 1997)
variation	3	smat (Russian; proper name) → [səmat]~[əsmat] (Tatar)

Similarly, the epenthesis pattern is not consistent for sibilant-lateral clusters, across languages or even in the same language. As shown in (14), there are six languages observed in the current survey in which sibilant-liquid clusters are repaired solely by internal epenthesis, two languages in which they are repaired solely by external epenthesis, and two languages in which both internal and external epenthesis are used for cluster repair.

(14) Epenthesis in sibilant-lateral clusters

epenthesis	#	example
internal	5	'slate' (En) → [ʃelet] (Mahato 1974, reported in Broselow, 1992)
external	3	ʃlʲapa (Russian) → [islaapaq] 'hat' (Yupik Eskimo; Koo, 1982)
variation	3	'Slovakia' → [solovaki] vs. 'Sleipnir' → [eslɛpnir] (Wolof; Fleischhacker, 2001)

Although there are only a few loanwords including the word-initial [sr] cluster, only internal epenthesis is observed in the current database. Here the [r] can be either a rhotic or a flap.

(15) Epenthesis in sibilant-[r] clusters

epenthesis	#	example
Internal	4	'Sri Lanka' → [seri laŋka] (Farsi; Fleischhacker, 2001)

Regarding epenthesis patterns in sibilant-initial clusters in word-initial position, Fleischhacker (2001) suggests an implicational universal, illustrated in (16). Specifically, in the hierarchy in

(16), if a cluster undergoes internal epenthesis, all the other clusters on its right undergo internal epenthesis too. For example, if internal epenthesis is allowed in [sm] in a language, it is also allowed in [sn], [sl], [sr], etc.

(16) Epenthesis patterns in sibilant-initial clusters (Fleischhacker, 2001)
(S: voiceless sibilant fricative, T: stop, Y: glide, R: sonorant)

ST Sm Sn Sl Sr (SY[5] TR)

← *external* *internal* →

In the present survey, there is no language that counter-exemplifies this implication universal, although it is also true that there are only a few languages that show the implicational universal within the language: Assamese, Kazakh, Wolof, and Farsi. Also, the ratio of internal epenthesis increases and the ratio of external epenthesis decreases as one moves to the right in (16), as shown in Figure 3-1.

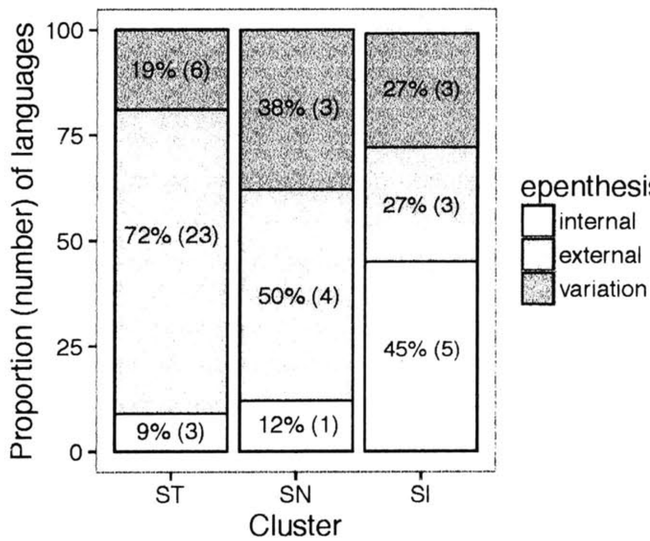


Figure 3-1: Ratios of epenthesis patterns in word-initial sibilant-initial clusters

Thus far we have seen that epenthesis takes place cluster-internally when the cluster features an intensity rise within the cluster, and cluster-externally when the cluster features no intensity rise within the cluster but features a rise outside the cluster. This cannot account for the implicational relationships in (16), as sibilant-sonorant clusters all involve an intensity rise, and thus with respect to intensity rise, there is no difference between Sm, Sn, Sl, and Sr clusters.

Here I suggest that the difference in epenthesis sites in sibilant-initial clusters is solid only between the sibilant-initial clusters without an intensity rise within the cluster, i.e., sibilant-stop and sibilant-non-sibilant, and the sibilant-initial clusters with an intensity rise within and outside the cluster, i.e., all of the other sibilant-initial clusters. This is primarily because the detailed differences between sibilant-initial clusters described in (16) do not have strong evidence. First, some of the languages that are reported to show the implicational relationship in (16) may not be an appropriate example. For example, Fleischhacker's (2001) only example language that shows the epenthesis asymmetry between Sm and Sn clusters is Hindi, described by Bharati (1994). According to the current survey, however, Hindi is one of the languages that shows variation

[5]Since C-glide clusters are not considered in this study, SY is excluded in the following discussion.

between internal and external epenthesis, as mentioned earlier, e.g., 'school' → [ɪskul] vs. 'score' → [sɪkor]. So the implicational universal in (16) does not hold if internal epenthesis is not allowed in Sm clusters. It is possible that the data reported in Bharati (1994) is partial and misses an example of internal epenthesis in Sm clusters. Furthermore, I have not observed any case that differentiates [m] and [n] in terms of epenthesis in the current database. So I tentatively conclude that there is no implicational relationship between Sm and Sn clusters.

Also, Fleischhacker (2001) reports that Kazakh is located between Sn and Sl in the hierarchy in (16), in which external epenthesis is allowed in sibilant-nasal clusters as well as in sibilant-stop clusters, and internal epenthesis only takes place in sibilant-liquid clusters, as shown in (17).

(17) Epenthesis in sibilant-sonorant clusters in Kazakh (Sulejmenova 1965, reported in Fleischhacker, 2001)

a. Sibilant-nasal clusters: variation between internal and external epenthesis

Russian	*Kazakh*	*gloss*
smorodina	[sɨmorodina]	'currant'
smena	[ɨsmen]	'change'
smat	[sɨmat]~[ɨsmat]	proper name

b. Sibilant-liquid clusters: internal epenthesis

Russian	*Kazakh*	*gloss*
slesar	[silesir]	'metalworker'
ʃleja	[ʃilija]	'breech'

It should, however, be mentioned that in Kazakh, sibilant-nasal and sibilant-lateral clusters are not allowed in the native phonology, and post-sibilant nasals and laterals undergo desonorization, as exemplied in (18).

(18) Post-sibilant desonorization in Kazakh (Davis, 1998)

	/kes-ma/	[kespe]	'noodles' + INT
	/koŋɯz-ma/	[koŋɯzba]	'bug' +INT
	/koŋɯz-lar/	[koŋɯzdar]	'bug' +PLURAL
cf.	/koŋɯz-ga/	[koŋɯzga]	'bug' +DIRECT

Previous studies account for the desonorization in Kazakh and the other Turkic languages based on Syllable Contact, a constraint prohibiting a sonority rise over a syllable boundary (Davis, 1998; Baertsch, 1998; Baertsch and Davis, 2001). In the case of Kazakh, consonant clusters across a syllable boundary are banned when the first consonant is less sonorous than the second. Therefore, external epenthesis in word-initial sibilant-sonorant clusters, as in [ɨsmen] and [ɨsmat] in (17), yields a critical violation of a phonotactic restriction in Kazakh. The same applies to the sibilant-lateral clusters. In the current survey, internal epenthesis in Kazakh sibilant-lateral is not counted as an example because external epenthesis, e.g., *[islesar] for *slesar*, is not allowed in the native phonology of Kazakh, and internal epenthesis may have been favored by the markedness restriction in the process of loanword adaptation. I suggest that the implicational relationship between sibilant-nasal and sibilant-lateral clusters in Kazakh vowel epenthesis may be driven by markedness. The ban on sibilant-nasal clusters may be weaker than the ban on sibilant-lateral clusters in Kazakh, because the degree of sonority rise in sibilant-nasal is smaller than that in sibilant-lateral clusters, which may indicate that the dispreference for the sibilant-nasal clusters is weak enough to allow the sibilant-nasal sequences that are illicit in native phonology in some loanwords. In contrast, it is possible that the avoidance of the sibilant-lateral clusters in Kazakh is strong enough to enforce internal epenthesis in order to avoid intervocalic sibilant-lateral clusters resulting from external epenthesis. In this case, the implicational relation between sibilant-nasal and sibilant-lateral clusters may not be real. It should also be mentioned, however, that Assamese shows the asymmetry in epenthesis sites between sibilant-nasal and sibilant-lateral clusters, although it is not reported in Fleischhacker (2001).

Another reason not to assume the detailed distinctions in the implicational hierarchy in (16) is that unlike the other type of clusters, sibilant-initial clusters show great variation in epenthesis sites, and so it is possible that the present generalization just describes limited data. For example, it is known that the word-initial [sl] cluster is repaired with internal epenthesis in Egyptian Arabic (e.g., 'slide' → [silajd]; Broselow, 1992). This may not be the only epenthesis repair because my Egyptian Arabic consultant indicates that 'slide' undergoes external epenthesis, resulting in [ʔeslajd]. The aforementioned example of Hindi is also a similar case. Whereas further investigation of data may also find counterexamples to the current generalizations about the other types of consonant clusters, most of the data in the survey that differs from the previous description involve sibilant-initial clusters in word-initial clusters. In conclusion, it is obvious that sibilant-sonorant clusters allow internal epenthesis more likely than sibilant-stop clusters, but it is not yet clear whether the detailed differences among sibilant-sonorant clusters are solid, which simplifies the implicational universal in (16) to the one in (19). The implicational relation with stop-sonorant cluster (TR) is vacuously established as there is no language showing external epenthesis before word-initial stop-sonorant clusters.

(19) Epenthesis site asymmetry in sibilant-initial clusters

ST SR (TR)

← *external* *internal* →

To sum up, external epenthesis is more readily in sibilant-initial clusters without an intensity rise within the clusters, i.e., the sibilant-stop and sibilant-non-sibilant clusters agreeing in voicing, whereas internal epenthesis is more likely in sibilant-initial clusters with an intensity rise inside, i.e., the voiceless sibilant-voiced non-sibilant clusters and sibilant-sonorant clusters. However, there are cases of internal epenthesis in sibilant-stop clusters and cases of external epenthesis in sibilant-sonorant clusters, which require further investigation.

3.1.3.2 Word-initial nasal-nasal clusters

Another type of consonant cluster that shows variable epenthesis sites is word-initial nasal-nasal clusters. As shown in (10a), when the following consonant is an obstruent, external epenthesis occurs before the initial nasal. When the following consonant is also a nasal, however, the epenthesis position differs depending on the language; epenthesis may occur before the initial nasal or between the two nasals, as shown in (20).

(20) Epenthesis in nasal-nasal clusters

epenthesis	#	example
internal	1	mnemosine (Greek; 'mnemosyne') → [minemosine] (Korean)
external	1	mnemonicheskij (Russian; 'mnemonic') → [ymnemonich-eskij] (Kirghiz; Gouskova, 2001)

Recall that word-initial nasal-nasal clusters may involve an audible release or not, indicating that they may involve an intensity rise or not within the cluster, depending on the phonetic realization. So epenthesis may happen before the initial nasal, as in Kirghiz, or between the two nasals, as in Korean.[6]

3.1.3.3 Word-final clusters involving a trill

Clusters involving the trill [r] show different epenthesis patterns from the clusters involving the lateral, although both [r] and [l] are categorized as liquids.. I was not able to find loanwords borrowing trill-initial clusters in word-initial position. In word-final position, when the cluster-initial consonant is the trill [r], the epenthesis position may show variation. That is, although it has been shown above that stop-final and fricative-final clusters undergo external epenthesis after the final stop or fricative as shown in (10b), internal epenthesis may occur when the preceding consonant is the trill, as shown in (21).[7]

(21) Epenthesis in word-final trill-initial clusters

cluster	epenthesis	#	example
rT#	internal	1	darb (Arabic, 'path') → [darub] (Nubian; Versteegh et al., 2006)
	external	2	park (Russian; 'park') → [parka] (Tatar)
rF#	internal	2	ʃarf (Russian; 'scarf') → [ʃarɨf] (Tatar)
rS#	internal	4	dars (Arabic; 'class') → [deris] (Uyghur)
	variation	1	borʃ (Russian; 'soup') → [borəʃ]~[borʃi] (Tatar)

Again, it is observed that epenthesis occurs where there is an intensity rise. As stated in the previous chapter, trills involve more than one intensity rise. Therefore, in word-final rT clusters, there exist two rises in intensity, one between the trill and the stop and the other after the stop, which is audibly released. So the positions of internal epenthesis and external epenthesis both involve an intensity rise in word-final rT clusters. On the other hand, when the final consonant is a fricative, internal epenthesis is predominant. In word-final rF clusters, where the final consonant is a non-sibilant fricative, both example languages show internal epenthesis. When the final fricative is a sibilant [ʃ], in Tatar, internal epenthesis and external epenthesis are in free variation. There is an intensity rise between the trill and the following fricative, so the cases of internal epenthesis in word-final trill-fricative clusters show the same tendency to have an epenthetic vowel at an intensity rise as the patterns in the other clusters we have seen thus far, with the exception of optional external epenthesis in Tatar.

Also, the epenthetic vowel may be positioned after the final trill, as shown in (22). The dominant pattern is internal epenthesis, but there is also a case of variation in Uyghur. Note that when the cluster-final consonant is a lateral, it is always internal epenthesis that occurs without exception, as shown in (9b).

[6] In Korean, a word-initial nasal geminate is adapted with external epenthesis (e.g., *nnaʒi* (Bantu) → [ɨnnaci] 'Nnaji' (name of a Nigerian Actress)), unlike internal epenthesis in heterorganic nasal-nasal clusters in (20). The initial nasal in a geminate is not likely to involve an audible release, and this asymmetry in epenthesis sites may be attributed to the different phonetic realizations of the nasal-nasal clusters.

[7] Clusters involving the trill are found in Russian and Arabic, but there are no clusters with the English approximant [ɹ]. In the database, there is no case in which epenthesis occurs before or after the English [ɹ], but the [ɹ] often deletes accompanied by compensatory lengthening of the preceding vowel.

(22) Epenthesis in word-final trill-final clusters

cluster	epenthesis	#	example
Tr#	internal	5	qabr (Arabic; 'grave') → [kabir] (Turkish; Kornfilt, p.c.)
	variation	1	qabr (Arabic; 'grave') → [kabre] vs. fikr (Arabic; 'opinion') → [fikir] (Uyghur)
Fr#	internal	3	siħr (Arabic; 'magic') → [sihir] (Uyghur)
Sr#	internal	2	misˤr (Arabic; 'Egypt') → [misir] (Uyghur)
Nr#	internal	1	ʕamr (Arabic; 'age') → [ømør] (Uyghur)

External epenthesis after the final trill may be due to the acoustic property of the trill. That is, an intensity rise is created by the last vibration of a trill, and the external epenthesis can occur at the rise, and internal epenthesis can also occur at the other rise in intensity between the preceding consonant and the trill.

In this section, it is shown that in the majority of cases, if there is an intensity rise within the cluster, internal epenthesis occurs at the rise, whereas if there is an intensity rise outside the cluster, external epenthesis occurs. Clusters showing variation between internal and external epenthesis either involve two rises in intensity, as in clusters including a trill, or may involve an intensity rise inside the cluster or not depending on the realization of the cluster, as in nasal-nasal clusters. The cases that do not fit this generalization, such as the complex epenthesis pattern of sibilant-initial clusters, will be discussed in more detail in the next section.

3.2 Hypothesis

The question is why vowel epenthesis occurs at an intensity rise. The working hypothesis of this dissertation is that the epenthesis site in consonant cluster adaptation is determined by the relative perceptual similarity between the consonant cluster and the corresponding sequences with epenthesis, and epenthesis occurs at an intensity rise because it makes for a perceptually smaller change to the consonant cluster, compared with epenthesis at a non-rise. This motivation is based on the P-map hypothesis (Steriade, 2008), which hypothesizes that perceptually smaller changes are preferred. In Chapter 2, we have seen that consonant clusters featuring an intensity rise inside are less accurately discriminated from their matching epenthesis forms than clusters featuring no intensity rise, which indicates that clusters with an intensity rise are perceptually more similar to their epenthesis forms than clusters without an intensity rise. Therefore, it is predicted that if there is an intensity rise in the input consonant clusters, an epenthetic vowel is inserted at the rise, because the resulting epenthesis form is perceptually similar to the cluster.

The current hypothesis can explain some asymmetries observed in the previous section. First, for word-final liquid-final clusters, the lateral [l] always has an epenthetic vowel before it whereas the trill [r] may allow epenthesis after it (external epenthesis) as well as before it (internal epenthesis). Lateral-final clusters involve an intensity rise whatever the consonant preceding the lateral is, because the lateral has greater intensity than the other consonants. So epenthesis occurs at the rise, before the final lateral [l]. On the other hand, the trill [r] yields two intensity rises, one at the beginning of the [r] and the other at the end of the [r], which allows both internal epenthesis and external epenthesis in trill-final clusters. Also, for word-final liquid-stop clusters, lateral-stop clusters always show external epenthesis, while trill-stop clusters show variation between internal and external epenthesis. This is because, in lateral-stop clusters, intensity is falling from the lateral to the following consonant, which means there is no intensity rise between the two, and intensity is rising when the final stop is released, resulting in external epenthesis at the stop release burst involving an intensity rise. In word-final trill-stop clusters, both internal epenthesis and external epenthesis may occur depending on the language, as shown in (21). Assuming the final stop is audibly released, which creates another rise in intensity after the stop closure, there are two intensity rises in the clusters, one between the two consonants and the other after the final stop, and either epenthesis pattern may occur.

Recall that external epenthesis is possible before word-initial sibilant-sonorant clusters, as shown in (13) and (14). This variable epenthesis pattern is not observed in clusters consisting of a non-sibilant fricative and a sonorant; as exemplified in (9a), only internal epenthesis between the [f] and the following liquid is observed cross-linguistically. This asymmetry can be explained by the difference in the intensity of the fricatives. As mentioned in the previous chapter, sibilants involve relatively large acoustic intensity, and thus a rise of intensity exists from the silence to the word-initial sibilant, as well as from the sibilant to the following sonorant. This results in two rises in intensity, and both internal and external epenthesis can take place in sibilant-sonorant clusters. On the other hand, the non-sibilant [f] has low intensity, comparable to that of stops in word-initial position, and so the intensity change from the silence to the initial [f] is not considered here as a rise. This is illustrated in (23). Therefore, in [f]-initial clusters, there is only one intensity rise between the [f] and the following sonorant, which results in internal epenthesis only.

(23) Intensity rise and epenthesis in word-initial fricative-initial clusters

	#		S		R ... vs.	#		F		R ...
(intensity)		/		/			—		/	
(epenthesis)		✓		✓			✗		✓	

The current hypothesis can also explain the positional asymmetry in sibilant-sonorant clusters; word-initially, both internal and external epenthesis may take place, as shown in (13) and (14), whereas only internal epenthesis appears word-finally, as shown in (9b). This can also be attributed to whether there is one intensity rise or two, as illustrated in (24). Unlike word-initial sibilant-sonorant clusters that involve two rises in intensity, word-final sibilant clusters only involve one intensity rise, between the two consonants, unless the sonorant is a trill, because intensity after the sonorant is decreasing to the silence.

(24) Intensity rise and epenthesis in word-initial and word-final sibilant-sonorant clusters

	#		S		R ... vs.	... S		R		#
(intensity)		/		/			/		\	
(epenthesis)		✓		✓			✓		✗	

Another variation in epenthesis sites that should be explained is found in word-initial nasal-nasal clusters. Although the examples are limited to only two, depending on the language, either internal epenthesis or external epenthesis can take place in nasal-nasal clusters, as shown in (20). I attribute the variable epenthesis sites to the presumed phonetic realization of the cluster-initial nasal. If the nasal involves an audible release, which creates a rise in intensity, internal epenthesis occurs, and if it does not, external epenthesis occurs. In the latter case, in which there is no rise in intensity, the intensity rise between the silence and the initial nasal is the only rise in intensity, and external epenthesis occurs at the rise, as illustrated in (25a). In the former case, however, there exist two rises in intensity, one between the silence and the initial nasal and the other between the two nasals, as illustrated in (25b), and in principle, both internal epenthesis and external epenthesis are possible. I assume that internal epenthesis is preferred in this case also has perceptual basis. To be specific, the cluster-initial consonant in word-initial position, the first nasal here, is the one that listeners here first in the word and thus are more sensitive to its perception compared to the perception of segments in non-initial positions. Even though there are two positions that involve an intensity rise, internal epenthesis will cause a perceptually smaller change to the cluster than external epenthesis because it is still the first nasal that listeners hear first while the epenthetic vowel is located where intensity is rising. This hypothesis is supported by the epenthesis pattern of the other word-initial clusters involving two intensity rises. Other nasal-sonorant clusters, [ml] and [mr] also involve two rises in intensity, but only internal epenthesis takes place in these clusters in loanword adaptation. Sibilant-sonorant clusters show variation between internal and external epenthesis, and the sensitivity to the initial sibilant does not play a role as much as in nasal-initial clusters. Still, however, we see internal epenthesis more often than external epenthesis in sibilant-sonorant clusters, and so we conclude that the epenthesis pattern in these clusters does not reject the hypothesis. Notice that maintaining the saliency of the word-initial consonant does not hold when there is no following rise in intensity. Therefore, clusters such as the nasal-obstruent [mb], which do not have a rise in intensity between the two consonants, are repaired by external epenthesis.

(25) Intensity rise and epenthesis in word-initial nasal-nasal clusters

	a. #		N⌝		N	vs.	b. #		N⌜		N
(intensity)		/		—				/		/	
(epenthesis)		✓		✗				(✗)		✓	

Intensity rise has an affinity with Fleischhacker's (2005) "perceptual break", which is assumed to be composed of the onset of vowel-like formant structure. She also assumes that the "perceptual break" is reinforced when the formant involves high intensity or a stop closure precedes the formant. According to this, stop-liquid clusters form the strongest "perceptual break", followed by stop-nasal, and that of sibilant-stop clusters is the weakest because there is no formant structure preceding stop closure. Fleischhacker argues that this is the reason why stop-liquid clusters are vulnerable to epenthesis, while sibilant-stop clusters are not. Although intensity rise also refers to acoustic properties that affect speech perception, it is a broader concept that makes more distinctions between consonant clusters than "perceptual break" does, explaining a wider range of data than in Fleischhacker (2005). For example, we cannot explain why initial stop-stop clusters always undergo internal epenthesis but sibilant-sonorant clusters show variation between internal and external epenthesis solely based on "perceptual break", because "perceptual break" may be stronger in sibilant-sonorant clusters that involve vowel-like formant structures than in stop-stop clusters in which the release burst of the first stop may not involve vowel-like formants. By adopting the two acoustic properties, intensity rise and C_1 voicing, we can explain the difference in epenthesis sites between the two clusters; stop-stop clusters involve an intensity rise between the two stops and undergo internal epenthesis, while sibilant-sonorant clusters involve two rises in intensity and may show either internal or external epenthesis.

So far we have seen that epenthesis sites correspond to where there is an intensity rise. If there is one intensity rise, one pattern of epenthesis, internal or external epenthesis, occurs, and if there are two intensity rises, internal epenthesis and external epenthesis are in variation. The remaining question is how we can explain the epenthesis patterns in consonant clusters involving no intensity rise. The clusters that involve no intensity rise are nasal-nasal, lateral-nasal, and nasal-sibilant clusters in word-final position. The two nasal-final clusters undergo internal epenthesis, as shown in (9b), while the nasal-sibilant clusters undergo external epenthesis, as shown in (10b). I hypothesize that what determines the epenthesis site in these clusters is whether the nasal in the cluster is audibly released or not. In word-final nasal-nasal and lateral-nasal clusters, the final nasal does not involve an audible release in the source languages, Russian and English. If external epenthesis occurs, the final nasal that was originally unreleased becomes released by the following epenthetic vowel. This change in the release of the nasal would bring about a perceptually noticeable change and thus is prevented. On the other hand, in nasal-fricative clusters, the final fricative is first not the subject of release as it is not an occlusive consonant. The cluster-initial nasal is unreleased before an obstruent, and if the cluster is repaired by internal epenthesis, the nasal becomes released. To avoid the release of the nasal that is not present in the cluster, external epenthesis would be preferred.

Summing up, the current hypothesis accounts for the cluster-dependent asymmetries in epenthesis sites based on perceptual similarity between clusters and epenthesis forms in terms of intensity rises. Attested epenthesis patterns depending on the number of intensity rises are summarized in Table 3.1. "↗" indicates the presence of intensity rise, and we see that the positions of intensity rise and those of the epenthetic vowels are matched when there exists an intensity rise. If there is one rise in intensity in the cluster, epenthesis occurs at the rise because an epenthetic vowel positioned at an intensity rise incurs a perceptually less salient change to the cluster than a vowel inserted at a non-rise position. Specifically, if there is an intensity rise within the cluster, internal epenthesis takes place, and if there is an intensity rise outside the cluster, external epenthesis takes place. If there are two intensity rises in a cluster, epenthesis sites may show variation between internal and external epenthesis. If there is no intensity rise in a cluster, the epenthesis site is determined so that it does not yield a perceptually salient change in the releasedness of the consonant in the cluster.

Table 3.1: Epenthesis sites depending on the number of intensity rises

# of intensity rise	*epenthesis site*
0 (CC#)	internal or external epenthesis
1 (within CC: #C↗C / C↗C#)	internal epenthesis (#CəC / CəC#)
1 (outside CC: #↗CC / CC↗#)	external epenthesis (#əCC / CCə#)
2 (#↗C↗C / C↗C ↗#)	variation between internal and external epenthesis (#CəC~#əCC / CəC#~CCə#)

Tables 3.2 and 3.3 below summarize the epenthesis sites in the attested consonant clusters presented in this section, along with the presence or absence of intensity rises (T: stop, F: non-sibilant fricative, S: sibilant fricative, N: nasal, L: liquid, O: obstruent, C: consonant).

Table 3.2: Summary of the epenthesis sites in word-initial cluster adaptation

epenthesis	clusters	intensity rise	
		#_C_1	C_1_C_2
internal	#TC, #FL	✗	✓
	#NL	✓	✓
external	#FT, #NO	✓	✗
variation	#SC, #NN	✓	✓/✗

Table 3.3: Summary of the epenthesis sites in word-final cluster adaptation

epenthesis	clusters	intensity rise	
		C_1_C_2	C_2_#
internal	TS#, CN#, Cl#	✓	✗
	lN#, NN#	✗	✗
external	CT#	✗	✓
	NS#	✗	✗
variation	rO#, Cr#	✓	✓

In Chapter 2, we have seen that consonant clusters involving an intensity rise within the cluster are more confusable with the corresponding internal epenthesis forms than clusters involving no intensity rise within the cluster. We have not tested, however, whether clusters without an intensity rise inside, most of which involve an intensity rise outside the cluster, are perceptually similar to the corresponding external epenthesis forms, which will serve as perceptual basis for external epenthesis in loanword adaptation. To test the hypothesis, additional perception experiments were conducted, which will be presented in the next section.

