THE ACQUISITION OF CONSONANT CLUSTERS BY JAPANESE LEARNERS OF ENGLISH: INTERACTIONS OF SPEECH PERCEPTION AND PRODUCTION

By

MIEKO SPERBECK

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the Doctor of Philosophy, the City University of New York

2010

This manuscript has been read and accepted for the
Graduate Faculty in Linguistics in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy

______________________	______________________________
Date	Winifred Strange, Ph.D. Chair of Examining Committee
______________________	______________________________
Date	Robert W. Fiengo, Ph.D. Executive Officer

Supervisory Committee:

Winifred Strange, Ph.D.

Charles Cairns, Ph.D.

Gita Martohardjono, Ph.D.

Outside Reader:

Lisa Davidson, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

THE ACQUISITION OF CONSONANT CLUSTERS

BY JAPANESE LEARNERS OF ENGLISH:

INTERACTIONS OF SPEECH PERCEPTION AND PRODUCTION

BY

MIEKO SPERBECK

ADVISER: PROFESSOR WINIFRED STRANGE

The primary aim of this dissertation was to investigate the relationship between speech perception and speech production difficulties among Japanese second language (L2) learners of English, in their learning complex syllable structures. Japanese L2 learners and American English controls were tested in a categorial ABX discrimination task of nonsense words sequences (e.g., /spani/ *vs* /səpani/) and included /sp, sk, pl, kl, bl, gl, spl, skl/ clusters. In the second study, production data on these same contrasts were collected by employing the delayed imitation task where speakers were asked to produce the target words in a short sentence. Productions were evaluated by American English listeners. In addition, the Versant test, a short test of English fluency by phone, was administered to see how Japanese participants' present English fluency level would correlate with their performance on the current experiments.

Results of the perception experiment showed that overall accuracy by the Japanese group was significantly poorer than for Americans (Median = 71 % and 100% correct, respectively). Certain clusters were harder than others for Japanese listeners.

Specifically, overall accuracy in the /bl/ clusters was significantly low (Median = 63% correct). The production experiment demonstrated that, as was the case with the perception experiment, the American group showed a ceiling effect for all types of consonantal sequences. In contrast, the Japanese group's performance was consistently lower (Mean = 64% correct). Specifically, Japanese participants had difficulty producing the voiced stops + ə + l tokens (e.g., /bəlani/) accurately. Interestingly, the major errors in these clusters were deletion of schwas. Corelational analyses between perception and production performance were conducted. Overall, the Japanese group's perception and production was correlated (rho = +7.82, $p < 0.01$, one-tailed). Additionally, overall Versant test score was correlated with perception performance (rho = + 0.470, $p < 0.01$) and production performance (rho = + 0.633, $p < 0.01$).

These results suggest that there is a link between L2 perception and L2 production at phonotactic level of acquisition. However, a picture of such link is much more complicated, as suggested by great variability among Japanese participants' performance. That is, the current study suggests that inaccuracy in L2 production is a product of interactions among inaccurate L2 perception, motor constraints of unfamiliar sequences of phonemes, and learners' individual differences in English fluency skills.

ACKNOWLEDGEMENTS

First of all, I would like to thank my adviser, Dr. Winifred Strange, for her dedicated support and guidance. She introduced me to the research of cross-linguistic speech perception. Her passion absolutely inspired me and left a life-long effect on me. I will always be thankful to her for spending enormous time on this project.

My gratitude also goes to my other committee members, Dr. Charles Cairns and Dr. Gita Martohardjono. Dr. Cairns guided me to critically think and apply phonological theory to the data. Dr. Martohardjono provided me with helpful comments and suggestions on the issues of second language acquisition.

Special thanks also to Dr. Lisa Davidson, my outside reader. Her insight on the issues of perception/production of non-native consonant sequences was extremely helpful for this dissertation.

Special thanks also to Dr. James J. Jenkins who gave me statistical advice even when he was far way from NY.

My thanks also go to colleagues and friends who are always inspiring and supportive at the Speech Acoustics and Perception Laboratory: Shari Berkowitz, Yana Gilichinskaya, Miwako Hisagi, Kikuyo Ito, Franzo Law II, Erika Levy, Marisa Monteleone, Kanae Nishi, and Puisan Wong. Especially, I have to admit that I owe the design of this study to “the mass mind” of this laboratory.

I would also like to thank my undergraduate teachers at Temple University, the late Dr. Bibhuti S. Yadav and Dr. Jeffery Kingston, who taught me the importance of “learning”.

My special thanks also go to Megumi Yoshida who researches the "non-sound" part of Linguistics. My conversations with Megumi on (mostly) non-linguistic topics were an absolute necessity to go through some difficult times.

I would also like to thank Linda Perrotta, a director at the Child Development Center at the Graduate Center, who runs the best child care program in the world. Thanks to her, I could focus on my research without feeling any sense of guilt.

My thanks also go to Ikuo and Sachiko Nakamura, my parents, and Yuuichi and Kazuhiro Nakamura, my brothers in Japan, who encouraged me to pursue higher education though they didn't seem to understand what I had been doing at a school in America. I would also like to thank Jack and Judy Sperbeck who were supportive through the graduate school years. My thanks go to my late-grandpa, Inami Nakamura, a hard worker, who just passed away two months before the completion of my dissertation.

My special thanks to my daughters, Marina and Emily, for their smile and laugh. Sometimes that was all I needed to go through my difficult times. Lastly but not the least, I thank Brad Sperbeck, my husband, for his unconditional support without ever complaining about countless baby-sitting on weekends. Without his encouragement, I could have never completed this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Late second language (L2) learners have a strong tendency to transform L2 syllable structures into native ones in producing L2 speech. For example, Japanese-speaking learners of English often break up English consonantal clusters by inserting a vowel [ɯ] (e.g., [sɯ.pɯ.las.shɯ] for “splash”. A period in the parenthesis stands for a syllable boundary). In Japanese phonology, consonantal clusters are not allowed and only single consonants are allowed as syllable onsets. Only the moraic nasal or the first part of geminate serves as a coda. In contrast, English allows for much more complicated syllable types. English word initial onsets may involve up to three consonants and English coda clusters may include up to four. Since complex onsets are not allowed in Japanese, an English word such as “splash” cannot be mapped into a single syllable. These loanwords with epenthesized vowels are often interpreted as being due to learners’ first language (L1) phonotactic restrictions. That is, Japanese speakers modify nonnative syllable forms in order to conform to their native ones.

The primary aim of this study was to investigate acquisition of L2 syllable structure in both perception and production by late learners of English whose native language was Japanese. Investigating both perception and production of L2 syllable structure may be key to understanding the systematicity underlying L2 learners’ behaviors. That is, the current study investigates whether accented productions of L2 syllable structure are due to the inaccurate perception of phonotactically illegal structure. Interwoven with these issues is the concept of *markedness*, in contrast to theories that focus on L1-L2 phonetic/articulatory interactions. That is, another goal of the current study is to investigate whether L2 learners’ overall performance in perception and production will be reflected in different degrees of difficulty predicted by markedness theory or by phonetic interaction theories.

CHAPTER 2: OVERVIEW OF SYLLABLE STRUCTURE

2.1. Basic Syllable Structure

The evidence of the syllable as the phonological domain is abundant. As per Kenstowicz (1994), the following reasons seem to serve for the legitimacy of the syllable as a basic unit of perception and production:

> First, the syllable is a natural domain for the statement of many phonotactic constraints. Second, phonological rules are often more simply and insightfully expressed if they explicitly refer to the syllable. Finally, several phonological processes are best interpreted as methods to ensure that the string of phonological segments is parsable into syllables (Kenstowicz, 1994, p. 250).

Among several models of syllable-internal structure, the "binary-branching with rime" model seems to be the best representation of syllable structure (Blevins, 1995). In this model, the syllable is seen as containing an obligatory nucleus, preceded by an optional consonantal onset and followed by an optional consonantal coda. The majority of languages contain the following basic syllable structure: CV, VC, V, and, CVC (Kenstowicz, 1994). An additional element is called a rime, which includes a nucleus and a coda. Rhyme often serves as an important unit in poetry or in language games.

Syllable structure and syllabification comprise the basics of syllable theory. Interwoven with these basic concepts is the concept of the mora. The mora, a unit of phonological weight, helps us to understand different types of syllables. A light syllable consists of one mora while a heavy syllable consists of two morae. For instance, "ka", meaning "mosquitoes" in Japanese, has one mora while "ka:" with a long vowel, meaning "car" in the Japanese loanword, has two morae. Such differences are also expressed in Japanese syllabary orthography, "*kana*". Each *kana* character represents one mora. Thus, [ka] has one *kana*, "か", while [ka:] requires an additional hyphen-like symbol for the second mora, “カー”. These long vowels are spelled differently in hiragana by adding a second vowel. For instance, “[ka:san]”, meaning “mother” in Japanese, is spelled as “かあさん” in hiragana, the second character stands for a second vowel. Thus, the mora is an essential rhythmic unit in Japanese (Shibatani, 1990).

Concerning the issue of the acquisition of nonnative consonantal sequences, looking into how a L2 word is syllabified in L1 provides a key to understanding the mental representation of L2 learners, especially those who are at the initial stage of acquisition. The accentedness of L2 learners often reflects the transfer of their native syllabification to the target language. Hence, Japanese-speaking learners of English might apply a unit of mora for syllabifying English complex onsets and codas.

2.2. The Sonority Sequence Principle

In the standard version of generative phonology, the creation of a syllable is regulated by the Sonority Sequencing Principle (the SSP), which states that sonority rises toward the nucleus of a syllable and lowers away from it (Clements, 1990). Each phoneme can be categorized in terms of a sonority scale where low vowels are considered the most sonorous, followed by mid vowels, high vowels, flaps, laterals, nasals, voiced fricatives, voiceless fricatives, voiced stops, and voiceless stops (Hogg and McCully, 1987, p. 33).

One of the examples where the SSP plays an important role is found in the issue of the reduplication of the prefix for the perfect tense, “CV(e)”, in Ancient Greek (Kenstowicz, 1994, p. 264–269). According to Kenstowicz, when the sonority of the initial stem is lower than that of the second one, the initial consonant is copied when adding the prefix (e.g., “ke-klop^{h}a” for /klop^{h}a/). However, such reduplication does not occur if consonant sequences lack a rising sonority toward a nucleus. Thus, Kenstowicz analyzes that the initial phoneme, /k/ in /ktona/ is not copied as "*ke-ktona" but rather appears as "e-ktona" because /kt/ is against the SSP. Hence, as per Kenstowicz (1994), this example shows how the SSP is related to the mechanism of reduplication in this language. However, a fair number of exceptions to the SSP has been reported and thus, Blevins (1995, p. 211) argues that the SSP should be considered as "the preference conditions of the syllable". For instance, a clear violation of the SSP is found in /sp, st, sk/ of English complex onsets. Sonority of voiceless stops is lower than /s/, according to the sonority scale described by Hogg and McCully (1987).

Related to the SSP, a Minimal Sonority Distance (MSD) parameter governs the minimal sonority distance between two phonemes of a complex onset in a given language. For instance, English does not necessarily require a large minimal distance between two phonemes and thus, allows several kinds of consonant sequences. On the other hand, Japanese with a smaller setting of the MSD allows single consonant onsets only. To summarize, the role of the syllable as a unit of analysis in phonology seems to be indisputable. However, the issues of the SSP, the internal structure of the syllable, and the syllabification are still ongoing debates among phonologists.

2.3. Markedness

The term "markedness' has been used in various ways in linguistic research. Markedness is related to frequency of a linguistic structure in languages of the world and is often applied to predict relative difficulty in acquisition (see Eckman, 2008 for reviews). Marked linguistic forms are less frequent and considered to be harder to acquire. With respect to syllabic onsets, a simple onset, Consonant + Vowel (hereafter, CV), is classified as less marked than more complex ones such as CCCV. If a given language allows complex syllable types, it also allows the simple CV syllable (e.g., Cairns and Feinstein, 1982). Hence, tri-consonantal (hereafter, CCC) onsets are regarded as more marked than bi-consonantal (hereafter CC) onsets.

Research in L1 acquisition often reports that young children show stronger preference toward CV syllables than consonant sequences in production (Martohardjono, 1989). Concerning the acquisition of consonant clusters, young children tend to produce CC clusters more accurately than CCC clusters (Gierut and Champion, 2001; Smit, 1993). With respect to the acquisition of CC sequences in English, markedness depends on the sonority distance between two phonemes of a complex onset. For instance, complex onsets such as /bl/ (e.g., "*bl*ow") are considered as more marked than /pl/ (e.g., "*pl*ay"). The sonority distance between /b/ and /l/ is smaller than that of /p/ and /l/ since voiced stops are more sonorous than voiceless stops and thus, closer to laterals in sonority. The studies that have explored sonority-based markedness are reviewed in Chapter 4. These studies have often tested whether universally more marked linguistic structures are harder to acquire than less marked ones regardless of learners' L1.

CHAPTER 3: THEORIES OF CROSS-LINGUISTIC SPEECH PERCEPTION AND PRODUCTION

3.1. The Contrastive Analysis Hypothesis and the Markedness Differential Hypothesis

. The first theory to account for L2 learners' relative difficulty in learning L2 speech sounds was called "Contrastive Analysis Hypothesis (CAH)". The central claim of the CAH is that the errors made by L2 learners are all due to the interference from their native language. To quote the main hypothesis:

The Contrastive Analysis Hypothesis

> We assume that the student who comes in contact with a foreign language will find some features of it quite easy and others extremely difficult. Those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult (Lado, 1957, P. 2; cited in Eckman, 1996, P. 196).

In short, the CAH predicts that L2 structures that are similar to those of the L1 should be easier to acquire whereas the structures that are different from those of L1 should be more difficult to acquire. However, several studies have shown that the CAH is not explanatorily adequate by itself (e.g., Broselow, 1983). Not only is the CAH unable to account for many errors that are not simply a result of L1-L2 differences, but also the hypothesis cannot predict what types of errors systematically occur among L2 learners.

Seeking a better explanation than the CAH of the difficulties that L2 learners face, Eckman (1977) proposed the Markedness Differential Hypothesis (MDH). To quote:

The Markedness Differential Hypothesis

> The areas of difficulty that a language learner will have can be predicted on the basis of a systematic comparison of the grammars of the native language (NL), the target language (TL), and the markedness relations stated in universal grammar (Eckman, 1996, P. 197).

As is the case with the CAH, the MDH assumes that L2 difficulty is predictable on the base of structural differences between the L1 and the L2. However, the MDH is advanced theoretically by adding the critical concept of "markedness". That is, the degree of difficulty among L2 structures directly corresponds to the degree of markedness. For instance, in Japanese, both voiceless stop-liquid and voiced stop-liquid clusters are not allowed. The CAH would predict that both are equally difficult since these two consonant clusters do not exist. The MDH would further predict that voiced stop-liquid clusters are more difficult than voiceless stop-liquid clusters since the former is more marked.

These theories based on L1-L2 structural differences do not predict difficulty based on articulatory-phonetic similarities between L1 and L2. For instance, the MDH will predict that difficulty in acquiring such sequences as /pl/ and /pr/ is expected to be the same among Japanese-speaking learners of English because the sonority distance between /pl/ and /pr/ is equal. Phonetic/gestural differences between English /r/ and /l/ are not taken into considerations. That is, the theory does not accommodate effects of phonetic/gestural differences of L2 (e.g., English /r/ and /l/) in reference to L1 (Japanese /r/).

3.2. Perceptual Assimilation Model

Best's Perceptual Assimilation Model (PAM) has been developed to account for perceptual difficulties that *naïve* listeners face from a Direct Realist perspective (Best, 1995). This model assumes that articulatory gestures are perceptual primitives for speech perception. PAM predicts how listeners will perceive nonnative contrasts between phones on the basis of the articulatory/phonetic similarities of nonnative segments to native phonological categories. It should be noted that PAM was further developed into PAM-

L2 (Best and Tyler, 2007) to accommodate the issues of cross-linguistic perception among *more experienced* L2 learners.

L2 speech segments are perceived in several patterns relative to the native language. For instance, in the two-category (TC) assimilation pattern, discrimination of L2 speech sounds is assumed to be excellent. In this pattern, contrasting L2 phones are assimilated to two different L1 categories. In the case of the category-goodness (CG) difference assimilation pattern, discrimination is expected to be moderate to very good. In this pattern, contrasting L2 phones are assimilated to the same L1 category, but differ in their perceived goodness as exemplars of the native category. That is to say, one is a better instance of the L1 category than the other. In the single category (SC) assimilation pattern, contrasting L2 phones are both assimilated as equally good or poor exemplars of a single L1 category and thus, discrimination of these segments is assumed to be very poor. In the uncategorized-categorized (UC) pattern, discrimination of contrasting L2 phones is expected to be very good since the "uncategorizable" segment is perceived as nonnative. Further assimilation patterns include an uncategorized (UU) pattern, where both L2 phones are uncategorizable within the native language inventory. Difficulty of discrimination depends on how close they are articulatorily to each other and to the listener's L1 segments. Lastly, in the nonassimilable (NA) pattern, neither non-native category is assimilated to the listener's phonological space at all and thus, both are perceived as nonspeech sounds.

Numerous studies of the perception of nonnative phones have provided evidence consistent with the PAM's predictions (e.g., Best, McRoberts, and Goodell, 2001). However, the model has rarely been tested on phonological structures larger than a segmental level of representation. Chang, Hong, and Hallé (2007) tested Taiwanese Mandarin speakers on perception of non-native consonant clusters. Taiwanese Mandarin speakers were asked to identify and discriminate pairs of English initial onsets; CCV (e.g., "blow") and CVCV (e.g., "below") where the second consonant was always a liquid, /l/ or /r/. Since consonant clusters are not allowed in their L1 phonology, Chang et al. predicted that Taiwanese Mandarin listeners would show a SC pattern of assimilation. That is, these speakers would perceptually assimilate non-native CCV into native CVCV. Results were divided into groups of good and poor performers; participants in the former group showed ceiling effects for both tasks. Results of the latter group showed that overall performance on the discrimination task was relatively good (mean accuracy = 93%; Table 4, p.798). However, overall performance on the identification task was poorer, although still above chance in this group (mean accuracy = 61%). Interestingly, 33% of CVL words (e.g., "below") and 21% of the CVR words (e.g., "correct") (Table 3, p.798) were identified as complex onsets. In other words, phonologically legal syllable structures in listeners' L1 were not necessarily parsed as they were. Chang and colleagues concluded that listeners' phonotactic knowledge alone could not account for these results and suggested that the unit of perception might be influenced by the Chinese Phonetic Symbol System where 1-3 phonetic symbols represent one Chinese character. In short, the predicted assimilation pattern, the SC pattern, was observed among some participants although it was not necessarily due to listeners' repair process into their L1 phonology.

3.3. Speech Learning Model

Flege (1995)'s Speech Learning Model (SLM) was designed to account for the difficulties which more experienced L2 learners might face in *producing* L2 speech contrasts. Compared to PAM, the model is more explicit about the relationship between L2 speech perception and production. Similar to PAM, SLM also predicts perceptual assimilation patterns based on L1-L2 phonetic interactions. According to SLM, the process called “equivalence classification” is at work in perceiving L2 segments. L2 phones that are similar to L1 phones are initially assigned to existing L1 phonetic categories. Even though L2 phones are acoustically different from the closest L1 phones, they may be perceived as instances of a L1 category, resulting in “accented” perception and production.

An important assumption about the L2 learning process in SLM is that the same perceptual learning abilities used for acquiring L1 are available over a learner’s lifespan and thus, learners’ phonetic categories will keep evolving as a function of language exposure and experience. Moreover, the model hypothesizes that L1 and L2 phonetic categories share a common phonological space, resulting in their influencing each other. This “bi-directionality” hypothesis between L1 phones and L2 phones was confirmed in Flege’s 1992 study. Examining the production of /u/ among Dutch speakers who were learning English as a second language, results showed that learners’ productions of L2 phones also affected production of L1 phones. English /u/ is phonetically more fronted than Dutch /u/. When Dutch speakers’ English proficiency advanced, productions of their native /u/ became more fronted.

Lastly, SLM hypothesizes age-related limits in the ability to produce L2 speech in a native-like fashion. That is, early learners tend to produce L2 sounds more accurately than late learners. This hypothesis has been tested and supported by several studies. For instance, Flege (1991) examined production of the word-initial English /p, t, k/ by native speakers of Spanish who had learned English as adults or as children. Results showed that early learners produced these English stops with the similar VOT values as native monolingual English speakers (53 ms; see Figure 1, p. 399) did. In contrast, late learners produced them with shorter VOT values (33 ms), which was intermediate between monolingual Spanish speakers (18 ms) and monolingual English speakers (51 ms).

To summarize, SLM claims that the more similar L2 phones are to L1 phones, the more difficult they are to acquire since L2 learners cannot perceive the differences and continue to categorize them as the existing native phones. Paradoxically, the greater the differences between L1 and L2 phones, the easier it is to eventually acquire the L2 phones since L2 learners can perceive the differences and establish new categories for the L2 phones. Moreover, as L2 learners gain more experience with L2 phones, they may come to discern the differences between the L2 and L1 phones, resulting in less accentedness in production.

3.4. Automatic Selective Perception Model

A recent theoretical framework proposed by Strange (2009) is called the Automatic Selective Perception Model (ASP). ASP focuses on the perceptual difficulties that L2 learners face in learning new segments and sequences. Like PAM, ASP takes a direct realist approach. That is, speech perception involves “an active, information-seeking process” (Strange, 2009). According to ASP, the online processing of native phonetic contrasts is automatic in adult monolingual listeners. Native phonetic contrasts are perceptually differentiated by over-learned “Selective Perceptual Routines”. In contrast, perceptual differentiation of nonnative phonetic contrasts requires more attentional resources. Specifically, ASP proposes that the mode of perception differs depending on interactions of stimulus types, task demands, and perceivers' knowledge (Strange, 2009). Two modes of speech perception are proposed: a phonological mode and a phonetic mode. In the phonological mode, perception is accomplished via automatic "Selective Perceptual Routines" and the detailed phonetic differences are not the primary focus of perception. In the phonetic mode, however, an act of perception requires more attentional resources as a result of focusing on phonetic details and thus, is less automatic. ASP hypothesizes that naïve listeners utilize the latter mode in listening for nonnative speech contrasts. As SLM proposes, ASP predicts that cross-language speech perception might improve over time but further points out that L2 speech perception in late learners may *never* be as automatic as L1 speech perception.

