Project Description: The relationship between the production and the perception of foreign language sequences

1. Introduction

Learning a new language involves developing the abilities to both hear and produce new sounds and sound sequences. The goals of the proposed research are to determine (1) the extent to which language learners’ problems in pronunciation originate from errors of perception; and (2) whether perception errors result from the failure of learners to hear the acoustic differences between different foreign language sounds, or simply from their failure to realize that these differences are relevant to linguistic categorization.

We will use two types of methodology to investigate listeners’ perception of foreign language contrasts: discrimination tasks in which subjects judge two stimuli as the same or different, and ERPs (event related potentials), which tap into involuntary, pre-conscious reactions of the auditory system to changes in stimuli. The results are relevant to the broader questions of how native language experience shapes perception, as well as to the plasticity of the neural structures responsible for language processing. The project also has implications for language teaching, as the results are expected to shed light on the question of whether training in the perception of a second language results in improvement in pronunciation (as claimed, for example, by Bradlow et al. 1997), and if so, what sorts of perceptual training are most effective.

2. Background

The assumption underlying most analyses of language contact phonology (second language acquisition and loanword adaptation) is that the learner approaches the foreign language from the perspective of the native language (NL) production grammar (which effects the mapping from lexical representation to phonetic representation). Foreign structures are altered to conform to the restrictions of the native language by transfer of NL-based processes, though increasing exposure to foreign language data may foster the development of interlanguage grammars which more closely approximate the grammar of the foreign language.

Many interlanguage pronunciation patterns, however, appear not to have their source in either the native language production grammar or in any developing grammar that could be motivated by the data to which the learner has been exposed (Broselow to appear, 2004a, b, Broselow and Finer 1991, Davidson, Jusczyk, and Smolensky 2004, Kabak and Idsardi 2007, Kenstowicz 2001, Peperkamp and Dupoux 2003, Peperkamp in press, Silverman 1992, Yip 2005, among others). While some of these have been argued to reflect universal markedness (e.g., Eckman 1977, Broselow, Chen, and Wang 1998) or input frequency (Broselow and Xu 2004), other patterns are not readily amenable to such analyses. We will investigate the hypothesis that such patterns have their origin not in the native language production grammar but rather in misperception of the foreign language forms.

The division of labor between the production and perception components of grammar in foreign language production is a topic of intense debate. At one end is the claim that all modification of foreign forms reflects misperception (e.g., Peperkamp, in press). On this view, listeners are simply not capable of accurately perceiving structures that do not occur in their native language. At the other end is the claim that foreign forms are accurately perceived, but modified by the native language production grammar—that is, in the mapping from lexical to phonetic representation (e.g., Jacobs and Gussenhoven 2000, Paradis and LaCharité 1997).

Each of these claims has been argued to be too strong (e.g., Kang 2003, Smith, in press). A number of researchers have proposed models that reflect both perceptual and articulatory factors, though differing in how perceptual information is encoded in the grammar.
Silverman (1992) argues for a Dual-Level Model in which modification may take place either in the mapping from the acoustic form to the phonological representation (i.e., in perception) or in the mapping from the lexical to the phonetic representation (i.e., in production). For example, Silverman cites the Cantonese adaptations *pinta* ‘print’ and *pilin* ‘printer’ as evidence that Cantonese listeners correctly perceive a liquid in the English sources, but that the liquid is deleted in the production grammar unless the resulting word falls below the bisyllabic minimum. In contrast, the failure to retain word-final voiceless stops after a nasal (such as the [t] in ‘print’), to be retained, even to meet the minimal word size, suggests these sounds are not perceived by the Cantonese listener. Most research within a Dual-Level framework (e.g., Broselow, to appear, Kenstowicz 2001, Yip 2005) assumes not only a conventional production grammar but also a perception grammar that directs the mapping from speech stream to phonological representations (see, e.g., Boersma 1998, 1999, 2000, Escudero and Boersma 2004 for explicit proposals concerning the form of the perception grammar). In contrast, the P-Map Model (e.g., Kenstowicz 2003, based on Steriade 2001a,b) assumes that perception of both native and foreign forms is in general fully accurate, but that considerations of perceptual similarity are built into the production grammar, which chooses the phonetic realization that is most similar to the lexical representation, subject only to NL restrictions on possible phonetic structures.