3.3 Experimental evidence

We have learned that a consonant cluster CC that involves an intensity rise and a voiced C_1 is perceptually more similar to its matching epenthesis form CəC than a cluster that does not. This section investigates the perceptual similarity between consonant clusters and the other matching epenthesis form with external epenthesis, either #əCC or CCə#, as well as the one with internal epenthesis, #CəC or CəC#. The working hypothesis is that a consonant cluster featuring an intensity rise within it, which undergoes internal epenthesis in loan cluster adaptation, is perceptually more similar to its matching form with internal epenthesis, and a cluster featuring no intensity rise within it, which tends to undergo external epenthesis, is perceptually more similar to the form with external epenthesis. C_1 voicing is not tested because it is intensity rise that is far more important in determining the site of epenthesis in loanword adaptation, as shown earlier. It is also hypothesized that consonant clusters that vary between internal and external epenthesis are perceptually similar to the internal and the external epenthesis forms to a similar degree. These hypotheses are tested with an ABX similarity judgment task in Section 3.3.1 and with a transcription task in Section 3.3.2.

3.3.1 Experiment 4

3.3.1.1 Methods

3.3.1.1.1 Participants

Participants of Experiment 4 were speakers of three different languages: (i) 11 speakers of English (allowing several consonant clusters word-initially and word-finally), (ii) 9 speakers of Korean (allowing several clusters intervocalically but not word-initially or finally), and (iii) 6 speakers of Japanese (allowing only a few homorganic clusters intervocalically but not word-initially or finally). The participants were all residents of the Boston Area, mostly students at MIT, and did not have a hearing problem. After the experiment, the participants were paid $10 for their participation.

3.3.1.1.2 Stimuli

Stimuli used in Experiment 4 were composed of nonce words containing a consonant cluster either in word-initial or in word-final position, and the two matching epenthesis forms, one with internal epenthesis and the other with external epenthesis. Specifically, for word-initial position, nonce words of the shape [C_1C_2áte] were compared with the internal epenthesis form [C_1əC_2áte] and the external epenthesis form [əC_1C_2áte], and for word-final position, nonce words [netáC_1C_2] were compared with [netáC_1əC_2] and [netáC_1C_2ə]. As listed in Table 3.4, 31 word-initial clusters and 31 word-final clusters formed from combinations of a stop (k), two fricatives (s, f), a nasal (m), and two liquids (l, r) were used in the stimuli. Stops in the stimuli were all clearly released. On the other hand, nasals were not audibly released except for the nasal [m] preceding the liquid [l]. The stimulus clusters in Table 3.4 are classified by the number of intensity rises in Table 3.5.

Table 3.4: Consonant clusters used in the stimuli

edge C	*word-initial clusters*	*word-final clusters*
stop	k⌜p, k⌜s, k⌜f, k⌜m, k⌜l, k⌜r	p⌜k⌜, sk⌜, fk⌜, mk⌜, lk⌜, rk⌜
fricative [s]	sk⌜, sf, sm, sl, sr	k⌜s, fs, m⌝s, ls, rs
fricative [f]	fk⌜, fs, fm, fl, fr	k⌜f, sf, m⌝f, lf, rf
nasal	m⌝k⌜, m⌝s, m⌝f, m⌝n, m⌜l, m⌜r	k⌜m⌝, sm⌝, fm⌝, m⌝n⌝, lm⌝, rm⌝
liquid [l]	lk, ls, lf, lm	k⌜l, sl, fl, m⌜l, rl
liquid [r]	rk, rs, rf, rm, rl	k⌜r, sr, fr, m⌜r

Table 3.5: Stimulus clusters by the number of intensity rises

# of intensity rise	clusters
0	#fk⌜, m⌝s#, ls#, sf#, m⌝f#, lf#, m⌝n⌝#, lm⌝#
1 (within CC)	#k⌜p, #k⌜s, #k⌜f, #k⌜m, #k⌜l, #k⌜r, #fs, #fm, #fl, #fr, k⌜s#, fs#, rs#, k⌜f#, rf#, k⌜m⌝#, sm⌝#, fm⌝#, rm⌝#, k⌜l#, sl#, fl#, m⌜l#, rl#
1 (outside CC)	#sk⌜, #sf, #m⌝k⌜, #m⌝s, #m⌝f, #m⌝n #lk, #ls, #lf, #lm, sk⌜#, fk⌜#, mk⌜#, lk⌜#
2	#sm, #sl, #sr, #m⌜l, #m⌜r, #rk, #rs, #rf, #rm, #rl, p⌜k⌜#, rk⌜#, k⌜r#, sr#, fr#, m⌜r#

The stimuli were recorded by a male Russian speaker, who also recorded the stimuli used in Experiment 1. The duration of epenthetic vowels ranged from 40 to 60 ms. [a] in the stimuli was intended to be a stressed vowel so that the epenthetic vowels would not be stressed, and the duration of vowel [a] was manipulated after recording by using Praat to have roughly the same duration in all the stimuli (approximately 150 ms). Also, part of the templates other than the consonant clusters and epenthetic vowels were all cross-spliced and thus were identical in all forms.

3.3.1.1.3 Procedure

The participants completed an AXB similarity judgment task, in which X refers to a form containing a consonant cluster, A refers to a matching form with internal epenthesis, and B refers to a matching form with external epenthesis. In particular, the participants heard the three nonce words in a row, and were asked whether the first one or the third one sounded more similar to the second one with a cluster. The cluster form, which was the reference, was flanked by the two epenthesized forms so that proximity to one of the epenthesis forms did not affect the similarity judgment, which might have happened in ABX or XAB orders. AXB and BXA orders were also counterbalanced. In each trial 500 ms of silence preceded the stimuli, which were played with inter-stimulus intervals of 500 ms. The participants of each language group were told that the words they were listening to were not words in their language. The participants were instructed to press A on the keyboard if they thought the second word sounded more similar to the first and to press L if they thought the second sounded more similar to the third. Also, they were asked to respond as quickly as possible after they heard the last word and to choose either first or third even if they did not know the answer. The experiment was conducted in the Experimental Syntax and Semantics Lab in the linguistics department at MIT.

372 trials in total were created for the experiment by using 62 types of clusters given in Table 3.4, two orders presenting the stimuli (AXB and BXA), and three repetitions. The trials were presented divided into eight blocks. The whole procedure lasted for 40–50 mins.

3.3.1.2 Results

Results show that the perceptual similarity judgments are mostly consistent with the loanword typology, supporting the main hypothesis of the thesis. In what follows, I present proportions of internal and external epenthesis responses by intensity rise within the cluster, for English (Figures 3-2 and 3-3), Korean (Figures 3-4 and 3-5), and Japanese participants (Figures 3-6 and 3-7). Here clusters with an intensity rise include clusters involving two intensity rises, and clusters with no intensity rise include clusters involving no intensity rise in either positions.

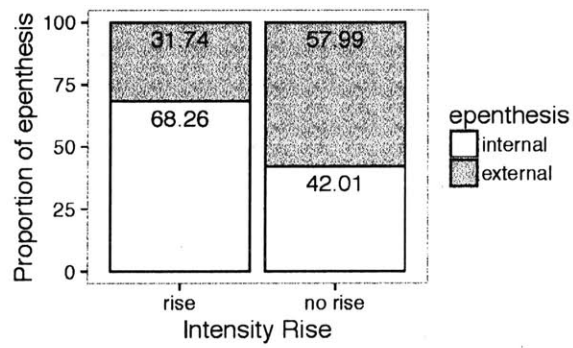


Figure 3-2: Proportions of epenthesis in word-initial clusters (English listeners)

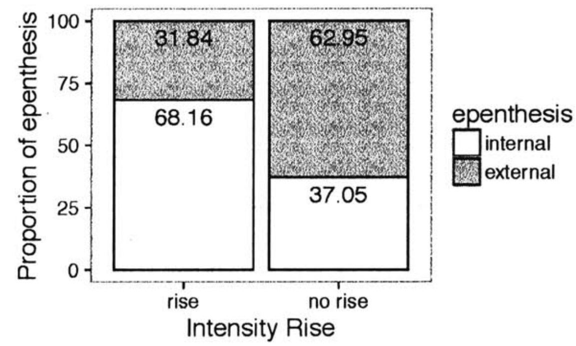


Figure 3-3: Proportions of epenthesis in word-final clusters (English listeners)

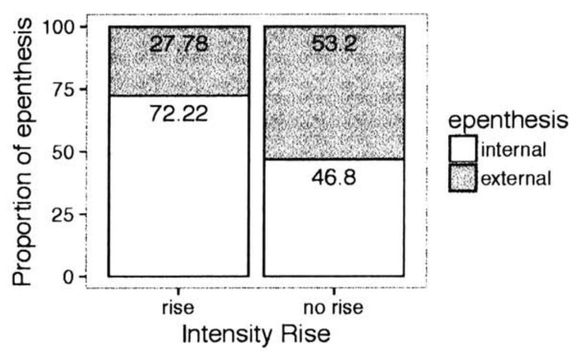


Figure 3-4: Proportions of epenthesis in word-initial clusters (Korean listeners)

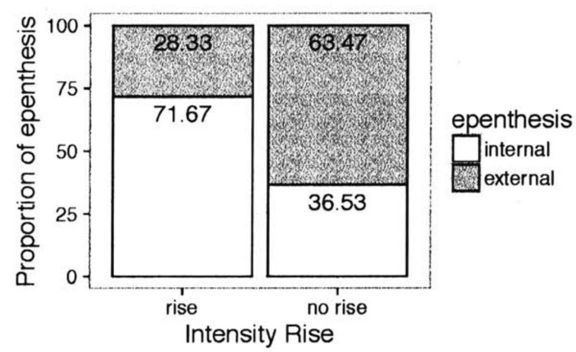


Figure 3-5: Proportions of epenthesis in word-final clusters (Korean listeners)

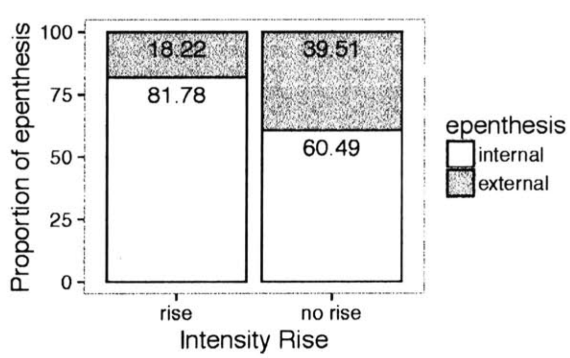


Figure 3-6: Proportions of epenthesis in word-initial clusters (Japanese listeners)

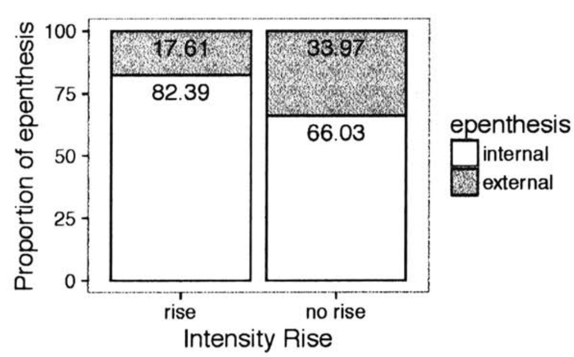


Figure 3-7: Proportions of epenthesis in word-final clusters (Japanese listeners)

The results for English and Korean participants show that internal epenthesis was preferred to external epenthesis for consonant clusters with an intensity rise within the cluster, and external epenthesis was preferred to internal epenthesis for consonant clusters with no intensity rise within the cluster. This tendency held for both word-initial and word-final positions. Mixed effects logistic regression models were fitted to the responses to internal or external epenthesis with the predictor of intensity rise within the cluster. Random intercepts for participant and random slopes for intensity rise by participant were also included in the models. It was shown that intensity rise was a significant predictor that affected perceptual similarity between consonant clusters and their epenthesis counterparts in all four groups (English, word-initial: $z=5.61$, $p<.001$; English, word-final: $z=7.47$, $p<.001$; Korean, word-initial: $z=4.82$, $p<.001$; Korean, word-final: $z=8.57$, $p<.001$).

Looking at the results for Japanese participants in Figures 3-6 and 3-7, on the other hand, internal epenthesis was favored over external epenthesis both in clusters involving an intensity rise and clusters involving no intensity rise inside the cluster. However, responses to external epenthesis were greater in non-rise clusters than in intensity rise clusters, and this difference was statistically significant (word-initial: $z=5.61$, $p<.001$; word-final: $z=3.21$, $p<.01$). Namely, it still holds that internal epenthesis is significantly preferred more often in clusters with an intensity rise than in clusters with no intensity rise, and external epenthesis is significantly preferred in clusters with no intensity rise than in clusters with an intensity rise. I attribute the tendency to prefer internal epenthesis in general to the phonotactics of Japanese; Japanese only allows some homorganic clusters word-medially, and most of the sequences of consonants resulting from external epenthesis in this experiment do not fit Japanese phonotactics. Japanese speakers may have difficulty in perceiving them correctly and tend to have an illusory vowel between the two. Unlike Japanese, Korean allows several consonant clusters in word-medial position, such as [kp], [mk], [lk], and [lm], although it does not allow any clusters at word edges. So for the word-initial stimulus cluster [#mk], for instance, both internal epenthesis ([#mək]) and external epenthesis ([#əmk]) produce sequences that are acceptable in Korean, and it is likely that Korean participants were less affected by the native phonology in deciding perceptual similarity between consonant clusters and their epenthesis counterparts than Japanese participants. Indeed, the percentage of internal epenthesis for Japanese listeners is higher than that for Korean listeners, which in turn is higher than that for English listeners.

As mentioned above, the consonant clusters that involve an intensity rise inside the cluster are composed of clusters that involve only one intensity rise inside and clusters that involve another intensity rise outside the cluster as well as the one inside. Also, the consonant clusters that involve no intensity rise inside the cluster are composed of clusters that involve no intensity rise in either position, i.e., inside and outside the cluster, and clusters that involve an intensity rise only outside of the cluster. The barplots below illustrate proportions of internal and external epenthesis responses according to the number of intensity rises in the consonant cluster: (i) no intensity rise in either position ("none"), (ii) one intensity rise cluster-externally ("onee"), (iii) one intensity rise cluster-internally ("onei"), and (iv) two intensity rises inside and outside the cluster ("two"). Clusters in the left two columns in Figure 3-8 to Figure 3-13 below correspond to the clusters without an intensity rise in the right column in Figures 3-2 to 3-7 above, and clusters in the right two columns in Figures 3-8 to 3-13 correspond to the clusters with an intensity rise in the left column in Figures 3-2 to 3-7.

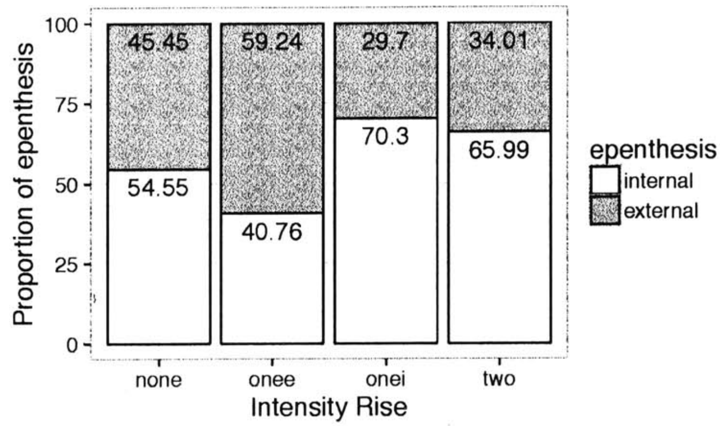


Figure 3-8: Proportion of epenthesis by the number of intensity rises in word-initial clusters (English listeners)

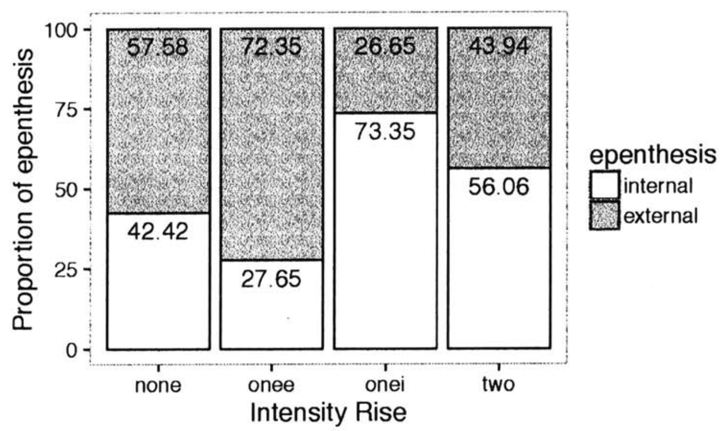


Figure 3-9: Proportion of epenthesis by the number of intensity rises in word-final clusters (English listeners)

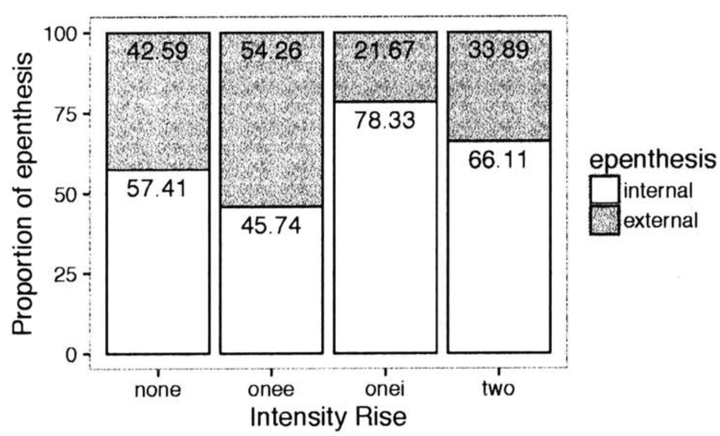


Figure 3-10: Proportion of epenthesis by the number of intensity rises in word-initial clusters (Korean listeners)

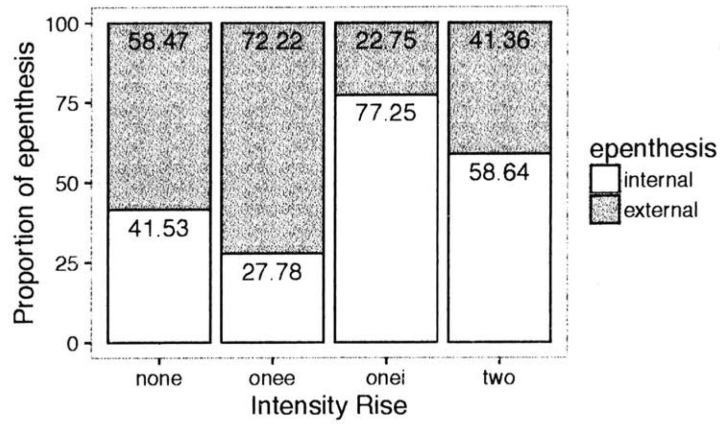


Figure 3-11: Proportion of epenthesis by the number of intensity rises in word-final clusters (Korean listeners)

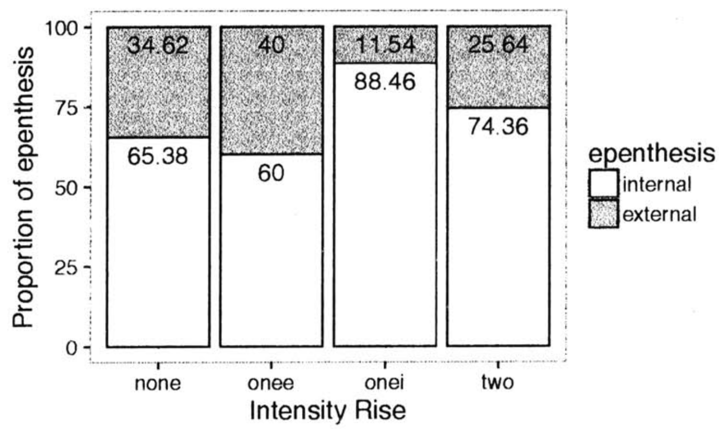


Figure 3-12: Proportion of epenthesis by the number of intensity rises in word-initial clusters (Japanese listeners)

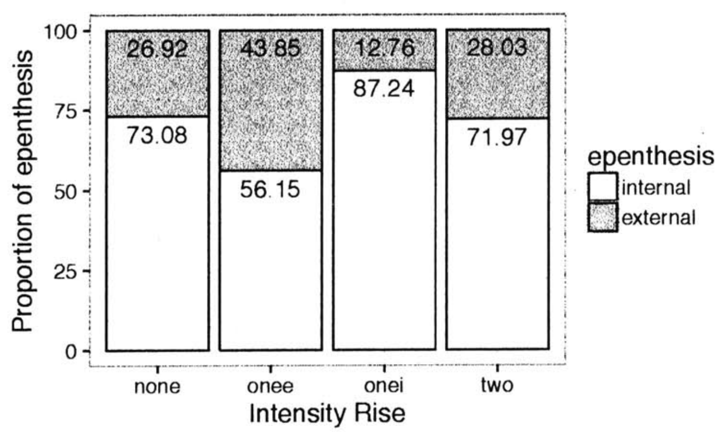


Figure 3-13: Proportion of epenthesis by the number of intensity rises in word-final clusters (Japanese listeners)

For all the participant groups, both in word-initial and in word-final position, the proportion of internal epenthesis is the highest in the clusters involving one intensity rise inside the cluster ("onei") and the proportion of external epenthesis is the highest in the clusters involving one intensity rise outside the cluster ("onee"). That is, the preference for internal or external epenthesis is more obvious when we compare only the clusters involving one intensity rise. So it can be said that these results are consistent with the current hypothesis: epenthesis is preferred where there is a single rise in intensity in the cluster, and may show either internal or external epenthesis when there is no intensity rise or there are two intensity rises.

Presented below are the ratios of each epenthesis pattern by the type of edge consonant, i.e., the first consonant in word-initial clusters and the second consonant in word-final clusters. First, mean responses of each epenthesis pattern in word-initial clusters for English, Korean, and Japanese listeners are illustrated in Figures 3-14, 3-15, and 3-16, respectively. "ir" and "no" in parentheses indicate the presence and the absence of intensity rise within the cluster. For the clusters beginning with a fricative (f, s) or a nasal (m), clusters are divided into two groups, i.e., intensity rise and non-rise clusters. The whole list of word-initial stimuli is provided by intensity rise within the cluster in Table 3.6.

Table 3.6: Word-initial clusters by intensity rise within the cluster and C_1

initial C	intensity rise	non-rise
stop	k⌜p, k⌜s, k⌜f, k⌜m, k⌝l, k⌝l	
fricative [s]	sm, sl, sr	sk⌜, sf,
fricative [f]	fs, fm, fl, fr	fk⌜
nasal	m⌝l, m⌜r	m⌝k⌜, m⌝s, m⌝f, m⌝n
liquid [l]		lk, ls, lf, lm
liquid [r]	rk, rs, rf, rm, rl	

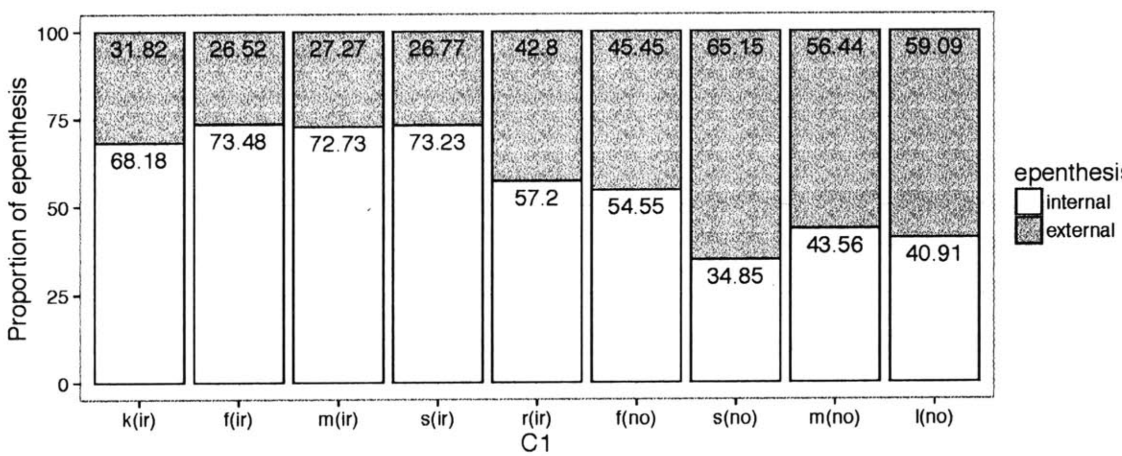


Figure 3-14: Proportions of epenthesis by the type of C_1 in word-initial clusters (English listeners)

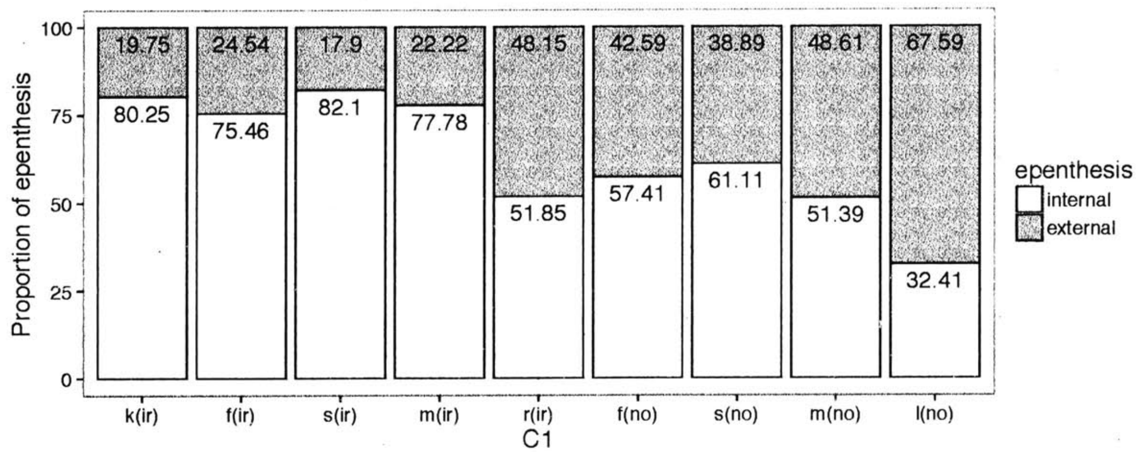


Figure 3-15: Proportions of epenthesis by the type of C_1 in word-initial clusters (Korean listeners)

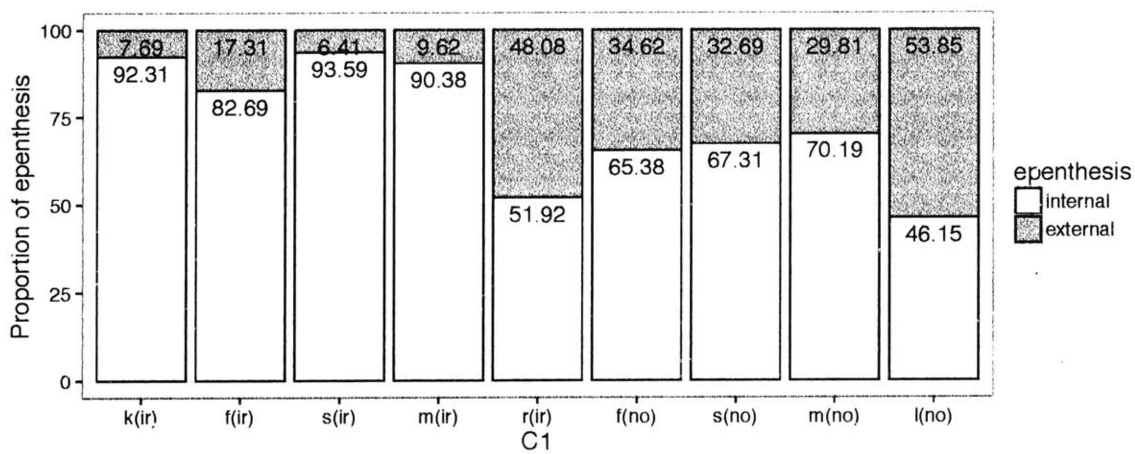


Figure 3-16: Proportions of epenthesis by the type of C_1 in word-initial clusters (Japanese listeners)

We have seen earlier that internal epenthesis is more favored in intensity rise clusters than in non-rise clusters, while external epenthesis is more favored in non-rise clusters than in intensity rise clusters. Here we see that the response of internal epenthesis was higher in all intensity rise clusters than in non-rise clusters, with the exception of trill-initial clusters for Korean and Japanese listeners. Note that [m]-initial intensity rise clusters here are [ml] and [mr], which also involve an intensity rise outside the cluster, and nevertheless intensity rise is much more preferred to external epenthesis. This is consistent with the loanword epenthesis data and the current hypothesis that listeners are also sensitive to the word-initial consonant, as well as the intensity rise. Out of intensity rise clusters, trill-initial clusters show the lowest proportion of internal epenthesis across the participants groups. While there are two intensity rises in clusters involving a trill, as in nasal-liquid clusters, the mean proportion of external epenthesis responses is much higher in trill-initial clusters than in nasal-liquid clusters. This is not surprising given that trill-initial clusters involve several rises in intensity; the first rise is followed by a fall in intensity, which is in turn followed by the second rise in intensity, it may be the case that listeners perceive the open phase following the first intensity rise as a vowel. This could particularly be true because the trill is not a phoneme in English, Korean, and Japanese.