For instance, Ito et al., (2007) tested whether non-native American English (AE) vowel contrasts were difficult to discriminate for adult late Japanese learners of English. A speeded-ABX discrimination task was employed and 8 AE vowels: [iː, ɪ, ɛ, æː, ɑː, ʌ, ʊ, uː] were embedded in /Vpə/ disyllables. Thirteen contrasts included control pairs that were Two-Category (TC) assimilation pairs for Japanese listeners such as [ɪ *vs* æː] and psychoacoustically similar Single Category (SC) pairs such as [ɑː *vs* æː]. Accuracy and (relative) reaction times were the dependent measures. Results showed that native AE listeners performed consistently better than Japanese counterparts and their reaction times across types of vowel contrasts varied relatively little. In contrast, Japanese listeners' overall performance on the SC contrasts such as [ɑː *vs* æː] was less accurate and their relative reaction times significantly slower even on correct trials. For other contrasts such as [ɪ *vs* ɛ], [ɑː *vs* ʌ], and [æː *vs* ɛ], discrimination accuracy did not differ between

Japanese and AE listeners (82%, 88%, and 90% correct, respectively for Japanese group), but reaction times on correct trials were slower for the L2 listeners. These results reflected the predictions made by ASP. Slower reaction times among Japanese listeners were interpreted as reflecting L2 learners' non-automaticity of perception of non-native vowel contrasts despite the fact that the overall performance on some contrasts was relatively good.

To summarize this chapter, the CAH predicts that L2 acquisition of all English complex onsets would be equally difficult for Japanese speakers, since Japanese phonology only allows a simple CV. The MDH would further predict that CC onsets with smaller sonority distances would be harder than those with larger sonority distances between the two consonants, since the former is more marked than the latter within a framework of generative phonology. In contrast, according to the PAM, the SLM, and the ASP model, predictions would be made based on the L1/L2 articulatory-phonetic similarities and differences between the segments and sequences of Japanese and English. For instance, consonant clusters including AE liquids /r-l/ which are known to be very difficult for Japanese L2 learners to perceive and produce, might be especially difficult, even in combinations with voiceless stops. The ASP would further predict that relative perceptual difficulties would be influenced by interactions of perceivers' language experience, stimulus materials, task types, and situations of listening. Of special interest here is that ASP is explicit about the importance of L1-L2 phonotactic influence on perception of non-native contrasts. In the current study, Japanese listeners are asked to discriminate consonantal clusters from structures with unstressed vowels between the initial and subsequent consonants (CCV *vs* CəCV; CCCV *vs* CəCCV). Even though

CəCV is phonotactically legal in their L1, ASP predicts that the schwa might not be necessarily parsed in the same way as native English speakers would do.

Chapter 4: CONSTRAINTS ON PRODUCTION AND PERCEPTION OF L2 SPEECH PHONES

4.1. Second Language Production of Nonnative Syllable Structure

Studies of the production of L2 syllable structure have often been focused on the concept of sonority-based markedness that was described in the previous chapter (e.g., Broselow and Finer, 1991; Broselow, 1995; Broselow, Chen, and Wang, 1998; Cardoso 2008; Carlisle 1997, 1998, 2006; Eckman 1977, 1981, 1991, 1996, 2004; Hancin-Bhatt and Bhatt, 1997; Major, 1996). These studies tested whether universally more marked linguistic structures are harder to acquire than less marked ones, regardless of learners' first language. On the basis of empirical studies, Eckman revised his MDH as the Structural Conformity Hypothesis, which states, "the universal generalizations that hold for primary languages hold also for interlanguages" (Eckman, 1991, p. 24).

Eckman (1991), for instance, examined L2 learners' productions of three sets of related onsets: /spr/, /sp/, and /pr/; /str/, /st/, and, /tr/; and /skr/, /sk/, and /kr/. The selection of these stimuli was based on the Resolvability Principle, which states "if a language has a consonantal sequence of length *m* in either initial or final position, it also has at least one continuous subsequence of length *m*-1 in this same position" (Eckman, 1991, p. 25). Participants were Japanese, Cantonese, and Korean whose first languages do not allow initial consonant clusters. The tasks that participants performed included reading words from a list, eliciting a word from pictures, and having a conversation with native speakers of English. Results showed that the error rate was correlated with markedness. That is, participants had more difficulties in producing marked sequences (e.g., /spr/) and thus, the author concluded that the Structural Conformity Hypothesis was confirmed.

With respect to the acquisition of CC sequences among L2 learners, Broselow and Finer (1991) also reported that more marked structures were harder to acquire than less marked ones. Production of six types of complex onsets by Japanese and Korean-speaking learners of English were examined: [pr], [br], [fr], [pj], [bj], [fj]. Based on the markedness principle of sonority, Broselow and Finer predicted relative difficulty to increase from least marked to most marked: [pj]-[pr]-[bj]-[br]-[fj]-[fr]. Participants were first asked to learn the pseudo-words that began with these six types of consonant clusters and then to produce them. The production data were transcribed by native speakers of English. Results showed that the more marked clusters had higher error rates than the less marked ones for both language groups. The authors concluded that universal markedness of sonority applied to the production of L2 syllable onsets.

However, Davidson (2001) and in her subsequent studies (2006; 2007) has shown that markedness of sonority failed to account for L2 production difficulties of consonantal sequences. For instance, Davidson (2006) tested 20 English speakers on production of pseudo-Czech words. These words began with, /s/, /z/, /v/, or /f/ and combined with stops and nasals. Two stimulus conditions were created: CC conditions (e.g., /zbano/) and CəC conditions (e.g., /zəbano/). English speakers were visually presented with a target word, asked to listen to two repetitions of the word, and repeat it into the microphone. The pattern of production accuracy was /s clusters/ (97% correct) followed by /f clusters/ (60% correct), /z clusters/ (39% correct) and /v clusters/ (22% correct) (Davidson, 2006, p. 113, Figure 2). Results of each sequence type showed that markedness alone could not account for production constraints on nonnative sequences.

For instance, both /zn/ clusters and /vn/ clusters have sonority distance of 1. However, the accuracy didn't reflect such markedness. That is, /zn/ clusters were produced significantly more accurately than /vn/ clusters (56% and 31%, respectively, Davidson, 2006, p. 113, Figure 3). Such results were also true for the /f/ + obstruent *vs* the /v/ + obstruent contrasts.

Davidson (2006) showed further that acoustic characteristics of schwas inserted between the first two consonants differed depending on the two conditions (e.g., /zbano/ *vs* /zəbano/). Regardless of consonantal context, the mean durations of epenthesized schwa were significantly shorter than those of lexical schwa. Additionally, both F1 and F2 midpoint values of epenthesized schwas were consistently lower than lexical schwa. Davidson interpreted these results from the viewpoint of the Articulatory Phonology (Browman and Goldstein, 1992). That is, vowel epenthesis in CC clusters was not due to speakers' intentions to modify illegal sequences into legal ones but rather due to "gestural mistiming", a failure to overlap two consonants (Davidson and Stone, 2004, p. 167-169). Hence, Davidson's studies on the production of nonnative sequences suggest that the articulatory constraints in producing nonnative sequences should be taken into considerations.

4.2. Second Language Perception of Nonnative Syllable Structures

The following section first reviews the perceptual studies of development of L1 phonotactics by infants, which will provide insight in understanding constraints of acquisition of L2 syllables. Then, perceptual studies of phonotactics on adult L2 learners are discussed; including what sort of errors they make in encountering nonnative sequences of segments.

4.2.1. Speech Perception of L1 Syllable Structure by Infants

Studies of infant speech perception of phonotactic structure have demonstrated that infants show sensitivity to the structure and organization of their L1 sound patterns during the latter half of the first year (Friederici, & Wessels, 1993; Gerken, & Aslin, 2005; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, A., 1993; Jusczyk, Luce, & Charles-Luce, 1994, Kajikawa, Fais, Mugitani, Werker, and Amano, 2006). For example, Jusczyk and colleagues (1993) tested whether English-learning infants and Dutch-learning infants aged 6 and 9 months had knowledge of phonetic and phonotactic properties of their L1, using a looking preference task. Dutch words are similar to English words *prosodically* (stress-timed) but differ *phonetically* and *phonotactically*. English does not allow phonetic sequences such as [kn] or [zw] to begin syllables, whereas Dutch does. English allows [d] in syllable-final position while Dutch does not (Jusczyk et al., 1993, p. 405). Thus, the stimulus materials included words that were not allowed in listeners' L1 in terms of its phonetic and phonotactic structure. Results showed that 6 month olds did not show any preference for words that occurred in their L1 whereas 9 month olds showed significant preferences for them. In a follow-up experiment, stimulus materials were low-pass filtered to remove most of the phonetic and phonotactic information but left prosodic properties intact. Neither group of infants showed a preference for L1 words; this confirmed that the results were not due to prosodic properties of words but rather, to their phonotactic structure.

Kajikawa, Fais, Mugitani, Werker, and Amano (2006) suggested that infants at the age of 6 months are already sensitive to surface input patterns in their L1. In Tokyo Japanese, there is a well-known phonological phenomenon called “vowel-devoicing” where high vowels tend to be devoiced when they occur between voiceless consonants. Thus, two forms of the surface input patterns are available for a word such as “sika (deer)”; one is with voiced [i] and the other is with devoiced [i̥].

English and Japanese monolingual infants at ages 6, 12, and 18 months were tested on discrimination of these types of words using the habituation-switch paradigm. In this paradigm, recovery of looking to a switch stimulus was interpreted as showing that infants could discriminate the switch stimulus from the habituated stimulus. Test stimuli were three nonsense words: *neek*, *neeks*, and *neekusu*. In English, all three nonsense words are phonotactically legal and distinct. In contrast, in Japanese, *neeks* and *neek* are illegal underlying word forms, although they are “possible forms in fluent speech, derived by vowel devoicing from canonical forms, *neekusu* and *neeku*, respectively” (Kajikawa et al., 2006, p. 2280). They hypothesized that Japanese-learning infants might not be able to discriminate between *neeks* and *neekusu* since they might be treated as variants of the same word. Results showed that discrimination accuracy on the *neeks-neekusu* contrast remained difficult regardless of infants’ age (70-75% of infants who showed recovery of looking time in the test items). In contrast, discrimination improved linearly for English-learning infants (70% at 6 months, 80% at 12 months, and 90% at 18 months; see Figure 2, p. 2281). Kajikawa and colleagues speculated that Japanese-learning infants could not discriminate variants that were not phonotactically distinct and thus, suggested infants’ sensitivity toward the phonotactic status of surface input in their L1.

4.2.2. Speech Perception of L2 Syllable Structure by Adults

Perceptual studies in the past have revealed that L2 learners’ inaccurate production of phonetic sequences in the target language is often a result of inaccurate perception of nonnative speech forms (see Strange, 1995; Strange and Shafer, 2008, for reviews). With respect to the perception of syllable structure, past studies have shown that L1 phonotactic properties seem to bias perception of nonnative phoneme sequences (e.g., Altenberg, 2005; Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi & Mehler, 2001; Fais, Kajikawa, Werker, & Amano, 2005; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kabak and Idsardi, 2007; Sperbeck and Strange, 2008). In other words, perception research has shown that L2 input is not necessarily parsed in a native-like manner by L2 learners.

First, studies have shown that L2 listeners have a tendency to perceptually assimilate illegal phoneme sequences in L1 to legal ones (Dehaene-Lambertz, Dupoux, and Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, and Mehler, 1999; Dupoux, Pallier, Kakehi, and Mehler, 2001; Fais, Kajikawa, Werker, and Amano, 2005; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kabak and Idsardi, 2003; Pitt 1998). For instance, Hallé and colleagues (1998) investigated how French speakers perceptually assimilated illegal clusters (/tl/ and /dl/) into legal ones (/kl/ and /gl/). This perceptual assimilation pattern was observed in both a transcription task (Experiment 1) and a forced-choice identification task (Experiment 2). Results of the transcription task showed that /dl/ and /tl/ items were transcribed as /gl/ or /kl/ 85% of the time. On the other hand, legal clusters were transcribed correctly 100% of the time (Hallé et al., 1998, p. 594).

Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999) tested Japanese listeners' perception of consonantal clusters that occurred word-medially (e.g., /a*bg*e/). Using a categorial ABX discrimination task, two types of contrasts were tested: an epenthesis contrast (e.g., "*abge-abuge*") and a vowel length control contrast (e.g., "*abuge-abuuge*"), the latter being phonologically distinctive in Japanese. Japanese participants heard three stimuli "ABX" and responded whether "X" was the same as "A" or "B". Results showed that Japanese participants had more difficulty with the epenthesis contrast than with the vowel length contrast (16% *vs* 4% errors, respectively). Dupoux et al. interpreted these results as showing that Japanese listeners perceived "*abge*" as the perceptually-modified form, "*abuge*". That is, the authors claimed that the modification of L2 syllable structures occurs at the level of L2 speech perception.

Such perceptual biases have also been reported in studies that test listeners' knowledge of rhythmic properties in on-line speech processing (Bradley, Sánchez-Casas, & García-Albea, 1993; Cutler, Mehler, Norris, & Segui, 1986; Cutler, & Otake, 2002; Mehler, Dommegues, Frauenfelder, & Segui, 1981; Otake, Hatano, Cutler, & Mehler, 1993). The major method of these studies is a Target Item Detection task where listeners are asked to respond to the particular target item promptly as well as accurately after detecting it. The hypothesis behind this method is that, if the detection of the target item is relatively easy, the miss rates should be lower and the response times (RTs) should be shorter. For instance, Mehler and colleagues (1981) examined the role of native syllable structure in word segmentation among French speakers. The authors hypothesized that, if the stimulus words were segmented according to their syllabic structure, RTs should be faster when the target phoneme sequence matched with the first segmented syllable of the stimulus word. The stimuli included 5 pairs of French nouns sharing the initial three phonemes (CVC) (e.g., such as (/pal/). In one member, the CVC sequence formed the first syllable (e.g., /pal.mier/) whereas in the other, the final C was the onset of the second syllable (e.g., /pa.la.ce/). Participants were presented with the target syllables written on cards, listened to the stimulus, and responded as soon as they detected the target in the word. Results showed that RTs were faster when the targets matched with the syllabic structure. For example, /pa/ in /pa.la.ce/ was detected faster than in /pal.mier/ (352 ms *vs* 371 ms) and /pal/ in /pal.mier/ was detected faster than in /pa.la.ce/ (356 ms *vs* 378 ms). Thus, the authors concluded that French speakers used a syllable-based strategy in speech processing.

The use of language-specific syllabic units in speech segmentation was further demonstrated by Otake and colleagues (1993). The purpose of their study was to see whether a rhythmic category such as the mora in Japanese would play a role in speech segmentation by native speakers. In the first experiment, eight pairs of Japanese real words were selected; within each pair, the initial and final morae were identical. However, the second mora was either nasal consonantal coda (e.g., "ta*N*shi") or a CV mora beginning with [n] (e.g., "ta*ni*shi"). Participants were instructed to listen for the target item (written in alphabetic script) and to press a response key as soon as they detected the target. Predictions were made differently for a syllable hypothesis and a mora hypothesis. According to the syllable hypothesis, authors predicted that CV targets (e.g., "ta") should be easily and quickly detected in CVCVCV (e.g., "ta*ni*shi") but harder in CVNCV words (e.g., "ta*N*shi"), since the target is a part of the syllable in CVNCV words. On the other hand, according to the mora hypothesis, there should be no differences in accuracy for spotting CV targets in both CVCVCV and CVNCV words since the targets match with the first mora. Likewise, the detection of CVN targets in

CVNCV words should be accurate, but RTs might be longer for this case since the targets correspond to the first two morae.

Results showed that Japanese speakers used a mora-based strategy in word processing. As predicted by the mora hypothesis, Japanese participants' miss rates were low for spotting CV targets in two types of words; CVCVCV and CVNCV as well as for detecting CVN targets in CVNCV words. Their RTs were fast for detecting CV targets regardless of types of words but slower for CVN targets in CVNCV words (Figure 1 & 2, p. 265). In their subsequent experiment, French speakers were tested with the same materials. Results showed that there were no significant differences in accuracy for detecting CV targets in CVCVCV and CVNCV words. However, unlike Japanese speakers, there were differences in RTs for spotting CV words. RTs for CV targets were slower for CVNCV than for CVCVCV words (Figure 7 & 8, p. 278). These results were interpreted as being compatible with the syllabic segmentation strategy proposed for French speakers. Thus, Otake and colleagues provided further evidence that naive adult listeners may apply their native rhythmic units for segmenting L2 words.

Fais, Kajikawa, Werker, and Amano (2005), however, suggest that listeners do not rely on phonotactic knowledge alone in perception. They investigated whether Japanese listeners perceive stimuli primarily in terms of phonological knowledge or in terms of acoustic differences in differing contexts. These authors assumed that, if phonotactics alone guided perception, Japanese listeners would perceive word-final vowels regardless of whether the vowel was fully realized (e.g., *neeku*) or devoiced (e.g., *neekụ̊*) even though word-final high vowels have a tendency to get devoiced in Japanese. On the other hand, if acoustic form played a major role, listeners should only "hear" a vowel when it was actually present in the signal, regardless of the context. Nonsense words that contained a vowel (e.g., neeku [ni:kɯ]) and those that lacked a vowel (e.g., neek [ni:k]) were created as test items. A category goodness-rating task was employed. In this task, each participant was asked to rate how good the stimulus was on the scale of 1 (bad) to 7 (good).

Results showed listeners' sensitivity toward the acoustics of the forms. For instance, even though /k/# in *neek* was an acceptable surface form due to the word-final devoicing rule, it was rated significantly lower than /ku/# in *neeku*. In addition, mean ratings for the nonsense words that ended with a single consonant (e.g., *neek*) were varied. For instance, *neek* was rated significantly higher than *keet*, suggesting that /k/# in *neek* was an acceptable surface form while /t/# was not. If phonotactic knowledge alone guided listeners' perception, mean ratings should be equally low for all words without a final voiced vowel.

Finally, a recent study by Berent, Steriade, Lennertz, and Vaknin (2007) tested whether perception of onset clusters was affected by the universal markedness of sonority. Three types of contrasts based on markedness were tested: sonority rises (e.g., "*bwif-bəwif*"), sonority plateaus (e.g., "*bdif-bədif*"), and sonority falls (e.g., "*lbif-ləbif*"). Participants were native speakers of American English, whose phonology does not allow any of these consonantal sequences, and native speakers of Russian, whose phonology allows all these clusters. Berent and colleagues hypothesized that accuracy in identifying onset clusters would correlate with the markedness of onsets. That is, the rate of vowel epenthesis would be higher for the more marked onset clusters (e.g., "*lbif*") than the less marked ones (e.g., "*bnap*") since the sonority fell in the former type of clusters. The task was a forced choice identification task where listeners selected whether the auditory stimulus had one syllable or two syllables by pressing the button.

Results showed that the overall percent correct for the three types of clusters reflected universal markedness of sonority among native speakers of English. That is, as the markedness increased, the accuracy rate fell. The overall percent correct was about 60% for sonority rise, 30% for sonority plateau, and 15% for sonority fall, respectively (Berent et al., 2007, p. 604, see Figure 1). Such effects of markedness were also obtained in the AX discrimination task where listeners had to respond whether two stimuli (e.g., "*lbif – lebif*") were the same or different. Berent et al. interpreted these results as showing that markedness of sonority operates in perceiving non-native onset clusters.

To summarize this section, infant studies have revealed that in the first year of life, infants start showing sensitivity to L1 phonotactic structure as well as to variations in surface input patterns. Adult studies have established that L1 phonotactic knowledge seems to bias the perception of nonnative phonemic sequences. Phonotactically illegal sequences in L1 tend to be perceptually assimilated to legal ones by adult listeners. Studies of on-line speech processing have shown that word segmentation is accomplished in a language-specific way using syllables and other rhythmic units. Lastly, the perception of nonnative initial complex syllable onsets might be influenced by universal markedness of sonority principles.

4.3. Studies of Perception-production Link among Adult L2 Learners

In general, perception of L2 speech segments is considered to precede production of those phones. Perceptual training studies on nonnative segments have shown that, with effective training techniques, perception of L2 segments by adult learners can improve significantly (e.g. Lively, Logan, & Pisoni, 1993; Strange and Akahane-Yamada, 1997; Sperbeck, Strange, & Ito, 2005) and the improvement of perception may lead to improvement of production without any explicit articulatory training (e.g., Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1996).

However, a mismatch between perception and production abilities in L2 acquisition has also been reported. For instance, Sheldon and Strange (1982) showed that Japanese participants' accuracy of production of English /r/ and /l/ exceeded their ability to accurately perceive recordings of their own and others' productions. In the production task, both American and Japanese participants produced test words that contained /r/ and /l/ in different word positions. Their productions were then judged by participants. Since participants were given which minimal pair (e.g., "road"-"load") would be presented before listening to the test item, they had only to decide whether the phoneme they heard was /r/ or /l/.

Results showed that overall production errors were only 1% for good Japanese speakers, as judged by American English listeners. In contrast, considerable individual differences were reported in terms of perceptual mastery (the overall percent errors in identification of American speakers' productions ranged from 5 to 19%; Sheldon and Strange, 1982, p. 251, Table 2). Furthermore, Japanese listeners demonstrated different degrees of perceptual difficulty in terms of contexts. "Perceptual difficulties were most marked for /r/ and /l/ in prevocalic stop + liquid clusters, whereas word-final postvocalic /r/ and /l/ were perceived very well" (Sheldon and Strange, 1982, p. 253-254). This study suggests that, unlike the acquisition of L1 speech phones, perceptual mastery does not necessarily precede production mastery in acquiring L2 segments.

Contrary to Sheldon and Strange's findings, Altenberg's study (2005) of Spanish-speaking learners of English showed that perceptual ability preceded production ability on nonnative consonantal sequences. In Spanish phonotactics, unlike English, initial clusters beginning with /s/ are not allowed. However, like English, the word-initial consonant clusters: /bl, dr, fl, kr/ are allowed (Alternberg, 2005, p. 60). Based on these phonotactic constraints, two types of consonant clusters were constructed. One was called "ES" in which an initial consonant cluster was legal in both English and Spanish. The second was called "E*S" where the initial consonant cluster was legal in English but not in Spanish. In the perception task, after hearing a stimulus such as [dras], participants were asked to write down the initial consonantal cluster. Results showed that the overall accuracy for both ES nonsense words and E*S nonsense words was about 90%.

In the production task, English real words were used as stimuli, and an elicitation task was employed. Participants saw pictures and produced the words. Two native speakers of English served as judges. Results showed more errors on E*S than ES words in the production task. The mean accuracy for ES words was 96%. In contrast, it was only 66% for E*S words. Most production errors were word-initial vowel epenthesis. In short, this study showed that Spanish-speaking learners of English could perceive illegal /s/ + C clusters but could not produce them accurately. Altenberg (2005) interpreted these results in terms of articulatory constraints, as suggested by Flege's SLM (1995). This study suggested that L2 production is not necessarily due to misperception but rather may be due to production constraints.

CHAPTER 5: METHODS

5.1. Overview

The current study investigates the production and perception of initial consonant clusters in English by Japanese late learners of English. Production and perception tasks were administered to each participant. As past studies of L2 perception-production suggest (e.g., Sheldon and Strange, 1982), the possibility of a mismatch between perception and production among L2 learners is of interest both theoretically and practically. There are three possible patterns of L2 perception-production that have been reported previously: 1) performance on the production task may correlate highly with performance on the perception task; 2) performance on the perception task will exceed performance on the production task; 3) performance on the production task will be better than performance on the perception task. The first pattern is predicted if there is a direct relationship between speech perception and production in acquiring L2 speech phones. The second and the third pattern are expected if the relationship between L2 perception and production is rather weak. The second pattern should be observed if L2 learners' problems are mainly due to articulatory constraints imposed by the L1 phonology. The third pattern should be expected if L2 learners' main problems are their inaccurate perception, but they have learned to produce the structures by means other than auditory feedback control of performance (i.e., direct articulatory training and tactile-kinesthetic monitoring). As previously mentioned, the current study also aims to test markedness constraints on acquisition of L2 syllable structure, in contrast to the theories of L1-L2 phonetic interactions. If markedness operates in L2 perception (Berent et al., 2007) as well as L2 production (e.g., Broselow and Finer, 1991), it is predicted that Japanese participants will show different degrees of difficulty in perceiving and producing complex onset clusters.

The choice of the stimulus materials was inspired by Eckman's 1991 study, which tested the production of sets of related onsets (e.g., /skr/ - /kr/ - /sk). However, since the production and perception of /r/ is known to be notoriously hard for Japanese speakers (e.g., Bradlow, 2008), /r/ was avoided in the stimulus materials. The perception of two sets of related onsets was examined: /spl/, /sp/, and /pl/ and /skl/, /sk/, and /kl/. Additionally, stimulus materials included /bl/ and /gl/ clusters in order to see effects of markedness within CC sequences (voiced *vs* voiceless stops).

In order to relate Japanese participants' L2 experience and mastery on performance in the perception and production tasks, a test of spoken English proficiency was administered. The Versant for English test is a useful tool for measuring Japanese participants' current spoken English proficiency (Bernstein, 2009). The test consists of several sections where participants are asked to repeat short sentences, answer simple questions, build a sentence after listening to words, and answer open questions about family life or personal choices. The entire test is administered by calling the computerized testing system and usually takes about 15 minutes to complete. The test score is reported on 20 to 80 point scale. The overall score is calculated on the basis of a combination of following four divided sections; sentence mastery, vocabulary, fluency, and pronunciation. Thus, the current study should reveal how L2 learners' fluency level correlates with overall performance on the experimental perception and the production task. It is expected that performance on the production and the perception task should be higher for those who scored well on the Versant test.

Each participant completed the perception experiment first. Then, he/she completed the Versant test. After the brief break, he/she completed the production experiment. Lastly, each participant was asked to fill out a language background and usage questionnaire. What follows is a description of the methods for the perception and production tasks.

5.2. Participants

Thirty native speakers of Japanese (19 female and 11 male; 22 to 52 yrs; Mean = 33 yrs; SD = 6.3) served as the experimental group while five native speakers of American English (3 female and 2 male; 23 to 34 yrs; Mean = 28 yrs) formed the control group (individual data is available in Appendix A for the Japanese group and Appendix B for the American group). Japanese participants' mean length of residence in English-speaking countries was 3.6 years (SD = 4.9; Range = 1 month to 19 yrs and 11 months) and their mean age of arrival was 29 years old (SD = 5.0; Range = 19 yrs to 39 years old). In addition, Japanese participants' information about their daily English usage: the overall amount of English use relative to Japanese; the amount of English use in listening and speaking; the amount of English use in reading and writing was collected via a questionnaire and interview. Participants were asked to report each amount as a percentage. As Table 5 shows, the majority of JP participants reported that they spoke Japanese and English in about equal proportions. The percentage on reading and writing in English was slightly higher than that on listening and speaking although both varied a great deal among individuals. All participants passed a pure-tone hearing screening prior to the experiment (ANSI standards 25dB HL at 500, 1000, 2000, and 4000 Hz) and were paid $20 for their participation.

An additional two Japanese participants' data were excluded from the study. One was born and lived in China until 3 years old and the other spent much of her childhood outside Japan (from 2 to 6 years old in Belgium and from 8 to 12 years old in Switzerland).

5.3. Stimulus Materials

A female speaker of American English who was a linguist produced stimulus materials for the experiment. Nonsense words were of the form /CC(C)áni/ and /CəC(C)áni/. An unstressed schwa served as the vowel between initial and subsequent consonants. There were 6 CC contexts (/sp, sk, pl, kl, bl, gl/) and 2 CCC contexts (/spl, skl/). These target words were produced in a short carrier sentence, "Say ___ now". Four tokens of each word were digitally recorded in a sound-attenuated room, using a Shure (SM48) microphone, which fed through a preamplifier (Earthworks LAB 101) into the A/D soundcard using SOUND FORGE 4.5 software. The digitization rate was 22.05 kHz with 16-bit resolution on a mono channel. The speaker was instructed to speak with normal conversational effort.

Acoustic measurements of the target words were made using Praat (Boersma & Weenink, 2005). The absence of an epenthetic vowel in the CC(C)ani materials (e.g., "splani") and its presence in the CəC(C)ani stimuli (e.g., "seplani") were verified. Durations of the schwa in the CəC(C)ani materials were measured. The average duration of schwa for all cluster types was 38 ms (SD = 11.3, Range = 18-57). In addition, formant frequencies of these schwas in the CəC(C)ani materials were measured. Since the vowels were short and steady, measurements for the first two formants, F1 and F2, were taken by highlighting the entire vowel section from a Formant object in Praat (see Table 1 for the detailed measurement).

Moreover, the durations of other distinct segments of the target words were measured. Mean durations of each segment are shown in milliseconds in Appendix C, D, and E. The measurement criteria for each cluster types were as follows:

1. /s + stop/ clusters (/sp, sk/): the measurement of /s/ began at the onset of frication and ended at the beginning of silence. The silent gap between /s/ and the following voiceless stop, which occurs due to the occlusion of the vocal tract, was also measured. The measurement of voiceless stops was characterized as the duration of Voice Onset Time (VOT). “The vertical spike marking the transient burst of stop release” was the onset of the interval and the onset of voicing for the following vowel marked the end of VOT (Raphael, Borden, and Harris, 1980, p. 131). An example of these measurements is displayed in Appendix F.
2. /stop + l/ clusters (/pl, kl, bl, gl/): stop consonants were measured with respect to VOT. However, the duration of /l/ plus the following vowel was measured as one segment as it was sometimes difficult to segment between these two phonemes. Measurement began at the onset of the phonation for [l] and ended at the onset of the nasal murmur, which was depicted as weak intensity in formants on the spectrogram. Examples of these measurements are displayed in Appendix G and H.
3. Tri-consonantal clusters (/spl, skl/): The duration of /s/ was characterized as frication displayed on the waveform. The following silent gap was depicted as absence of voicing energy on the spectrogram and waveform. The measurement

for following /voiceless stops + l/ clusters was taken in the same manner as section 2. Examples of the measurements are displayed in Appendix I.

Acoustic characteristics of the stimuli should be noted as they serve as important cues for the perceptual discrimination task. Since participants' perceptual abilities are measured by a categorial ABX task, comparisons of distinct segments between the CVC(C)ani stimuli and the CəC(C)ani materials were made. Polka (1991) classified acoustic parameters as either *differentiating* or *correlated*. "Differentiating acoustic parameters were defined as those acoustic parameters that did not overlap for the contrasted consonants and thus alone could support category differentiation. Correlated acoustic parameters were defined as acoustic parameters that differed significantly in central tendency for the contrasted syllables, but also had some overlapping variability, so that, as a single cue, the parameter could not reliably support category differentiation" (Polka, 1991, p. 2968-2969). Following Polka (1991), durational differences of each distinct segment were classified as either differentiating or correlated.

5.3.1. Acoustic Characteristics of /s / + Stops Clusters

Table 2A shows durations of each distinct segment for the /s/ + voiceless stops (sC) clusters *vs* the /s/ + /ə/ + voiceless stops (səC) clusters. The durations of /s/ were significantly shorter in /səC/ sequences (Mean = 125 ms) than in /sC/ sequences (mean = 145 ms). Statistically, such durational differences were significant as tested by a nonparametric Mann-Whitney U test of independent groups ($z = -2.812$, $p = 0.005$). In addition, the duration of the VOTs were significantly longer in /səC/ sequences (mean = 46 ms) than in /sC/ sequences (mean = 12 ms) ($z = -2.892$, $p = 0.004$) and did not overlap. However, the durations of silent gap between the /səC/ sequences (mean = 77 ms) and the /sC/ sequences (mean = 67 ms) overlapped significantly and means were not significantly different. Thus, three segments: schwa presence/absence, /s/ frication duration, and stop VOT values were distinctive acoustic cues for these clusters.

5.3.2. Acoustic Characteristics of Stops + Laterals Clusters

Table 2B and 2C shows a summary of durational differences of each distinct segment in stops + laterals clusters: VOT, schwa, [la]. First, the presence/absence of schwa was completely consistent across syllables with both voiced and voiceless stops. Second, mean durations of VOT in the voiceless stops + /l/ clusters were longer than those in the voiceless stops + /ə/ + /l/ clusters (71 ms *vs* 48 ms, respectively) but overlapped slightly. A Mann-Whitney U test showed that distributions of duration differences were significantly different ($z = -2.656$, $p = 0.008$) and thus, VOT was regarded as a correlated cue. Statistical significance was not observed for VOT in their voiced stop + /l/ tokens ($z = -.321$, $p = 0.748$) and distributions overlapped almost completely for voiced stop + /l/ clusters and voiced stop + /ə/ + /l/ sequences. Thus, VOT cued syllable structure differences only in voiceless stop contexts. Finally, overlap was not observed in the durations of /la/; thus, this durational difference was categorized as a differentiating parameter. A Mann-Whitney U test showed that durational differences of [la] in voiceless stop + /l/ *vs* voiceless stop + /ə/ + /l/ clusters and in voiced stop + /l/ *vs* voiced stop + /ə/ + /l/ clusters were significant ($z = -2.882$, $p = 0.004$ and $z = -2.887$, $p = 0.004$, respectively). To summarize, for voiced and voiceless stop + /l/ clusters, schwa presence/absence was the primary cue, and /la/ duration another differentiating cue. VOT was a correlated cue (with a little overlap) only for voiceless stops. Thus, voiceless stop contexts could be somewhat more advantageous than voiced stop contexts from a phonetic perspective.

5.3.3. Acoustic Characteristics of Triconsonantal Clusters

Table 2D shows a summary of durational differences of distinct segment for triconsonantal pairs: [sCC] *vs* [səCC]. Neither the durations of [s] frication nor those of stop silences were significantly different between the [sCC] clusters and the [səCC] clusters. However, the average VOTs were significantly longer in [səCC] clusters (mean = 46 ms) than in [sCC] clusters (mean = 16 ms). In contrast, the durations of [la] were shorter in [səCC] clusters (mean = 167 ms) than in [sCC] clusters (mean = 185 ms). A Mann-Whitney U test showed that VOT (non-overlapping) and [la] durations were significantly different in [sCC] clusters *vs* [səCC] clusters ($z = -2.882$, $p = 0.004$ and $z = -1.992$, $p = 0.05$, respectively). These data suggest that two differentiating cues: VOTs and schwa are available for differentiating [sCC] *vs* [səCC].

To summarize, the acoustic analysis revealed the following: 1) three cues: frication, schwa, and VOTs were available in the [s(ə)C] contexts; 2) for both voiced and voiceless stop + /(ə)l/ clusters, schwa, and [la] were available as differentiating cues. However, durational difference in VOT values was only seen in voiceless stop + /(ə)l/ clusters. Hence, voiceless stop + /(ə)l/ clusters might be easier to differentiate than their voiced counterparts; 3) with respect to the [s(ə)CC] contexts, in addition to the presence/absence of schwa, other cues such as VOTs and the duration of [la] might serve as distinctive cues.

5.4. Perceptual Task Procedures

Participants were tested individually in a sound-attenuated room at the CUNY

Graduate Center. The perception task was created by the Paradigm beta-5d (written by Bruno Tagliaferri, 2008). A categorial ABX task was employed in which participants heard three short sentences in a row that contained the target words (e.g., “say spani now” – “say sepani now” – “say spani now”) and answered whether the third target word was the same as the first or the second one by pressing a button (the first one is correct in this example). The tokens of each word in a triad were different productions so that there was no physical match between X and either A or B. The inter-trial interval (ITI) was self-paced. The new trial was set up to start when a participant clicked the designated place on the computer screen.

The perception experiment consisted of seven blocks. The first block was for familiarization and was not scored. Each block contained 32 trials (8 consonantal clusters (/sp, sk, pl, bl, kl, gl, spl, skl/) × 4 combinations (ABA, BAB, ABB, BAA) where A = CC(C) and B = CəC(C). Thus, there were 224 trials (32 trials × 7 blocks) in total; 24 trials (4 trials × 6 blocks) for each comparison were scored. The presentation of the trials was randomized by the computer program for each participant, thus controlling for possible order effects.

5.5. Production Task Procedures

The stimuli that were used for the perception task also served as stimulus materials for the production task. There were 8 pairs of nonsense words: /spani-səpani, skani-səkani, plani-pəlani, klani-kəlani, blani-bəlani, glani-gəlani, splani-səplani, sklani-səklani/. A delayed imitation task was used to assess production. First, participants heard a native speaker’s productions (e.g., “Say *blani* now”) twice (50 ms ISI). Second, they had to produce the target word in isolation (e.g., “*blani*”), and then, embedded it in the carrier sentence (e.g., “I said *blani* now”). The latter was scored in this study. Participants were aided with the sign that said, “I said ------- now” placed on the glass wall in front of the microphone. No written materials for the target words were given to participants in order to avoid any orthographic effects. All recordings were made in the sound-attenuated room.

5.6. Transcription Task Procedures

5.6.1. Participants

Two phonetically-trained native English speakers perceptually transcribed the productions. One was a graduate student in the department of Linguistics while the other was a graduate student in the department of the Speech-Language-Hearing Sciences. Neither reported any history of hearing problems. They received payment for their participation.

5.6.2. Design and Procedure

The transcription task was constructed in Microsoft PowerPoint. The experiment consisted of seven blocks. Each block contained 112 trials (16 consonant clusters (/s(ə)p, s(ə)k, p(ə)l, b(ə)l, k(ə)l, g(ə)l, s(ə)pl, s(ə)kl/) × 7 speakers (1 American English speaker and 6 Japanese speakers). Thus, there were 560 trials (112 trials × 5 blocks) in total. Each block always began with an American English speaker’s productions, followed by 6 Japanese speakers. Within each block, the trials were blocked by speakers. Transcribers clicked the number on the computer screen and heard a stimulus (e.g., “I said *blani* now”) through headphones. After listening to each stimulus, they wrote down what they heard on an answer sheet. The inter-trial interval was self-paced. Participants were allowed to repeat the audio stimulus if they were unsure of what they had just heard.

5.6.3. Coding

Each answer was categorized into the following codes: 1) Correct (C) if the transcribed word matched the target word. All the responses written as [r] for the stop + /l/ tokens (e.g., [prani] for /plani/) were included in this category. This is due to the fact that production and perception of the English /r-l/ contrast is notoriously hard for Japanese learners of English; 2) Vowel Deletion (VD) if words were written without a vowel for the CəC(C) tokens (e.g., [spani] for /səpani/); 3) Vowel Epenthesis (VE) for the CC(C)V tokens where words were written with a vowel (e.g., [belani] for /blani/); 4) Segment Change (SC) if the intended consonants were replaced with others (e.g., [skani] for /spani/);; 5) Others (O) included any modifications other than the above categories. The cases where more than one error was made were included in this category (e.g., [skani] for /səklani/)

5.7. Predictions

Assuming that markedness operates in L2 perception and L2 production, it is predicted that CCC sequences will be harder than CC sequences since the former has more complex linguistic structures than the latter. In addition, based on the sonority-based markedness on CC sequences, performance should be better on voiceless stop + /l/ (hereafter "−V stop + /l/"] clusters than the voiced stop + /l/ (hereafter "+V stop + /l/") clusters since the former has larger sonority distance between two phonemes. Strictly following the minimal sonority distance parameter, the sC clusters are predicted to be more difficult than other CC clusters due to the smaller sonority distance between two phonemes in the initial onsets. Although the status of the initial /s/-clusters as a complex onset has been controversial (e.g., Cairns, 2009; Vaux and Wolfe, 2009) since these clusters do not follow the Sonority Sequencing Principle, it is assumed here that there is no structural distinction between /s/- and non-/s/ clusters, following the analysis of Boyd (2006).

In contrast, different predictions were made from theories that take into account L1-L2 articulatory-phonetic interactions (hereafter, L1 phonetic interference hypothesis as named by Monteleone, 2009). For instance, consonant clusters including English liquids /r-l/, which are known to be very difficult for Japanese L2 learners to perceive and produce, might be especially difficult, even in combinations with stops. Thus, it was predicted that Japanese participants would have more difficulty perceiving and producing all stops + /l/ clusters than [sC] clusters. Among the clusters containing /l/, +V stop + /l/ clusters might be more difficult than −V stop + /l/ and [CCC] clusters, especially in perception since the former has fewer distinctive acoustic cues than the latter. The VOT cue is not available in +V stop + /l/ *vs* +V stop + /ə/ + /l/ comparison. In addition, there shouldn't be any difference in relative difficulty between −V stop + /l/ and [CCC] clusters since both cluster types have the same types/numbers of acoustic cues; two distinctive cues and one correlated cue.

To summarize, based on the markedness, CCC clusters are predicted to be harder than CC clusters. Within CC clusters, the following is the predicted relative difficulty from the most to the least marked: sC (2) > +V stop + /l/ (4) > −V stop + /l/ (5) where the number in parenthesis represents the sonority distance between two phonemes according to Hogg and McCully (1987). In contrast, the L1 phonetic interference hypotheses predict that sC clusters are less difficult than all other clusters containing /l/. Within the clusters with /l/, the predicted relative difficulty based on phonetic distinctiveness is as follows:

+V stop + /l/ > −V stop + /l/ = CCC.

Chapter 6: RESULTS

6.1. Results of Perception Experiment

First, overall discrimination accuracy for the Japanese (hereafter, JP) and the Americans (hereafter, AE) groups was compared. As predicted, overall accuracy of perception by the JP group (Mean = 72%, Median = 71%, Range = 56-96%) was poorer than for the AE group (Mean = 99%, Median = 100%, Range = 96-100%). There was far greater variability in the JP group (SD = 16.03) than in the AE groups (SD = 2.46); individual data is available in Appendix J for the AE group and Appendix K for the JP group. However, all but one JP participant performed more poorly overall than the worst AE participant.

Next, the scores on the 8 contrasting syllable structures for AE and JP groups were computed. As Table 3 shows, the AE group showed ceiling effects for all types of consonantal sequences (Median = 96-100% correct). In contrast, the JP group's performance for all eight consonant types was consistently lower (Median = 63-79% correct) although a few JP individuals performed at or near ceiling on some types of syllable structures.

Since the AE group's performance was at ceiling on all stimulus types, within-group comparisons were performed only for the JP data. Figure 1 plots the median, inter-quartile ranges, and total ranges of discrimination scores (percent correct) across the eight types of clusters for the JP group. As can be seen in these box and whisker plots, there was great variability in individual performance on all eight cluster types. Overall accuracy was the highest for /sp/ clusters (Median = 79% correct) and lowest for /bl/ clusters (Median = 63% correct).

A nonparametric repeated measures Friedman test (Siegel and Castellan, 1988) was conducted using ranked differences among these eight types of clusters. The test revealed that JP listeners' performance differed significantly by cluster types (χ^2 (7df) = 18.21, *p* = 0.011). Percent correct scores were then grouped by the complexity of initial onsets: CC *vs* CCC. In this analysis, the scores for voiced stop + /l/ clusters were excluded since they didn't belong to CCC. Overall accuracy was numerically higher for CC clusters than CCC clusters (Median = 75% correct and 69% correct, respectively); however, a Wilcoxon signed-ranks test revealed that this difference was not statistically reliable across participants significant (z = −1.041, *p* = .298). Note also that JP listeners' performance on the CCC sequences was more variable (SD = 18.99), compared to those of CC sequences (SD = 14.50).

Within CC cluster types, the perception scores were further grouped into three different cluster types: "sC (= /sp/ and /sk/)", "−V stop + /l/ (= /pl/ and /kl/)" and "+V stop + /l/ (= /bl/ and /gl/)". A Friedman test revealed that JP listeners' performance differed significantly across these three types of CC sequences (χ^2 (2df) = 12.25, *p* = 0.02). Pairwise comparisons using a Wilcoxon signed-ranks tests revealed that there were significant differences between −V stop + /l/ and +V stop + /l/ (z = −3.26, *p* = .001) as well as between +V stop + /l/ and sC (z = −2.94, *p* = .003). However, there was no significant difference between sC and −V stop + /l/ clusters (z = −.238, *p* = .812). This shows that, within CC conditions, +V stop + /l/ clusters were significantly more difficult than other clusters.

To summarize, the perception experiment clearly demonstrated that Japanese late L2 learners of English have difficulty perceiving complex syllable onsets. The results showed that there was no difference in discrimination accuracy between CCC *vs* CC clusters or between sC *vs* other CC clusters. The observed order of relative difficulty was, thus, from easy to difficult: sC easier than –V stop + /l/ and CCC easier than +V stop + /l/. Hence, relative difficulty based on the sonority-based markedness was confirmed only for the comparison of −V stop + /l/ and +V stop + /l/ clusters.

6.2. Results of Production Experiment

6.2.1. Overall Performance

Inter-rater reliability was 94% between two judges (29 discrepancy cases out of 480 JP speakers' productions and 1 discrepancy case out of 80 AE productions) on the presence/absence of a schwa between first and subsequent consonants. Among the 29 discrepancy cases, the two judges disagreed more on the productions of two JP speakers (6 cases for JP12 and 5 cases for JP30; see Appendix L for the proportion of disagreements among individuals). However, no specific type of clusters yielded more discrepancy cases than others (see Appendix M for frequency of disagreement by cluster types). These discrepancy cases were thus regarded as incorrect productions based on the assumption that such productions were less intelligible than the ones that both judges scored as correct.

As predicted, overall production accuracy showed ceiling effects for the AE group (99% accuracy). There was only one error found for a /gl/ cluster due to the discrepancy between judges on the presence/absence of a vowel. In contrast, for the JP group, overall production accuracy was 64% (308 correct out of 480 tokens; individual data can be found in Appendix N1 and N2). The 16 stimulus types were divided into 2 large groups: the CC(C) types and the CəC(C) types. It was originally predicted that the

JP group would repair the CC(C) tokens by inserting a vowel in order to conform to native syllable structure. Thus, production errors would be greater for these members of each pair.

Contrary to this expectation, overall accuracy was higher for the CC(C) tokens than for CəC(C) tokens (83% *vs* 45% correct, respectively). Figure 2A shows percent correct for each CC(C) type. As can be seen, the JP group did relatively well in producing all 8 types of CC(C) onsets (77% to 90% correct). The /sk/ clusters were the least problematic whereas the /kl/ clusters was the most problematic. Since the production data were in the form of a frequency count (i.e., either "accurate" or "inaccurate"), the chi-square was chosen to examine whether there was a significant difference in the observed frequencies of correct productions of each cluster type. Observed frequencies of correct productions (see Appendix O for the detailed tabulation) were grouped by the following four cluster types: sC (/sp/ and /sk/), −V stop + /l/ (/pl/ and /kl/), +V stop + /l/ (/bl/ and /gl/), CCC (/spl/ and /skl/). However, the chi-square test revealed no significant differences in accuracy rate among these four cluster types (χ^2 (3df) = 1.21).