Next, in Table 3.7, the word-final clusters are listed, classified by the final consonant and intensity rise. Figures 3-17, 3-18, and 3-19 illustrate the responses of internal and external epenthesis by the type of word-final consonant for English, Korean, Japanese listeners, respectively.

Table 3.7: Word-final clusters by intensity rise within the cluster and C_2

final C	intensity rise	non-rise
stop	p⌝k⌜	sk⌜, fk⌜, mk⌜, lk⌜, rk⌜
fricative [s]	k⌜s, fs, rs	m⌝s, ls
fricative [f]	k⌜f, rf	sf, m⌝f, lf
nasal	k⌜m⌝, sm⌝, fm⌝, rm⌝	m⌝n⌝, lm⌝
liquid [l]	k⌜l, sl, fl, m⌜l, rl	
liquid [r]	k⌜r, sr, fr, m⌜r	

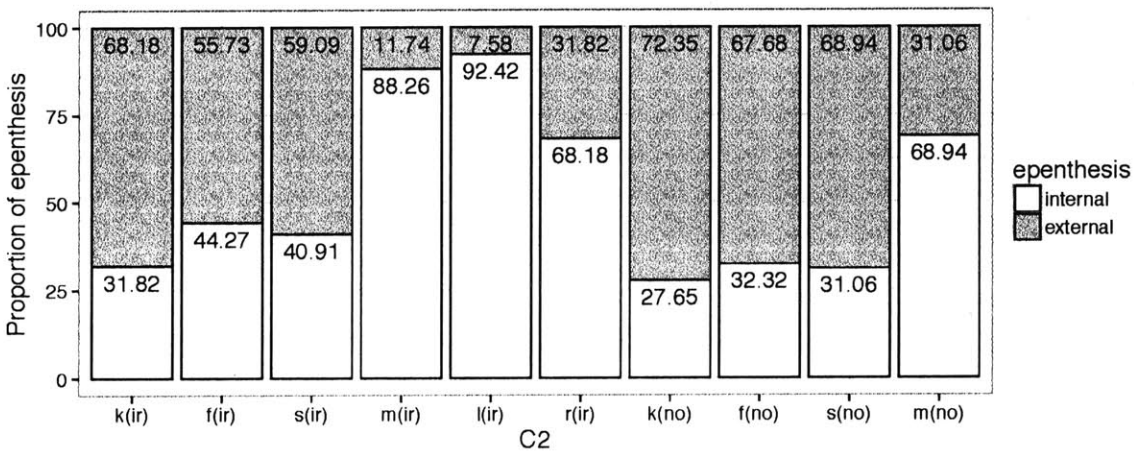


Figure 3-17: Proportions of epenthesis by the type of C_2 in word-final clusters (English listeners)

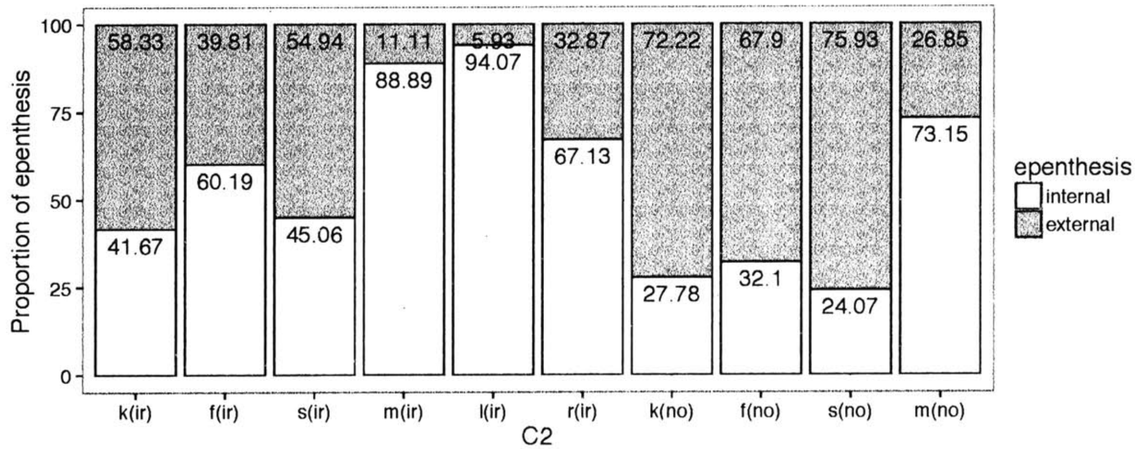


Figure 3-18: Proportions of epenthesis by the type of C_2 in word-final clusters (Korean listeners)

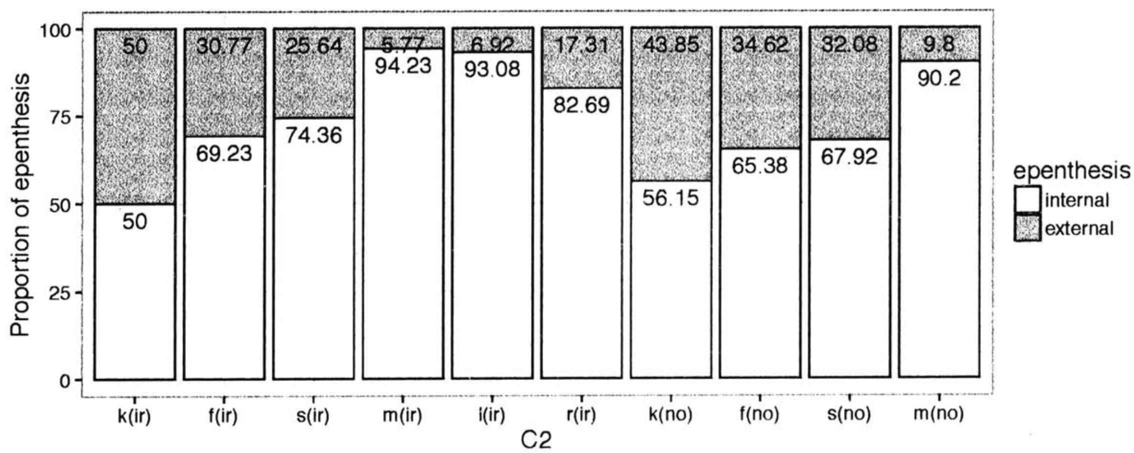


Figure 3-19: Proportions of epenthesis by the type of C_2 in word-final clusters (Japanese listeners)

Let us first see the results of non-rise clusters presented in the last four bars. Except for [m]-final clusters, in all non-rise clusters for English and Korean listeners, external epenthesis is preferred to internal epenthesis, confirming the hypothesis. For Japanese listeners, the proportions of external epenthesis are slightly higher in the non-rise clusters than in the intensity rise clusters. The pattern in [m]-final non-rise clusters, i.e., [m⌝n⌝] and [lm⌝], in which internal epenthesis is preferred to external epenthesis, is consistent with the cross-linguistic pattern of epenthesis; we have seem in (9b) that those clusters, which do not involve an intensity rise both inside and outside of the cluster, are repaired with internal epenthesis (e.g., *gimn* (Russian) → [gimun] (Kirghiz); 'film' → [filim] (Uyghur)). Notice that in the other word-final cluster involving no intensity rise, the nasal-sibilant cluster, which is repaired by external epenthesis in loanword adaptation, external epenthesis is far more preferred in English and Korean listeners. While the response of internal epenthesis is higher than that of external epenthesis for Japanese listeners, the proportion of external epenthesis in [s]-final non-intensity-rise clusters is significantly higher than that in [m]-final non-intensity-rise clusters.

Considering the results of the stop-final intensity rise cluster, i.e., [p⌜k⌜], in the first bars, the response of internal epenthesis is comparable to that of external epenthesis. This is not surprising because both of the stops in the cluster were clearly released and there were two intensity rises, between the two stops and after the final stop. So participants preferred the second intensity rise for the position of epenthesis. For the fricative-final clusters, external epenthesis was somewhat preferred to internal epenthesis, not consistent with the hypothesis, except that Korean listeners preferred internal epenthesis in [f]-final clusters. The response of internal epenthesis was slightly higher in intensity rise clusters than in non-rise clusters, but the differences were only significant in English listeners' [f]-final clusters ($z=2.32$, $p<.05$) and in Korean listeners' [f]-final clusters ($z=4.34$, $p<.001$) and [s]-final clusters ($z=2.94$, $p<.001$).

Recall that the epenthesis pattern in word-initial sibilant-initial clusters are somewhat complicated; internal epenthesis is more likely in sibilant-trill clusters than in the other sibilant-sonorant clusters (i.e., sibilant-nasal and sibilant-lateral), which in turn show a higher internal epenthesis rate than sibilant-stop clusters. Let us see whether the present similarity judgment results reflect the cross-linguistic epenthesis patterns. Figures 3-20, 3-21, and 3-22 illustrate the proportion of each epenthesis pattern in word-initial [s]-initial clusters for English, Korean, and Japanese listeners, respectively.

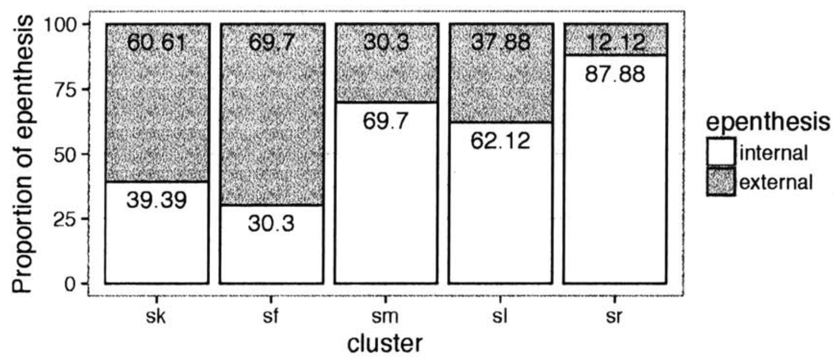


Figure 3-20: Proportions of epenthesis in word-initial [s]-initial clusters (English listeners)

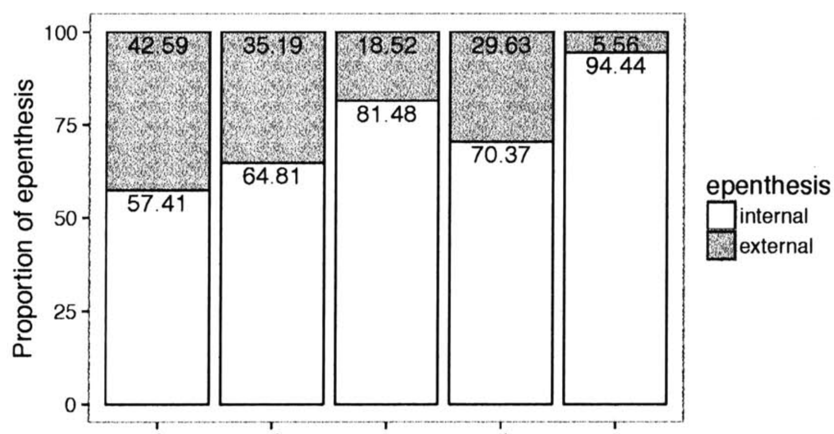


Figure 3-21: Proportions of epenthesis in word-initial [s]-initial clusters (Korean listeners)

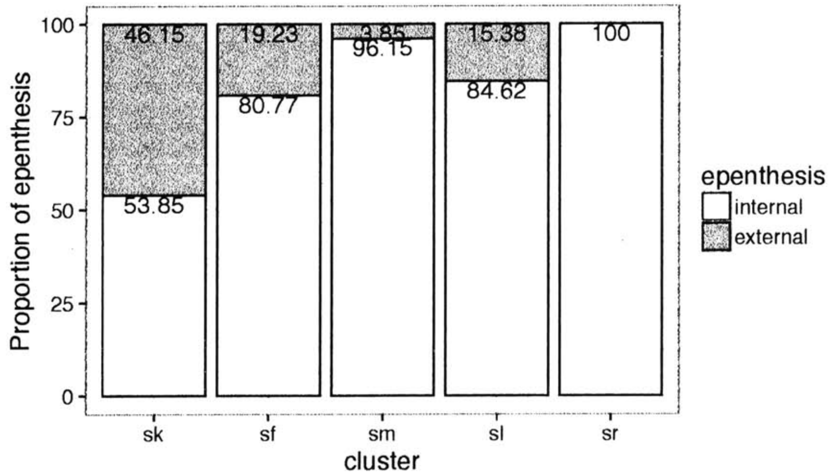


Figure 3-22: Proportions of epenthesis in word-initial [s]-initial clusters (Japanese listeners)

For all the participant groups, internal epenthesis was preferred to external epenthesis more in [sm] and [sl] clusters than in [sk] clusters. A mixed effects logistic regression model was fit to the responses with the predictor of C_2 in [s]-initial clusters with a random intercept of participant, and the differences were mostly statistically significant (English: [sm] $z=3.8$, $p<.001$, [sl] $z=2.88$, $p<.01$ / Korean: [sm] $z=2.9$, $p<.01$, [sl] $p=0.1$ *n.s.* / Japanese: [sm] $z=3.26$, $p<.01$, [sl] $z=2.45$, $p<.05$). There was no significant difference in the responses of internal and external epenthesis between [sm] and [sl] clusters for all the participant groups. Internal epenthesis was also preferred more in [sr] clusters than in [sm] and [sl] clusters, and the difference was statistically significant for English ([sm] $z=2.7$, $p<.01$, [sl]: $z=3.59$, $p<.001$) and Korean listeners ([sm]: $z=2.31$, $p<.05$, [sl]: $z=3.4$, $p<.001$), but not for Japanese listeners. To conclude, we can say that the results of the similarity judgment task correspond to the cross-linguistic patterns of epenthesis in word-initial sibilant-initial clusters.

Unlike in word-initial clusters, no variation in epenthesis sites is observed in word-final sibilant-sonorant clusters. I explain this positional asymmetry based on intensity rise in (24); whereas there are two intensity rises and thus two possible epenthesis positions in word-initial sibilant-sonorant clusters, word-final sibilant-sonorant clusters (except for trill-final clusters) involve only one intensity rise between the two consonants and only internal epenthesis occurs. The similarity judgment results for word-final sibilant-initial clusters support this hypothesis. The responses for similar epenthesis sites in word-final sibilant clusters are presented in Figures 3-23, 3-24, and 3-25 for English, Korean, and Japanese listeners, respectively.

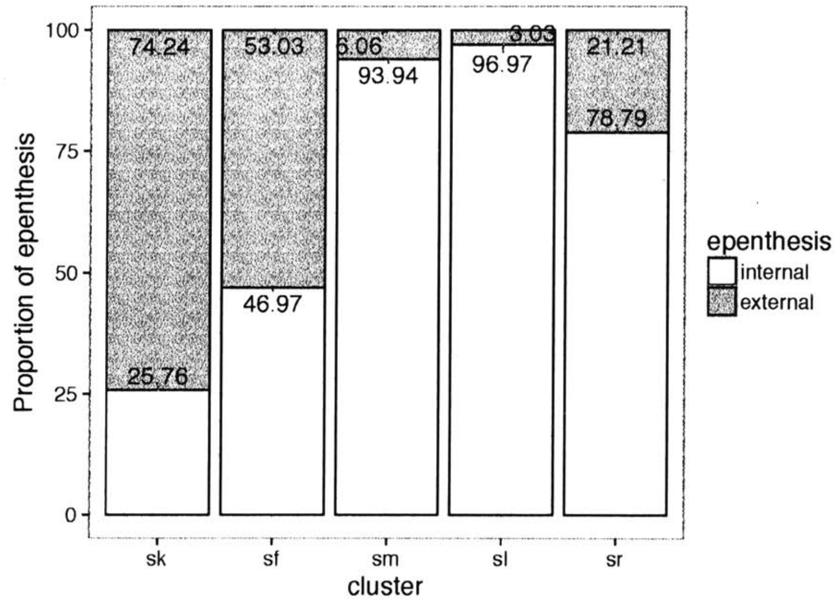


Figure 3-23: Proportions of epenthesis in word-final [s]-initial clusters (English listeners)

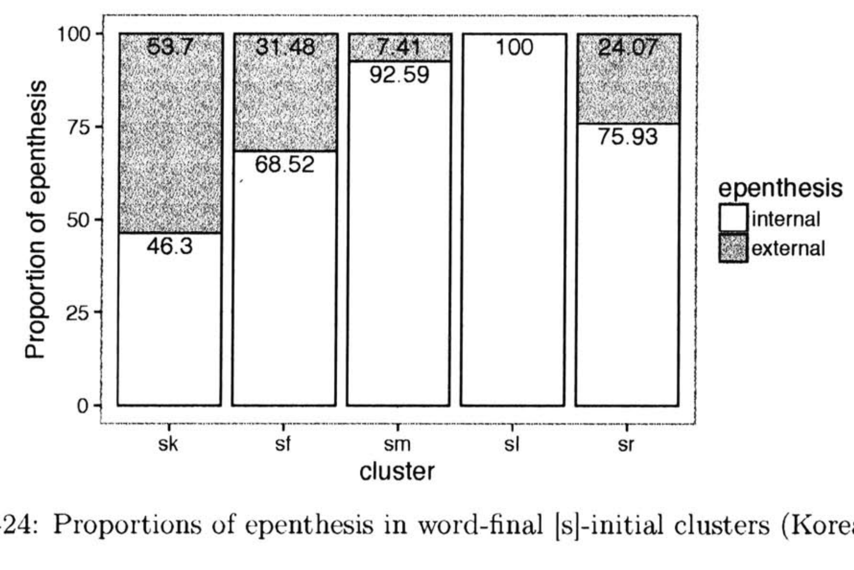


Figure 3-24: Proportions of epenthesis in word-final [s]-initial clusters (Korean listeners)

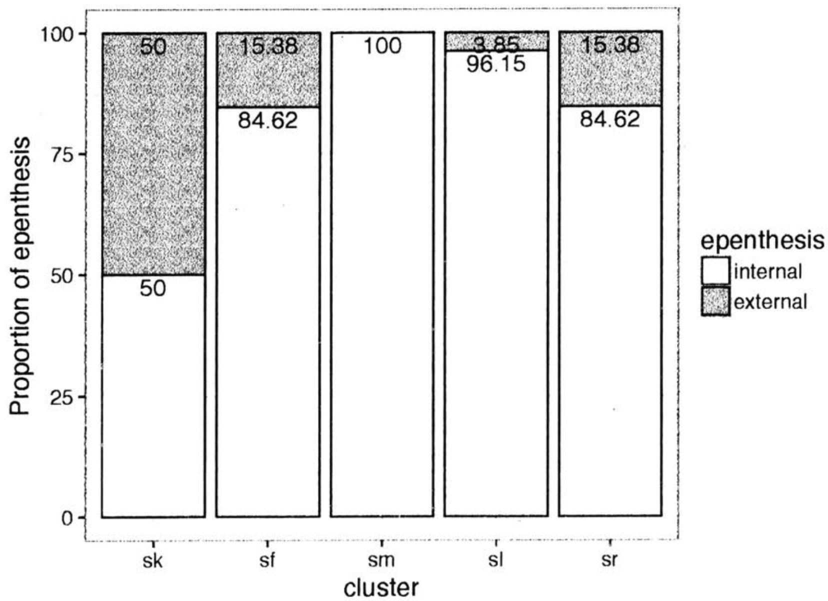


Figure 3-25: Proportions of epenthesis in word-final [s]-initial clusters (Japanese listeners)

Let us focus on [sm] and [sl] clusters first. In word-final position, the percentages of internal epenthesis are higher than 90% across the participant groups. Comparing the results of [sm] and [sl] clusters in word-initial and in word-final position, there were more responses of internal epenthesis in word-final position than in word-initial position. This reflects the asymmetry between word-initial and word-final sibilant-sonorant clusters in epenthesis sites. For [sr] clusters, the rate of internal epenthesis is lower than [sm] and [sl] clusters as well as lower than [sr] clusters in word-initial position. This is also attributed to intensity rise; in word-final [sr] clusters, an intensity rise at the end of the final trill exists, as well as the one between [s] and [r], which may be the reason for the higher external epenthesis rate.

3.3.1.3 Discussion

The results of Experiment 4 show that epenthesis sites in loanword adaptation are largely consistent with the higher response of epenthesis type; participants judged clusters that involve an intensity rise and undergo internal epenthesis in loan adaptation as more similar to the form with internal epenthesis than to the form with external epenthesis, and clusters that involve no intensity rise inside and undergo external epenthesis as more similar to the form with external epenthesis in loanword adaptation than to the form with internal epenthesis. Notice that the patterns observed in the results across participant groups are extremely similar, which indicates that the results are not derived from language-specific knowledge but are derived from universal preference for a perceptually similar change.

The results of the current experiment are consistent with the results of two similarity rating experiments in Fleischhacker (2001). In these experiments, English-speaking participants rated on a 7-point scale similarity or preference for English words beginning with a consonant cluster to their corresponding forms with either internal epenthesis or external epenthesis. For example, 'stoke' was compared with [sətok] and [əstok]. While the first experiment used stimulus words beginning with sibilant-stop, obstruent-sonorant (sm, sl, sw, fl, and stop-sonorant), and triconsonantal clusters such as [spl], the second experiment focused on sibilant-sonorant clusters (sm, ʃm, sn, ʃn, sl, ʃl, and sw). It was shown in the first experiment that in sibilant-stop clusters, external epenthesis was preferred to internal epenthesis, and internal epenthesis was preferred to external epenthesis in all the other clusters, i.e., obstruent-sonorant, while no statistical significance was found in [sm] for similarity raters and [sm] and [fl] for preference raters. The mean rating of internal epenthesis for [sl] clusters was higher than that for [sm] clusters, although the statistical significance of this difference was not reported. In the second experiment, however, no significant differences in ratings were observed across sibilant-sonorant clusters. Unlike in the first experiment, there was no significant difference in ratings between sibilant-nasal and sibilant-lateral clusters.

In the current experiment, we have seen that external epenthesis was preferred in sibilant-stop clusters, consistent with Fleischhacker's (2001) first experiment. Also, for [sm] and [sl] clusters, internal epenthesis was preferred and there was no significant difference between the two clusters, which is consistent with Fleischhacker's second experiment and possibly with her first experiment too. Although Fleischhacker did not test [#sr], it was clear in the current experiment that the proportion of internal epenthesis was significantly higher in [#sr] than in [#sm] and [#sl]. To conclude, the increasing proportion of internal epenthesis in [#sk], [#sm]/[#sl], and [#sr] corresponds to the different epenthesis patterns between #ST, #SN/#Sl, and #Sr. What remains to be determined by future research is the motivation for the asymmetry between #SN/#Sl vs. #Sr both in the typology of epenthesis and in the similarity judgments.

3.3.2 Experiment 5

While listeners' choices were limitied to internal or external epenthesis in Experiment 4, vowel epenthesis is not the only possible repair of the illegal clusters. Experiment 5 was conducted employing a transcription task to see whether listeners' perception of consonant clusters with regard to epenthesis is consistent with the epenthesis site typology, even when no choice is forced.

3.3.2.1 Methods

3.3.2.1.1 Participants

Twenty seven English speakers participated in the transcription experiment. The participants were recruited from the subject pool of the MIT Behavioral Research Lab, as native speakers of English who had not learned a second language involving many consonant clusters (e.g., Russian, Arabic, German, Greek, etc.) and had no speaking or hearing problems. The participants were compensated $10 for their participation.

3.3.2.1.2 Stimuli

The stimuli used for the transcription task consisted of the 58 types of nonce words beginning with a consonant cluster used in Experiment 2. The matching epenthesis forms were not used.

3.3.2.1.3 Procedure

A transcription task was conducted in the Behavioral Research Lab at MIT. Each stimulus word was played once following 350 ms of silence after the participants moved on to the next trial by clicking the mouse on the "Click to proceed" button on the screen. After listening to the stimlus, the participants were asked to write down on paper what they heard using the English alphabet.

The participants were told that the words they were about to hear were nonsense words and were allowed to add explanation if they thought English orthography was not sufficient to transcribe the sounds they heard.

There were four practice items, which were also used for practice in Experiment 2, before the actual experiment started. The 58 stimuli were played twice in a randomized order. It took 25–30 mins to complete the task.

3.3.2.2 Results

Since the stimulus clusters included native clusters in English and participants showed almost 100% accuracy for those clusters, results of non-native clusters in English are considered in what follows.

Proportions of correct responses by intensity rise, i.e., intensity rise and C_1 voicing, are illustrated in Figure 3-26. Here we see a clear difference in percentage correct between clusters beginning with a voiced consonant and clusters beginning with a voiceless-consonant. The difference between clusters with an intensity rise and clusters with no intensity rise was significant only when C_1 was voiceless, not when C_1 was voiced.

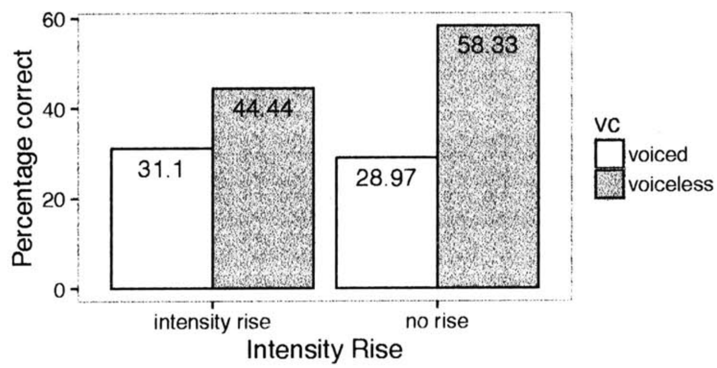


Figure 3-26: Percentage correct on non-native trials

Now let us see whether the transcription errors are similar to the epenthesis patterns in loanword adaptation. Table 3.8 summarizes the typology and the most common epenthesis error in transcription by word-initial consonant and intensity rise inside the cluster. The parentheses in epenthesis pattern under the typology column means that loan epenthesis for the corresponding clusters is not observed but is expected according to the intensity rise hypothesis. We see here that the typological pattern of epenthesis largely corresponds with the most frequent epenthesis error in transcription.

Table 3.8: Epenthesis typology and the most common epenthesis error in the transcription results

	intensity rise	typology	transcription error
[k]-initial	✓	internal	internal
[g]-initial	✓	internal	internal
[f]-initial ([fs], [fm])	✓	internal	internal
[v]-initial ([vm], [vl], [vr])	✓	internal	internal
[m]-initial ([m⌜k], [m⌜f], [m⌜s], [m⌜n], [m⌜l], [m⌜r])	✓	internal	internal
[m]-initial ([m⌝k], [m⌝f], [m⌝s], [m⌝n])	✗	external	internal
[f]-initial ([fk])	✗	(external)	internal
[v]-initial ([vd], [vz])	✗	external	external
[l]-initial	✗	(external)	external
[r]-initial	✓	(variation)	internal
[s]-initial ([sr])	✓	internal	internal
[z]-initial ([zm], [zl], [zr])	✓	(internal)	internal
[z]-initial ([zd], [zv])	✗	(external)	internal

In what follows, proportions of errors for each cluster type are presented in a barplot. For the type of errors, 'internal' stands for internal epenthesis, 'external' for external epenthesis, 'both' for both internal epenthesis and external epenthesis, 'vocalization' for vocalization of one of the consonants in the cluster (e.g., [lk] → [ik]), 'deletion' for deletion of one of the consonants (e.g., [km] → [m]), and 'change' for feature change in one of the consonants (e.g., [km] → [pm], [gm], [kn], etc.).

First, stop-initial clusters always undergo internal epenthesis in word-initial position in loan adaptation. As shown in Figures 3-27 and 3-28 for [k]-initial and [g]-initial clusters, respectively, the error of internal epenthesis is extremely high compared to the other types of errors involving external epenthesis, which is consistent with the typology in which only internal epenthesis occurs in stop-initial clusters.

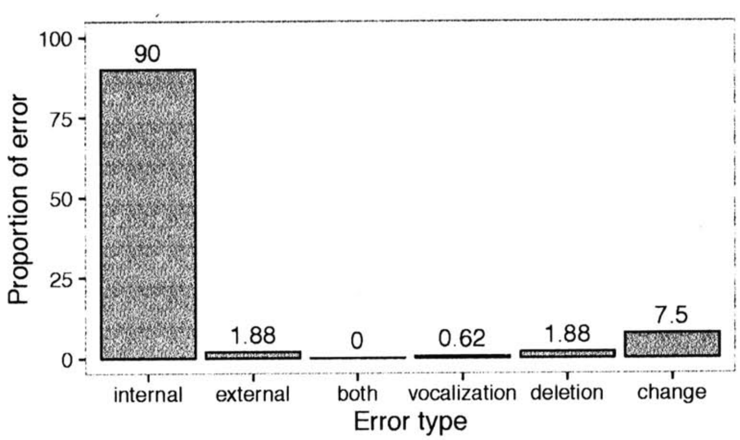


Figure 3-27: Error types on [k]-initial clusters

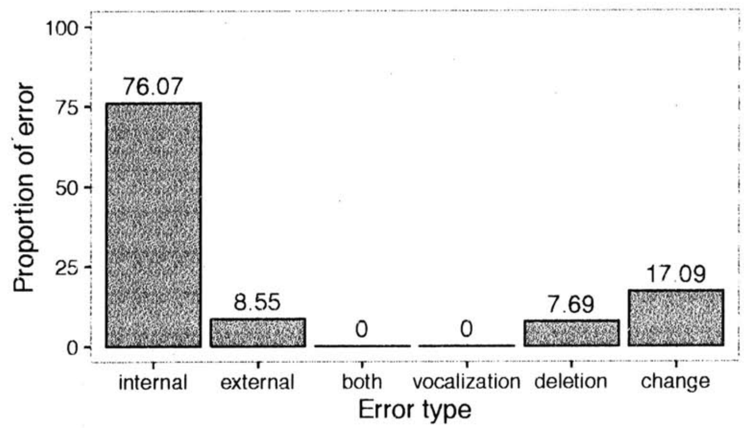


Figure 3-28: Error types on [g]-initial clusters

Moving on to the fricative-initial clusters, [f]-initial clusters involve an intensity rise except for the fricative-stop, [fk], cluster. The [f]-initial intensity rise cluster [fl] showed internal epenthesis in the loanword typology, whereas there is no real example of epenthesis in the other [f]-initial clusters. As illustrated in Figure 3-29, the proportion of internal epenthesis errors was about eightfold higher than that of external epenthesis errors. Even though the errors of deletion and change are more frequent than internal epenthesis in this case, the preferrence for internal epenthesis is obvious when comparing the errors of internal epenthesis and external epenthesis. On the other hand, the non-rise cluster [fk] shows a high proportion of internal epenthesis errors but no external epenthesis errors. Although there is no actual example of epenthesis, this result is unexpected because the [fk] cluster does not involve an intensity rise.

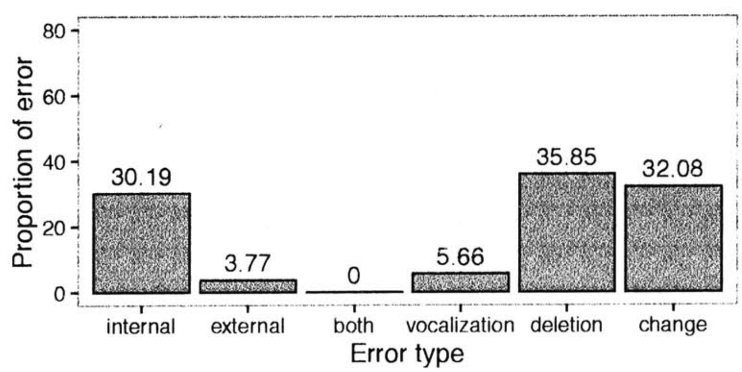


Figure 3-29: Error types on [f]-initial intensity rise clusters

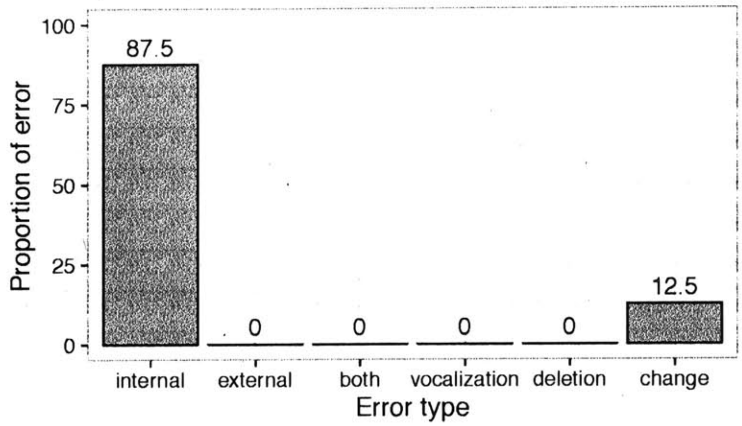


Figure 3-30: Error types on [f]-initial non-rise clusters

Figures 3-31 and 3-32 present the type and proportion of transcription errors in [v]-initial clusters. As expected, the ratio of internal epenthesis is the highest in intensity rise clusters, which undergo internal epenthesis in the typology, as shown in Figure 3-31, and in non-rise clusters, which undergo external epenthesis in the typology, the ratio of external epenthesis is higher than that of internal epenthesis, although the most frequent error was change. Although the difference between internal epenthesis and external epenthesis errors is small, the ratio of external epenthesis with respect to internal epenthesis is much higher in non-rise clusters than in intensity rise clusters.

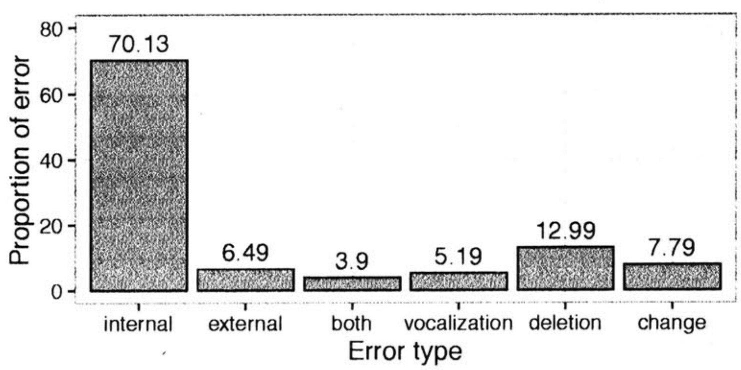


Figure 3-31: Error types on [v]-initial intensity rise clusters

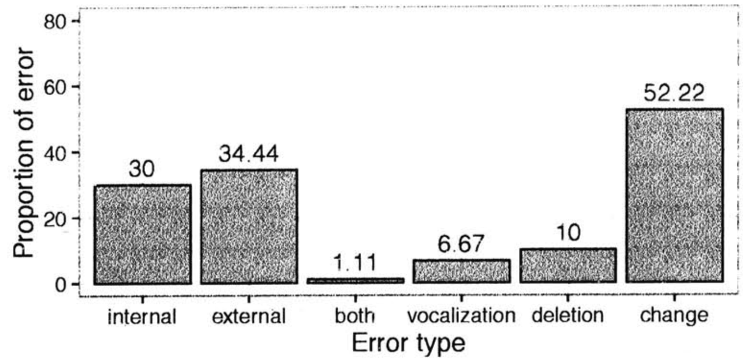


Figure 3-32: Error types on [v]-initial non-rise clusters

The results of [s]-initial clusters were not as informative because all the clusters except for [sr] were native clusters of English and participants rarely made an error. For the [sr] clusters, which are illustrated in Figure 3-33, the internal epenthesis rate was close to 90% but there was no error of external epenthesis. This is consistent with the epenthesis pattern in loanword adaptation that shows internal epenthesis only across languages.

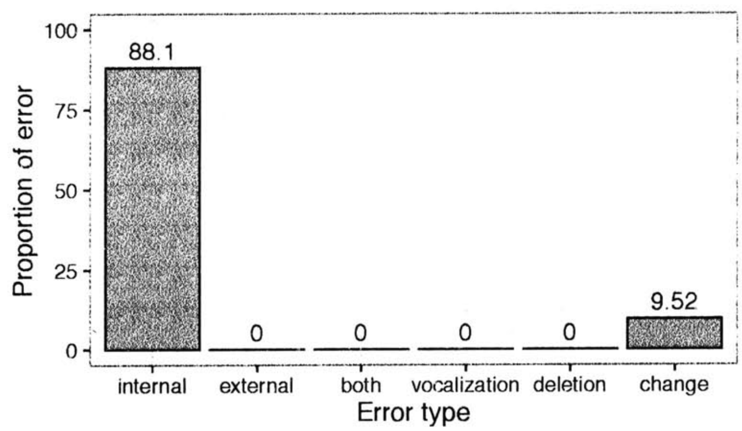


Figure 3-33: Error types on intensity rise [s]-initial intensity rise clusters

Just like the [sr] clusters, [z]-initial clusters with an intensity rise show a very high percentage of internal epenthesis errors, as illustrated in Figure 3-34, while there were also about 8% of external epenthesis errors. For [z]-initial non-rise clusters, which are expected to undergo external epenthesis although there is no actual example, internal epenthesis is still the most common error, as shown in Figure 3-35. However, the proportion of external epenthesis errors is somewhat comparable with that of internal epenthesis, and crucially, it is far higher than that in [z]-initial intensity rise clusters in Figure 3-34.

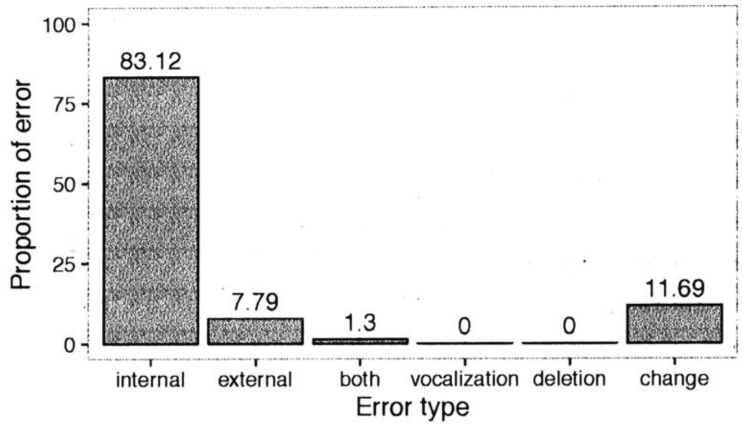


Figure 3-34: Error types on [z]-initial intensity rise clusters

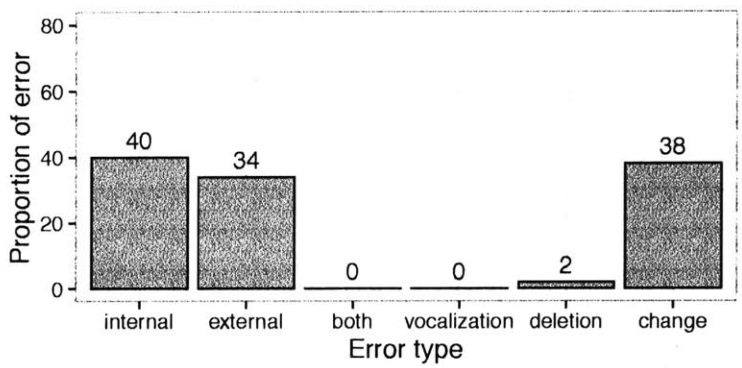


Figure 3-35: Error types on [z]-initial non-rise clusters

Consider the results of nasal-initial clusters now. Presumably, when there is an intensity rise in nasal-initial clusters, internal epenthesis occurs, and when there is no intensity rise, external epenthesis occurs in loanword adaptation. The transcription results in Figures 3-36 and 3-37 show that in both cases internal epenthesis is the most frequent error. However, the external epenthesis error is much more frequent in non-rise clusters than in intensity rise clusters. So although the ratio of internal epenthesis error is higher than that of external epenthesis in [m]-initial non-rise clusters that undergo external epenthesis in loanword adaptation, the difference in the ratio of external epenthesis between [m]-initial intensity rise and non-rise clusters is noticeable and consistent with the similarity judgment results in Experiment 4.

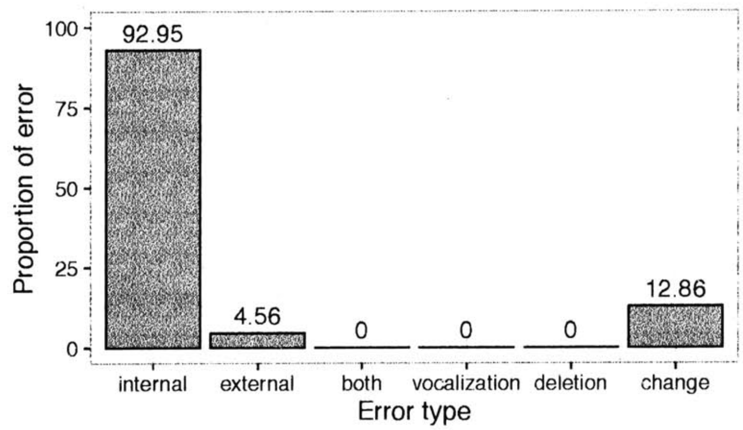


Figure 3-36: Error types on [m]-initial intensity rise clusters

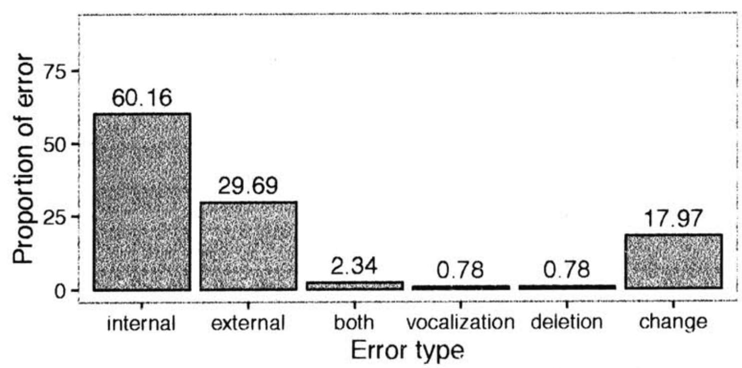


Figure 3-37: Error types on [m]-initial non-rise clusters

In the current loanword database, no example of epenthesis is observed in word-initial lateral-initial clusters. It is prediceted, however, by the intensity rise hypothesis that external epenthesis would be preferred before the lateral-initial cluster because lateral-initial clusters do not feature an intensity rise between the lateral and the following consonant. As expected, external epenthesis was the most frequent transcription error in [l]-initial clusters, as shown in Figure 3-38.

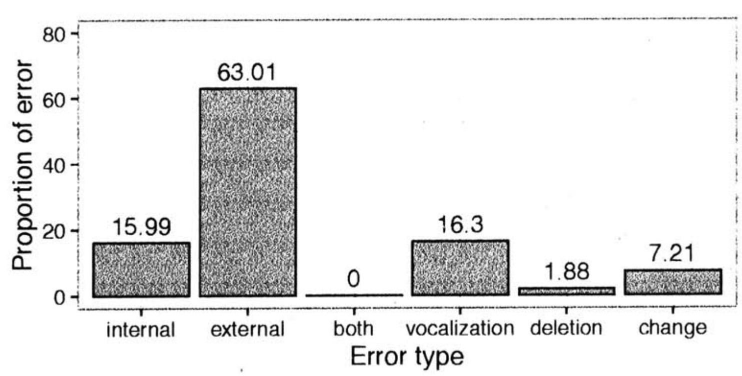


Figure 3-38: Error types on [l]-initial clusters

As in lateral-initial clusters, there is no actual example of epenthesis in trill-initial clusters word-initially. Unlike lateral-initial clusters, however, trill-initial clusters involve an intensity rise between the trill and the following consonant, as well as from the silence to the trill, and it is expected that internal epenthesis is possible. The results in Figure 3-39 show a much higher rate of internal epenthesis than that of external epenthesis, which is also much higher than in lateral-initial clusters involving no intensity rise..

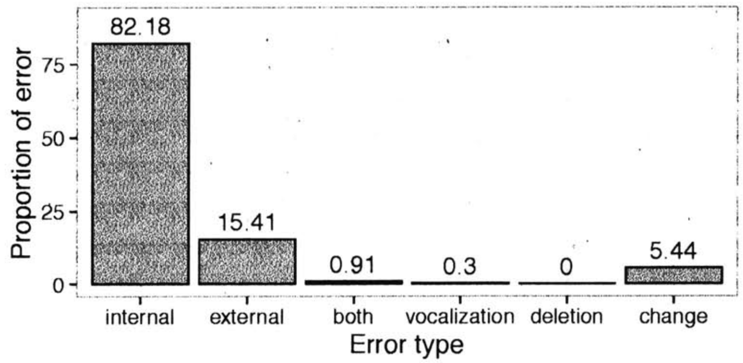


Figure 3-39: Error types on [r]-initial clusters

To summarize, the transcription errors are overall in accord with the epenthesis patterns in loanword cluster adaptation. Figure 3-40 summarizes the ratios of internal epenthesis to external epenthesis by the initial consonant in the cluster. It is shown that the ratio of internal epenthesis is higher than 84% in all intensity rise clusters on the left side of the red vertical line, with the exception of the [fk] cluster with non-rise. In non-rise clusters on the right, the ratio of external epenthesis is higher than 33%.

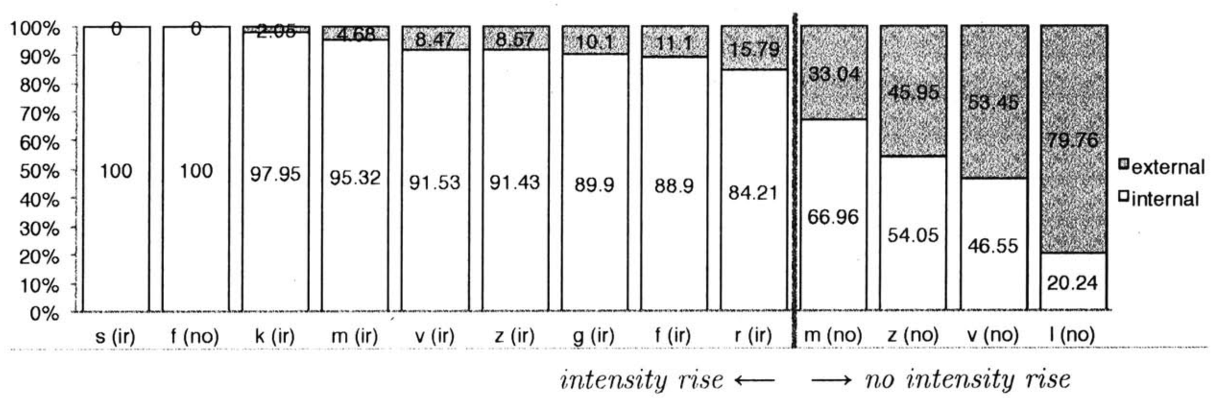


Figure 3-40: Proportions of internal and external epenthesis

3.3.2.3 Discussion

The transcription results of Experiment 5 reflect the typological patterns of epenthesis sites in loanword adaptation. In general, the most frequent epenthesis error corresponded to the epenthesis pattern. There were two cases, however, in which the epenthesis pattern and the transcription error did not agree: [m]-initial non-rise clusters and [z]-initial non-rise clusters. Both of these clusters were expected to have more external epenthesis responses than internal epenthesis responses, but the most frequent error was internal epenthesis in the experiment.

Importantly, however, the ratio of external epenthesis responses to internal epenthesis responses was much higher in these clusters than in their matching clusters with an intensity rise. So the asymmetry between intensity rise and non-rise clusters is still observed in the transcription results. These results support the current hypothesis that epenthesis sites are determined based on perceptual similarity between consonant clusters and their corresponding epenthesis forms.

3.4 Analysis

This section provides a formal analysis of the typology of epenthesis sites in loan adaptation based on the P-map hypothesis (Steriade, 2008), within the framework of Optimality Theory (Prince and Smolensky, 2004). It will be shown that interactions between suggested perceptually-driven constraints and their rankings can correctly derive the typology introduced above.

Before embarking on the analysis, we need to clarify what exactly the input represents in loan adaptation. Here I assume that in loanword adaptation the input is an auditory representation of the source language that includes phonetic details. So a given phoneme of the source language can be adapted differently in different segmental contexts depending on its surface phonetic characteristics in the source language (cf. Kang (2003), Shinohara (2006)). Therefore, information about acoustic cues such as intensity rises discussed earlier should be available in the input.

Let me introduce constraints we need to analyze the epenthesis sites in loan adaptation. First of all, there would be a markedness constraint prohibiting consonant clusters, which motivates epenthesis. I assume here that *COMPLEX defined in (26) penalizes consonant clusters in question in the borrowing language, although in fact different markedness constraints might be working in each language. When vowel epenthesis takes place, a faithfulness constraint prohibiting vowel insertion, DEP-V in (27), will be violated. So if *COMPLEX outranks DEP-V, epenthesis may occur. Additionally, internal epenthesis causes a split of the two consonants in a cluster, violating CONTIGUITY in (28), and external epenthesis causes the modification at the word edge, violating ANCHOR in (29).

(26) *COMPLEX (*CC; cf. Prince and Smolensky, 2004): No tautosyllabic consonant sequences

(27) DEP-V (McCarthy and Prince, 1995): No insertion of a vowel.