Figure 2B shows the results for each type of CəC(C) syllables. Overall, the production accuracy was lower in CəC(C) tokens for all eight syllable types (range = 33-63%). Visual inspection of this figure shows that the JP group made more errors on the +V stop + ə + l clusters. As was the case for the CC(C) tokens, the eight consonant contexts were grouped into four cluster types (see Appendix O for the tabulation): səC (/səp, sək/), −V stop + /ə/ + /l/ (/pəl, kəl/), +V stop + /ə/ + /l/ (/bəl, gəl/), and CəCC (/səpl, səkl/). A chi-square test yielded no significant overall difference (χ^2 (3df) = 2.16).

Another way to examine the production data was to examine the differences in production accuracy, grouping the CC(C) and the CəC(C) tokens for 8 syllable clusters: SP (/sp, səp/), SK (/sk, sək/), PL (/pl, pəl/), KL (/kl, kəl/), BL (/bl, bəl/), GL (/gl, gəl/), SPL (/spl, səpl) and SKL (/skl, səkl/). For this test, a "correct" score was tabulated when a participant produced both words of each pair correctly (e.g., /spani/ and / səpani/ for the "SP" score; see Appendix P for the detailed tabulation). A chi-square test again failed to reveal any significant differences among the eight syllable types (χ^2 (7df) = 10.41).

Finally, scores were grouped into four cluster types: SC (/sp, səp, sk, sək/), −V stop + L (/pl, pəl, kl, kəl/), +V stop + L (/bl, bəl, gl, gəl/), CCC (/spl, səpl, skl, səkl/). The observed frequencies of the correct productions of both the CC(C) and the CəC(C) tokens were tabulated and submitted to the chi-square test (see Appendix Q for the detailed tabulation). For instance, a correct production was tabulated if a participant produced four "SC" types correctly: /spani/, /səpani/, /skani/, and /səkani/. A series of 2 x 2 chi-square tests, performed to determine whether there was a difference in accuracy among these 4 cluster types, revealed a significant difference between the +V stop + L and the SC clusters (χ^2 (1df) = 7.95, $p < 0.01$) only. These results suggest that, for the JP group, productions of the SC clusters were the least problematic, while those of the +V stop + L clusters were the most problematic.

6.2.2. Types of Errors

As previously mentioned, the JP group performed relatively well on delayed imitation of the CC(C) tokens. In fact, there were only 41 errors out of 240 CC(C) tokens. No error types seemed to be more prominent than others (14 cases of epenthesis, 11 cases of segment change, and 16 cases of discrepancies, respectively). However, it should be noted that 3 JP participants made more errors than others (5 errors for JP7, 4 errors for JP15 and JP30) in producing these CC(C) tokens (see Appendix N1 for the individual data).

In contrast, the JP group produced more errors on the CəC(C) tokens. There were 131 errors in total. The majority of errors were vowel deletion (100 cases in total). There seemed to be no differences in occurrence of other error types in the CəC(C) tokens (9 cases of segment change, 9 cases of others, and 13 cases of discrepancy; see Appendix N2 for the individual data). With respect to vowel deletion, 13 JP participants exhibited a strong tendency to delete the intended schwa in the CəC(C) tokens in more than 50% of the tokens (see Appendix R for the proportion of vowel deletion among individuals). Vowel deletion occurred the most for the /bəl/ syllables (18 cases), followed by the /gəl/ and /səpl/ contexts (15 cases each; see Appendix S for the proportion of vowel deletion by each consonant context). In order to see whether there were effects of cluster type on production accuracy, frequencies of vowel deletion were tabulated and submitted to a chi-square test. The test showed no significant differences in frequency over the eight consonant clusters (χ^2 (7df) = 10.15). However, the general tendency to delete a vowel in the initial onset seems to have occurred in the +V stop + ə + /l/ tokens more often than the rest of three cluster types.

6.2.3. Productions of Intended /l/

Recall that transcriptions of [r] for intended /l/ were scored as “correct” for JP productions. Although the productions of /l/ were not a main focus of the current study, the frequency of producing [r] was examined to see whether it was related to the vowel deletion phenomena seen in the productions of the CəC(C) tokens.

With respect to the productions of /l/, inter-rater reliability was 90%. There were 34 discrepancy cases out of 360 tokens containing /l/ where one judge transcribed the intended /l/ as [r] or [w] while the other judge heard it as [l]. The overall proportion of productions of [l] for the intended /l/ in both CC(C) and CəCC tokens was only 54% (195 out of 360 tokens). These results suggest that almost half of the productions of the intended /l/ by JP speakers were heard as [r] by American English listeners.

With respect to the CC(C) tokens, 44% of the intended /l/s were produced as [r] (e.g., [prani] for /plani/; see Appendix T for the frequency of the [r] productions by cluster types). The [r] replacement seemed to have occurred more frequently in CCC contexts (18 cases for /spl/ and 17 cases for /skl/; see Appendix U). A one-sample chi-square test was performed to see whether the /r/ replacement varied significantly across cluster types. The test showed no significant differences over 6 consonant clusters with intended /l/ (χ^2 (5df) = 4.75, p = .447) or 3 cluster types (χ^2 (2df) = 4.38, p = .112). Thus, there appeared to be no interaction of the /l/ production errors and consonant cluster types.

Table 4 shows the summary of the production errors on syllables containing intended /l/ in CəCC tokens. Almost half of the tokens were produced without a vowel (78 out of 180 tokens). Among them, 40 tokens were produced with [r] (e.g., [prani] for /pelani/). The observed frequencies of such cases seemed to be especially high in +V stop + /ə/ + /l/ and CəCC syllable types (15 and 17 cases, respectively). A one-sample chi-square test was performed to see the effects of cluster type on observed frequencies of vowel deletion plus replacing [r] for the intended /l/. The test showed no significant differences across six cluster types (χ^2 (5df) = 4.40, p = .493) or 3 cluster types (χ^2 (2df) = 3.35, p = .187). These results suggest that JP speakers' tendency to delete a vowel and to replace /l/ with [r] did not interact with cluster content or its complexity.

Additionally, it should be noted that the frequencies of the [r] replacement for both CC(C) and CəC(C) were not normally distributed among individuals. The average frequency of the [r] replacement was 5.5 among 30 JPs (SD = 3.95, range = 0-12). Specifically, almost half of the JP speakers exhibited a strong tendency for the [r] replacement (see Appendix V for the observed frequencies of the [r] replacement among 30 JP speakers).

To summarize, JP speakers showed a strong tendency to produce intended /l/ as [r] for CC(C) tokens (44% of all tokens). In addition, the JP speakers often deleted a vowel and produced either [l] or [r] for the CəC(C) tokens (43% of all the tokens). In other words, the JP speakers' productions of clusters with the intended /l/ were often transcribed as either [stops + r] (e.g., [brani]) or [s + stop + r] (e.g., [skrani]), regardless of the presence of /ə/ in the target stimuli.

The following results were revealed in this section: 1) AE productions were accurately transcribed as the intended utterances by both judges (99 % correct overall); 2) JP productions were judged to be incorrect much more often (64 % correct overall); 3) for JP speakers, there were significantly more errors on CəC(C) tokens than on CC(C) tokens (45% and 83% correct, respectively); 4) among the four cluster types: SC, −V stop + Ll/, +V stop + L, and CCC, the productions of the +V stop + L clusters were the most problematic for the JP group while those of the SC clusters were the least problematic; 5) almost half the productions of intended /l/ were heard by native listeners as [r]. However, there were no apparent interactions with syllable structure errors; 6) large individual differences in production accuracy across the 30 JP participants were observed.

6.3. Correlations between L2 Perception and L2 Production

In order to investigate whether JP participants' overall performance on the perception task was associated with their production performance, Spearman rank-order correlations were performed. Overall percent correct scores on the perception and production tasks served as the variables. Figure 3 shows the scatterplot of the JP group's overall performance. The correlation was positive and relatively strong (rho = +0.78, $p < 0.01$, one-tailed), with over 60% of the variance accounted for. This result shows that perception performance was highly predictive of production performance.

6.4. Correlations of Overall Perception and Production Performance with English Experience and Proficiency

Spearman rank order correlations were performed to determine how JP participants' language experience and proficiency were associated with their overall performance on the perception and production tasks. Overall percent correct on the perception and production tasks were again chosen as performance variables. Two separate sets of correlation analyses: one with perception performance and the other with production performance were computed. This section reports the results in the following order: 1) correlation between JP participants' LOR and their perception and production performance; 2); correlation between JP participants' self-reported amount of L2 use and their perception and production performance 3) correlation between JP participants' Versant test score and their perception and production performance.

6.4.1. Correlations with Length of Residence (LOR)

JP participants' overall performance on the perception task did not correlate significantly with their LOR in an English speaking country (rho = + 0.213, *ns*). The correlation of LOR with performance on the production task was significant (rho = +0.325, $p < 0.05$) but only moderate (only 11% of the variance accounted for). See Appendix W and X for the scatterplots. In the present study, Japanese participants' LOR was not systematically sampled. Whereas LOR varied from 1 month to 19 years, there were 13 participants who had resided in English speaking countries less than 1 year at the time of testing. One possible explanation for the lack of correlation between Japanese listeners' performance and their LOR might have been due to the small sample size for those with longer LOR. However, the great variability in performance for recent arrivals argues against this hypothesis.

Another explanation might be due to differences in the amount of exposure to native English among those with shorter LOR. Specifically, two participants (JP18 and JP29; see Appendix A for individual data) had 80% accuracy for both perception and production tasks despite the fact that they had resided in the U.S. less than 1 year. These participants reported that they had had considerable exposure to native speakers' English while they lived in Japan. Thus, in the present study, LOR might not serve as a good index for English conversational experience.

6.4.2. Correlations with L2 Use

Recall that the majority of JP participants reported that they spoke Japanese and English in about equal proportions (Median = 50%; see Table 5). The percentage on reading and writing in English (Median = 90%) was slightly higher than that on listening and speaking (Median = 80%) although both varied a great deal among individuals.

Spearman rank order correlations of the overall amount of English use with performance on the perception task (rho = − 0.154) and production task (rho = − 0.161) were non-significant and very low. Correlations of the amount of reading and writing in English with performance on the perception task (rho = + 0.185) and production task (rho= + 0.067) were positive but, again, non-significant and very low. The correlation of the amount of speaking and listening with performance on the perception task (rho = − 0.129) was also negative and low. However, there was a moderate negative correlation of the amount of speaking and listening with performance on the production task (rho = −. 338, $p < 0.05$). An inspection of individual subject data suggests great variability in the self-reported amount of listening and speaking. For instance, there were 11 ESL students who attended private language schools in the New York City area. Despite the fact that their daily English experience at school must have been similar, the self-reported amount of listening and speaking varied from 25 to 100% (SD = 27.39). These data suggest that, in the present study, the self-reported amount of English usage might not serve as a good index for accounting for overall performance in both perception and production task.

6.4.3. Correlations with Versant Test Performance

As mentioned earlier, each JP participant completed the Versant test via telephone. Table 6 shows the summary statistics for the Versant test scores of the JP group. Overall, JP participants' abilities to understand and speak English were considered as low-intermediate (Median = 45 points). The following is the description for those who score at 45 points;

"Test-taker can handle short utterances using common words and simple structures, but has difficulty following a native-paced conversation. Pronunciation may sometimes not be intelligible; test-taker speaks slowly and pauses, but can convey basic information to a cooperative listener." (http://www.ordinate.com).

The Versant test score was further subdivided into four types of English skills; sentence mastery, vocabulary, fluency, and pronunciation. As Table 6 shows, the mean score for each skill ranged from the low 40s to the low 50s. Specifically, JP participants' overall fluency was evaluated as rather low (Median = 42 points), characterized as "an irregular speech rate and some disconnected phrases" (http://www.ordinate.com). As was the case for the overall performance on the current perception and the production tasks, the Versant test scores were also highly variable across JP participants as suggested by the high standard deviation and ranges.

Figure 4 illustrates the relationship between overall performance on the Versant test and performance on the perception (A) and production (B) tasks. The correlation of the overall Versant score with production performance (rho = + 0.633, $p < 0.01$) was slightly higher than those with perception performance (rho = + 0.470, $p < 0.01$). Separate Spearman-rank order correlations were performed to determine which of the four English subtests scores were associated most strongly with performance on perception and production tasks. All correlations were positive. However, rather low correlations of some subtests with the overall performance on the perception task were observed (rho = + 0.394 for fluency skill and + 0.396, $p < 0.05$ for pronunciation skill, respectively). The correlations of fluency and pronunciation skills with overall performance on the production task were higher than those with the perception task (rho = + 0.541 for fluency skill and + 0.493 for pronunciation skill, both $p < 0.01$). Stronger correlations were found between vocabulary skills and overall performance on the experimental tasks (rho = + 0.528 with perception and + 0.657 with production, both p <

0.01) as well as sentence mastery skills and the overall performance on these tasks (rho = + 0.544 with perception and + 0.656, with production, both $p < 0.01$). In short, the overall Versant scores were correlated with both production and perception scores, with slightly stronger correlations of production scores with measures of English spoken language proficiency. Specifically, overall vocabulary and sentence mastery skills were most highly correlated with overall performance on the two tasks. These findings suggest that those who performed well on the perception and the production tasks had larger English vocabularies and good command of syntactic structure.

The following is the summary of the findings on the correlations of the JP participants' language background and proficiency with their performance in the experiment: 1) only the production scores were correlated significantly with LOR; 2) the self-reported amount of English use was not correlated significantly with either production or perception scores; 3) overall Versant scores were correlated significantly with overall performance on both production and perception tasks. Specifically, vocabulary and sentence mastery subscores predicted overall performance on the two experimental tasks, with 31% to 45% of the variance accounted for.

CHAPTER 7: GENERAL DISCUSSION

7.1. Results of the Perception Test and their Implications

Results of the perception experiment clearly showed that JP listeners had difficulty accurately perceiving complex syllable onsets. Overall accuracy by the JP group was significantly poorer than that of the AE group. Specifically, JP listeners showed most difficulty perceiving +V stop + /l/ clusters. In contrast, overall performance on sC clusters was relatively good for the JP group. Overall performance on –V stop + /l/ and CCC cluster types was intermediate.

One possible explanation of the observed order of difficulty in the current data would be that the syllable structure of sC clusters is linguistically different from other CC clusters. As mentioned earlier, the status of the /s/ in English initial onsets has been controversial. Some say that /s/-clusters have the same branching onset as other complex onsets (e.g., Boyd, 2006) while others claim that the initial /s/ is not a part of an onset (e.g., Kenstowicz, 1994). Specifically, Vaux and Wolfe (2009) propose that the initial /s/ is "an appendix", directly connected to the higher prosodic nodes (however, see Cairns, 2009 for his critique of Vaux and Wolfe and his analysis based on Cairns and Feinstein's [1982] syllable structure theory). Suppose that the initial /s/ in /spani/ is an appendix. Then, the remaining syllable structure is CV for /s.pa.ni/ and CCV for /s.pla.ni/. Thus, /s.CV/ might be structurally more complex than /CV/, a syllable with the simple onset but less complex than /CCV/, a syllable with the branching onset. If so, the predicted relative difficulty based on the /s/ as an appendix theory might be as follows: /V/ > /CV/ > /s.CV/ > /CCV/ > /s.CCV/. This analysis would accommodate the current perceptual data of sC clusters in the JP group, yet it wrongly predicts sCC clusters as the most difficult.

Perhaps, regardless of differences in the remaining syllable structure after the initial /s/, what listeners have to pay attention to is whether there is a schwa between /s/ and C for both sC and sCC clusters under the current paradigm (L. Davidson, personal communication, April 9, 2010).

Another possible account might be derived from a consideration of listeners' L1 phonetic realization rules. Vowel devoicing in Japanese phonology is a well-known phenomenon. In general, Japanese high vowels /i/ and /u/ are devoiced when surrounded by voiceless consonants (e.g., Kondo, 2005). Suppose that Japanese perceptually epenthesize a devoiced vowel between /s/ and the following voiceless stop in sC clusters, as suggested by Dupoux et al (1999). If so, perceptual discrimination between /səpani/ heard as "sepani" with a voiced vowel and /spani/ heard as "supani" with a devoiced vowel might have been relatively easy for Japanese listeners. Along this line of argument, discrimination of sCC *vs* səCC (e.g., /splani/ *vs* /səplani/) clusters should also be easy, since again, listeners could differentiate them on the basis of a voiced *vs* devoiced perceptually-epenthesized vowel. However, overall performance on sCC *vs* səCC clusters was lower than that on sC *vs* səC clusters, arguing against this hypothesis. The presence of the liquid in the sCC *vs* səCC clusters may have prevented them from attending to the nature of the vowel between the /s/ and stop. As some JP participants informed the experimenter, Japanese-speaking learners of English are often aware of their difficulties in perceptually discriminating English /r/ and /l/. The presence of the "notoriously difficult" liquid might have distracted some JP listeners, even though it was not involved in the contrastive cues for the syllable types.

With respect to discrimination accuracy between −V stop + /l/ and +V stop + /l/ clusters, the predictions made by both markedness and L1 phonetic interference hypotheses were consistent with the results. According to the former hypothesis, the sonority distance between two phonemes plays a major role in perceptual discrimination. The sonority distance between the initial two phonemes was smaller in +V stop + /l/ clusters than their voiceless counterparts and thus, more marked. However, from a phonetic perspective, the availability of distinctive acoustic cues in the stimuli was also reflected in these results. Specifically, −V stop + /l/ clusters had an additional correlated VOT cue whereas for +V stop + /l/ clusters, VOT values almost completely overlapped for the contrasting CCV *vs* CəCV tokens. In the current experiment, the mean VOT values of voiceless stops in the stimuli were categorized as a "correlated cue" because of a slight overlap in durations (mean VOT range = 55-96 ms for the /pl and kl/ *vs* VOT range = 37-56 ms for the /pəl and kəl/).

Whether this VOT cue accounted for the better discrimination performance on −V stop + (ə) + /l/ distinctions remains to be determined. Ito and Strange (2009), who explored effects of listeners' L1 on word segmentation (e.g., " keep*s t*alking" *vs* " keep *st*alking"), reported that Japanese L2 learners could not take advantage of the VOT stop cue for word segmentation (aspirated for syllable initial stops – unaspirated for stops following /s/). One can speculate that the difference in VOT values in the voiceless stop stimuli used in the present experiment might not have been perceived as a distinctive cue by the Japanese listeners. In other words, the additional VOT cue in the −V stop + /l/ stimuli might have been *not salient enough* for Japanese listeners. If so, this phonetic realization difference cannot account for the better perception of syllable structure differences in C(ə)C stimuli with initial voiceless stops.

A recent study by Riney, Takagi, Ota, and Uchida (2007) reported that the mean VOT values of Japanese word-initial voiceless stops: /p, t, k/ ranged from 28.5 ms to 56.7 ms. Preliminary acoustic results of the productions analyzed in the present study revealed that Japanese speakers did not make a distinction in VOT values when producing /plani/ *vs* /pəlani/. The mean VOT values for /plani/ was 54 ms *vs* 46 ms for /pəlani/. Such durational differences were not statistically significant ($z = -1.331$, $p = .193$). Thus, in production, Japanese apparently did not produce the allophonic variations in VOT (aspiration) that native English listeners make for initial voiceless stops *vs* voiceless stops following /s/.

Interestingly, there is yet another possible account of perception of consonant clusters, which might be relevant to the current study. Fleischhacker (2005) showed cross-linguistic tendencies to "split" stop-liquid clusters but "reserve" fricative-stop clusters based on several perceptual experiments. For instance, native English speakers were asked to perform a similarity-rating task on CCVC *vs* CəCVC comparisons using a 7-point Likert scale (7 as the most similar). Results showed that English speakers rated [plʌk] *vs* [pəlʌk] pairs as more perceptually similar (mean rating = 5.36) than [stok] *vs* [sətok] comparisons (mean rating = 3.10). In discussing this study, Zuraw (2007) stated, "Fleischhacker assumes that splitting a cluster at a stronger perceptual break creates a smaller perceptual departure from the unsplit original" (p. 281). That is, "a perceptual break" is stronger in stop-liquid sequences because formant structures are strong in liquids whereas it is weaker in fricative-stop sequences because of the absence of formant structure (i.e., silent gap) (see Zuraw, 2007, p. 280-281 for a summary of Fleischhacker's study). Thus, according to Fleischhacker (2005), cross-linguistically, speakers have a strong tendency to modify fricative-stop clusters by prothesis (e.g., espain for "Spain"). If the perceptual break account is also true for Japanese listeners in the current study, the [plani - pəlani] comparison must have been perceptually similar to each other.

To summarize, contrary to the results of past research that have demonstrated effects of universal markedness in L2 perception (Berent et al., 2007), the predictions made by markedness in the present study were not generally confirmed. Only one prediction concerning the comparison of −V stop + /l/ *vs* +V stop + /l/ cluster types was supported. The results of the current perception experiment suggest that phonetic distinctiveness might have played a major role in perception of non-native consonant clusters. The observed order of perceptual difficulty among JP listeners: sC easier than −V stop + /l/ equal to CCC easier than +V stop + /l/ was consistent with the predictions made by L1 phonetic interference hypotheses.

7.2 Results of the Production Test and their Implications

JP speakers' productions were judged by native English listeners as incorrect much more often than AE speakers' productions (64% *vs* 99% correct, respectively). Among the four cluster types, the most errors were found on +V stop + /(ə) l/ clusters whereas the least errors were found on s-(ə)-C clusters overall. Most of all, JP speakers' productions were heard as making significantly more errors on CəC(C) tokens than on CC(C) tokens (45% and 83% correct, respectively).

Recall that the current production study employed the delayed imitation task. In this task, participants were asked to produce the target word after the presentation of auditory stimuli as models. No orthographic aids were provided. Thus, JP participants' production errors might have been due to their misperception of the target stimuli. Peperkamp and Dupoux (2003), who explored the loanword adaptation mechanism, states "a good approach to obtain 'clean' loanword data" is to present nonce words auditorily and ask them to produce what they heard (Peperkamp and Dupoux, 2003, p.369-370). The approach in the present production study was this "clean" form of loanword adaptation. If L2 production is a mirror image of L2 perception as suggested by Peperkamp and Dupoux (2003), relatively good performance on non-native CC(C) tokens might have been because some JP participants were able to perceive CC(C) sequences accurately. In contrast, poorer performance on CəC(C) tokens might have been due to their misperception as CC(C) structures. Consequently, these interpretations on the current production data based on Peperkamp and Dupoux (2003) would obscure the claims that L1 phonological knowledge constrains L2 perception. Dupoux et al. (1999) suggested that Japanese speakers' strong tendency to add a vowel between consonant sequences in production originates in their perception, which is influenced by their L1 phonotactic knowledge. Repairing CVCV syllable structures (legal in L1) into CCV structures (illegal in L1) contradicts this phonological process. However, that does not necessarily deny perceptual effects. In other words, JP speakers' production errors might not have been due to their repair strategy. Then, why would some JP speakers repair CəC(C) tokens as CC(C)?