(28) CONTIGUITY (McCarthy and Prince, 1995):
Elements adjacent in the input must be adjacent in the output.

(29) ANCHOR (McCarthy and Prince, 1995):
A segment at the periphery of the output has a correspondent at the periphery of the input.

In the previous section, I argued that epenthesis at an intensity rise would be a perceptually smaller modification of the original cluster than epenthesis at a non-rise. By the P-map, this perceptual difference is projected to correspondence constraints and their fixed ranking. I suggest two types of DEP-V constraints in (30). Henceforth, IR stands for intensity rise in the constraint.

(30) a. DEP-V/IR: No insertion of a vowel in intensity rises
b. DEP-V/¬IR: No insertion of a vowel in non-rises

DEP-V/IR penalizes vowel epenthesis that occurs where there is an intensity rise, and DEP-V/¬IR penalizes vowel epenthesis in the other contexts. Based on the perceptual disparity that epenthesis in a non-rise in intensity is perceptually more salient than epenthesis in a rise in intensity, DEP-V/¬IR always dominates DEP-V/IR, as in (31). This universal ranking is based on the P-map hypothesis (Steriade, 2008), whose gist is that perceptually smaller changes are preferred, by ranking faithfulness constraints that prohibit perceptually more salient changes over faithfulness constraints that prohibit less salient changes.

(31) Universal ranking: DEP-V/¬IR ≫ DEP-V/IR

Notice that there are two possible epenthesis sites in word-initial and word-final clusters, cluster-internally and cluster-externally. If one position involves an intensity rise and the other does not, vowel epenthesis takes place at the AD since DEP-V/¬IR always outranks DEP-V/IR, as shown in (33). Let us see how these constraints interact and derive cross-linguistic patterns of epenthesis, based on the ranking in (32).

(32) Partial ranking for the epenthesis typology:
*COMPLEX ≫ DEP-V/¬IR ≫ DEP-V/IR

First, we have seen in (9) that clusters that undergo internal epenthesis only involve an intensity rise between the two consonants in the cluster. In the tableau in (33), it is shown that how internal epenthesis is derived in word-initial stop-initial clusters. In the following input forms, I will use "↗" to represent intensity rise for readers' convenience.

(33) Internal epenthesis in stop-initial clusters: #T↗C → #TəC

#T↗C		*CC	DEP-V/¬IR	DEP-V/IR
a.	#TC	*!		
b. ☞	#TəC			*
c.	#əTC		*!	

The cluster form without epenthesis in (33a) is ruled out, violating the highest-ranked markedness constraint *COMPLEX (*CC). The candidate with internal epenthesis in (33b) beats the candidate with external epenthesis (33c) because in the case of external epenthesis, a vowel is inserted at a non-rise, critically violating DEP-V/¬IR, which is universally ranked higher than DEP-V/IR violated in the candidate with internal epenthesis.

In contrast, the universal ranking derives external epenthesis when there is no intensity rise between the two consonants in the cluster. Recall that word-initial nasal-obstruent clusters are repaired by external epenthesis. As shown in (34), external epenthesis in candidate (34c) is favored because there is an intensity rise from the initial silence to the initial nasal and thus DEP-V/IR is violated in the case of external epenthesis. Since there is no intensity rise between the nasal and the following stop, internal epenthesis in (34b) incurs a violation of DEP-V/¬IR, which outranks DEP-V/IR.

(34) External epenthesis in word-initial nasal-stop clusters: #↗N⌝T → #əNT

#↗NT		*CC	DEP-V/¬IR	DEP-V/IR
a.	#NT	*!		
b.	#NəT		*!	
c. ☞	#əNT			*

On the other hand, if both of the possible epenthesis sites are the same with respect to intensity rise, i.e., both involve an intensity rise or both involve no intensity rise, what determines the epenthesis site? I suggest that the relavant ranking between CONTIGUITY and ANCHOR determines the epenthesis site. Let me take an example of variable epenthesis sites in word-initial sibilant-sonorant clusters. As seen in (13) and (14), sibilant-sonorant clusters, except for [sr] that undergoes internal epenthesis only, show internal epenthesis only, external epenthesis only, or variation between the two depending on the language. In the previous section, we have assumed that word-initial sibilant-sonorant clusters involve two intensity rises, one from the silence to the sibilant and the other from the sibilant to the following sonorant. Therefore, both internal epenthesis and external epenthesis occur at an intensity rise, violating DEP-V/IR. The tableau in (35) illustrates constraint interaction in sibilant-sonorant clusters.

(35) Variable epenthesis sites in word-initial sibilant-sonorant clusters:
#↗S↗R → #SəR~#əSR

#↗S↗R	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. #SR	*!				
b. ☞ #SəR			*	*	
c. ☞ #əSR			*		*

Since candidate (35b) with internal epenthesis and candidate (35c) both violate DEP-V/IR, the universal ranking in (31) does not differentiate the two candidates. I assume that the epenthesis pattern is decided depending on the ranking of the other faithfulness constraints, CONTIGUITY and ANCHOR. In (35), the two constraints are freely ranked, and internal epenthesis and external epenthesis show a variation in the language. If CONTIGUITY outranks ANCHOR, external epenthesis is preferred, and if ANCHOR outranks CONTIGUITY, internal epenthesis is preferred.

Let us see how these constraints can derive the implicational universal in word-initial sibilant-initial clusters shown in (19). As previously mentioned, there are only a few languages that show the implicational relationship within a language. Table 3.9 below summarizes the epenthesis patterns in such languages.

Table 3.9: Summary of epenthesis patterns and languages in sibilant-initial clusters

language	#ST	#SN, #Sl	#Sr
Farsi	external	external	internal
Wolof	external	variation	internal
Egyptian Arabic	external	internal	internal
Uyghur	variation	internal	internal

First, in Farsi, external epenthesis occurs regardless of the type of post-sibilant consonant. External epenthesis in word-initial sibilant-stop clusters is prevalent cross-linguistically and is predicted by the universal ranking in (31). In particular, there is no intensity rise between the initial sibilant and the following stop, and thus internal epenthesis causes a critical violation of DEP-V/¬IR. On the other hand, we assume an intensity rise before the word-initial sibilant, and external epenthesis with a violation of DEP-V/IR, which is lower ranked than DEP-V/¬IR, is favored. This is demonstrated in the tableau in (36). Here the ranking between CONTIGUITY and ANCHOR is not relevant.

(36) External epenthesis in #ST in Farsi

#↗ST	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. #ST	*!				
b. #SəT		*!		*	
c. ☞ #əST			*		*

In sibilant-nasal and sibilant-lateral clusters, however, there are two intensity rises and the output form with internal epenthesis and the form with external epenthesis are the same in terms of intensity rise. External epenthesis in these clusters in Catalan can be explained by ranking CONTIGUITY over ANCHOR, as illustrated in (37).

(37) External epenthesis in #SN, #Sl in Farsi

#↗S↗N	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. #SN	*!				
b. #SəN			*	*!	
c. ☞ #əSN			*		*

In this case, both candidate (37b) with internal epenthesis and candidate (37c) with external epenthesis violate DEP-V/IR. Candidate (37c) with external epenthesis is chosen because CONTIGUITY outranks ANCHOR in this language and breaking the cluster is considered a more critical violation than having an edge epenthesis.

Next, in Wolof, external epenthesis occurs in sibilant-stop clusters but external epenthesis and internal epenthesis show lexical variation in sibilant-nasal and sibilant-lateral clusters. External epenthesis in sibilant-stop clusters is explained in the same way as in Farsi in (36), which is repeated below in (38).

(38) External epenthesis in #ST in Wolof

#↗ST	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. #ST	*!				
b. #SəT		*!		*	
c. ☞ #əST			*		*

The difference between Farsi and Wolof is that in Wolof, the ranking between CONTIGUITY and ANCHOR is not fixed, unlike in Farsi. Therefore, both internal epenthesis and external epenthesis can appear, as shown in (39).

(39) Variable epenthesis sites in #SN, #Sl in Wolof

#↗S↗N	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. #SN	*!				
b. ☞ #SəN			*	*	
c. ☞ #əSN			*		*

In Egyptian Arabic, external epenthesis occurs in sibilant-stop clusters as in Farsi and Wolof, but not only the epenthetic vowel but also the glottal stop [ʔ] is inserted before the epenthetic vowel because Egyptian Arabic does not allow onsetless syllables. This is illustrated in (40).

(40) External epenthesis in #ST in Egyptian Arabic

#↗ST	*CC	ONSET	DEP-V/¬IR	DEP-V/IR	DEP-?
a. #ST	*!				
b. #SəT			*!		
c. ☞ #?əST				*	*
d. #əST		*!		*	

Candidate (40b) with internal epenthesis fatally violates DEP-V/¬IR and is ruled out. Both candidtate (40c) and candidate (40d) involve external epenthesis, but candidtate (40d) is ruled out because it begins with a vowel and violates ONSET. Consequently, candidate (40c) with an epenthetic [?] and a vowel preceding the cluster becomes the winner, despite the additional violation of DEP-[?].

In sibilant-nasal and sibilant-lateral clusters, only internal epenthesis is observed in Egyptian Arabic. This suggests that in this language ANCHOR dominates CONTIGUITY, as opposed to Farsi in which ANCHOR is dominated by CONTIGUITY. The higher ranked ANCHOR constraint favors internal epenthesis, as demonstrated in (41).

(41) Internal epenthesis in #SN, #Sl in Egyptian Arabic

#↗S↗N	*CC	ONSET	DEP/¬IR	DEP/IR	ANCHOR	CONTIG
a. #SN	*!					
b. ☞ #SəN				*		*
c. #?əSN				*	*!	
d. #əSN		*!		*	*	

Again, both candidate (41b) with internal epenthesis and candidate (41c) with external epenthesis violate DEP-V/IR, and the universal ranking between DEP-V/¬IR and DEP-V/IR does not prefer one of them. However, candidate (41c) violates ANCHOR, which outranks CONTIGUITY violated by candidate (41b), and thus candidate (41b) with internal epenthesis is selected as optimal.

Since the universal ranking in (31) favors external epenthesis in sibilant-stop clusters by default, ANCHOR plays a role to derive internal epenthesis for the languages that allow internal epenthesis in sibilant-stop clusters as well as sibilant-nasal and sibilant-lateral clusters. In Uyghur, internal epenthesis and external epenthesis both can appear depending on the lexical item, whereas only internal epenthesis is observed in sibilant-nasal clusters. We assume that in Uyghur ANCHOR is ranked as high as DEP-V/¬IR, deriving variation between internal epenthesis and external epenthesis. This is demonstrated in (42).

(42) Variable epenthesis sites in #ST in Uyghur

#↗ST	*CC	DEP-V/¬IR	ANCHOR	DEP-V/IR	CONTIGUITY
a. #ST	*!				
b. ☞ #SəT		*			*
c. ☞ #əST			*	*	

In this case, candidate (42c) with external epenthesis at an intensity rise is as bad as candidate (42b) with internal epenthesis at a non-rise because ANCHOR violated by candidate (42c) outranks DEP-V/IR. Thus, both internal epenthesis and external epenthesis are possible.

The same ranking predicts internal epenthesis in sibilant-sonorant clusters, as shown in (43). Since both internal epenthesis and external epenthesis occur at an intensity rise, both candidate (43b) and candidate (43c) violate DEP-V/IR. Candidate (43c) is ruled out because it violates ANCHOR ranked higher than DEP-V/IR, and candidate (43b) is chosen as the winner.

(43) Internal epenthesis in #SN in Uyghur

#↗S↗N	*CC	DEP-V/¬IR	ANCHOR	DEP-V/IR	CONTIGUITY
a. #SN	*!				
b. ☞ #SəN				*	*
c. #əSN			*!	*	

To conclude, the asymmetry in sibilant-initial clusters can be accounted for by the universal ranking in (31) and the relevant ranking of CONTIGUITY and ANCHOR.

There are also clusters that involve no intensity rise in either possible epenthesis position. One type is word-final lateral-nasal clusters, which undergo internal epenthesis, and the other is word-final nasal-fricative clusters, which undergo external epenthesis. First, the word-final lateral-nasal cluster [lm] does not involve an intensity rise between [l] and [m], and the final [m] is not released in the example source languages in the database and thus does not create an intensity rise. Therefore, both internal epenthesis and external epenthesis trigger a violation of DEP-V/¬IR. Internal epenthesis is forced if ANCHOR outranks CONTIGUITY as seen earlier, and the ranking in (44 successfully derives internal epenthesis in [lm#].

(44) Possible constraint interaction for internal epenthesis in [lm#]

lm#	*CC	DEP-V/¬IR	DEP-V/IR	ANCHOR	CONTIGUITY
a. lm#	*!				
b. ☞ ləm#		*			*
c. lmə#		*		*!	

However, there is no motivation for ANCHOR outranking CONTIGUITY in the languages that show internal epenthesis in [lm#] clusters. Crucially, this ranking wrongly predicts internal epenthesis in word-final nasal-fricative clusters. In Uyghur, for example, 'film' is borrowed with internal epenthesis as [filim], and 'dance' is borrowed with external epenthesis as [tansa]. Under the ranking in which ANCHOR outranks CONTIGUITY as in (42) and (43), the candidate with internal epenthesis becomes the winner, which is counterfactual. This is demonstrated in (45), in which the black hand indicates the candidate selected by the constraint ranking and the normal hand in parenthesis indicates the attested output that is supposed to be selected.

(45) Failure to derive external epenthesis in NS# in Uyghur

NS#	*CC	DEP-V/¬IR	ANCHOR	DEP-V/IR	CONTIGUITY
a. NS#	*!				
b. ☛ NəS#		*			*
c. (☞) NSə#		*	*!		

I suggest a kind of IDENT constraint (McCarthy and Prince, 1995), which requires the releasedness of an occlusive in the input. This constraint, IDENT(release) in (46) is ranked higher than DEP-V/¬IR and plays a role in determining the epenthesis sites in word-final lateral-nasal and nasal-fricative clusters.

(46) IDENT(release):
Corresponding input and output segments have the same value of release.

IDENT(release) is violated either when a released stop/nasal in the input is unreleased in the output or when an unreleased stop nasal in the input is released in the output. In (47), the candidate with external epenthesis in (47b) fatally violates IDENT(release) because a prevocalic nasal is always released while the final nasal in the input does not involve an audible release. Thus, candidate (47a) with internal epenthesis is successfully selected as the winner.

(47) Internal epenthesis in lm# in Uyghur

lm˺#	*ID(rel)	DEP-V/¬IR	ANCHOR	DEP-V/IR	CONTIG
a. ☞ ləm˺#		*			*
b. lm⌜ə#	*!	*	*		

The current ranking can also derive external epenthesis in word-final nasal-fricative clusters, as shown in (48). Candidate (48a) with internal epenthesis violates the highest ranked IDENT(release) since the vowel inserted after the cluster-initial nasal releases the nasal, which is not audibly released in the input. So candidate (48b) with external epenthesis wins, even though it violates ANCHOR, which outranks CONTIGUITY in this language.

(48) External epenthesis in NS# in Uyghur

N˺S#	*ID(rel)	DEP-V/¬IR	ANCHOR	DEP-V/IR	CONTIG
a. N⌜əS#	*!	*			*
b. ☞ N˺Sə#		*	*		

Let us now see how the cluster-dependent asymmetries in epenthesis sites can be analyzed by the suggested constraint ranking. The first asymmetry that will be explained is the lateral-trill asymmetry in word-final clusters. Consonant clusters that end with the lateral [l] always undergo internal epenthesis, while clusters that end with the trill [r] show variation between internal and external epenthesis. As shown in (49), word-final stop-lateral clusters involve an intensity rise between the two consonants, and thus the form with internal epenthesis in (49b) violating the lower-ranked DEP-V/IR beats the form with external epenthesis in (49c) violating the higher-ranked DEP-V/¬IR. On the other hand, word-final stop-trill clusters in (50) have two intensity rises and equally violate DEP-V/IR, allowing both internal epenthesis and external epenthesis.

(49) Internal epenthesis in T⌜l#: T↗l# → Təl#

T↗l#	*CC	DEP-V/¬IR	DEP-V/IR	CONTINUITY	ANCHOR
a. Tl#	*!				
b. ☞ Təl#			*	*	
c. Tlə#		*!			*

(50) Variable epenthesis sites in T⌜r#: T↗r↗# → Trə#~Tər#

T↗r↗#	*CC	DEP-V/¬IR	DEP-V/IR	CONTINUITY	ANCHOR
a. Tr#	*!				
b. ☞ Tər#			*	*	
c. ☞ Trə#			*		*

When the liquid is cluster-initial, we have also seen an asymmetry in epenthesis sites. Whereas lateral-stop clusters always undergo internal epenthesis, trill-stop clusters show a variation between internal epenthesis and external epenthesis. Here we assume that the word-final stop is audibly released. The asymmetry in epenthesis sites comes again from the asymmetry in terms of intensity rise; there is no intensity rise after the [l], while there is an intensity rise after the [r]. As demonstrated in (51), internal epenthesis in candidate (51b) occurs at a non-intensity rise position, fatally violating DEP-V/¬IR. Candidate (51c), however, has a vowel inserted after the released stop, i.e., at an intensity rise, and becomes the winner. Unlike in lateral-stop clusters, there are two intensity rises in word-final trill-stop clusters, one between the trill and the stop and the other after the released stop. As shown in (52), thus, candidate (52b) with internal epenthesis and candidate (52c) with external epenthesis equally violate DEP-V/IR, and either one can appear depending on the ranking between CONTIGUITY and ANCHOR.

(51) External epenthesis in lT^{r}#: lT$^{r\nearrow}$# → lTə#

lT$^{r\nearrow}$#	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. lT#	*!				
b. ləT#		*!		*	
c. ☞ lTə#			*		*

(52) Variable epenthesis sites in rT^{r}#: r$^{\nearrow}$T$^{r\nearrow}$# → rəT#~rəT#

r$^{\nearrow}$T$^{r\nearrow}$#	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. rT#	*!				
b. ☞ rəT#			*	*	
c. ☞ rTə#			*		*

The present ranking also accounts for the sibilant-non-sibilant asymmetry in epenthesis sites, in which word-initial sibilant-sonorant clusters may show both internal and external epenthesis, whereas non-sibilant-sonorant clusters are always subject to internal epenthesis. As shown in (53), the forms with internal (53b) and external (53c) epenthesis in sibilant-sonorant clusters equally violate DEP-V/IR, without any other violations. The output with external epenthesis in non-sibilant-sonorant clusters (54c), however, violates higher-ranked DEP-V/¬IR, because low intensity of the non-sibilant fricative does not form an intensity rise in word-initial position, and thus the form with internal epenthesis (54b) wins.

(53) Epenthesis in #SR: #$^{\nearrow}$S$^{\nearrow}$R → #SəR~#əSR

#$^{\nearrow}$S$^{\nearrow}$R	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. #SR	*!				
b. ☞ #SəR			*	*	
c. ☞ #əSR			*		*

(54) Epenthesis in #FR: #F$^{\nearrow}$R → #FəR

#F$^{\nearrow}$R	*CC	DEP-V/¬IR	DEP-V/IR	CONTINUITY	ANCHOR
a. #FR	*!				
b. ☞ #FəR			*	*	
c. #əFR		*!			*

The next asymmetry to be explained is positional asymmetry in sibilant-initial clusters; sibilant-sonorant clusters show inter- or intra-language variation between internal and external epenthesis in word-initial position, but only undergo internal epenthesis in word-final position. The word-initial variation is already explained in (53). Internal epenthesis in word-final position is again explained by the universal ranking of DEP-V/¬IR over DEP-V/IR, as shown in (55); external epenthesis in (55c) violates DEP-V/¬IR, and thus internal epenthesis in (55b) is selected to be optimal.

(55) Epenthesis in SR#: S↗R# → SəR#

S↗R#	*CC	DEP-V/¬IR	DEP-V/IR	CONTIGUITY	ANCHOR
a. SR#	*!				
b. ☞ SəR#			*	*	
c. SRə#		*!			*

In this section, the intensity rise hypothesis about the position of epenthesis, laid out in section 3.2 and experimentally supported in section 3.3, has been formalized in the framework of Optimality Theory. The perceptual distinctness shown in section 3.3 was projected to the universal ranking of DEP-V constraints based on the P-map hypothesis, and the universal ranking successfully derives the typology of epenthesis sites in loan adaptation with the help of certain other faithfulness constraints, such as IDENT(release), CONTIGUITY, and ANCHOR.

Appendix 3: The list of borrowing languages

Abbreviations
T: stop, F: non-sibilant fricative, S: sibilant fricative, N: nasal: L: liquid, C: consonant, Ru: Russian, En: English, Ar: Arabic

Epenthesis in word-initial clusters

Epenthesis in stop-initial clusters

#TT

Language	*Example*	*Reference*
English	gdansk (Polish) → [gədænsk] city name	Riggs (2014)
Korean	gdansk (Polish) → [kɨədansɨkʰɨ] city name	my elicitation
Even	tkanʲ (Ru) → [tikan] 'fabric'	Choi (2010)
Evenki	tkanʲ (Ru) → [tikan] 'fabric'	Choi (2010)
Egyptian Arabic	gdansk (Polish) → [gedanisk] city name	my elicitation
Thai	kbaːl (Khmer) → [kabaːn] 'head'	Suthiwan (2009)

#TF

Language	*Example*	*Reference*
Farsi	kvadrat (Ru) → [kuwaːdraːt] 'types'	Bashiri (1994)

#TN

Language	*Example*	*Reference*
Alaskan Eskimo	knut (Ru) → [knuːtak]~[kənuːtaq] 'whip'	Hammerich (1954)
Even	kniga (Ru) → [kiniga] 'book'	Choi (2010)
Evenki	kniga (Ru) → [kiniga] 'book'	Choi (2010)
Dolgan	kniga (Ru) → [kiniga] 'book'	Stachowski (2010)
Kirghiz	kniʃka (Ru) → [kineʃke] 'book'	Gouskova (2001)
Uyghur	kniʃka (Ru) → [kiniʃkɛ] 'license'	my elicitation

#TL

Language	*Example*	*Reference*
Alaskan Eskimo	plita (Ru) → [plita]~[pəlitaq] 'stovetop'	Hammerich (1954)
Farsi	kliʃe (Ru) → [kəliːʃeh] 'cut'	Bashiri (1994)
Basque	cross/crucifix (En) → [gurutze]	Riggs (2014)
Tuvan	pl̪an (Ru) → [pɨlaan] 'plan'	Harrison (2000)
Tatar	klub (Ru) → [kəlub] 'club'	my elicitation
Uyghur	klub (Ru) → [kulub] 'club'	my elicitation
Even	klop (Ru) → [kolop] 'bedbug'	Choi (2010)
Evenki	klop (Ru) → [kəlapɨ]~[kelapɨ] 'bedbug'	Choi (2010)
Egyptian Arabic	plastic (En) → [bilastik]	Broselow (1992)
Bengali	gram (Sanskrit) → [geram] 'village'	Kar (2009)

Inuktitut	Clause (En) → [kalasi]	Pollard (2008)
Hindi	please (En) → [pɪlɪz]	Singh (1985)
Sinhalese	tribidə (Sanskrit) → [tirividə] 'triple'	Samarajiwa and Abeysekera (1964)
Ket	krʲuk (Ru) → [kuruk] 'hook'	Vajda and Nefedov (2009)
Wolof	klas (French) → [kalas] 'class'	Ka (1985)
Farsi	traffic (En) → [terafik]	Shademan (2002)
Aleut Eskimo	prʲistol (Ru) → [pɨlisɨtooluq] 'alter'	Koo (1980)
Uzbek	tramvaj (Ru) → [tiromvaj] 'tram'	Akiner (1997)
Dolgan	Prósʲka (Ru) → [poroskuo]	Stachowski (2010)
Indonesian	glas (Dutch) → [gelas] 'glass'	Tadmor (2009a)
Kirghiz	trupka (Ru) → [turupke] 'pipe'	Gouskova (2001)
Turkish	'prince' (En) → [pirɛns]	Beel and Felder (2013)

Epenthesis in non-sibilant fricative-initial clusters

#FT

Language	*Example*	*Reference*
Yakut	vdova (Ru) → [ogdo:bo] 'widow'	Pakendorf and Novgorodov (2009)

#FL

Language	*Example*	*Reference*
Yupik Eskimo	xlʲep (Ru) → [kɨliipaq] 'bread'	Koo (1982)
Farsi	xlʲep (Ru) → [xelab] 'bread'	Bashiri (1994)
Even	xlʲep (Ru) → [kilep] 'bread'	Choi (2010)
Evenki	xlʲep (Ru) → [kilep] 'bread'	Choi (2010)
Egyptian Arabic	floor (En) → [filoor]	Broselow (1992)
Hindi	fruit (En) → [fɪrut]	Singh (1985)
Aleut Eskimo	xlʲep (Ru) → [x1leebaq] 'bread'	Koo (1980)
Dolgan	xlʲep (Ru) → [kïläp] 'bread'	Stachowski (2010)
Kirghiz	frunze (Ru) → [boronzo] 'Frunze'	Gouskova (2001)
Tatar	vrac (Ru) → [vəratʃ] 'doctor'	my elicitation

Epenthesis in sibilant fricative-initial clusters

#ST

Language	*Example*	*Reference*
Aleut Eskimo	stakan (Ru) → [sitakkaanaq] 'glass cup'	Koo (1980)
Archi	stol (Ru) → [ustul] 'table'	Chumakina (2009)
Assamese	'school' (En) → [ɪskul]	Baishya (2010), reported in Dutta (2015)
Bengali	'school' (En) → [ɪʃkul]	Gouskova (2001)

Bezhta	ʃkola (Ru) → [iʃkola] 'school'	Comrie and Khalilov (2009)
Catalan	'stop' (En) → [əstɔp]	Bonet and Lloret (1998), ?
Central Pahari	'stool' → [istuːl]	Sharma (1980), reported in Broselow (1992)
Dolgan	skalo (Ru) → [yskaːla]	Stachowski (2010)
Egyptian Arabic	'study' (En) → [ʔistadi]	Broselow (1992)
Even	stol (Ru) → [ostol] 'table'	Choi (2010)
Evenki	stol (Ru) → [ostol] 'table'	Choi (2010)
Farsi	ʃkaf (Ru) → [ʃekap] / [ʔeʃkap] 'cupboard'	Kambuziya and Hashemi (2011)
Fula	staty (French) → [istati] 'statue'	PL 1997
Haitian Creole	'statue' (En) → [estati]	Tinelli 1981, Flei 2001
Hausa	'scale' (En) → [sakeli]	Han (2007)
Hindi	'school' (En) → [ɪskul]	Singh (1985)
Kazakh	stantsʲiya (Ru) → [ïstansa] 'train station'	
Kirghiz	stakan (Ru) → [ɯstakan] 'glass cup'	Gouskova (2001)
Indonesian	stri (Sanskrit) → [istri] 'woman'	Yuliati (2014)
	'stamp' (En) → [sətæmp]	
Manange	'school' (En) → [eskul]	Hildebrandt (2009)
Miami Spanish	'store' (En) → [estorear]	Fernández (1983)
Pahari	'school' (En) → [səkuːl]	Khan and Bukhari (2011)
Sinhalese	'school' (En) → [iskul]	Samarajiwa and Abeysekera (1964), Broselow (1992)
Spanish	'stop' (En) → [estop]	Gibson (2012)
Tatar	ʃkola (Ru) → [ʃikola] 'school'	my elicitation
	'ski' (En) → [iski]	
Telugu	'station' (En) → [isteʃənu]	Broselow (1992)
Turkish	'studio' (En) → [sʉtʉdiɔ]	Beel and Felder (2013)
	'skeleton' (En) → [iskelet]	
Tuvan	stakanʲ (Ru) → [ɨstaqaan] 'glass cup'	Harrison (2000)
Uyghur	statistika (Ru) → [istatistika]	my elicitation
	'Stockholm' → [sɨtokholom]	my elicitation
Uzbek	ʃkaf (Ru) → [iʃkop] 'cupboard'	Akiner (1997)
Wolof	'statue' (En) → [estati]	Ka (1985)
Yakut	stol (Ru) → [ostuol] 'table'	Pakendorf and Novgorodov (2009)
Yupik Eskimo	'school' (En) → [sɨkuulaq]	Koo (1982)
	ʃkola (Russian; 'school') → [iskuuluq]	

#SF

Language	*Example*	*Reference*
Alaskan Eskimo	svinka (Ru) → [sitinkaq] 'pig'	Hammerich (1954)
Dolgan	sverlo (Ru) → [sibärliä]	Stachowski (2010)

Even	svinʲja (Ru) → [tɕivindʲa] ‘pig’	Choi (2010)
Evenki	svʲetɕka (Ru) → [siveska] ‘candle’	Choi (2010)
Hindi	‘sphere’ (En) → [ɪsfiʌr]	Singh (1985)
Kirghiz	zveno (Ru) → [uzvana] ‘chain link’	Gouskova (2001)
Tatar	svəta (Ru) → [siviəta] female name (dim.)	my elicitation
Yakut	svinʲja (Ru) → [sibinɲa] ‘pig’	Pakendorf and Novgorodov (2009)
Yupik Eskimo	svinka (Ru) → [cɨtiinkaaq] ‘pig’ svʲitʲir (Ru) → [ɨswataq]/[ɨsvataq] ‘sweater’	Koo (1982)

#SN

Language	*Example*	*Reference*
Assamese	‘smile’ (En) → [ismail]	Baishya (2010), reported in Dutta (2015)
Uyghur	‘Smith’ (En) → [sɨmis]	my elicitation
Catalan	‘snack’ (En) → [əznak]	BL, Fleischhacker
Kazakh	smena (Ru) → [ismen] ‘change’ smorodina (Ru) → [simorodina] ‘currant’ smat (Ru) → [simat]~[ismat] proper name	Sulejmenova (1965) Fleischhacker (2001)
Spanish	‘smoking’ (En) → [esmoquin]	Gibson (2012)
Tatar	smat (Ru) → [səmat]~[əsmat] proper name	my elicitation
Uzbek	Smolny (Ru) → [ismolnin] proper name	Akiner (1997)
Wolof	‘smoke’ (En) → [esmok] ‘smoking’ (En) → [somokɪŋ]	Fleischhacker (2001)

#Sl

Language	*Example*	*Reference*
Aleut Eskimo	ʃlʲapa (Ru) → [ʃiʎaabaq] ‘hat’	Koo (1980)
Assamese	‘slate’ (En) → [silet]	Dutta (2015)
Bengali	‘slate’ (En) → [ʃelet]	Mahato (1974), reported in Broselow (1992)
Catalan	‘slam’ (En) → [əzlam]	Bonet & Lloret (1998), reported in Fleischhacker (2001)
Egyptian Arabic	‘slide’ (En) → [silayd] ‘slide’ (En) → [ʔeslajd]	Broselow (1992) my elicitation
Farsi	‘Slav’ → [eslav] ʃlaŋ (Ru) → [ʃelaŋ] ‘hose’	Kambuziya and Hashemi (2011)
Hindi	‘slipper’ (En) → [sɪlɪpʌr]	Singh (1985)
Indonesian	slot (Dutch) → [selot] ‘lock’	Tadmor (2009a)
Spanish	‘slogan’ (En) → [eslogan]	Gibson (2012)
Wolof	‘Slovakia’ → [solovaki]	Fleischhacker (2001)
Yupik Eskimo	ʃlʲapa (Ru) → [islaapaq] ‘hat’	Koo (1982)

#Sr

Language	*Example*	*Reference*
Egyptian Arabic	'Sri Lanka' → [seri lanka]	my elicitation
Farsi	'Sri Lanka' → [seri laŋka]	Fleischhacker (2001)
Indonesian	srigala → [sedrigala] 'wolf'	Tadmor (2009a)
Wolof	'Sri Lanka' → [siri laŋka]	Fleischhacker (2001)

Epenthesis in nasal-initial clusters

#NT

Language	*Example*	*Reference*
English	mbira (Shona) → [əmbirə]	my elicitation
Korean	mbira (Shona) → [ɨmbira]	my elicitation
Malagasy	mbe (Bantu) → [òmby] 'ox'	Adelaar (2009)

#NF

Language	*Example*	*Reference*
Korean	nʒamena (French) → [ɨncamena] capital of Chad	my elicitation

#NN

Language	*Example*	*Reference*
Kirghiz	mnemonicheskij (Ru) → [ymnemonicheskij] 'mnemonic'	Gouskova (2001)
Korean	mnemosine (Greek) → [mɨnemosine] nnaʒi (Bantu) → [ɨnnaci] 'Nnaji'	my elicitation

#NL

Language	*Example*	*Reference*
Bengali	mlan (Sanskrit) → [melan] 'gloomy'	Kar (2009)
English	mladic (Serbian) → [məladitʃ]	my elicitation
Egyptian Arabic	mraz (En) → [maraz]	my elicitation

Epenthesis in word-final clusters

Epenthesis in stop-final clusters

TT#

Language	*Example*	*Reference*
Indonesian	al-sabt (Ar) → [sabtu] 'Saturday'	Tadmor (2009a)
Kazakh	abyekt (Ru) → [äbyekti] 'object'	Krippes (1994)
Korean	'Egypt' (En) → [iciptʰɨ]	my elicitation
Tatar	'Egypt' (En) → [igipta]	my elicitation

ST#

Language	*Example*	*Reference*
Dolgan	most (Ru) → [mosta] 'floor'	Stachowski (2010)
Even	most (Ru) → [mosta] 'floor'	Choi (2010)
Evenki	most (Ru) → [mosta] 'floor'	Choi (2010)
Egyptian Arabic	'list' (En) → [lesta]	my elicitation
Irish	'toast' (En) → [tosta]	Riggs (2014)
Yakut	most (Ru) → [muosta] 'floor'	Pakendorf and Novgorodov (2009)
Tatar	'list' (En) → [lista]	my elicitation

NT#

Language	*Example*	*Reference*
Arabic	'lamp' (En) → [lamba]	Versteegh et al. (2006)
Ewenki	bank (Ru) → [banka] 'bank'	Choi (2010)
Kanuri	'bank' (En) → [bánkì]	Löhr et al. (2009)
Korean	'paint' (En) → [pʰɛintʰɨ]	my elicitation
Uyghur	bank (Ru) → [banka] 'bank'	my elicitation
Yakut	vint (Ru) → [biːnte] 'screw'	Pakendorf and Novgorodov (2009)

lT#

Language	*Example*	*Reference*
Even	ʃolk (Ru) → [xolka] 'silk'	Choi (2010)
Evenki	ʃolk (Ru) → [ʃɔlkɔ] 'silk'	Choi (2010)
Ket	ʃolk (Ru) → [solga] 'silk'	Vajda and Nefedov (2009)
Tatar	ʃolk (Ru) → [ʃelki] 'silk'	my elicitation
Yakut	ʃolk (Ru) → [solko] 'silk'	Pakendorf and Novgorodov (2009)

rT#

Language	*Example*	*Reference*
Nubian	darb (Ar) → [darub] 'path'	Versteegh et al. (2006)
Tatar	park (Ru) → [parka] 'park'	my elicitation
Yakut	park (Ru) → [paarka] 'park'	Vasilyeva (2010)

Epenthesis in fricative-final clusters

TS#

Language	*Example*	*Reference*
Georgian	'paradox' (En) → [paradoksi]	Riggs (2014)
Tuvan	gips (Ru) → [giipis] 'cast'	Harrison (2000)
Uyghur	qudsˤ (Ar) → [qutus] 'holy'	my elicitation

NS#

Language	*Example*	*Reference*
Indonesian	'dance' (En) → [dansa]	my elicitation
Uyghur	ʒins (Ar) → [ʒinsi] 'gender'	my elicitation
Tatar	'France' (En) → [fransi]	my elicitation

rF#/rS#

Language	*Example*	*Reference*
Nubian	ðˤarf (Ar) → [zarif] 'envelope'	Versteegh et al. (2006)
	dars (Ar) → [dérs], [dérís] 'lesson'	
Tatar	xarf (Ar) → [xäref] 'letter'	Poppe (1968)
	borʃ (Ru) → [borəʃ]~[borʃi]	my elicitation
Tuvan	farʃ (Ru) → [faariʃ] 'ground meat'	Harrison (2000)
Uyghur	dars (Ar) → [deris] 'lesson'	my elicitation

Epenthesis in nasal-final clusters

TN#

Language	*Example*	*Reference*
Assamese	'cotton' (En) → [koton]	Dutta (2015)
Indonesian	ħukm (Ar) → [hokum] 'law'	Tadmor (2009a)
Tatar	ritm (Ru) → [ritəm] 'rhythm'	my elicitation
Uyghur	ħukm (Ar) → [hokum] 'law'	Versteegh et al. (2006)

FN#

Language	*Example*	*Reference*
Portuguese	rahn (Ar) → [refem] 'hostage'	Versteegh et al. (2006)

SN#

Language	*Example*	*Reference*
Assamese	'prism' (En) → [prijɔm]	Dutta (2015)
Bulgarian	komunizm (Ru) → [komunisəm] 'communism'	my elicitation
Kirghiz	komunizm (Ru) → [komunisɯm] 'communism'	my elicitation
Indonesian	ʒism (Ar) → [ʒisim] 'body'	Tadmor (2009b)
Tatar	ʒism (Ar) → [ʒisəm] 'body'	my elicitation

LN#

Language	*Example*	*Reference*
Hebrew	'film' (En) → [filim]	Cohen (2009)
Indonesian	'film' (En) or film (Dutch) → [filim]/[filem]	
Korean	'film' (En) → [tʰillim]	my elicitation
Sinhalese	'film' (En) → [pilm]/[pilim]	Samarajiwa and Abeysekera (1964)
Uyghur	'film' (En) → [filim]	my elicitation

Epenthesis in [l]-final clusters

Tl#

Language	*Example*	*Reference*
Arabic	'cable' (En) → [keːbil]	Versteegh et al. (2006)
Assamese	'bubble' (En) → [babul]	Dutta (2015)
Bengali	'people' (En) → [pɪupol]	Islam (2004)
Even	spektaklʲ (Ru) → [spektakəl] 'play'	Choi (2010)
Egyptian Arabic	'shuttle' (En) → [ʃatəl]	my elicitation
Indonesian	ʕaql (Ar) → [akal] 'wisdom'	Tadmor (2009a)
Javanese	ʕaql (Ar) → [akal] 'wisdom'	Versteegh et al. (2006)
Moroccan Arabic	imæbl (French) → [mubəl] 'building'	Versteegh et al. (2006)
Pahari	'cycle' (En) → [saɪkʊl]	Khan and Bukhari (2011)
Tatar	korablʲ (Ru) → [karabɨl] 'ship'	my elicitation
Uyghur	ʕaql (Ar) → [äqɨl] 'wisdom'	Lindblad
Uzbek	ʃakl (Ar) → [ʃakil] 'form'	Versteegh et al. (2006)

Fl#

Language	*Example*	*Reference*
Indonesian	'waffle' (En) → [wafel]	my elicitation
Nubian	sahl (Ar) → [sahal] 'easy'	Versteegh et al. (2006)
Dutch	buffle (French) → [buffel] 'buffalo'	van der Sijs (2009)

Nl#

Language	*Example*	*Reference*
Bengali	'personal' (En) → [parsonal]	Islam (2004)
Bulgarian	kreml^j (Ru) → [kreməl] 'the Kremlin'	my elicitation
Kirghiz	kreml^j (Ru) → [kɯremɯl] 'the Kremlin'	my elicitation
Tatar	kreml^j (Ru) → [kreməl] 'the Kremlin'	my elicitation

Epenthesis in [r]-final clusters

Tr#

Language	*Example*	*Reference*
Even	tigr (Ru) → [t^{j}igər] 'tiger'	Choi (2010)
Indonesian	fikr (Ar) → [fikir] 'opinion'	Tadmor (2009a)
Javanese	s^{ʕ}abr (Ar) → [sabar] 'patience'	Versteegh et al. (2006)
Turkish	qabr (Ar) → [kabir] 'grave'	Jaklin Kornfilt, p.c.
Uyghur	fikr (Ar) → [fikir] 'opinion'	my elicitation

Fr#

Language	*Example*	*Reference*
Javanese	ð^{ʕ}uhr (Ar) → [luhur] 'midday'	Versteegh et al. (2006)
Nubian	baħr (Ar) → [bahar] 'ocean'	Versteegh et al. (2006)
Uyghur	siħr (Ar) → [sihir] 'magic'	my elicitation

Sr#

Language	*Example*	*Reference*
Nubian	mas^{ʕ}r (Ar) → [masr]/[masir] 'Egypt'	Versteegh et al. (2006)
Uyghur	mis^{ʕ}r (Ar) → [misir] 'Egypt'	my elicitation

Nr#

Language	*Example*	*Reference*
Uyghur	ʕamr (Ar) → [ømør] 'age'	my elicitation

Chapter 4

Sonority sequencing in consonant cluster adaptation and perception

In the previous chapters, I have provided an analysis of the epenthesis typology and consonant cluster perception in terms of auditory factors, intensity rise and C_1 voicing. There is, however, an alternative analysis based on the sonority sequencing of the consonant clusters. Previous research on vowel epenthesis in consonant cluster adaptation and of illusory vowel perception in consonant clusters has crucially relied on the sonority profile of the cluster. So the purpose of this chapter is to assess whether the sonority-based hypotheses can explain the typology of vowel epenthesis in loan adaptation (Section 4.1) as well as the current experimental results reported in Chapter 2 (Section 4.2).

4.1 Sonority-based analyses of epenthesis sites

Previous studies, such as Gouskova (2001), Steriade (2006), and Flemming (2008), have used the sonority profile between the two consonants in the cluster to account for the cluster-dependent asymmetry in epenthesis sites. Although they all employ sonority profile for the analysis, they have different points of view about how sonority profile plays a role in vowel epenthesis; Gouskova (2001) argues that epenthesis sites are determined so that the resulting form is not phonologically marked in terms of syllable structure, whereas Steriade (2006) and Flemming (2008) assume that it is the faithfulness of the sonority profile that determines the epenthesis sites.

4.1.1 Markedness-based account

One markedness-based view attributes this asymmetry between sibilant-stop and stop-sonorant clusters to the inherent unsplittability of sibilant-stop clusters (Selkirk, 1982; Broselow, 1992, a.o.). That is, sibilant-stop clusters are considered a single constituent sharing a linked laryngeal node, while stop-sonorant clusters are not. For this structural reason, sibilant-stop clusters cannot be broken by an epenthetic vowel and have to undergo external epenthesis, whereas stop-sonorant clusters can undergo internal epenthesis allowing a split by the epenthetic vowel. This view, however, cannot cover all the cases of vowel epenthesis in sibilant-initial clusters, as well as the present typology of vowel epenthesis including the cases of external epenthesis in clusters that are not sibilant-stop clusters, such as nasal-nasal clusters. There are cases, such as Indonesian and Hindi, in which internal epenthesis and external epenthesis show a variation in repairing sibilant clusters, and it is not plausible that some of these sibilant-stop clusters form a contituent while the others do not. In addition, external epenthesis in sibilant-sonorant clusters, which could be in free variation with internal epenthesis, needs an explanation not based on the internal structure.

Gouskova (2001) employs a markedness constraint based on sonority contour. Gouskova's (2001) analysis of epenthesis sites in Russian loanwords in Kirghiz adopts the Syllable Contact Law (Murray and Vennemann, 1983; Vennemann, 1988), a cross-linguistic observation that sonority tends to fall across a syllable boundary, as a markedness constraint which bans a sonority rise across syllable boundaries, defined in (56).

(56) SYLLABLE CONTACT: Sonority must not rise across a syllable boundary.

Gouskova argues that an epenthetic vowel is placed so that the resulting sequence does not violate SYLLABLE CONTACT. For example, a cluster with rising sonority /#bl.../ undergoes internal epenthesis [#bəl...] because if it undergoes external epenthesis, the resulting form [#əb.l...], having a sonority rise across a syllable boundary, will violate SYLLABLE CONTACT. In contrast, a cluster with flat/falling sonority such as /#st.../ undergoes external epenthesis because even though neither internal epenthesis [#sə.t...] nor external epenthesis [#əs.t...] involve a sonority rise across syllable boundaries, internal epenthesis additionally violates the faithfulness constraint CONTIGUITY by breaking the input cluster. This can be derived when SYLLABLE CONTACT outranks CONTIGUITY, as illustrated in the tableaus in (57) and (58).

(57) Non-rising sonority clusters → external epenthesis

#sta	SYLLABLE CONTACT	CONTIGUITY
a. #səta		*!
b. ☞ #əsta		

(58) Rising sonority clusters → internal epenthesis

#bla	SYLLABLE CONTACT	CONTIGUITY
a. ☞ #bəla		*
b. #əbla	*!	

The SYLLABLE CONTACT account of the cluster-dependent asymmetry in epenthesis sites is also not sufficient to explain the entire typology, particularly when considering epenthesis patterns in word-final position. As seen earlier, the word-final lateral-nasal cluster [lm] undergoes internal epenthesis (e.g., 'film' → [filim] (Uyghur)), even though external epenthesis would not result in a sonority rise across the syllable boundary (*[fil.mi]). In Kirghiz, the word-final nasal-nasal cluster also undergoes internal epenthesis (e.g., 'gimn' → [gimun]), as seen in the previous chapter, and the ranking between SYLLABLE CONTACT and CONTIGUITY suggested for epenthesis in word-initial clusters does not derive the correct epenthesis pattern in word-final position, as shown in (59).

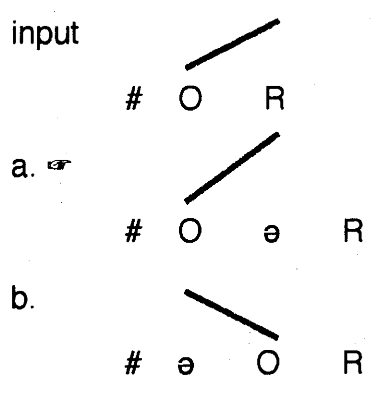
(59) Word-final non-rising sonority sonorant-sonorant clusters → internal epenthesis

gimn#	SYLLABLE CONTACT	CONTIGUITY
a. (☞) gimun#		*!
b. ☛ gimnu#		

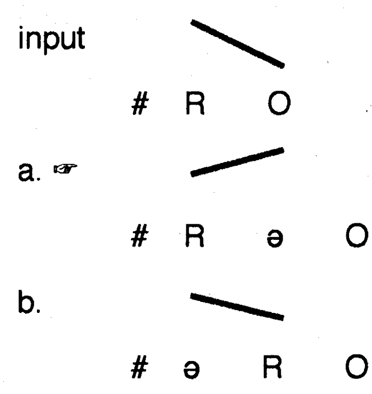
4.1.2 Faithfulness-based account

Steriade (2006) and Flemming (2008) adopt the notion of sonority contour in formulating faithfulness constraints, based on perceptual similarity between clusters and epenthesis forms, proposed by Fleischhacker (2001, 2005). Recall that Fleischhacker (2001, 2005) suggests a hierarchy in consonant cluster splittability such that stop-sonorant (TR) clusters are most likely to be split by an epenthetic vowel, followed by sibilant-liquid (SL), sibilant-nasal (SN), and sibilant-stop (ST) clusters in order, showing the implicational relationship mentioned in the previous chapter. She argues that this hierarchy originates from perceptual similarity between the cluster and the resulting form with internal epenthesis, based on experimental evidence. That is, #TR and #TəR are perceptually most similar to each other, and #ST and #SəT are the least. Building on this, Steriade (2006) suggests faithfulness constraints requiring correspondence for sonority contour between the input cluster and the epenthesized output, arguing that epenthesis sites are determined so that the sonority contour of the input cluster is perceptually similar to that of the output sequence with an epenthetic vowel. This is illustrated in (60) and (61), in which the blue line indicates the sonority contour between the two consonants. O stands for obstruent and R stands for sonorant. An obstruent-sonorant cluster, such as /#bl.../, involves a sonority rise, and its corresponding form with internal epenthesis [#bəl...] also involves a sonority rise, which is slightly sharper. So the change from the shallow sonority rise in /#bl/ into the steep sonority rise in [#bə] is perceptually small, resulting in internal epenthesis. In a clusters with level or falling sonority, such as /#mb.../, however, the sonority fall in /#mb/ is perceptually very different from the sonority rise in [#məb] with internal epenthesis, and thus internal epenthesis is not likely in clusters involving a sonority plateau or a sonority fall.

(60) Sonority contour correspondence in word-initial obstruent-sonorant clusters



(61) Sonority contour correspondence in word-initial sonorant-obstruent clusters



Flemming (2008) develops a metric to calculate the sonority contour; abstract terms such as shallow or steep sonority rise are replaced with a specific ratio derived by dividing the sonority contour in the input consonant cluster C_1C_2 by the contour in the output sequence of the initial consonant and the following epenthetic vowel C_1ə, based on the sonority scale given in (3), which is repeated in (62) below. By subtracting the ratio from 1, the more similar the corresponding sonority contour is, the smaller distance value it receives. Internal epenthesis is preferred in the clusters having a small sonority rise distance to the clusters having a large distance, by ranking a faithfulness constraint requiring preservation of a larger distance value over the one requiring preservation of a smaller value.

(62) Sonority hierarchy (e.g., Clements, 1990)

obstruents	nasals	liquids	glides	vowels
1	2	3	4	5

← *less sonorous* *more sonorous* →

The faithfulness analyses based on sonority contour also cannot explain the full range of epenthesis data presented in the previous chapter. The unexplainable cases are not specifically problems of the faithfulness-based account but are problems of any sonority-based account in general. I discuss these problems in the next section.

4.1.3 General limitations of sonority-based accounts

The hypothesis based on sonority contour, regardless of whether it employs markedness or faithfulness, does not seem to be sufficient to explain the whole range of the epenthesis data in loan adaptation. First of all, not all clusters involving the same sonority profile pattern together in epenthesis. For example, both word-initial stop-stop clusters and nasal-nasal clusters involve level sonority, but the nasal-nasal clusters may be repaired with external epenthesis (e.g., Kirghiz) or with internal epenthesis (e.g., Korean), as seen in (20), while stop-stop clusters are always repaired with internal epenthesis cross-linguistically, as shown in (9a).

Second, the same sonority class may also show different epenthesis patterns. In the case of liquids, the epenthetic vowel is always positioned before the word-final lateral [l] as shown in (9b) (e.g., korabl^{j} (Russian; 'ship') → [karabəl] (Tatar)), whereas the vowel may go before or after the final trill [r] as shown in (22) (e.g., qabr (Arabic; 'grave') → [kabre] (Uyghur) vs. [kabir] (Turkish)), even though both the lateral and the trill are classified as liquids. This cannot be fixed by refining the sonority hierarchy such that it differentiates laterals from trills. For example, in Parker's (2002) sonority hierarchy, laterals are more sonorous than trills and flaps. However, it does not change the fact that the sonority is rising in both lateral-final and trill-final clusters. In other words, it cannot be explained why only trill-final clusters may allow external epenthesis, even though both lateral-final and trill-final clusters are of rising sonority. If one assumes that the steeper the sonority rise is, the more likely internal epenthesis is, it still does not explain why then nasal-final clusters, which involve a smaller sonority rise than trill-final clusters, always show internal epenthesis.

A similar asymmetry is found in fricatives too; word-initially, clusters beginning with a non-sibilant fricative always have the epenthetic vowel after the fricative, as shown in (9a), whereas clusters beginning with a sibilant fricative may undergo either internal or external epenthesis. In terms of sonority, sibilant and non-sibilant fricatives are not differentiated, and thus the sonority profiles of sibilant-initial and non-sibilant-initial clusters are identical, but their behaviors in epenthesis are different from each other.

Third, even the same cluster in a language may behave differently in terms of epenthesis depending on its position in the word. The adaptation of sibilant-nasal clusters in Tatar is a good example. As presented earlier, in Tatar, epenthesis in word-initial sibilant-nasal clusters shows variation; for example, both [əsmat] and [səmat] are used for Russian *smat* (proper name). The same cluster, however, always undergoes internal epenthesis word-finally, e.g., ʒism (Arabic; 'body') → [ʒisəm], and external epenthesis, e.g., *[ʒismə], is not possible. Not only in Tatar but also in the typology, word-initial sibilant-sonorant clusters show a variation in epenthesis sites in several languages, but there is no language that shows the variation in word-final sibilant-sonorant clusters.

Fourth, the same type of cluster may be adapted with different epenthesis patterns in the same language depending on its phonetic realization. For instance, word-final stop-stop clusters $T_1T_2\#$ from English are borrowed into Korean with a single external epenthesis (e.g., 'compact' → [k^{h}əmp^{h}ɛkt^{h}i]), whereas $T_1T_2\#$ clusters in Russian are optionally borrowed with double epenthesis, i.e., both internal and external epenthesis (e.g., *relikt* (Russian; name of a vocal trio) → [lellikt^{h}i]~[lellik^{h}it^{h}i]).