As Strange (2009) claims in her ASP model, to predict perceptual difficulties in a phonetic mode of processing, one needs to consider the specific phonetic realizations of target segments, which vary as a function of the phonetic and prosodic context in which they are produced, as well as the register and rate of speech used by the speakers. The schwas in our current stimuli were short and unstressed. Thus, it is possible that these schwas were not parsed by Japanese listeners in the same manner as AE listeners did, despite the fact that the target stimuli had the legal syllable structure in L1 phonology. Hence, as Chang et al. (2007, p. 800) suggested, L1 phonotactic constraints might not be the only "decisive factor" in perception and production of non-native consonant sequences for the current study.

Another possible account would be that AE judges did not hear JP speakers' epenthesized schwas in CC clusters as "V" even though there was a schwa-like gesture between the initial and the second consonant. Davidson (2007) points out that "degraded" schwas might not be transcribed as vowels by native English speakers. In her 2007 study, AE speakers were asked to transcribe pseudo-Slavic words. Of interest here was how AE speakers transcribed the C^{ə}C tokens (e.g., /z^{ə}vaba/). The schwas in these items were called "transitional" and were shorter in duration than those of lexical schwas (48 ms *vs* 68 ms, respectively; Table 1, p. 265). The results of the transcription task showed that the C^{ə}CVCV tokens were often transcribed as [CCVCV]. In a subsequent AX discrimination task, listeners often confused clusters and clusters with transitional schwas. If such confusion occurred for the AE judges in the current study, it is possible that Japanese produced "transitional schwas" (e.g., [b^{ə}lani])" in intended CC clusters, but they were nevertheless transcribed as “correctly” not containing a vowel (e.g., heard as [blani] or [brani]).

A recent study by Davidson and Shaw (2009) shows the complicated nature of the link between native English speakers’ perception and production of nonnative sequences. In their study, English speakers were asked to perceptually discriminate between CCV clusters (e.g., /spano/) *vs* a variety of modified clusters such as an epenthesis (e.g., /səpano/), prothesis (e.g., /əspano/), and C1 deletion (e.g., /pano/). Results showed that English listeners’ perception was affected by the type of initial consonant. For instance, English listeners showed perceptual confusion between CCV *vs* əCCV structures, a prothetic modification, for fricative-stop clusters. On the other hand, for stop-stop clusters, English listeners’ perceptual confusion occurred on both the prothetic modifications (e.g., /tkano/ - /ətkano/) and the epenthetic modifications (e.g., /tkano/ - /təkano/). These data show that English speakers do not necessarily perceive non-native clusters as they are, and that the phonetic details of the consonants affect the types of perceptual confusions made.

Preliminary results of acoustic analysis on /spani/ and /səpani/ tokens show that, as the transcription data revealed, some JP speakers “deleted” the schwas completely in /səpani/ tokens (N = 11). That is, there was no voiced vocalic segment visible on the spectrograms of the productions. Interestingly, a comparison of the mean durations of visible segments in productions by the JP and the AE groups showed that the consonant sequences were not always produced with the same temporal structure. For instance, with respect to the intended /spani/ tokens, the durations of the silent gap between /s/ frication offset and stop burst onset were longer on average for the JP group (mean = 97 ms) than for the AE group (mean = 70 ms). The mean durations of /s/ and VOTs were similar for two language groups. However, there was great variability in JP speakers' productions (Range = 5-49 ms for JP speakers' VOTs). These preliminary results suggest that the JP speakers might have inserted "unvoiced schwa" between [s] and [p], as indicated by the longer period of silence. That is, some JP speakers may have repaired illegal consonant sequences by epenthesizing unvoiced schwa. However, AE judges transcribed these productions as CC because they did not perceive a voiced vowel between the first two consonants. Additionally, a strong tendency to de-aspirate the voiceless stops after /s/ by JP speakers might have affected AE judges' transcriptions. That is, in the absence of an aspirated stop, AE listeners heard /s/ + unvoiced vowel + de-aspirated stop as /s/ + stop clusters. These preliminary acoustic results suggest that, even though productions of sC tokens by the JP group were perceived as "correct" (around 90%), they were not produced in a native-like manner. Follow-up studies comparing native listener judgments with the acoustic structure of the productions will be needed to further determine how the L2 speakers produced non-native syllable structures and how specific production deviations led to perceived misproductions *vs* "correct" productions. In addition, productions by Japanese speakers of these syllable types when presented with orthographic representations of the target syllables will be needed to tease apart their production and perception difficulties in the present study. In the next section, the link between perception and production is discussed further.

7.3 The Link between L2 Perception and L2 Production

Flege's Speech Learning Model explicitly mentions the link between L2 perception and L2 production, claiming that the accuracy in L2 perception affects that of

L2 production. For instance, Flege, Mackay, and Meador (1999) tested Italian-English bilinguals on production and perception of English vowels and demonstrated that there was a significant correlation between these two abilities. The current results further confirm such a relationship in the acquisition of L2 phonotactic structures. There was a strong correlation between overall perception and production accuracy by these Japanese L2 learners. The correlation between perception and production performance was also reflected in the observed order of difficulty for the two experimental tasks: sC easier than −V stop + /l/ and CCC which were easier than +V stop + /l/. However, unlike the results of past studies on L2 phonotactics (e.g., Dupoux et al., 1999), the results of the current study show that the inaccuracy in L2 production is not necessarily due to the phonological repair process. That is, L2 learners with a limited L1 syllable inventory do not always repair L2 CCV clusters into productions that are heard as CVCV structures, as shown in the current JP participants' productions of CCVCV items (perceived as correct 87% of the time).

As discussed briefly above, preliminary results of acoustic analysis detected some cases of transitional unvoiced schwas in /s/ + stop clusters, which would indicate a repair process that was not perceived by native listeners. Other examples were found for stop + /l/ clusters. For instance, a Japanese production of intended /glani/ was transcribed as [glani] by native speakers of English and thus, was scored as "correct". However, the spectrogram showed about 40 ms of voicing before [l] whereas that of a native AE speaker did not (see Appendix Y for JP4 and Appendix Z for AE4's production of [glani]). These "transitional schwas" in consonant clusters are probably due to "gestural mistiming" (Davidson, 2006), and though undetected as vowels, might lead native listeners to hear the consonant clusters as "correct" but "accented." Further follow-up work asking native listeners to judge the accentedness of the Japanese productions will shed light on the extent to which differences in gestural timing of consonant clusters affects native language perception. These examples demonstrate that the relationship between L2 perception and L2 production is a complicated one. Does it mean that these transitional schwas occurred because of JP speakers' difficulty in timing gestures in unfamiliar consonant sequences (motor constraints), rather than their misperception of the intended structures? Future research is needed to tease apart the relative contributions of misperception and production mistiming on L2 production patterns. If the former factor seems more dominant, L2 learners of English need more training on improving their perception. If the latter factor is dominant, L2 learners need more training on articulatory temporal coordination.

7.4 Individual Performance and English Experience/ Proficiency

Following Jenkins' Tetrahedral Model (Jenkins, 1979, cited in Strange 1992), Strange (1992) states that the outcome of a study is a result of interactions among variables such as participants, types of tasks, and types of materials. With respect to variability among subjects, the following four variables are typically mentioned in studies of L2 phonological learning: L1 experience, L2 experience, age, and "talent" (see Table 1, Strange 1992, p. 200). The first three variables are often measured in terms of participants' AOA, LOR, and self-reported amount of L1/L2 use. In general, it is assumed that, as AOA increases, L2 production/perception performance will decrease (e.g., Munro, Flege, & MacKay, 1996) and the less L2 is used, the less accurate L2 perception/production will be (e.g., Flege, MacKay, & Meador, 1999). However, mixed results have been reported with respect to LOR. Some research showed no significant correlations between LOR and L2 perception/production (e.g., McAllister, 2001) while others have demonstrated positive correlations between these variables (e.g., Flege, 1988).

Japanese participants in the current study were considered “late-learners of English (mean AOA = 29 years old)” who had passed the so-called “a sensitive period” for learning L2. Some researchers infer that such a period ends around the age 12 years (e.g., Lenneberg, 1967) or even before the age 12 (e.g., Flege, 1988), while others doubt the existence of such a period based on “exceptional late L2 learners” who exhibit native-like performance (e.g., Ioup, Boustagi, El Tigi, & Moselle, 1994). In the current study, no significant correlations between AOA and overall perception performance (rho = − 0.245) or AOA and production performance (rho = − 0.233) were found, confirming that, when past the sensitive period, AOA does not predict L2 perception/production performance. A replication of the current study with those whose AOA were younger might further reveal whether AOA serves as a predictor for acquiring phonological units larger than segments. However, as Moyer (2009, p. 160-161) points out, AOA might be, nonetheless, an unreliable measure of L2 experience since “socio-psychological orientation factors (e.g., motivation, attitudes toward the target language, etc.)” might be different, depending on whether one is exposed to L2 at an early age or at a later age. Those who are exposed to L2 at an early age might have different attitudes toward the language “culturally and socially” (Moyer, 2009, p. 161). Thus, it is unclear whether AOA or “socio-psychological orientation factors” play a major role.

With respect to effects of the self-reported L2 use on overall performance, correlations of the overall amount of English use with performance on the perception task (rho = − 0.154) and production task (rho = − 0.161) were non-significant and very low. As mentioned earlier, large individual differences in reporting the amount of L2 use on a daily basis were observed (Median = 50%; Range = 5-97%). Japanese participants were asked to indicate the amount of L2 use in percentage but some participants might have been conservative while others might not have. This variability in the self-reported amount indicates the difficulty in accurately measuring this variable. Hence, the self-reported L2 use might not be a good index to account for L2 experience in the current study.

Concerning effects of LOR on performance in the two tasks, there was no significant correlation between JP participants' overall performance on the perception task and LOR (rho = + 0.213, *ns*). The correlation of LOR with performance on the production task was only moderate (rho = +0.325, $p < 0.05$; only 11% of the variance accounted for). Flege and Liu (2001) suggested that the "quality" of L2 input might be different for student *vs* non-student participants. In their study, LOR was strongly correlated with overall performance only for participants who were in a school setting at the time of testing. Another Spearman rank order test of the present data was performed to determine whether there was a difference in the strength of the correlations between overall performance and LOR for student *vs* non-students. In the current study, there were 9 participants who were non-students and 19 participants who were students at the time of testing. The data of 2 participants (JP18 and JP29) who reported that they had had considerable English exposure to native speakers while living in Japan were excluded from this analysis. The results for those who were in school showed that their LOR was significantly correlated with performance on the perception task (rho = + 0.511, $p < 0.05$) but not with performance on the production task (rho = + 0.375, $p > 0.05$). For non-students, neither perception or production performance correlated significantly with LOR (rho = −0. 234 and +0.135, respectively) but the small sample size might have affected the outcome. Non-significant correlations between overall performance and LOR for non-students suggest that the amount of L2 input might have been different between students and non-students, which are consistent with the findings of Flege and Liu (2001). Future research with systematic sampling of participants' LOR might further clarify such relationships.

Finally, the Versant test appeared to be a promising index to assess L2 learners' English conversation skills. Overall scores on this test, as well as all four sub-scores, pronunciation, fluency, vocabulary, and sentence mastery, were correlated significantly and strongly with overall performance on both production and perception tasks. Specifically, vocabulary and sentence mastery sub-scores predicted overall performance on the two experimental tasks most strongly. Of particular interest here are the correlations between vocabulary and performance on the two experimental tasks. Such strong correlations between vocabulary size and L2 perception/production is consistent with the claims made by PAM-L2 (Best and Tyler, 2007), which suggest that the size of the L2 vocabulary affects L2 perception. Bundgaard-Nielsen, Best, and Tyler (2009) recently demonstrated that JP listeners' overall performance on identification and discrimination of Australian English vowels was highly correlated with the size of their L2 vocabulary; those who had larger English vocabularies showed better performance on the perception tasks. However, vocabulary size may not be a direct cause in improving JP participants' perception or production skills. Rather, it may indirectly influence L2 perception/production because those who have large vocabularies are probably those individuals who actively seek additional L2 input by reading in L2 and/or listening to L2. Thus, L2 learners' amount of reading and listening will be reflected in their perception/production abilities, as well as in their knowledge of the lexicon.

7.5. Conclusions

The main goal of the current study was to investigate how L2 perception related to L2 production in acquisition of non-native consonant sequences. The second goal was to examine what constraints (i.e., markedness *vs* L1/L2 phonetics) were at work in this process. Japanese participants clearly showed difficulty perceiving and producing English complex syllable onsets. The observed order of relative difficulty was more consistent with the predictions made by a L1 phonetic interference hypothesis. The results demonstrated that the JP group had less difficulty perceiving/producing /s/ plus stop clusters than other clusters containing English /l/. Second, among clusters containing /l/, overall accuracy on voiced stop + /l/ clusters was significantly lower, with no difference in performance accuracy between voiceless stop + /l/ clusters and /s/ + voiceless stop + /l/ clusters. In contrast, relative difficulty based on the sonority-based markedness was confirmed only for the comparison between voiceless stop + /l/ *vs* voiced stop + /l/ clusters. Contrary to the predictions made by markedness constraints, there was no significant difference between CC *vs* CCC clusters or between sC clusters and other CC clusters. These findings suggest that markedness cannot, by itself, account for relative perception difficulty of English complex onsets.

Great variability among Japanese participants was also observed. Versant test scores served as a reliable measurement of English fluency in the JP group. Specifically, in accordance with the PAM-L2, those who scored high on vocabulary on the Versant test did well on the two experimental tasks, which suggest that the amount and quality of L2 input by reading and listening influence L2 learners' perception and production acquisition patterns.

Lastly, another important finding of the current research was that there was a link between L2 perception and L2 production in acquisition of L2 phonotactic structures. However, although productions of CCVCV tokens were, by and large, heard as accurate by native listeners, their productions were not necessarily accent-free. As preliminary acoustical measures of these productions suggest, longer periods of silence between /s/ and stops for sC clusters might have been due to "devoiced vowel epenthesis". Moreover, there might have been transitional schwas in other CC clusters, which were not detected by AE listeners. Hence, the picture of L2 perception-production link is not a simple matter. Future research with acoustic analysis of L2 productions, in comparison with native productions will further reveal which factors (i.e. misperceiving or misarticulating) are more dominant in production of non-native clusters.

8. TABLES

Table 1

Acoustic measurement of schwas in the /C + ə + C(C)/ stimuli – mean duration (in ms) and formant frequencies (in Hz).

Cluster type	Mean duration	(Range)	Mean F1	(Range)	Mean F2	(Range)
/səp/	24	(23-26)	514	(511-519)	1860	(1842-1887)
/sək/	33	(31-37)	456	(452-464)	1950	(1906-2031)
/pəl/	30	(18-46)	611	(532-729)	1263	(1230-1328)
/kəl/	48	(36-57)	684	(651-738)	1346	(1306-1407)
/bəl/	44	(42-45)	620	(693-636)	1322	(1278-1368)
/gəl/	52	(49-56)	640	(601-669)	1452	(1425-1485)
/səpl/	30	(25-33)	493	(484-503)	1758	(1719-1790)
/səkl/	45	(41-49)	502	(474-516)	2020	(1986-2050)

Table 2

Summary of duration differences across four contrast types. D = differentiating acoustic parameters, C = Correlated acoustic parameters, NS = nonsignificant difference.

A.	/s/ + voiceless stops		*vs*	/s/ + /ə/ + voiceless stops		
Duration (ms)	Mean	Range		Mean	Range	Parameters
Frication	145	(132-153)		125	(113-132)	C
Schwa	0			29	(23-37)	D
Silence	67	(53-85)		77	(60-98)	NS
VOT	12	(7-18)		46	(41-50)	D

B.	Voiceless stops + /l/		*vs*	Voiceless stops + /ə/ + /l/		
Duration (ms)	Mean	Range		Mean	Range	Parameters
VOT	71	(55-96)		48	(37-56)	C
Schwa	0			39	(18-57)	D
[la]	196	(188-203)		241	(218-275)	D

C.	Voiced stops + /l/		*vs*	Voiced stops + /ə/ + /l/		
Duration (ms)	Mean	Range		Mean	Range	Parameters
VOT	16	(8-25)		20	(7-35)	NS
Schwa	0			48	(42-56)	D
[la]	228	(221-237)		257	(246-268)	D

D.	CCC		*vs*	CəCC		
Duration (ms)	Mean	Range		Mean	Range	Parameters
Frication	116	(98-130)		115	(101-127)	NS
Schwa	0			38	(25-49)	D
Silence	74	(54-98)		72	(57-91)	NS
VOT	16	(8-23)		46	(37-56)	D
[la]	185	(140-215)		167	(153-174)	C

Table 3

Results of the perception task by each consonant cluster for two language groups: American English (AE) and Japanese (JP). IQ stands for interquartile.

	AE Group				JP Group			
	Mean	SD	Median	IQ Range	Mean	SD	Median	IQ Range
sp	100	0	100	100-100	80	12.48	79	70-92
sk	98	2.19	96	96-100	71	16.89	71	58-88
pl	98	2.19	100	96-100	76	13.64	75	63-89
kl	98	2.19	100	96-100	73	13.45	75	66-84
bl	100	0	100	100-100	64	14.33	63	54-72
gl	98	2.19	100	96-100	70	15.16	71	57-79
spl	98	5.37	100	94-100	70	18.93	67	54-93
skl	99	1.79	100	98-100	74	19.18	71	54-92

Table 4

Summary of frequency of the /l/ replacement by three cluster types: −VəL = voiceless stop + ə + l and +VəL = voiced stop + ə + l. Others included all the errors that were not related to the replacement of the intended /l/.

	−VəL	+VəL	CəCC	Total #	Proportion
Correct (e.g., [pel]	27	19	17	63	35%
Replacement with R (e.g., [per]	8	4	14	26	14%
Deletion-L (e.g., [pl]	12	18	8	38	21%
Deletion-R (e.g., [pr])	8	15	17	40	22%
Others	5	4	4	13	7%

Table 5

Summary of Japanese participants' amount of English usage on a daily basis. Participants reported these numbers in percentage.

	Overall L2 use	Speaking & Listening	Reading & Writing
Mean	52	69	78
SD	23.2	29.4	26.2
Median	50	80	90
IQ Range	38-70	50-90	70-96
Range	5-97	5-100	5-100

Table 6

Summary of Japanese Participants' Versant Test Score. The score is computed on a scale of 20-80. IQ stands for interquartile.

	Overall	Sentence Mastery	Vocabulary	Fluency	Pronunciation
Mean	46	47	51	42	45
SD	10.9	10.8	13.5	12.5	10.8
Median	45	45	51	42	46
IQ Range	36-52	39-53	39-58	31-52	34-53
Range	30-67	29-73	22-80	23-71	30-63

9. FIGURES

Figure 1

Japanese group's overall percent correct on the perception task by each cluster type.

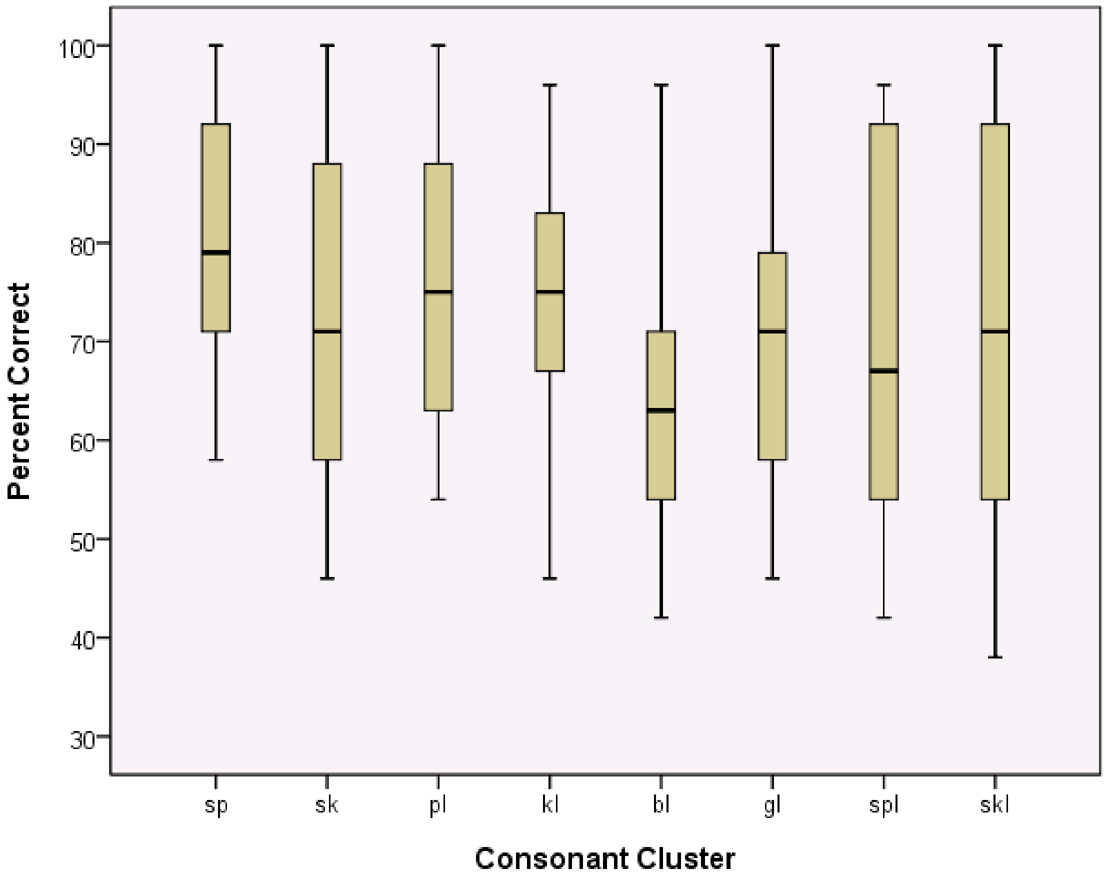


Figure 2A

Overall percent correct of the JP group's production data for the CC(C) tokens evaluated by AE listeners. Percent correct was calculated by the number of correct productions summed over 30 JP participants.

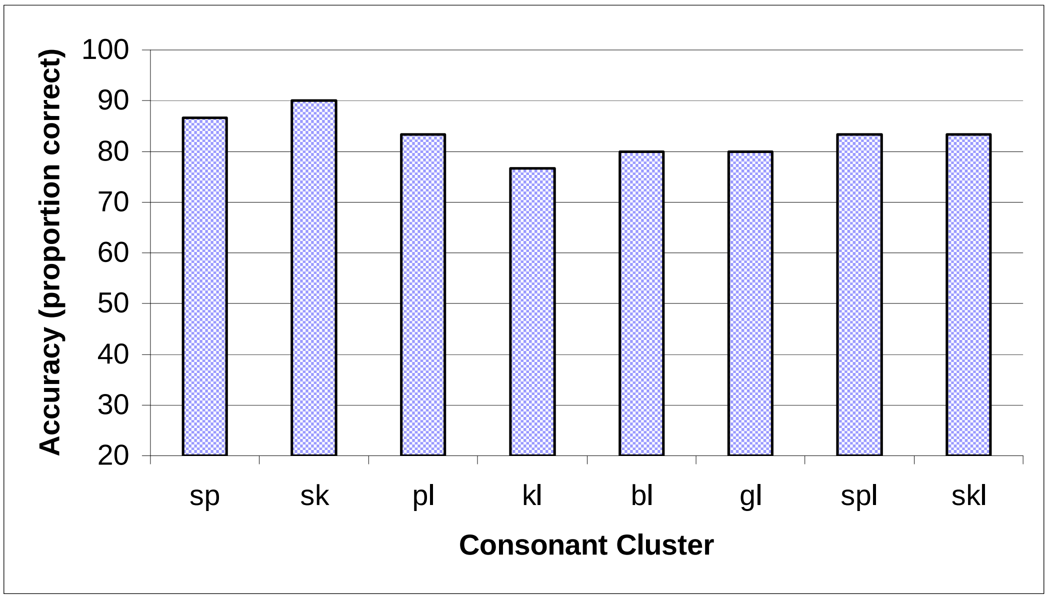


Figure 2B

Overall percent correct of the JP group's production data for the CVC(C) tokens evaluated by AE listeners.

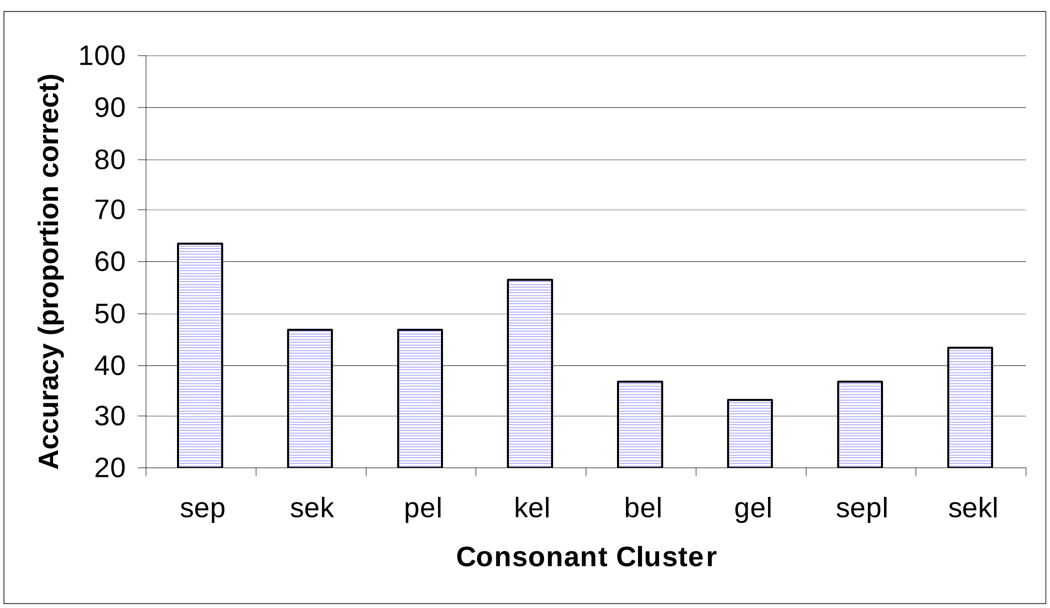


Figure 3

Scatterplot illustrating the relationship between the JP group's overall percent correct on the perception task and their overall score on the production data evaluated by transcribers. The line indicates Fit Line for Total.

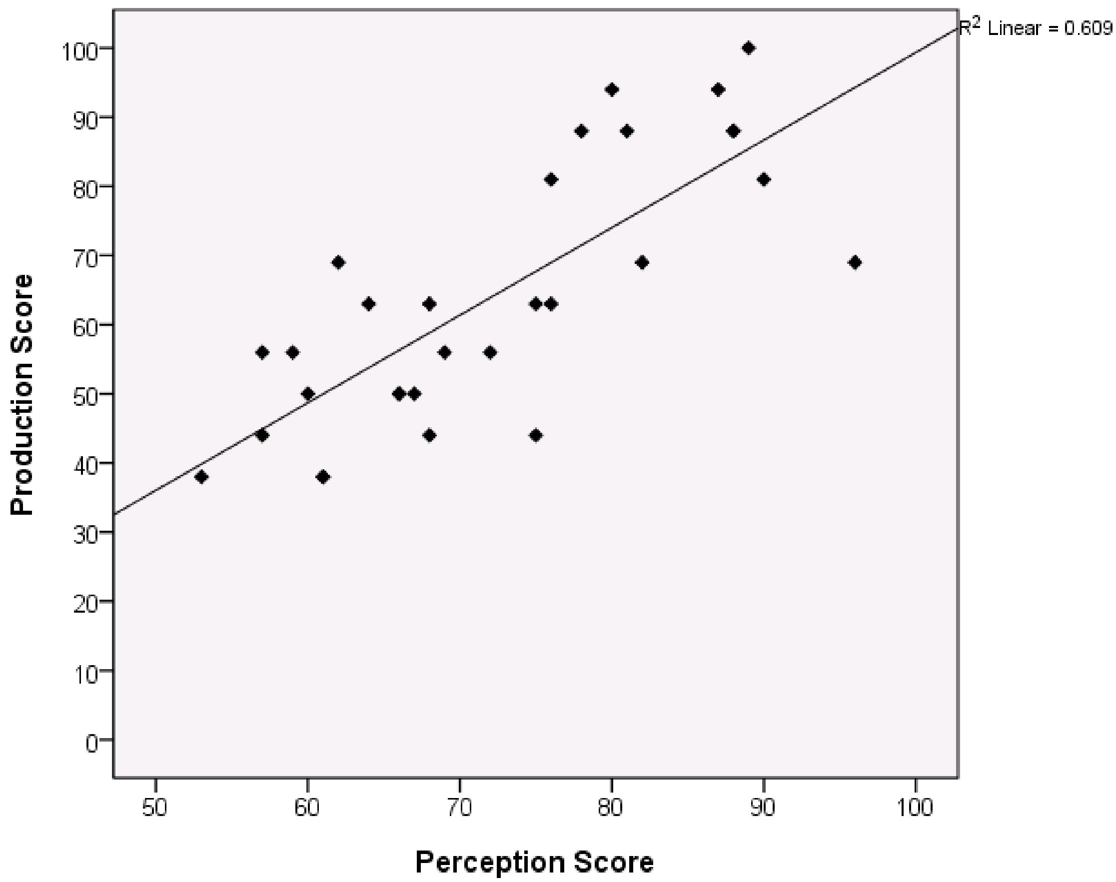
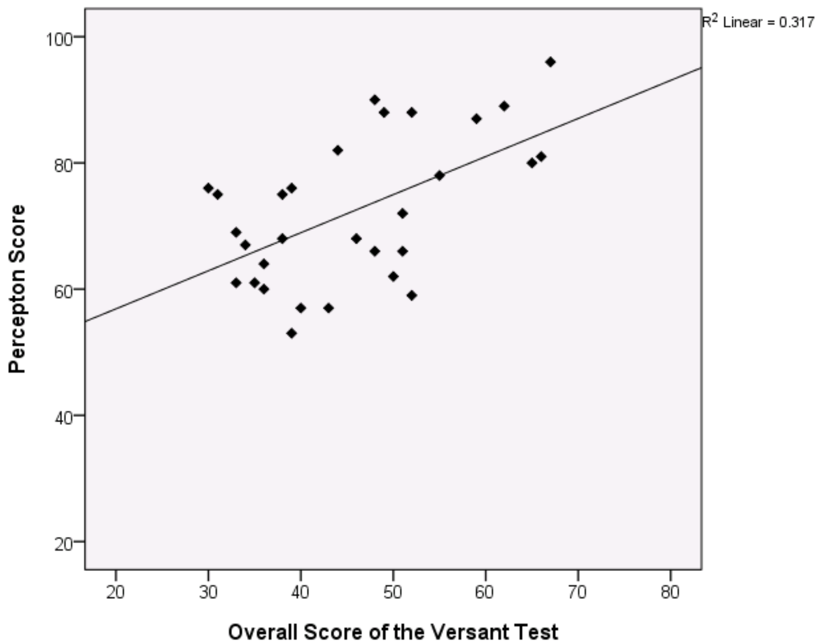
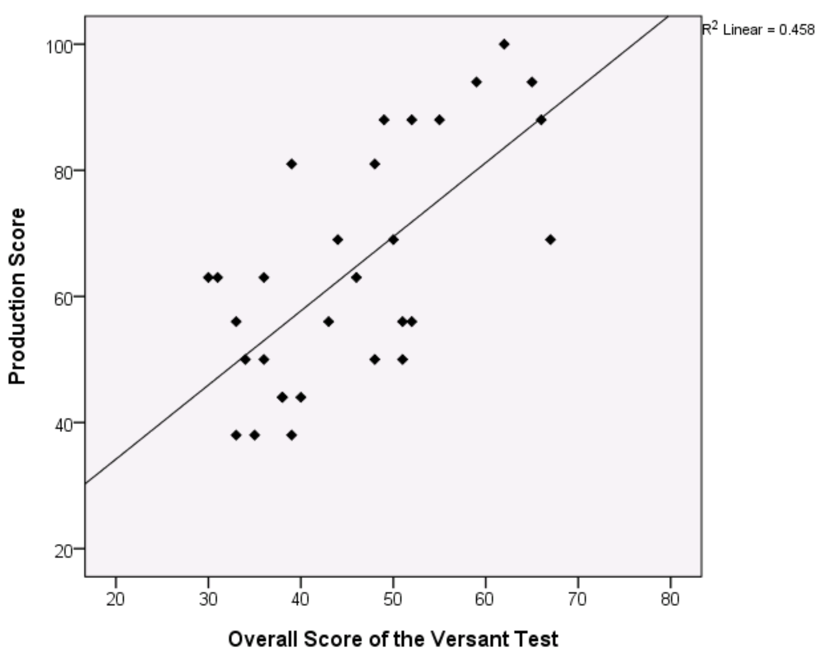


Figure 4

A. Scatterplot illustrating the relationship between the JP group's overall percent correct on the perception task and their overall Versant test score.



B. Scatterplot illustrating the relationship between the JP group's overall percent correct on the production task and their overall Versant test score.



10. APPENDECIS

Appendix A

Japanese Participant Characteristics and Overall Scores of the Versant Test, the Perception & the Production Test. V = the Versant test score; PC = overall percent correct in the Perception Test; PD = overall percent correct in the Production test.

ID	Occupation	M/F	Age	LOR	AOA	L2 use	V	PC	PD
JP3	ESL student	F	33	0;1	33;1	70	40	57	44
JP18	College student	F	27	0;1	27;7	70	48	90	81
JP19	Office worker	M	39	0;1	39;6	90	48	66	50
JP23	ESL student	F	26	0;2	26;4	30	36	60	50
JP1	ESL student	M	37	0;4	37;7	20	35	61	38
JP13	ESL student	M	25	0;4	25;1	60	50	62	69
JP17	ESL student	F	33	0;4	33;5	50	30	76	63
JP15	ESL student	F	33	0;6	33;1	30	38	68	44
JP27	College student	M	27	0;6	26;8	97	39	53	38
JP29	Grad student	F	24	0;9	23;9	50	62	89	100
JP22	ESL student	F	30	0;11	29;2	50	33	69	56
JP9	Grad student	F	26	1	25;9	50	38	75	44
JP25	ESL student	M	27	1	26;9	50	31	75	63
JP5	College student	F	28	1;11	27;3	50	49	88	88
JP21	Office worker	F	32	1;3	31;5	50	39	76	81
JP30	ESL student	M	31	1;5	29;7	70	33	61	38
JP14	ESL student	M	32	2;4	30;4	60	52	88	88
JP11	Housewife	F	42	2;7	39;7	5	51	72	56
JP24	ESL student	M	31	3	28;1	80	43	57	56
JP6	College student	M	22	3;1	19;7	20	44	82	69
JP16	Housewife	F	37	3;11	33;3	30	59	87	94
JP12	College student	M	28	3;6	25;3	50	34	67	50
JP26	Grad student	M	34	4;2	30;2	50	46	68	63
JP20	Grad student	F	27	5;1	22;6	80	67	96	69
JP8	Office worker	F	37	6;10	31;6	40	55	78	88
JP32	Office worker	F	29	8;3	21;1	10	36	64	63
JP4	Grad student	F	37	9;11	28;1	50	66	81	88
JP31	Office worker	F	38	13;5	24;8	95	51	66	50
JP2	Office worker	F	41	16;11	25	50	65	80	94
JP7	Housewife	F	52	19;10	33;5	50	52	59	56

Appendix B

Characteristics of 5 American Participants

Subject ID	Occupation	M/F	Age
AE1	Office Worker	M	34
AE2	Grad Student	F	25
AE3	Grad Student	F	28
AE4	Grad Student	M	23
AE5	Grad Student	F	32

Appendix C

Summary of acoustic measurement data for stimuli - mean duration of each distinct segment for the stops + laterals pairs (in ms).

	[blani]	*vs*	[bəlani]	
Duration (ms)	Mean	Range	Mean	Range
VOT	9	(8-10)	8	(7-9)
Schwa	0		44	(42-45)
[la]	228	(221-237)	258	(246-268)

	[glani]	*vs*	[gəlani]	
Duration (ms)	Mean	Range	Mean	Range
VOT	23	(21-25)	33	(31-35)
Schwa	0		52	(49-56)
[la]	229	(225-231)	255	(246-264)

	[plani]	*vs*	[pəlani]	
Duration (ms)	Mean	Range	Mean	Range
VOT	59	(55-65)	47	(37-56)
Schwa	0		30	(18-46)
[la]	200	(198-203)	257	(241-275)

	[klani]	*vs*	[kəlani]	
Duration (ms)	Mean	Range	Mean	Range
VOT	83	(77-96)	49	(47-54)
Schwa	0		48	(36-57)
[la]	191	(188-194)	225	(218-233)

Appendix D

Summary of acoustic measurement data for stimuli - mean duration of each distinct segment for /s/ + voiceless stops pairs (in ms).

	[spani]	*vs*	[səpani]	
Duration (ms)	Mean	Range	Mean	Range
Frication	139	(132-147)	126	(123-132)
Schwa	0		24	(23-26)
Silence	79	(73-85)	85	(76-98)
VOT	7	(7-8)	44	(41-50)

	[skani]	*vs*	[səkani]	
Duration (ms)	Mean	Range	Mean	Range
Frication	150	(147-153)	123	(113-129)
Schwa	0		33	(31-37)
Silence	55	(53-57)	68	(60-73)
VOT	16	(14-18)	47	(45-49)

Appendix E

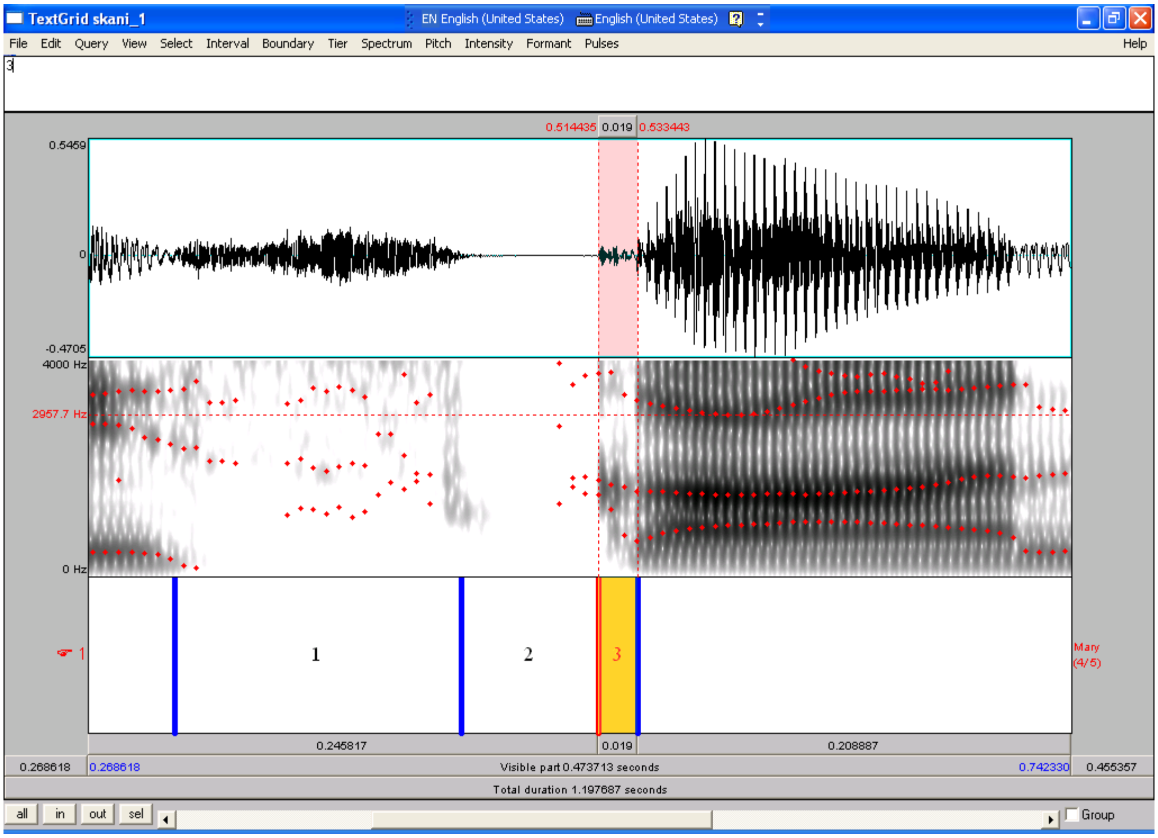
Summary of acoustic measurement data for stimuli - mean duration of each distinct segment for tri-consonantal sequences.

	[splani]	*vs*	[səplani]	
Duration (ms)	Mean	Range	Mean	Range
Frication	106	(98-116)	120	(115-127)
Schwa	0		30	(25-33)
Silence	94	(91-98)	86	(82-91)
VOT	11	(8-15)	39	(37-43)
[la]	198	(185-215)	163	(153-174)

	[sklani]	*vs*	[səklani]	
Duration (ms)	Mean	Range	Mean	Range
Frication	126	(124-130)	111	(101-117)
Schwa	0		45	(41-49)
Silence	55	(54-55)	59	(57-61)
VOT	22	(20-23)	52	(48-56)
[la]	171	(140-187)	172	(171-173)

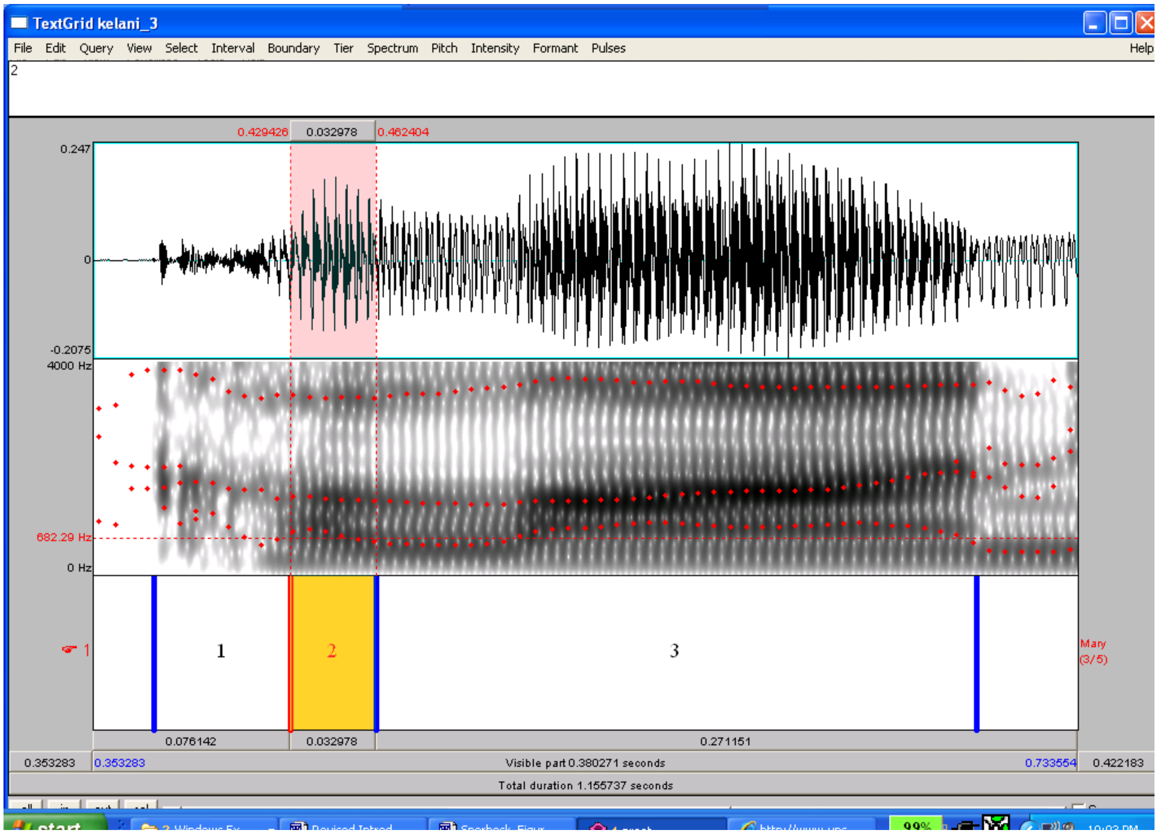
Appendix F

The figure below shows both the waveform and spectrogram of [skani] extracted from the sentence ‘say skani now’. Regions are as follows: 1) Fricative duration, 2) Silence duration, 3) VOT.