Hence, I conclude that the sonority contour hypothesis is not sufficient to cover the entire typology of epenthesis sites in loan adaptation. The current hypothesis, however, can explain the aforementioned problematic cases. The first asymmetry in clusters with the same sonority profile originates from the asymmetry in the presence of intensity rise. In word-initial stop-stop clusters, the initial stop is always released creating an intensity rise, and thus epenthesis takes place at the intensity rise, i.e., internal epenthesis. In contrast, word-initial nasal-nasal clusters may involve an audible vocalic release or not, as mentioned previously. So depending on the phonetic realization of the initial nasal in the cluster, either internal epenthesis or external epenthesis can occur, as demonstrated in (??) and (??) in the previous chapter.

The intensity rise hypothesis can also account for the epenthesis asymmetry in the same sonority classes, liquids and fricatives. For the liquids, in [l]-final clusters, the intensity is rising from the preceding consonant to the [l], and is falling from the [l] to the post-[l] silence. Therefore, internal epenthesis takes place at the intensity rise. In trill-final clusters, in contrast, we have two intensity rises, one between the preceding consonant and the trill and the other after the trill, and we observe both internal epenthesis and external epenthesis (cf. (49), (50)). Recall also that the asymmetry between sibilants and non-sibilants is also explained by the difference in intensity rise. In sibilant-sonorant clusters, we assume an additional intensity rise from the silence to the sibilant, and external epenthesis, as well as internal epenthesis, is possible. For non-sibilant-sonorant clusters, intensity rise from silence to [f] is not sufficient to create an intensity rise, so there is one intensity rise between the initial [f] and the following sonorant, and only internal epenthesis occurs (cf. (53), (54)).

Also, the intensity rise hypothesis has accounted for the asymmetry in which sibilant-sonorant clusters show a variation in epenthesis sites in word-initial position but not in word-final position in (53) and (55). That is, word-finally, there is only one intensity rise between the sibilant and the following sonorant and internal epenthesis occurs, whereas word-initially there is an additional rise in intensity from the silence to the word-initial sibilant allowing both internal epenthesis and external epenthesis.

Lastly, the donor asymmetry between English and Russian stop-stop clusters may be attributed to different phonetic realizations of the cluster in each language. That is, in English word-final stop-stop clusters, the first stop may be unreleased, but in Russian stop-stop clusters, the first stop is always clearly released (cf. Kochetov, 2008). This indicates that in English stop-stop clusters there is only one intensity rise at the release of the final stop, but Russian stop-stop clusters involve two intensity rises, one between the two stops and the other at the release of the final stop.

In summary, we have seen several cluster-dependent asymmetries in epenthesis sites that cannot be explained by the sonority-based hypothesis. The asymmetries stem from phonetic properties of the clusters, and the intensity rise hypothesis can successfully provide an explanation for them.

4.2 Sonority profile in non-native cluster perception

4.2.1 Sonority sequencing in non-native cluster perception

It has been argued that the sonority profile of consonant clusters has an influence on consonant cluster perception. Specifically, the likelihood of perceptual epenthesis may differ depending on the type of consonant cluster. Berent et al. (2007, 2008, 2009, *et seq.)* argue that the likelihood of perceptual epenthesis is determined by the universal markedness of onset clusters, particularly by the sonority profile of the cluster. The markedness is based on the Sonority Sequencing Principle (SSP). According to the SSP, a good syllable involves sonority increasing throughout the onset to the nucleus, and decreasing from the nucleus and throughout the coda. Berent et al. conducted several tasks (syllable count, AX discrimination, transcription), across speakers of different native languages, English, Korean, and Spanish, who have some or no relevant clusters in their native languages. The results showed that universally less marked clusters with rising sonority, e.g., *blif*, are more accurately discriminated from their matching forms with epenthesis, e.g., *bəlif*, than universally more marked clusters with level sonority, e.g., *bdif* from *bədif*, which in turn are more accurately discriminated than even more marked clusters with falling sonority, e.g., *lbif* from *ləbif*. Based on these results, Berent et al. argue that speakers possess universal linguistic preferences regarding sonority, and extend them to non-native cluster perception.

Several questions arise, however, as to whether the perception of a variety of consonant clusters can be simply generalized in terms of the three sonority profiles. First of all, several different clusters belong to each sonority profile, and their phonetic characteristics may not be classified together as the same category in perception. For example, consonant clusters with falling sonority in Berent et al.'s (2007, 2008) stimuli include clusters beginning with the lateral /l/, the trill /r/, the flap /ɾ/, and the nasal /m/. The accuracy of the perception of these clusters, however, is likely to differ from one another. As for the liquids, as mentioned and experimentally proved earlier, laterals that do not involve an intensity rise would be more accurately perceived than trills and flaps that involve an intensity rise. Also for nasals, if the nasal [m] is audibly released, its perception would be less accurate than when the [m] is unreleased. Crucially, Berent et al's results conflict with the current experimental results. The current results suggest that a consonant cluster C_1C_2 and its epenthesis form are more confusable with each other when intensity rises from C_1 to C_2, whereas Berent et al. suggest that they are less confusable when sonority rises in the cluster. Since the phonetic correlate of sonority is intensity, these results stand in contradiction. In addition, voicing of the first consonant in clusters, which also turned out to have a significant effect on the consonant cluster perception in Chapter 2, is not considered in the sonority-based hypothesis. To sum up, the sonority hypothesis assumes that the more marked the cluster is, the more likely epenthesis is, whereas the current hypothesis assumes that the more similar the cluster is to the epenthesis form, the more likely epenthesis is. The notions of "marked" and "similar" may not agree with each other. The next section will investigate the contributions of phonological factors such as the SSP as well as phonetic factors such as intensity rise and C_1 voicing in non-native cluster perception.

4.2.2 Sonority effects in the current experiments

This section reanalyzes the results of Experiments 1, 2, and 3, the current experiments on non-native cluster perception, in terms of sonority sequencing, and see whether the same sonority effect observed in the previous studies is found in the current study too. The sonority profiles were coded based on the sonority hierarchy suggested by Clements (1990) in (62), because it was used in the previous studies of sonority sequencing in non-native clusters (e.g., Berent et al. 2007, et seq., Daland et al., 2011), as well as because it is the best sonority hierarchy that works for all available studies.

4.2.2.1 Experiment 1

Recall that Experiment 1 was an AX discrimination task between a word-initial cluster CC and its matching sequence with internal epenthesis CəC.

4.2.2.1.1 Korean listeners

Figures 4-1 and 4-2 show the percentage correct by sonority profile and by sonority distance, respectively. It is shown that clusters with falling sonority are perceived with the highest accuracy. These show the lowest accuracy in Berent et al.'s (2008) results, who also used Korean participants. Clusters with rising sonority, which were best discriminated in Berent et al., showed a slightly lower mean percentage correct than clusters with falling sonority and a slightly higher percentage correct than clusters with level sonority. A mixed effects logistic regression model was fit to the percentage correct, in which the fixed effect was sonority profile and the random effects were participant and cluster. Sonority profile was treatment-coded with sonority plateau as the baseline. It turned out that the difference in percentage correct between sonority fall clusters and sonority plateau clusters was statistically significant ($z=3.98$, $p<.001$), and so was the difference between sonority plateau clusters and sonority rise clusters ($z=2.04$, $p<.05$). The latter difference is consistent with Berent et al.'s, even though it is a much smaller difference compared to Berent et al.'s results, but the former difference between sonority fall and sonority plateau is opposed to Berent et al.'s results. Another logistic regression model with sonority profile coded with sonority rise as the baseline showed that the difference between sonority fall clusters and sonority rise clusters was not significant, which was also not expected by the sonority hypothesis. Sonority distance differentiates clusters with falling sonority of −2 (e.g., [lk]) and of −1 (e.g., [mk]), as shown in Figure 4-2, and there was no significant difference in the accuracy between the two.

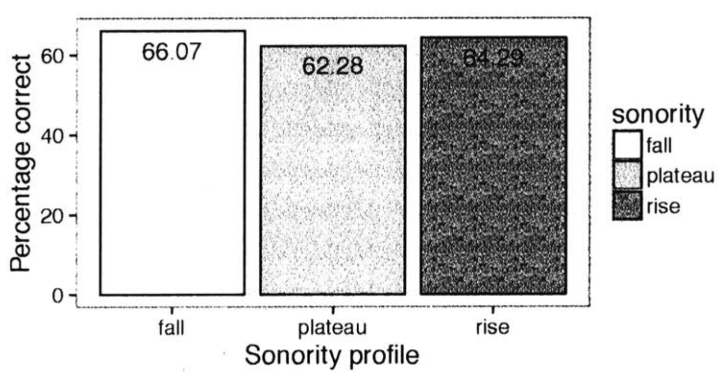


Figure 4-1: Percentage correct by sonority profile (Korean listeners)

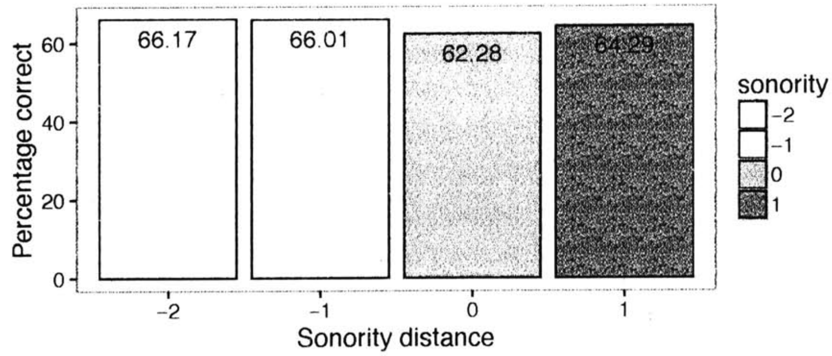


Figure 4-2: Percentage correct by sonority distance (Korean listeners)

4.2.2.1.2 Chinese listeners

Now we analyze the results for the Chinese listeners in terms of sonority profile and sonority distance. Figure 4-3 shows a completely opposite pattern compared to Berent et al.'s studies; the percentage correct was the highest in clusters with falling sonority, followed by clusters with level sonority, and was the lowest in clusters with rising sonority. However, a mixed effects logistic regression model showed that no difference between clusters was statistically significant. Figure 4-4 presents the difference in the accuracy between clusters with falling sonority of −2 and of −1, which was also not significant.

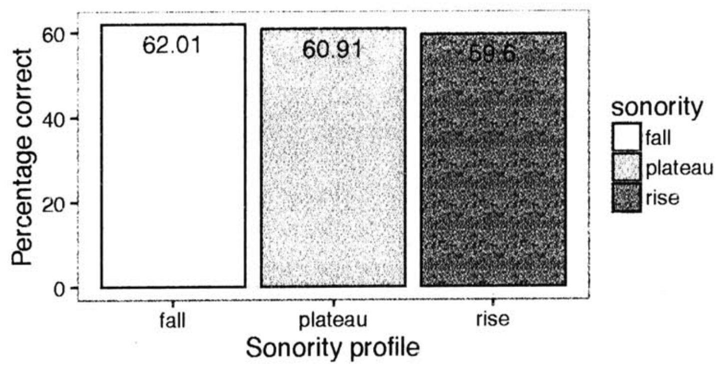


Figure 4-3: Percentage correct by sonority profile (Chinese listeners)

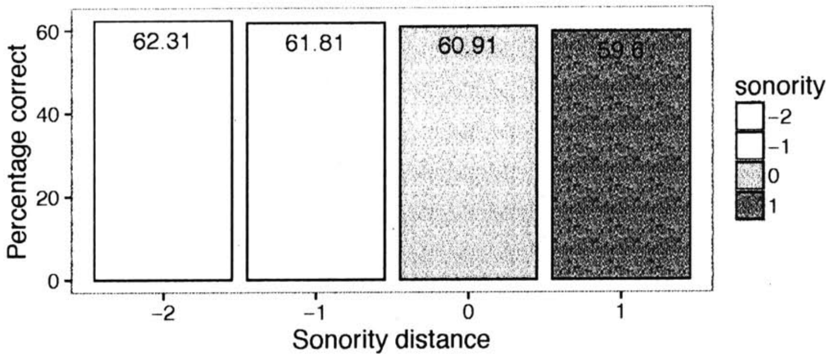


Figure 4-4: Percentage correct by sonority distance (Chinese listeners)

Summing up, the results of Experiment 1 did not show the expected sonority effect. The only predicted difference was found between clusters with rising sonority and clusters with level sonority for Korean participants, but the results for Korean participants also included a significant difference between clusters with falling sonority and clusters with level sonority in the opposite direction. This is totally unexpected by Berent et al.'s previous research, which also conducted an AX discrimination task using stimuli recorded by a Russian speaker and employed Korean and Mandarine Chinese speakers as participants.

4.2.2.2 Experiment 2

Experiment 2 was also an AX discrimination task but used more varied types of clusters for stimuli. Two speaker groups, speakers of Korean and English participated in the experiment.

4.2.2.2.1 Korean listeners

Figures 4-5 displays percentage correct results for each sonority profile. The percentage correct was the highest in clusters with falling sonority, contrary to the sonority hypothesis. A mixed effects logistic regression model, however, showed that there were no significant differences in the accuracy between the sonority profiles.

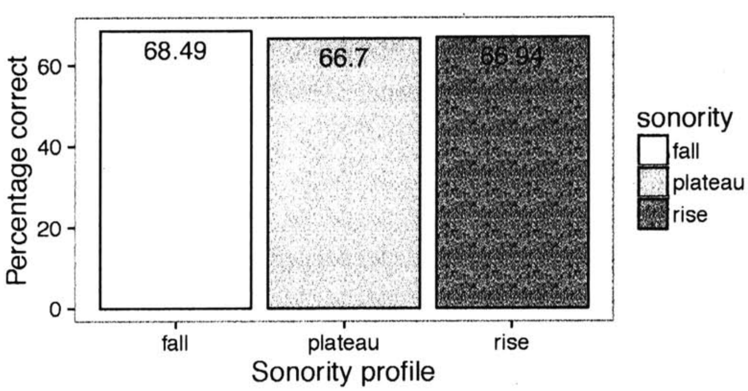


Figure 4-5: Percentage correct by sonority profile (Korean listeners)

Figures 4-6 illustrates percentage correct results for each sonority distance. Again, no significant differences were found between the sonority distances.

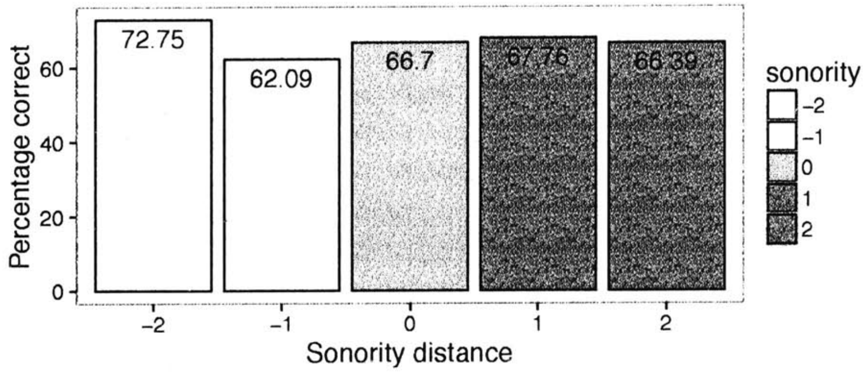


Figure 4-6: Percentage correct by sonority distance (Korean listeners)

4.2.2.2.2 English listeners

Here are results for English listeners. As in the analysis in Chapter 2, the results of native clusters in English were excluded because participants were significantly better at discriminating native clusters than non-native clusters. Figures 4-7 and 4-8 illustrate the percentage correct by sonority profile and by sonority distance, respectively. Although we see small differences in the percentage correct between sonority profiles and between sonority distances, mixed effects logistic regression models showed that no differences were statistically significant.

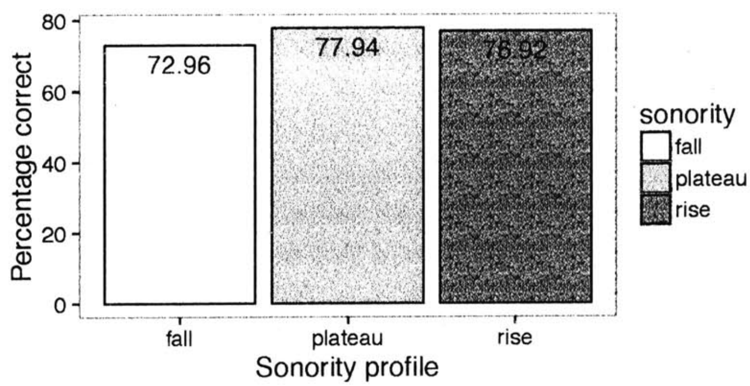


Figure 4-7: Percentage correct by sonority profile (English listeners, non-native clusters only)

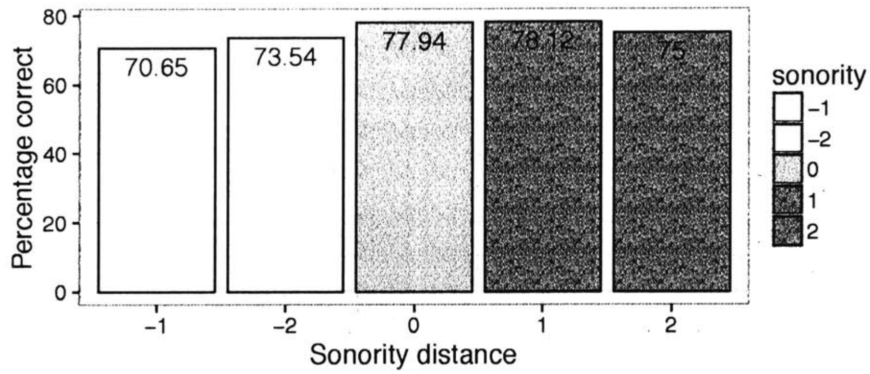


Figure 4-8: Percentage correct by sonority distance (English listeners, non-native clusters only)

In Experiment 2, it is shown that sonority profile does not make any significant differences in the accuracy of consonant cluster discrimination. This is again unexpected given the sonority hypothesis and contradicts previous results in which sonority profile was a significant predictor for the accuracy of discrimination between consonant clusters and epenthesis forms.

4.2.2.3 Experiment 3

Recall that Experiment 3 was an AX discrimination on word-final clusters and their matching forms with internal epenthesis with Korean listeners. There has been no discrimination experiment on word-final consonant clusters, to my knowledge, but the same sonority hypothesis can apply to word-final clusters. According to the SSP, falling sonority is the most desirable in coda clusters, followed by flat sonority and rising sonority, exactly opposite to the sonority markedness in onset clusters. Therefore, we can assume the opposite hypothesis of sonority profile in word-final consonant cluster perception; the accuracy of discrimination between word-final clusters and their matching forms with epenthesis would be higher in clusters with falling sonority than in clusters with level sonority, which in turn shows a higher accuracy than clusters with rising sonority. The percentage correct results by sonority profile are presented in Figure 4-9. It is shown that the percentage correct was the highest in clusters with a sonority plateau, which was slightly higher than in clusters with a sonority fall. The difference between these two was not statistically significant. On the other hand, the participants were significantly less accurate in clusters with a sonority rise than clusters with a sonority plateau/fall ($p<.001$), which partially confirmed the hypothesis.

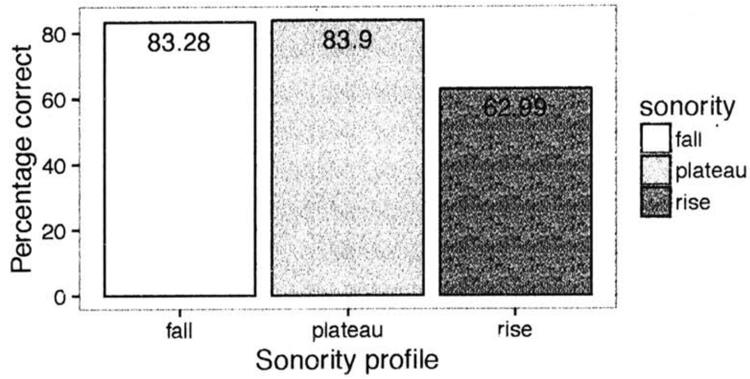


Figure 4-9: Percentage correct by sonority profile

Figure 4-10 shows the percentage correct results by sonority distance. Here we see that the percentage correct was highest in clusters with a sonority distance of −1, but there were no statistical differences between sonority distances of −2, −1 and 0. As in Figure 4-9, clusters with rising sonority, i.e., sonority distances of 1 and 2, showed significantly less accuracy than the others, and there was no difference between the two sonority distances.

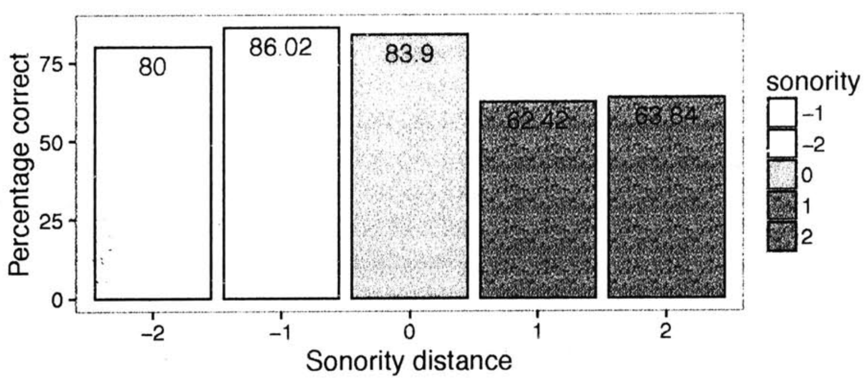


Figure 4-10: Percentage correct by sonority distance

In sum, the sonority hypothesis is only partially confirmed in Experiment 3 between clusters with sonority rises and clusters with sonority non-rises. No significant difference was observed between clusters with sonority falls and with sonority plateaus.

It is shown that in the current discrimination experiments the sonority hypothesis pursued and confirmed in the previous studies does not hold. Some differences were consistent with the prediction of the sonority hypothesis, but there were also other differences that contradicted the hypothesis and most of the differences were statistically insignificant.

4.2.3 Deriving sonority effects

Then why do we get different results from Berent et al.'s (2007, 2008, 2009, et seq.)? I believe that this is because sonority profiles may be confounded with the acoustic factors that were found to affect consonant cluster perception, i.e., intensity rise and C_1 voicing, in the previous experiments. This means that the results of Berent et al.'s experiments are just one possible result we may get from some combinations of initial consonant clusters, and thus some combinations of the stimuli with different phonetic realizations may derive the same results as in Berent et al., but other combinations may not.

Let us take a closer look at the stimuli used in Berent et al. (2007, 2008). The stimuli were recorded by a native speaker of Russian. First, 30 clusters with a sonority rise were composed of 18 voiceless stop-sonorant clusters and 12 voiced stop-sonorant clusters, all of which involve an intensity rise. For 30 sonority plateau clusters, 18 clusters were combinations of voiceless stops and 12 were combinations of voiced stops. Since the first stop in word-initial stop-stop clusters is always audibly released, all of the clusters with flat sonority involve an intensity rise too. Thirty clusters with falling sonority consisted of 15 trill/tap-stop clusters, 7 lateral-obstruent clusters, and 8 nasal-stop clusters. All the trill/tap-initial clusters must have involved an intensity rise and all the lateral-initial clusters must have involved no intensity rise within the cluster, but it is not clear how many initial nasals in the nasal-stop clusters were audibly released and involved an intensity rise within the cluster Focusing on C_1 voicing, the ratio of voiceless consonant to voiced consonant in C_1 position was 3:2 in the clusters with rising sonority and with flat sonority, and all C_1 consonants were voiced in the clusters with falling sonority. We have already seen in Chapter 2 that when there is a rise in intensity, the cluster and the epenthesis form is significantly more confusable when beginning with a voiced consonant than when beginning with a voiceless consonant. So the reason for the higher discriminability of clusters with sonority rises in Berent et al. might be because of the many voiceless consonant initial clusters, although this does not explain the difference between sonority rise and plateau. Also, we found out in the current experiment that clusters with a rise in intensity were more confusable with their epenthesis forms than clusters with no rise in intensity when the C_1 is voiced, which is the case for the clusters with falling sonority here. Although there was no information about phonetic realizations of nasal-initial clusters, which makes their intensity rise factor unavailable to us, we can still expect that there would be more clusters with a rise in intensity (15 trill-initial + α nasal-initial) than clusters without a rise in intensity (8 lateral-initial + β nasal-initial). So by hypothesis, the low discrimination accuracy of clusters with sonority falls in Berent et al. may come from a greater number of clusters with an intensity rise.

In order to see whether this is true, out of my own stimuli used in Experiment 2, I constructed a subset, which had a similar distribution of clusters to Berent et al.'s stimuli, in terms of intensity rise and C_1 voicing. The clusters included in Subset 1 are listed in Table 4.1.

Table 4.1: Clusters used in Subset 1

#	sonority	C_1 voiced	C_1 voiceless
5	rise	#g⌜m, #g⌜l	#k⌜m, #k⌜l, #k⌜r
2	plateau	#g⌜d	#k⌜p
12	fall	#lb, #lk, #ls, #m⌝f, #m⌝k, #m⌜k, #rf, #rg, #rk, #rs, #rv, #rz	

For clusters with rising sonority, two clusters beginning with the voiced stop [g] and three clusters beginning with the voiceless stop [k] were used. No fricative-initial clusters with rising sonority were included because such clusters were not employed in Berent et al.'s (2007, 2008) experiments. Two stop-stop clusters, one voiced and the other voiceless, were included in the subset. In fact, the ratio of clusters with a voiced C_1 to clusters with a voiceless C_1 stood at 2:3 in Berent et al.'s stimuli, just as in the sonority rise clusters. Since these two clusters, [#g⌜d] and [#k⌜p], were the only sonority plateau clusters used in Experiment 2, the C_1 voiced-voiceless ratio in Subset 1 was not identical with but not too different from Berent et al.'s . Next, clusters with falling sonority consisted of three nasal-obstruent clusters, three lateral-obstruent clusters, and six trill-obstruent clusters. This shows a very similar distribution to the stimulus clusters with falling sonority in Berent et al. (2007, 2008), although information about the releasedness of the cluster-initial nasal was not available and I included two clusters with an unreleased nasal and one cluster with a released nasal at random. Percentage correct results of Subset 1 by the sonority profile are illustrated in Figure 4-11.