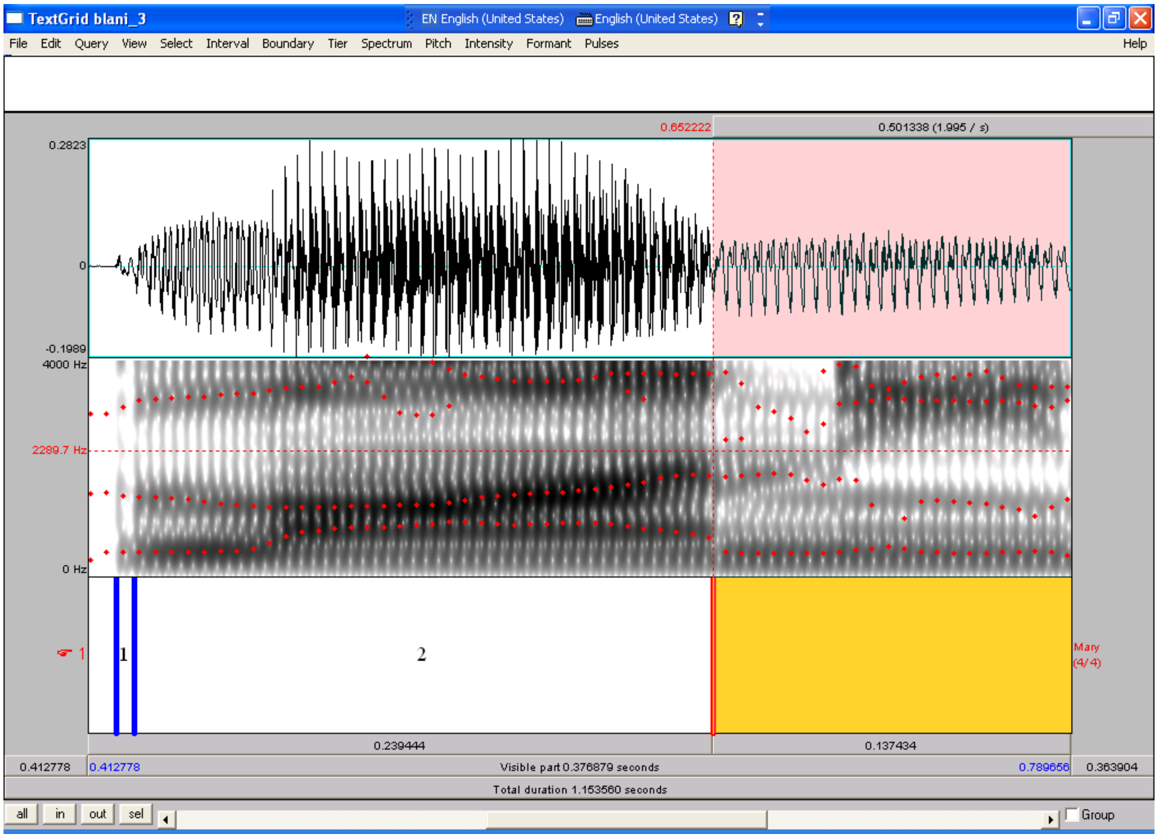
Appendix G

The figure below shows both the waveform and spectrogram of [kelani] extracted from the sentence ‘say kelani now’. Regions are as follows: 1) VOT, 2) Schwa, 3) Duration of [la].



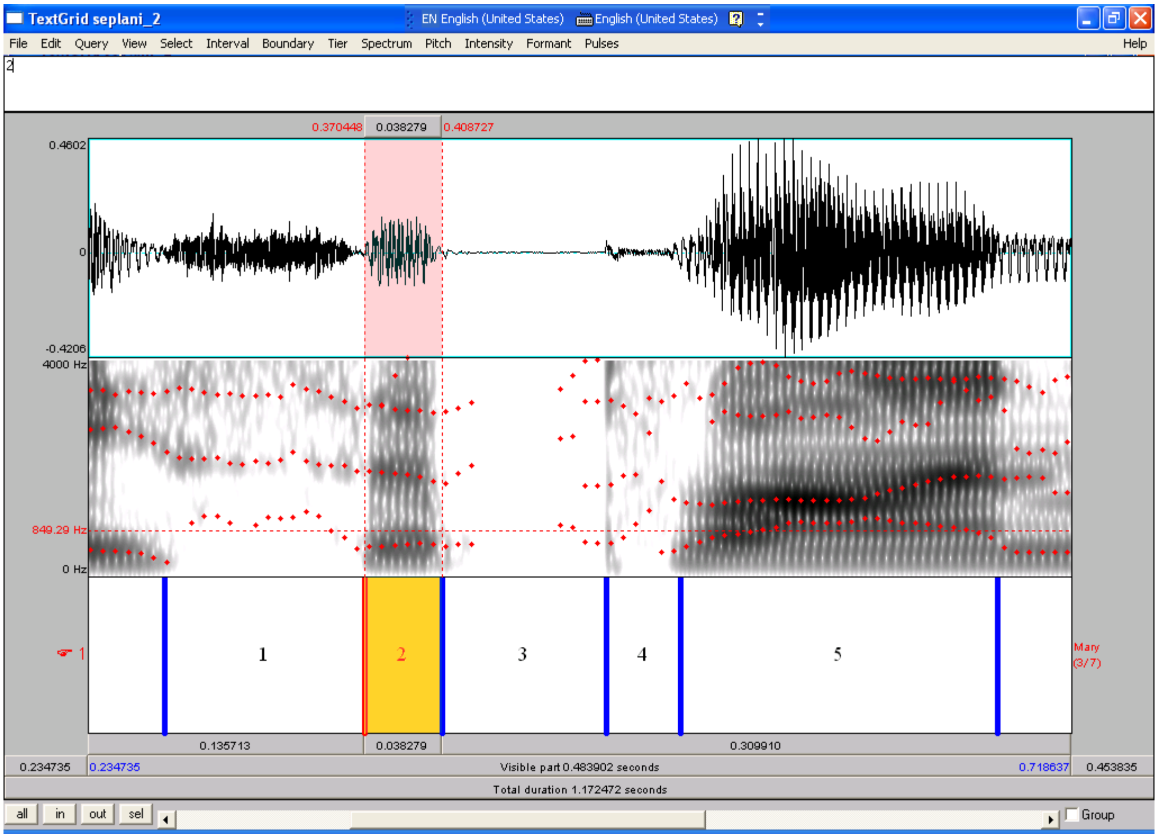
Appendix H

The figure below shows both the waveform and spectrogram of [blani] extracted from the sentence ‘say blani now’. Regions are as follows: 1) VOT, 2) Duration of [la].



Appendix I

The figure below shows both the waveform and spectrogram of [seplani] extracted from the sentence ‘say seplani now’. Regions are as follows: 1) Fricative duration, 2) Schwa, 3) silence, 4) VOT, 5) Duration of [la].



Appendix J

Summary of individual results of 5 AE participants in the perception task.

	Mean	Median	Range	SD
AE1	99	100	96-100	1.85
AE2	100	100	96-100	1.41
AE3	100	100	100	0.00
AE4	97	96	88-100	3.96
AE5	99	100	96-100	2.07
Overall	99	100	88-100	2.46

Appendix K

Summary of individual results of 30 JP participants in the perception task.

	Mean	Median	Range	SD
JP1	62	65	46-79	12.9
JP2	80	88	50-96	17.1
JP3	57	56	42-75	11.1
JP4	81	84	58-92	11.6
JP5	88	92	75-96	8.8
JP6	83	84	71-92	7.5
JP7	59	58	46-71	8.0
JP8	78	78	58-96	16.1
JP9	75	75	50-92	15.5
JP11	72	71	58-88	9.2
JP12	67	69	46-79	12.4
JP13	62	63	54-75	7.7
JP14	88	94	67-100	13.3
JP15	68	69	46-96	17.8
JP16	87	94	63-100	14.1
JP17	76	75	58-96	14.9
JP18	90	94	67-100	11.5
JP19	66	63	50-88	12.0
JP20	97	96	88-100	4.0
JP21	76	75	58-88	10.0
JP22	69	69	58-83	8.5
JP23	60	56	42-88	14.7
JP24	57	56	46-71	7.8
JP25	75	77	54-88	10.9
JP26	68	61	54-96	16.3
JP27	53	52	38-71	11.0
JP29	89	94	63-100	12.3
JP30	61	61	46-79	12.4
JP31	66	63	58-79	6.5
JP32	64	65	50-75	8.9
Overall	72	71	38-100	16.0

Appendix L

The proportion of disagreement on the vowel deletion cases in JP participants' productions.

Participant	Frequency	Proportion
JP12	6	0.21
JP30	5	0.17
JP15	3	0.10
JP27	2	0.07
JP18	2	0.07
JP9	2	0.07
JP3	2	0.07
JP1	1	0.03
JP7	1	0.03
JP11	1	0.03
JP13	1	0.03
JP16	1	0.03
JP23	1	0.03
JP25	1	0.03

Appendix M

Frequency and proportion of disagreement between the judges on the presence of a vowel in the initial onset by each consonant context.

Context	Frequency	Proportion
sp	2	0.07
sk	3	0.10
pl	1	0.03
kl	1	0.03
bl	3	0.10
gl	3	0.10
spl	1	0.03
skl	2	0.07
sep	2	0.07
sek	2	0.07
pel	2	0.07
kel	1	0.03
bel	1	0.03
gel	3	0.10
sepl	1	0.03
sekl	1	0.03

Appendix N1

Results of the transcription on the productions of CC(C) words for the JP group. E stands for epenthesis; D for deletion; SC for segment change; DA for disagreement between the judges; O for others; R for the transcribed [r] for the intended /l/; R/L or R/W for discrepancy between the judges.

Subject	**sp**	**sk**	**pl**	**kl**	**bl**	**gl**	**spl**	**skl**
JP1			R/L	R	SC, R/L	R	SC, R	R
JP2					E			
JP3			R	R	R	R	DA	R
JP4	E		E					R/L
JP5							R	
JP6							R	R
JP7			E	E	DA	E	SC	R
JP8				R			R	R
JP9				SC		DA	R	E, R
JP11					E		R	
JP12	DA	DA	R	R/L	R	R/L	R	R
JP13		DA	E	E			R	R
JP14			R/L	R/L	R/L	R	R/L	
JP15			E, R	DA	R	DA	R	DA
JP16					DA			
JP17			R	SC	R	SC	R	R
JP18				R		DA		DA
JP19				SC				SC
JP20								
JP21				R		R	R	R
JP22						R	R	R
JP23			R/L	R/L	R/L		R/L	R/L
JP24	E							
JP25			R/L	R	R/L	R	SC	R/W
JP26			R	R	R		R/L	R
JP27	DA	DA	R	R	R	R	R	R
JP29			R	R			R	R
JP30			DA	E, R	DA	E, R	R	R
JP31				R		R	SC	
JP32			R	R	R	R	R	SC
	sp	**sk**	**pl**	**kl**	**bl**	**gl**	**spl**	**skl**
# of Errors	4	3	5	7	6	6	5	5
# of Corrects	26	27	25	23	24	24	25	25
# of E	2	0	4	3	2	2	0	1
# of D	0	0	0	0	0	0	0	0
# of SC	0	0	0	3	1	1	4	2
# of O	0	0	0	0	0	0	0	0
# of DA	2	3	1	1	3	3	1	2
# of R/L, R/W	0	0	4	3	4	1	3	3
# of R	0	0	8	12	7	10	16	15

Appendix N2

Results of the transcription on the productions of CəC(C) words for the JP group. E stands for epenthesis; D for deletion; SC for segment change; DA for disagreement between the judges; O for others; R for the transcribed [r] for the intended /l/; R/L or R/W for discrepancy between the judges.

Subject	sep	sek	pel	kel	bel	gel	sepl	sekl
JP1	D	O	DA	D, R	D, R	D, R	D, R	D, R
JP2								
JP3	DA	D	O	D, R	D, R	D, R	D, R	D, R
JP4								
JP5			D					SC
JP6		D	O, R		D	D	R	SC, R
JP7							SC	SC
JP8				R	D	D	R	R
JP9		D	D	D	D	DA	D	R
JP11	D			D	DA	D	D, R	D, R
JP12	DA	DA	DA, R/L	R	D, R/L	D, R/L	DA, R/L	R
JP13		D	R	R/L	D, R	R	R	R
JP14			D, R/L			D, R		R
JP15		D	SC, R	R	R	SC, R	D, R	D, R
JP16								
JP17	D	D	R/L	R/L	R/L	R/L	D, R	O
JP18		D					R	R
JP19	D	D			D	O	D	D
JP20			SC	D	D	D	D	
JP21					D	R	SC, R	SC
JP22	D	D	D	O, R	D	D, R	D, R	R
JP23	D	D	D	DA	D	D	D	D
JP24			D	D	D	D	D	D
JP25				D, R/L	D, R	DA, R/L	D, R/L	O
JP26	D	D	D	D		D		D, R
JP27	D	D	D, R	D, R	D, R	D, R	D, R	D, R
JP29							R	R
JP30		DA	R	O, R	D	DA, R	D, R	DA, R
JP31	D	D	D		D, R/L	D	D, R/L	D, R/L
JP32			D, R	D, R	D, R	D, R	O	R
	sep	**sek**	**pel**	**kel**	**bel**	**gel**	**sepl**	**sekl**
Errors	11	16	16	13	19	20	19	17
Corrects	19	14	14	17	11	10	11	13
# of E	0	0	0	0	0	0	0	0
# of D	9	13	10	10	18	15	15	10
# of SC	0	0	2	0	0	1	2	4
# of O	0	1	2	2	0	1	1	2
# of DA	2	2	2	1	1	3	1	1
# of R/L, R/W	0	0	3	3	3	3	3	1
# of R	0	0	6	9	7	10	14	17

Appendix O

Summary of the transcription results of those who produced both nonsense words in each cluster type accurately. For instance, JP1 produced both /spani/ and /skani/ in “sC” cluster type accurately.

	Cluster type							
	sC	pl, kl	bl, gl	CCC	səC	pəl, kəl	bəl, gəl	CəCC
Subject	JP1	JP1	JP3	JP2	JP2	JP2	JP2	JP2
	JP2	JP2	JP4	JP4	JP4	JP4	JP4	JP4
	JP3	JP3	JP5	JP5	JP5	JP7	JP5	JP8
	JP5	JP5	JP6	JP6	JP7	JP8	JP7	JP13
	JP6	JP6	JP8	JP8	JP8	JP13	JP16	JP14
	JP7	JP8	JP12	JP11	JP14	JP16	JP17	JP16
	JP8	JP11	JP13	JP12	JP16	JP17	JP18	JP18
	JP9	JP12	JP14	JP13	JP20	JP18	JP29	JP29
	JP11	JP14	JP19	JP14	JP21	JP19		
	JP14	JP16	JP20	JP16	JP24	JP21		
	JP15	JP18	JP21	JP17	JP25	JP29		
	JP16	JP20	JP22	JP20	JP29			
	JP17	JP21	JP23	JP21	JP32			
	JP18	JP22	JP24	JP22				
	JP19	JP23	JP25	JP23				
	JP20	JP24	JP26	JP24				
	JP21	JP25	JP27	JP26				
	JP22	JP26	JP29	JP27				
	JP23	JP27	JP31	JP29				
	JP25	JP29	JP32	JP30				
	JP26	JP31						
	JP29	JP32						
	JP30							
	JP31							
	JP32							
Total N	25	22	20	20	13	11	8	8

Appendix P

Summary of the transcription results of those who produced both CC(C) and CəC(C) tokens for each consonant type accurately. For instance, JP2 produced both /spani/ and /səpani/ in “SP” context correctly.

	Cluster type							
	SP	SK	PL	KL	BL	GL	SPL	SKL
Subject ID	JP2	JP2	JP2	JP2	JP4	JP2	JP2	JP2
	JP5	JP4	JP8	JP4	JP5	JP4	JP4	JP4
	JP6	JP5	JP11	JP5	JP14	JP5	JP5	JP8
	JP7	JP7	JP16	JP6	JP15	JP13	JP6	JP12
	JP8	JP8	JP17	JP8	JP17	JP29	JP8	JP13
	JP9	JP11	JP18	JP12	JP18	JP21	JP13	JP14
	JP13	JP14	JP19	JP14	JP26	JP16	JP14	JP16
	JP14	JP16	JP21	JP16	JP29		JP16	JP20
	JP15	JP20	JP25	JP18			JP18	JP22
	JP16	JP21	JP29	JP21			JP26	JP29
	JP18	JP24		JP29			JP29	
	JP20	JP25		JP31				
	JP21	JP29						
	JP25	JP32						
	JP29							
	JP30							
	JP32							
Total N	17	14	10	12	8	7	11	10

Appendix Q

Summary of the transcription results of those who produced both the CC(C) and the CəC(C) tokens in each cluster type with accuracy. For instance, JP2 produced all the followings for "SC" type correctly: /spani/, /səpani/, /skani/, /səkani/.

	Cluster type			
	SC	−VL	+VL	CCC
Subject ID	JP2	JP2	JP4	JP2
	JP5	JP8	JP5	JP4
	JP7	JP16	JP29	JP8
	JP8	JP18		JP13
	JP14	JP21		JP14
	JP16	JP29		JP16
	JP20			JP29
	JP21			
	JP25			
	JP29			
	JP32			
Total N	11	6	3	7

Appendix R

The proportion of the vowel deletion cases among JP participants.

Participant	Frequency	Proportion
JP27	8	0.08
JP23	7	0.07
JP31	7	0.07
JP1	6	0.06
JP22	6	0.06
JP24	6	0.06
JP26	6	0.06
JP3	6	0.06
JP11	5	0.05
JP19	5	0.05
JP9	5	0.05
JP20	4	0.04
JP32	4	0.04
JP15	3	0.03
JP17	3	0.03
JP25	3	0.03
JP6	3	0.03
JP12	2	0.02
JP13	2	0.02
JP14	2	0.02
JP30	2	0.02
JP8	2	0.02
JP18	1	0.01
JP21	1	0.01
JP5	1	0.01

Appendix S

Frequency and proportion of vowel deletion by each consonant context.

Context	Frequency	Proportion
sep	9	0.09
sek	13	0.13
pel	10	0.1
kel	10	0.1
bel	18	0.18
gel	15	0.15
sepl	15	0.15
sekl	10	0.1

Appendix T

Summary of frequency and proportion of the /l/ replacement in CC(C) tokens. Others included all the errors that were not related to the replacement of the intended /l/.

	−VL	+VL	CCC	Total #	Proportion
Correct (e.g., [pl])	23	28	15	66	37%
Replacement with R (e.g., [pr])	25	20	35	80	44%
Epenthesis (e.g., [pel])	5	3	0	8	4%
Epenthesis-R (e.g., [per])	2	1	1	4	2%
Others	5	8	9	22	12%

Appendix U

Frequency of the /l/ replacement by each cluster in CC(C) words. “R-Replacement” means that the intended /l/ was transcribed as [r]. “Epenthesis-R” means that a schwa was inserted between the first two consonants and the intended /l/ was transcribed as [r]. Others included all the errors that were not related to the replacement of the intended /l/.

	pl	kl	bl	gl	spl	skl	Total #
Correct (e.g., [pl])	14	9	14	14	7	8	66
R-Replacement (e.g., [pr])	11	14	10	10	18	17	80
Epenthesis (e.g., [pel])	3	2	2	1	0	0	8
Epenthesis-R (e.g., [per])	1	1	0	1	0	1	4
Others	1	4	4	4	5	4	22

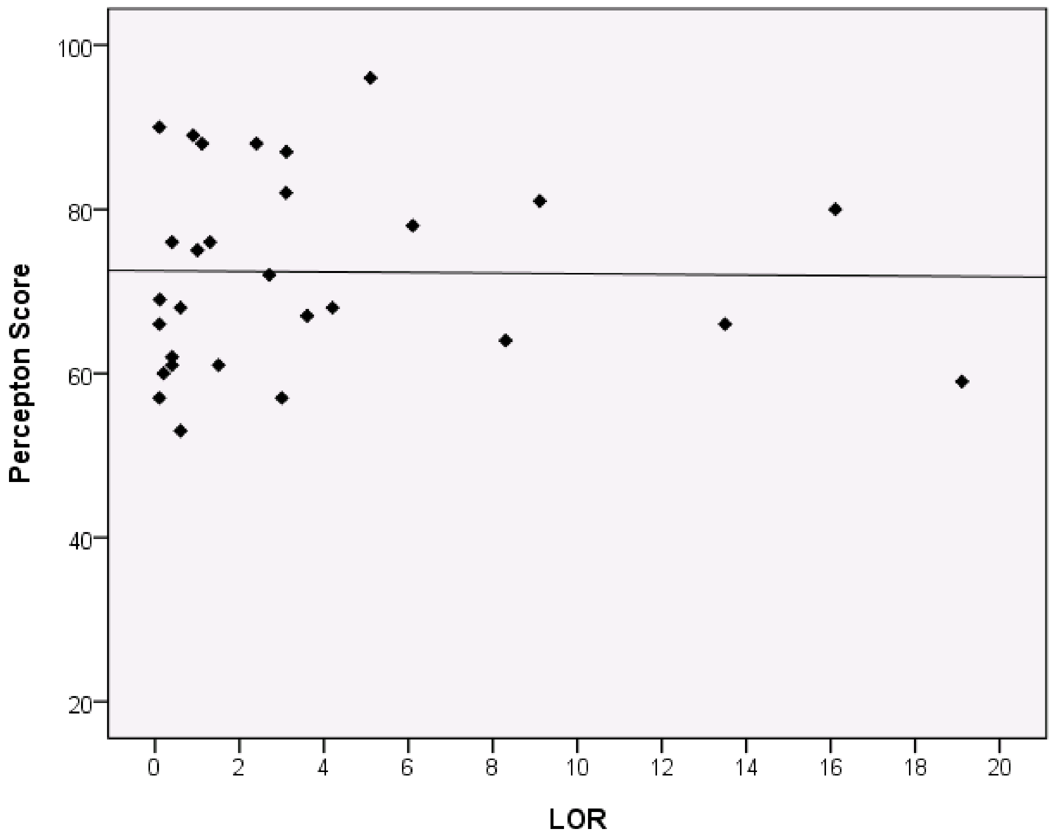
Appendix V

Frequency of the [r] productions for 30 JPs. Frequency rate was computed for each participant by dividing the total number of the [r] occurrence by 12 (12 clusters with /l/).

Subject	N of /r/ in CC	N of /r/ in CVC	Total	Frequency Rate
JP1	6	5	11	0.92
JP2	0	0	0	0.00
JP3	5	5	10	0.83
JP4	1	0	1	0.08
JP5	1	0	1	0.08
JP6	2	3	5	0.42
JP7	1	0	1	0.08
JP8	3	3	6	0.50
JP9	2	1	3	0.25
JP11	1	2	3	0.25
JP12	6	6	12	1.00
JP13	2	6	8	0.67
JP14	5	3	8	0.67
JP15	3	6	9	0.75
JP16	0	0	0	0.00
JP17	4	5	9	0.75
JP18	1	2	3	0.25
JP19	0	0	0	0.00
JP20	0	0	0	0.00
JP21	4	2	6	0.50
JP22	3	4	7	0.58
JP23	5	0	5	0.42
JP24	0	0	0	0.00
JP25	5	4	9	0.75
JP26	5	1	6	0.50
JP27	6	6	12	1.00
JP29	4	2	6	0.50
JP30	4	5	9	0.75
JP31	2	3	5	0.42
JP32	5	5	10	0.83
Overall	86	79	165	0.46

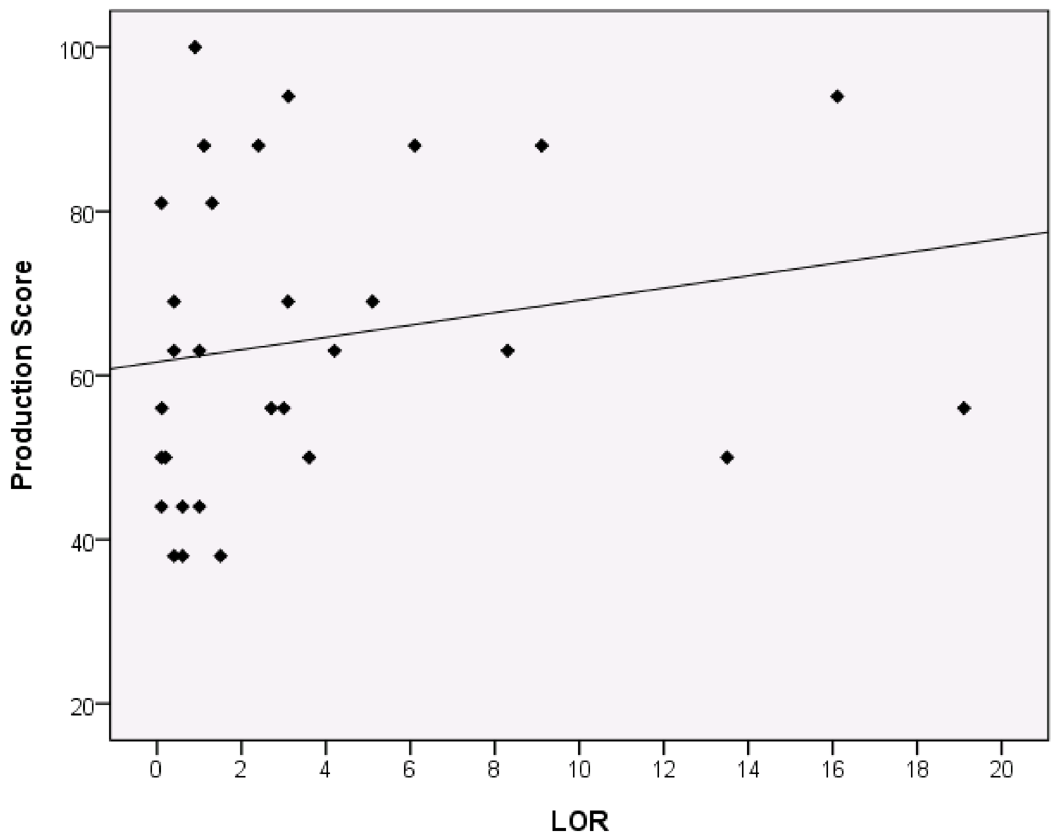
Appendix W

Scatterplot illustrating the relationship between the JP group's overall percent correct on the perception task and their LOR.



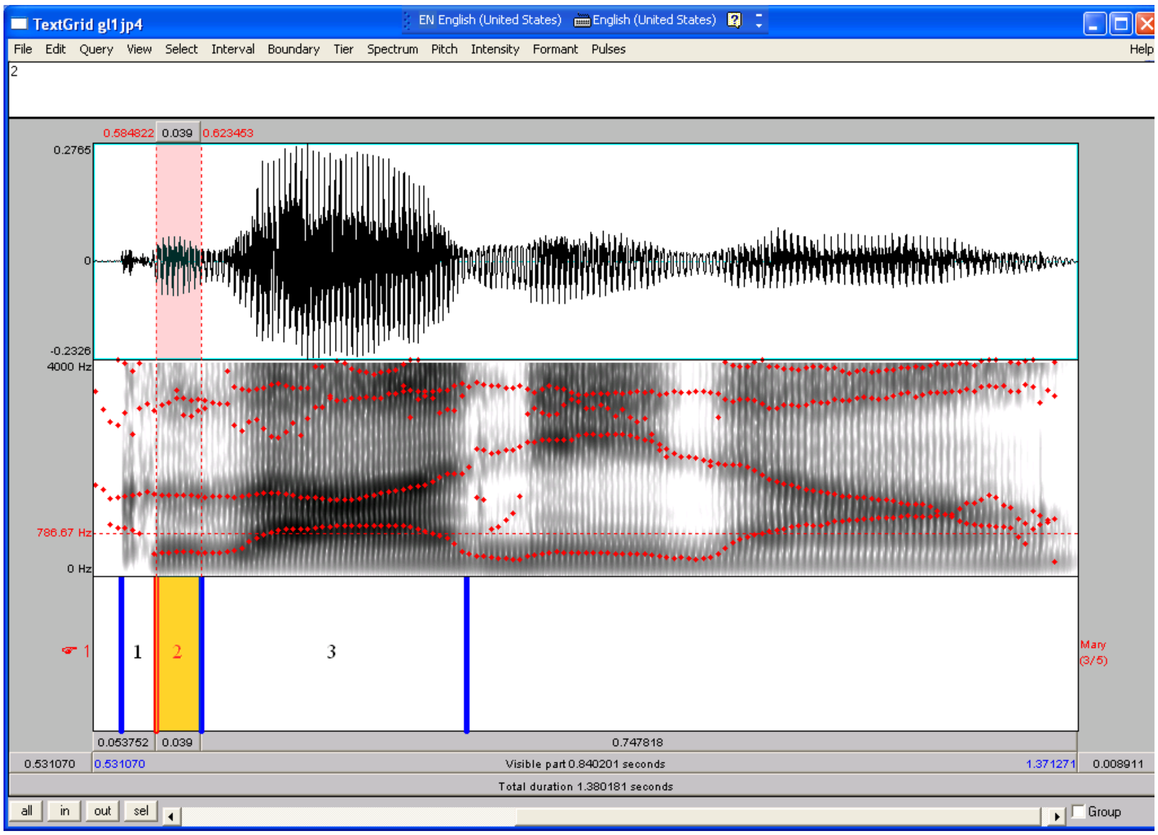
Appendix X

Scatterplot illustrating the relationship between the JP group's overall percent correct on the production task and their LOR.



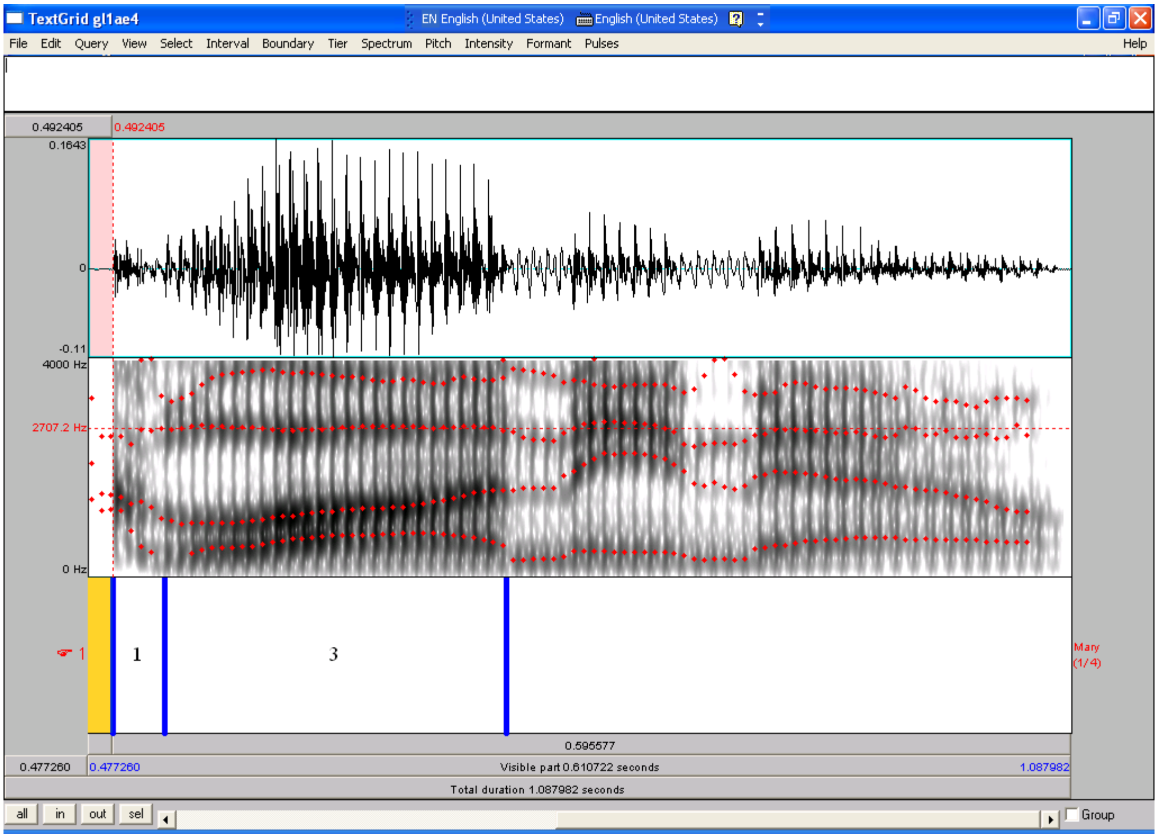
Appendix Y

The figure below shows an example of transitional schwa in [gəlani] produced by JP4. That is, the schwa was epenthesized between C1 and C2. However, this schwa was not detected and transcribed as [glani]. Regions are as follows: 1) VOT, 2) Transitional schwa, 3) Duration of [la].



Appendix Z

The figure below shows an example of [glani] produced by AE4. Regions are as follows: 1) VOT, 3) Duration of [la].



11. REFERENCES

Altenberg, E. (2005). The judgment, perception, and production of consonant clusters in a second language. *IRAL, 43*, 53-80.

Altenberg, E. and Vago, R. (1983). Theoretical implications of an error analysis of second language phonology production. *Language Learning, 33,* 427-447.

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

Bernstein, J. (2009). Proficiency instrumentation for cross language speech perception studies. *Journal of the Acoustical Society of America, 125,* 2753.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109,* 775-794.

Best, C. T. & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins.

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 206-244). Cambridge: Blackwell.

Boersma, P. & Weenick, D. (2005). Praat: Doing phonetics by computer (Version 4.3.14) [Computer Program]. Retrieved April 01, 2008, from: http://www.praat.org/

Boyd, J. (2006). On the representational status of /s/-clusters. *San Diego Linguistic Papers, 2*, 39–84.

Bradley, D. C., Sanchez-Casas, R. M., & Garcia-Albea, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes*, *8*, 197-233.

Bradlow, A. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/ - /l/ contrast. In Hansen Edwards, J. G. & Zampini, M. L. (eds.), *Phonology and Second Language Acquisition* (pp. 287-308). Amsterdam and Philadelphia: John Benjamins.

Bradlow, A., Pisoni, D., Akahane-Yamada, R., and Tohkura, Y. (1996). Training Japanese listeners to identify English /r/ and /l/:IV. Some effects of perceptual learning on speech production. *Journal of the Acoustic Society of America 101 (4)*, 2299-2310.

Broselow, E. (1983). Non-obvious transfer: On predicting epenthesis errors. In S. Gass & L. Selinker (Ed.), *Language transfer in language learning* (pp. 269-280). Rowley, MA: Newbury House.

Broselow, E. (1995). Skeletel positions and moras. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 175-205). Cambridge: Blackwell.

Broselow, E., Chen, S., & Wang, C. (1998). The Emergence of the unmarked in Second Language Phonology. *Studies in Second Language Acquisition*, *20*, 261-280.

Broselow, E. & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research, 7*, 35-59.

Broselow, E., & Park, H. S. (1995). Mora conservation in second language prosody. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 151-168). Hillsdale, NJ: Erlbaum.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2009). Increased second-language vocabulary size is associated with more consistent second-language vowel categorization and better discrimination. *Journal of the Acoustic Society of America 125 (4)*, 2773.

Cairns, C. E. (2009). Phonological representations and the Vaux-Wolfe proposal. In E. Raimy & C. E. Cairns (eds), *Contemporary Views on Architecture and Representations in Phonology* (pp. 145-164). Cambridge: MIT Press.

Cairns, C. E. & Feinstein, M. (1982). Markedness and the theory of syllable structure. *Linguistic Inquiry, 13,* 193-225.

Cardoso, W. (2008). The development of sC onset clusters in interlanguage: markedness vs. frequency effects. In Roumyana Slabakova et al. (eds), *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)* (pp. 15–29). Somerville, MA: Cascadilla Proceedings Project.

Carlisle, R. S. (1997). The modification of onsets in a markedness relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Language Learning, 47*, 327-361.

Carlisle, R. S. (1998). The acquisition of onsets in a markedness relationship. *Studies in Second Language Acquisition, 20,* 245-260.

Carlisle, R. S. (2006). The role of sonority cycle in core syllabification. In Baptista, B. and Watkins, M. (eds), *English with a Latin Beat – Studies in Portuguese/Spanish English Interphonology* (pp. 105–138). Amsterdam: John Benjamins.

Chang, Y., Hong, G., and Hallé, P. (2007). English cluster perception by Taiwanese Mandarin speakers. *Proceedings of the 16th International Congress of Phonetic Sciences (Saarbrucken, Germany, August 5-10, 2007),* 797-799.

Clements, G. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (eds.), *Papers in laboratory phonology 1: between the grammar and physics of speech* (pp. 283-333). Cambridge: Cambridge University Press.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1986). The syllable's differing role in segmentation of French and English. *Journal of Memory and Language, 25*, 385-400.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33*, 824-844.

Cutler, A. & Otake, T. (2002). Rhythmic Categories in Spoken-Word Recognition. *Journal of Memory and Language, 46*, 5296-322.

Davidson, L. (2001). Hidden rankings in the final state of the English grammar. In G. Horwood and S. –K. Kim (eds.). *Ruling Papers II* (pp. 21-48). New Brunswick: Rutgers University.

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

Davidson, L. (2007). The relationship between acoustic detail and underlying representations in loanword adaptation. *Phonology, 24*, 261-186.

Davidson, L. (to appear). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception & Performance.*

Davidson, L., & Shaw, J. (2009). A closer look at perceptual epenthesis in cross-language perception. *Journal of the Acoustical Society of America, 125,* 2752.

Davidson, L., & Stone, M. (2004). Epenthesis versus gestural mistiming in consonant cluster production: An ultrasound study. In M. Tsujimura & G. Garding (eds)., *Proceedings of the West Coast Conference on Formal Linguistics (WCCFL) 22, (University of California, San Diego, March 21-23, 2003),* 165-178. Somerville, MA: Cascadilla Press.

.

Davis, S. & Hammond, M. (1995). On the status of onglides in American English. *Phonology, 12*, 159-182.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*, 635-647.

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

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes, 16 (5/6)*, 491-505.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*, 315-330.

Eckman, F. R. (1981). On predicting phonological difficulty in second language acquisition. *Studies in Second Language Acquisition, 4*, 18-30.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*, 23-41.

Eckman, F. R. (1996). A functional-typological approach to second language acquisition theory. In W. C. Ritchie & T. K. Bhatia (eds.) *Handbook of second language acquisition* (pp. 195-211). New York: Academic Press.

Eckman, F. R. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second language Acquisition, 26*, 513-549.

Eckman, F. R. (2008). Typological markedness and second language phonology. In Hansen Edwards, J. G. and Zampini, M. L. (eds.) *Phonology and Second Language Acquisition* (pp. 95-115). Amsterdam and Philadelphia: John Benjamins.

Eckman, F. R., Elreyes, A., & Iverson, G. K. (2003). Some principles of second language phonology. *Second Language Research 19,* 169-208.

Eckman, F. R. & Iverson, G. K. (1993). Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research*, *9*, 234-252.

Fais, L., Kajikawa, S., Werker, J., & Amano, S. (2005). Japanese listeners' perception of phonotactic violations. *Language and Speech, 48 (2)*, 185-201.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 89,* 395-411.

Flege, J. E. (1995). Second language speech learning: Theory findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America, 91,* 370-389.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25,* 169-186.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527-552.

Flege, J. E., MacKay, I., & Meador, E. (1999). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America, 106*, 2973-2987.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41,* 78-104.

Friederici, A., & Wessels, J. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics, 54*, 287-295.

Gerken, R., & Aslin, R. (2005). Thirty years of research on infant speech perception: the legacy of Peter W. Jusczyk. *Language Learning and Development I (1),* 5-21.

Gierut, J. & Champion, A. H. (2001). Syllable onsets II: Three-element clusters in phonological treatment. *Journal of Speech, Language, and Hearing Research, 44,* 886–904.

Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 592-608.

Hancin-Bhatt, B. (2000). Optimality in second language phonology: codas in Thai ESL. *Second Language Research, 16*, 201-232.

Hancin-Bhatt, B., & Bhatt, R. M. (1997). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition, 19,* 331-378

Hogg, R. & McCully, C. (1987). *Metrical phonology: A Coursebook*. Cambridge: Cambridge University Press.

Ioup, G., Boustagi, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition, 16*, 73-98.

Ito, J. & Mester, R.A. (1995). Japanese phonology. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 817-838). Cambridge: Blackwell.

Ito, K. & Strange, W. (2009). Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *Journal of the Acoustical Society of America, 125(4)*, 2348-2360.

Ito, K., Law II, F., Sperbeck, M., Berkowitz, S., Gilichinskaya, Y., Monteleone, M., & Strange, W. (2007). Speech discrimination of American vowels by experienced Japanese late L2 learners. *Journal of the Acoustical Society of America*, *121*, 3073.

Jenkins, J. J. (1979). Four points to remember: a Tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. Craik (eds.), *Levels of processing in human memory* (pp. 429-466). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jusczyk, P. & Aslin, R. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1-23.

Jusczyk, P., Friederici, A., Wessels, J., Svenkerud, V., & Jusczyk, A. (1993). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language*, *32*, 402-420.

Jusczyk, P., Luce, P., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630-645.

Kabak, B. & Idsardi, W. (2003). Syllabically conditioned perceptual epenthesis. *Proceedings of the Berkeley Linguistics Society 29*, 233-245.

Kajikawa, S., Fais, L., Mugitani, R., Werker, J., & Amano, S. (2006). Cross-language sensitivity to phonotactic patterns in infants. *Journal of Acoustical Society of America*, *120 (4)*, 2278-2284.

Kenstowicz, M. (1994) *Phonology in Generative Grammar*, Cambridge, MA: Blackwell.

Kondo, M. (2005). Syllable structure and its acoustic effects on vowels in devoicing environments. In J. van de Weijer, Nanjo, K., & Nishihara, T. (eds.), *Voicing in Japanese* (pp. 229-245). Berlin: Mouton de Gruyter.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech, 32*, 249-278.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor: University of Michigan Press.

Lively, S. E., Logan J. S., & Pisoni, D. B. (1983). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242-1255.

MacKay, I., Flege, J., Piske, T., & Schirru, C. (2001). Category restructuring during second-language (L2) speech acquisition. *Journal of the Acoustical Society of America*, *110*, 516-528.

Major, R. (1996). Markedness in second language acquisition of consonant clusters. In R. Bayley, R. and Preston, D. (eds.), *Variation and Second Language Acquisition*, (pp. 75–96). Amsterdam: John Benjamins.

Martohardjono, G. (1989). The sonority cycle in the acquisition of phonology. In *Papers and Reports on Child Language Development, 28*, 131-139. Stanford University.

Martohardjono, G. & Flynn, S. (1995). Is there an age factor for Universal Grammar? In D. Singleton and Z. Lengyel (eds.) *Multilingual Matters* (pp. 135-153). Clevedon: Avon.

Massaro, D. & Cohen, M. (1983). Phonological constraint in speech perception. *Perception & Psychophysics, 34*, 338-348.

Mehler, J., Dommegues, J., Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal learning and verbal behavior, 20*, 298-305.

Monteleone, M. (2009). Effects of first language voicing rules on the perception and production of English obstruent sequences by adult Hungarian and Polish learners of English. Unpublished doctoral dissertation. City University of New York, the Graduate Center, USA.

Munro, M., Flege, J., & MacKay, I. (1996). The effects of age of second-language learning on the production of English vowels. *Applied Psycholinguistics, 17*, 313-334.

Nagano-Madesen, Y. (1994). Vowel devoicing in Japanese. *Lund University Working Papers, 42*, 119–129.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 258-278.

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

Pitt, M. (1998). Phonological processes and the perception of phonetically illegal consonant clusters. *Perception & Psychophysics, 60*, 941-951.

Polka, L. (1991). Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. *Journal of Acoustical Society of America, 89(6)*, 2961-2977.

Prince, A. & Smolensky, P. (2002). *Optimality Theory: Constraint interaction in a generative grammar*. New Brunswick, NJ: Rutgers University Center for Cognitive Science.

Raphael, L., Borden, G., & Harris, K. (1980). *Speech Science Primer*. *Physiology, Acoustics and Perception of Speech.* Baltimore: Lippincott Williams & Wilkins.

Raimy, E. & Cairns, C. E. (2009). *Contemporary Views on Architecture and Representations in Phonology*. Cambridge: MIT Press.

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degeree of VOT in Japanese initial voiceless stops. *Journal of Phonetics, 35(3)*, 439-443.

Sheldon, A. & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243-261.

Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.

Siegel, S. & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences.* New York, NY: McGraw-Hill.

Smit, A. (1993). Phonological error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research, 36,* 931–947.

Smith, J. (2006). Loan phonology is not all perception: Evidence from Japanese loan doubles. In T. J. Vance & K. A. Jones, (eds.), *Japanese/Korean Linguistics 14* (pp. 63-74). Stanford: CSLI.

Sperbeck, M. & Strange, W. (2008). An epenthetic vowel between consonantal sequences in perception and production by Japanese. *Journal of the Acoustical Society of America, 124(4)*, 2595.

Sperbeck, M., Strange, W., & Ito, K. (2005). Training Japanese L2 learners to perceive difficult American vowel contrasts. *Journal of the Acoustical Society of America, 117*, 2400.

Strange, W. (1992). Learning non-native phoneme contrasts: interactions among subject, stimulus, and task variables. In Tohkura, T., Vatikiotis-Bateson, E., & Sagisaka, Y. (eds), *Speech perception, production, and linguistics structure* (pp. 197-220). Burke, VA: IOS Press.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 3-45). Timonium, MD: York Press.

Strange, W. (2009). Automatic selective perception (ASP) of first-language (L1) and second-language (L2) speech: A working model. *Journal of Acoustical Society of America, 125 (4)*, 2769.

Strange, W., & Akahane-Yamada (1997). Effects of identification training on Japanese adults' perception of American-English vowels. *Journal of the Acoustical Society of America, 102*, 3137

Strange, W. & Dittman, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics, 36 (2),* 131-145.

Strange, W. & Shafer, V. L. (2008). Speech perception in second language learners: the re-education of selective perception. In Hansen Edwards, J. G. & Zampini, M. L. (eds.), *Phonology and Second Language Acquisition* (pp. 153-191). Amsterdam and Philadelphia: John Benjamins.

Tagliaferri, B. (2008). Paradigm (Version Beta 5d) [Computer Program]. Retrieved April 01, 2008, from: http://perceptionresearchsystems.com/

Tsuchida, A. (2001). Japanese vowel devoicing: Cases of consecutive devoicing environments. *Journal of East Asian Linguistics, 10,* 225–245.

Vance, T. J. (1987). *An introduction to Japanese Phonology*. New York: State University of New York Press.

Weber, A. & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of Acoustical Society of America, 119 (1)*, 597-607.

Werker, J. F. (1995). Age-related changes in cross-language speech perception: Standing at the crossroads. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research.* Baltimore: York Press.

Werker, J. F. & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development*, *1(2)*, 197-234.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49-63.

Young-Scholten, M. & Archibald, J. (2000). Second language syllable structure. In. J. Archibald (Ed.), *Second language acquisition and linguistic theory*. Massachusetts: Blackwell.