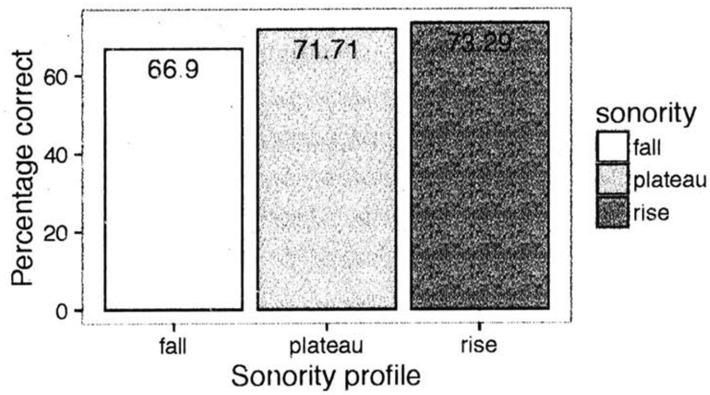


Figure 4-11: Percentage correct by sonority profile (Subset 1)

Here we see the same tendency in the proportion of accuracy observed in Berent et al's experiments. Namely, consonant clusters with a sonority rise showed the highest percentage correct, followed by clusters with a sonority plateau, and clusters with a sonority fall showed the lowest, indicating that clusters with a sonority fall were the least discriminable from the epenthesis, and clusters with a sonority rises were the best discriminable. The same mixed effects logistic regression model that was run in Section 4.2.2 was fit to the percentage correct results in Subset 1, and the difference was statistically significant only between the sonority falls and rises ($p<.01$). So we can say that the anaylsis of Subset 1 replicated Berent et al.'s results albeit partially.

Next, I created another subset, consisting of the consonant clusters listed in Table 4.2. This subset was intended to have all intensity rise clusters for sonority fall clusters and all voiceless-consonant-initial clusters for sonority rise clusters, expecting that the accuracy decreases in the falling sonority clusters under the effect of intensity rise and increases in the rising sonority clusters under the effect of C_1 voicing, while maintaining the same sonority profiles. All of the clusters were taken from the stimuli used in Experiment 2, as in Subset 1.

Table 4.2: Clusters used in Subset 2

#	sonority	C_1 voiced	C_1 voiceless
6	rise		#k^{r}m, #k^{r}l, #fm, #fl, #sm, #sl
2	plateau	#g^{r}d	#k^{r}p
6	fall	#rf, #rg, #rk, #rm, #rv, #rz	

Subset 2 involves clusters with a sonority rise, all of which began with a voiceless obstruent, and clusters with a sonority fall, all of which began with a trill, involving an intensity rise. The clusters with a sonority plateau were the same as in Subset 1. It is predicted in this subset that the accuracy in sonority rise clusters is higher than that in Subset 1, because the clusters all began with a voiceless C_1 that turned out to make the cluster more discriminable from the epenthesis form than a voiced C_1. Conversely, for clusters with falling sonority, it is predicted that the accuracy in Subset 2 is poorer than that in Subset 1, because trill-initial clusters involve an intensity rise, which makes the cluster perceptually more similar to the epenthesis form. The accuracy results by sonority profile are presented in Figure 4-12.

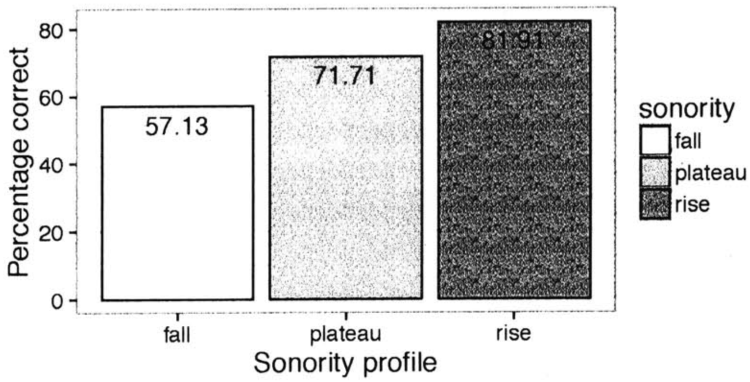


Figure 4-12: Percentage correct by sonority profile (Subset 2)

As predicted, the accuracy in sonority rise clusters became higher, and the accuracy in sonority fall clusters became lower than that in Subset 1. This results in more evident differences in percentage correct between the sonority profiles than those in Subset 1, showing the same tendency as in Berent et al. All the differences here were statistically significant ($p<.001$).

Lastly, I constructed Subset 3, using the clusters given in Table 4.3. Again, all the clusters were the same ones used in Experiment 2. This subset consists of all falling sonority clusters without an intensity rise and all rising sonority clusters beginning with a voiced consonant.

Table 4.3: Clusters used in Subset 3

#	sonority	C_1 voiced	C_1 voiceless
7	rise	#g˹m, #g˹l, #vm, #vr, #zm, #zl, #zr	
2	plateau	#g˹d	#k˹p
6	fall	#lb, #ls, #lk, #lm, #lv, #lz	

In Subset 3, all the clusters with rising sonority began with a voiced obstruent, and it is thus predicted that the accuracy in these clusters is lower than in sonority rise clusters in Subset 1, which included both clusters beginning with a voiced C_1 and clusters beginning with a voiceless C_1, and than in sonority rise clusters in Subset 2, which consisted of clusters beginning with a voiceless C_1 only. Clusters with falling sonority in Subset 3 consisted solely of lateral-initial clusters, featuring no intensity rise. Therefore, it is expected that the accuracy in these lateral-initial clusters in Subset 3 will be higher than the accuracy in clusters with falling sonority in Subset 1, which involved nasal-initial and trill-initial clusters as well as lateral-initial clusters, and than in clusters with falling sonority in Subset 2, which were all trill-initial. There was no change in clusters with a sonority plateau. Figure 4-13 shows the accuracy results of Subset 3.

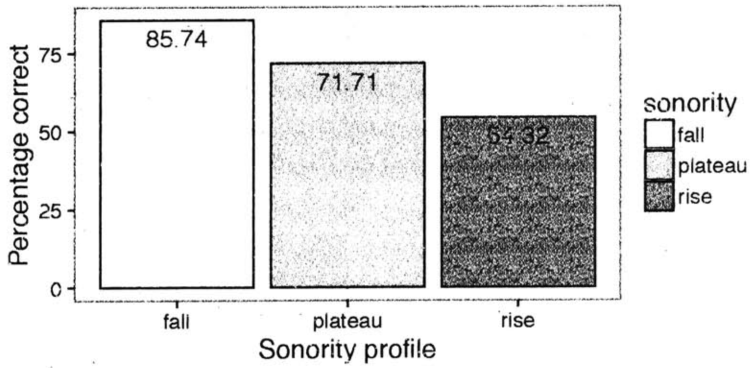


Figure 4-13: Percentage correct by sonority profile (Subset 3)

Now we see the reverse pattern, in which clusters with a sonority fall were the most discriminable from the epenthesis forms, followed by clusters with a sonority plateau, and clusters with a sonority rise were the least discriminable. All these differences were statistically highly significant ($p<.001$). This suggests that the sonority effect in non-native cluster perception is one possible result that derives from one possible combination of particular consonant clusters, not the one that holds universally.

Unlike the sonority profiles, which showed different results depending on the subset, we see a consistent effect of intensity rise regardless of the subsets. As illustrated in Figure 4-14, both in Subset 1 and in Subset 3, the participants discriminated clusters with no intensity rise better than clusters with an intensity rise, from the epenthesis forms, at a statistically significant level ($p<.001$). There were no clusters without an intensity rise in subset 2.

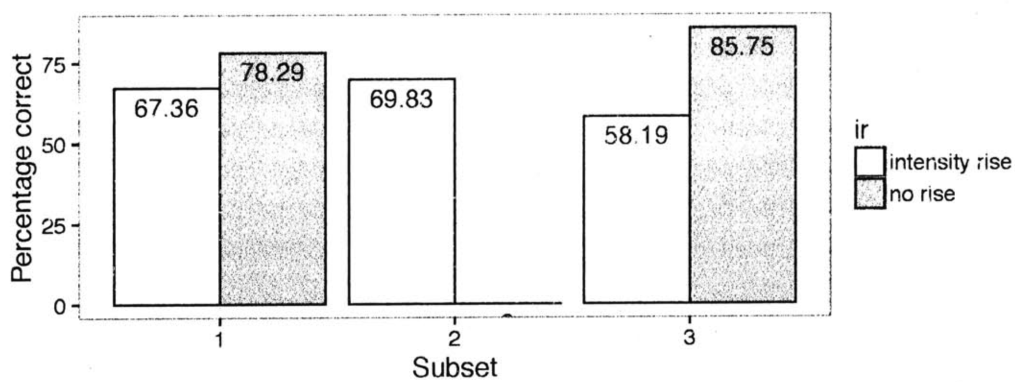


Figure 4-14: Percentage correct by intensity rise

Also, across the subsets, the effect of C_1 voicing was significant. As shown in 4-15, in all three subsets, clusters beginning with a voiced consonant were harder to discriminate from the epenthesis forms than clusters beginning with a voiceless consonant. The difference was statistically highly significant in every subset ($p<.001$).

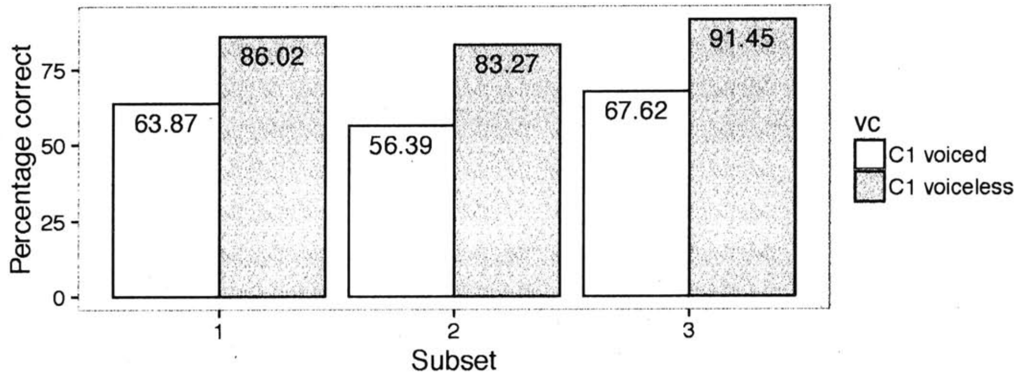


Figure 4-15: Percentage correct by C_1 voicing

In this chapter, we have examined the effect of sonority profile in vowel epenthesis in loan-word adaptation and in non-native cluster perception. First, the larger typology of loan vowel epenthesis described in this dissertation is not fully covered by the sonority-based hypothesis because it includes several asymmetries observed in the same sonority profiles, the same sonority classes, and even the same consonant clusters. Those asymmetries can be explained when the intensity rise and C_1 voicing are taken into account. Illusory vowel epenthesis in non-native cluster perception, which has been explained based on the sonority profile, is also best explained by the auditory factors, i.e., intensity rise and C_1 voicing. The results of current perception experiments involving stimuli that systematically vary in the sonority profile, intensity rise and C_1 voicing show that in discrimination between a consonant cluster CC and its matching epenthesis form CəC, the significant predictors are intensity rise and C_1 voicing, and sonority profile is not a significant predictor. An apparent effect of the SSP, like the one reported in the previous studies, can arise if sonority profile is confounded with intensity rise and C_1 voicing.

Chapter 5

Conclusion

This dissertation has examined cluster-dependent asymmetries in vowel epenthesis in loanword adaptation and in non-native cluster perception. In Chapter 2, I discussed two auditory factors that crucially affect the relative perceptual similarity between consonant clusters and their corresponding vowel epenthesis forms: intensity rise and C_1 voicing. The working hypothesis was that consonant clusters that involve a rise in intensity or begin with a voiced consonant are perceptually more similar to the corresponding epenthesis forms than clusters that involve no rise in intensity or begin with a voiceless consonant. A series of discrimination experiments, which employed stimuli including various types of consonant clusters in both word-initial and word-final positions, largely confirmed this hypothesis.

In Chapter 3, I provided novel generalizations about epenthesis sites in consonant cluster adaptation, based on an extensive cross-linguistic survey of loanwords including a variety of clusters both word-initially and word-finally. It was shown that the cluster-dependent asymmetries in epenthesis sites are best explained by the auditory property of consonant clusters, i.e., intensity rise. Specifically, if a cluster involves an intensity rise inside the cluster, epenthesis occurs inside the cluster; if a cluster involves an intensity rise outside the cluster, epenthesis occurs outside the cluster; and if a cluster involves two intensity rises, either internal or external epenthesis can occur. I proposed that this is because the epenthetic vowel insertion where there is an intensity rise makes a perceptually less salient change from the original cluster than epenthesis where there is no intensity rise, based on the P-map hypothesis (Steriade, 2008) that an output involving a perceptually smaller change is more optimal. The results of the similarity judgment and transcription experiments supported this hypothesis.

In Chapter 4, I reviewed the alternative analyses of cluster-dependent asymmetries in vowel epenthesis and in non-native cluster perception based on the sonority profile. It was shown that the sonority-based accounts fail to generalize to the full set of data. For the epenthesis site typology in loanword adaptation, I pointed out several asymmetries that cannot be explained relying on the sonority profile. Clusters with the same sonority profile or with the same sonority class may behave differently in loan epenthesis across languages and even within the same language. Also, even the same consonant cluster may show different epenthesis patterns depending on the position in the word or on the phonetic realization of the cluster in the source language. For non-native cluster perception, the results of my perception experiments involving the same clusters or the same sonority categories with different phonetic realizations in terms of the auditory properties, intensity rise and C_1 voicing, showed that it is the auditory factors that play a more important role in non-native cluster perception than the sonority profile. The difference between my results and the results of the previous research could be due to the fact that the sonority profile variable was confounded in their experiments with different phonetic realizations.

There are, of course, remaining questions. First, there were several results in the results of the perception experiments that cannot be fully explained by the current hypothesis. There may be other auditory factors that affect the vowel similarity of the cluster, which will need further investigation. Also, although the epenthesis data in this dissertation is limited to loanwords and second language acquisition, epenthesis patterns in L1 phonology and first language acquisition are possibles area in which to further explore the cluster-dependent asymmetries and to see whether the current auditory hypothesis holds.

Bibliography

Adelaar, A. (2009). *Malagasy.* Max Planck Institute for Evolutionary Anthropology, Leipzig.

Akiner, S. (1997). Survey of the lexical influence of Russian on Modern Uzbek (1870-1990). *The Slavonic and East European Review*, 75(1):1–35.

Baertsch, K. (1998). Onset sonority distance constraints through local conjunction. In Gruber, M. C., Higgins, D., Olson, K. S., and Wysocki, T., editors, *CLS 34, Part 2: The Panels*, pages 1–15. Chicago Linguistic Society, Chicago.

Baertsch, K. and Davis, S. (2001). Turkic C+/l/(uster) phonology. In *CLS 37: The Main Session. Papers from the 37th Meeting of the Chicago Linguistic Society*, pages 29–43.

Bashiri, I. (1994). Russian loanwords in Persian and Tajiki languages. In Marashi, M., editor, *Persian Studies in North America: Studies in Honor of Mohammad Ali Jazayery*, pages 109–140. Ibex Publishers, Inc.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Becker, M. and Levine, J. (2013). Experigen – an online experiment platform. Available at http://becker.phonologist.org/experigen.

Beel, R. and Felder, J. (2013). Phonological adaptations of English loanwords in Turkish. In *Other Undergraduate Scholarship.*

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., and Smolensky, P. (2008). Language universals in human brains. *PNAS*, 105:5321–5325.

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

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

Bharati, S. (1994). *Aspects of the Phonology of Hindi and English.* Arnold, New Delhi.

Boersma, P. and Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. version 6.0. retrieved 28 October 2015 from http://www.praat.org/.

Broselow, E. (1983). Nonobvious transfer: On predicting epenthesis errors. In Gass, S. and Selinker, L., editors, *Language Transfer in Language Learning*, pages 269–280. Newbury House, Rowley, MA.

Broselow, E. (1992). Transfer and universals in second language epenthesis. In Gass, S. M. and Selinker, L., editors, *Language Transfer in Language Learning*, pages 71–86. John Benjamins, Amsterdam.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24:209–244.

Chitoran, I., Goldstein, L., and Byrd, D. (2002). Gestural overlap and recoverability: articulatory evidence from Georgian. In Gussenhoven, C. and Warner, N., editors, *Laboratory Phonology 7*, pages 419–447. de Gruyter, Berlin.

Choi, M. (2010). The Russian loanwords in Ewen and Ewenki. *Altai Hakpo*, 20:1–40.

Chumakina, M. (2009). *Archi*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In Kingston, J. and Beckman, M., editors, *Papers in Laboratory Phonology 1: Between the grammar and physics of speech*, pages 283–333. Cambridge University Press, New York.

Cohen, E.-G. (2009). *The role of similarity in phonology: Evidence from loanword adaptation in Hebrew*. PhD thesis, Tel Aviv University.

Comrie, B. and Khalilov, M. (2009). *Bezhta*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., and Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, 28(2):197–234.

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

Davis, S. (1998). Syllable Contact in Optimality Theory. *Korean Journal of Linguistics*, 23(2):181–211.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25(6):1568–1578.

Durvasula, K. and Kahng, J. (2016). The role of phrasal phonology in speech perception: What perceptual epenthesis shows us. *Journal of Phonetics*, 54:15–34.

Dutta, H. (2015). Constraints, phonotactics and loan words: Evidence from Assamese phonology. Presented at Harvard, May 14, 2015.

Fernández, R. G. (1983). English loanwords in Miami Cuban Spanish. *American Speech*, 58(1):13–19.

Fleischhacker, H. (2001). Cluster-dependent epenthesis asymmetries. In Albright, A. and Cho, T., editors, *UCLA Working Papers in Linguistics*, volume 7, pages 77–116.

Fleischhacker, H. (2005). *Similarity in Phonology: Evidence from Reduplication and Loan Adaptation*. PhD thesis, University of California, Los Angeles.

Flemming, E. (2008). Asymmetries between assimilation and epenthesis. Manuscript, MIT.

Fry, D. B. (1979). *The Physics of Speech. Cambridge Textbooks in Linguistics.* Cambridge University Press, Cambridge.

Gibson, M. (2012). A gestural-based analysis of /e/ prosthesis in word-initial /sc/ loanwords in Spanish. *Revista Philologica Romanica*, 12:35–56.

Gouskova, M. (2001). Falling sonority onsets, loanwords, and syllable contact. In Andronis, M., Ball, C., Elston, H., and Neuvel, S., editors, *CLS 37: The Main Session. Papers from the 37th Meeting of the Chicago Linguistic Society*, volume 1, pages 175–186.

Hammerich, L. L. (1954). The Russian stratum in Alaskan Eskimo. *Word*, 10(4):401–428.

Han, K.-I. (2007). Vowel epenthesis in hausa loanwords. *The Journal of Linguistic Science*, 41:133–147.

Harrison, K. D. (2000). *Topics in phonology and morphology of Tuvan.* PhD thesis, Yale University.

Hildebrandt, K. (2009). *Manange.* Max Planck Institute for Evolutionary Anthropology, Leipzig.

Hirano, A. (1965). *A linguistic study of the assimilation of English loanwords into Japanese.* PhD thesis, The University of British Columbia.

Howitt, A. W. (2000). *Automatic syllable detection for vowel landmarks.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Hwang, J. (2011). *Non-native perception and production of foreign sequences.* PhD thesis, Stony Brook University.

Islam, A. (2004). L1 influence on the spoken English proficiency of Bengali speakers. C-Essay in English, Hogskolan Dalarna University.

Jany, C., Gordon, M., Nash, C. M., and Takahara, N. (2007). How universal is the sonority hierarchy?: A cross-linguistic acoustic study. In Trouvain, J. and Barry, W. J., editors, *Proceedings of the International Congress of Phonetic Sciences 16*, pages 1401–1404. Saarbrucken.

Ka, O. (1985). Syllable Structure and Suffixation in Wolof. *Studies in the Linguistic Sciences*, 15(1):61–90.

Kambuziya, A. K. Z. and Hashemi, E. S. (2011). Russian loanword adaptation in Persian; optimal approach. *Iranian Journal of Applied Language Studies*, 3(1):77–96.

Kang, Y. (2003). Perceptual similarity in loanword adaptation: English post-vocalic word-final stops to Korean. *Phonology*, 20(2):219–273.

Kar, S. (2009). *The syllable structure of Bangla in Optimality Theory and its application to the analysis of verbal inflectional paradigms in Distributed Morphology.* PhD thesis, Tubingen University.

Kara, D. S. (2003). *Kyrgyz.* Lincom Europa, Muenchen.

Khan, A. Q. and Bukhari, N. H. (2011). Phonological adaptation of English loan words in Pahari. *Language in India*, 11(1).

Knoblauch, K. (2014). psyphy: Functions for analyzing psychophysical data in R. https://CRAN.R-project.org/package=psyphy.

Kochetov, A. (2008). Phonology and phonetics of loanword adaptation: Russian place names in Japanese and Korean. *Toronto Working Papers in Linguistics*, 28:159–174.

Koo, J. H. (1980). On russian loanwords in aleut eskimo. *Russian Language Journal*, 34(117):83–100.

Koo, J. H. (1982). Russian loanwords in Yupik Eskimo. *Gengo Kenkyu*, 82:91–105.

Krippes, K. A. (1994). Russian loan-words and orthography in the post-soviet Turkic languages. *General Linguistics*, 34(2):107–116.

Ladefoged, P. (1975). *A course in phonetics*. Harcourt Brace Jovanovich, New York.

Lavoie, L. M. (2000). *Phonological patterns and phonetic manifestations of consonant weakening*. PhD thesis, Cornell University.

Löhr, D., Wolff, H. E., and Awagana, A. (2009). *Kanuri*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

MacMillan, N. A. and Creelman, C. D. (2005). *Detection theory: A user's guide*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2nd edition edition.

McCarthy, J. and Prince, A. (1995). Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory*, volume 249–384, Amherst, MA. University of Massachusetts, GLSA.

Murray, R. and Vennemann, T. (1983). Sound change and syllable structure in Germanic phonology. *Language*, 59:514–528.

Padgett, J. (2002). Russian voicing assimilation, final devoicing, and the problem of [v] (or, The mouse that squeaked). Ms. Santa Cruz.

Pakendorf, B. and Novgorodov, I. (2009). *Sakha*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Parker, S. (2002). *Quantifying the sonority hierarchy*. PhD thesis, University of Massachusetts, Amherst, Amherst, MA.

Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics*, 36:55–90.

Phillips, D. P., Hall, S. E., and Boehnke, S. E. (2002). Central auditory onset responses, and temporal asymmetries in auditory perception. *Hearing Research*, 167:192–205.

Pollard, M. G. (2008). Epenthetic vowel quality in English loanwords in Inuktitut. Master's thesis, University of Toronto.

Poppe, N. (1968). *Tatar Manual: Descriptive Grammar and Texts with a Tatar-English Glossary.* Indiana University Publications, Bloomington.

Prince, A. and Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar.* Blackwell, Malden, Mass. & Oxford.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Riggs, D. (2014). *Sound sequence adaptation in loanword phonology.* PhD thesis, University of Southern California.

Samarajiwa, C. and Abeysekera, R. M. (1964). Some pronunciation difficulties of Sinhalese learners of English as a foreign language. *Language Learning*, 14:45–50.

Selkirk, E. (1982). *The Syntax of Words.* MIT Press, Cambridge, MA.

Shademan, S. (2002). Epenthetic vowel harmony in farsi. Master's thesis, UCLA.

Shinohara, S. (2006). Perceptual effects in final cluster reduction patterns. *Lingua*, 116:1046–1078.

Sievers, E. (1893). *Grundzüge der Phonetik zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen.* Originally published 1881. Breitkopf and Härtel.

Singh, R. (1985). Prosodic adaptation in interphonology. *Lingua*, 67:269–282.

Stachowski, K. (2010). Quantifying phonetic adaptations of Russian loanwords in Dolgan. *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, (127):101–177.

Steriade, D. (2006). Contour correspondence: Evidence from cluster interruption. Handout from Old World Conference of Phonology 3.

Steriade, D. (2008). The phonology of perceptibility effects: the P-Map and its consequences for constraint organization. In Hanson, K. and Inkelas, S., editors, *The Nature of the Word: Studies in Honor of Paul Kiparsky*, pages 151–179. MIT Press, Cambridge.

Suthiwan, T. (2009). *Thai.* Max Planck Institute for Evolutionary Anthropology, Leipzig.

Tadmor, U. (2009a). *Indonesian.* Max Planck Institute for Evolutionary Anthropology, Leipzig.

Tadmor, U. (2009b). Loanwords in indonesian. In Haspelmath, M. and Tadmor, U., editors, *Loanwords in the World's Languages: A Comparative Handbook*, pages 686–716. De Gruyter Mouton.

Tsuchida, A. (1997). *Phonetics and phonology of Japanese vowel devoicing.* PhD thesis, Cornell University.

Vajda, E. and Nefedov, A. (2009). *Ket.* Max Planck Institute for Evolutionary Anthropology, Leipzig.

van der Sijs, N. (2009). *Dutch.* Max Planck Institute for Evolutionary Anthropology, Leipzig.

Vasilyeva, L. (2010). Vowel harmony for Russian loanwords into Yakut. Master's thesis, California State University, Fresno.

Vennemann, T. (1988). *Preference Laws for Syllable Structure.* Mouton de Gruyter, Berlin, New York, Amsterdam.

Versteegh, K., Eid, M., Elgibali, A., Woidich, M., and Zaborski, A., editors (2006). *Encyclopedia of Arabic Language and Linguistics.* Brill.

Wilson, C., Davidson, L., and Martin, S. (2014). Effects of acoustic-phonetic detail on cross-language speech production. *Journal of Memory and Language*, 77:1–24.

Wright, R. (2004). A review of perceptual cues and cue robustness. In Hayes, B., Kirchner, R., and Steriade, D., editors, *Phonetically-Based Phonology*, pages 34–55. Cambridge University Press.

Yuliati (2014). Final consonant clusters simplification by Indonesian learners of English and its intelligibility in international context. *International Journal of Social Science and Humanity*, 4(6).

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog infixation. *Language*, 83(2):277–316.