Thus, an ongoing controversy concerns the extent to which listeners can accurately perceive a foreign language, and the extent to which misproduction reflects misperception. The effect of NL phoneme contrasts on perception is well established. For example, Miyawaki et al. (1975) compared the ability of native speakers of English and native speakers of Japanese to discriminate pairs from a continuum of syllables ranging from [ra] to [la]. The r/l contrast is significant in English (signaling lexical distinctions such as ‘red’ vs. ‘led’) but not in Japanese. While the English speakers perceived syllables on the [ra]-to-[la] continuum as falling into two discrete categories corresponding to the phoneme categories of English, Japanese speakers listeners perceived a continuum. Subsequent research, however, demonstrated that the failure to perceive a phoneme contrast does not necessarily imply failure to produce this contrast; Goto (1971) and Sheldon and Strange (1982) presented evidence that many Japanese speakers who cannot perceive the r/l contrast can nonetheless produce a native-like contrast.

Alongside experiments in which listeners are asked to make conscious similarity judgements of speech stimuli, recent years have seen a growing literature demonstrating that language experience affects perception even in the absence of conscious attention. Changes in an acoustic signal elicit a characteristic response in the electrical activity of the brain known as the MMN (mismatch negativity) response (Näätänen et al. 1978), even when attention is directed elsewhere. MMN responses (or the corresponding MMF (mismatch field) recorded using magetoencephalography, MEG) are typically evoked by contrasts that are linguistically significant in the subject’s native language, but not by non-significant acoustic changes such as the r/l difference for Japanese speakers (e.g., Buchwald et al. 1994, Näätänen et al. 1997, Winkler, Kujala, et al.1999, Winkler, Lehtokoski, et al. 1999, Sharma and Dorman 2000, Zhang et al. 2005). Similarly, Phillips et al. (2000) have shown that MMF responses are evoked by changes in stimuli that cross native language phoneme boundaries, but not by changes that fall within a native language phoneme boundary, even though in both cases the acoustic properties of the standard and deviant stimuli differ to the same degree. Furthermore, responses to foreign vs. native sound contrasts tend to show different regions of activation in the brain, leading Zhang et al. (2005,73), following Kuhl (2004), to hypothesize that “early exposure to a particular language produces a “neural commitment” to the acoustic properties of that language” which “interferes with foreign language processing, making it less efficient” (Zhang et al. 2005, 703).
There is good evidence that the effect of early language experience may persist despite extensive exposure to a foreign language system. Many studies have uncovered the persistence of a foreign accent in speakers who acquired the foreign language after the age of 6 (Long 1990), or in some cases as early as age 4 (Flege et al. 1995). A similar effect was found in perception by Pallier et al. (1997), who assessed the ability of participants who were bilingual in Spanish and Catalan to discriminate [e] and [ɛ], two sounds which contrast in Catalan but not in Spanish. The ability to successfully discriminate the Catalan-only contrast depended on the language of first exposure. Although the subjects with Spanish-speaking parents had been exposed to Catalan by or before the age of 6 and were highly proficient speakers of Catalan, they still were unable to reliably perceive the e/ɛ distinction, in contrast to those bilinguals whose first exposure was to Catalan. Similarly, Cutler, Weber, and Otake (2006) found that even highly proficient L2 listeners “may maintain a distinction between two phonetic categories of the L2 in their lexical representations, even though their phonetic processing is incapable of delivering the perceptual discrimination required for correct mapping to the lexical distinction.” (See also Weber and Cutler 2004, Curtin, Goad, and Pater 1998.)

The perceptual effects of native language phonotactics (constraints on possible sequences and possible syllables) have received considerably less attention than the effects of crosslinguistic differences in phoneme categories. In one of the few studies of the cross-language perception of sound sequences, Dupoux et al. (1999) demonstrated that when presented with stimuli on a continuum ranging from ebzo to ebu佐, Japanese-speaking listeners (whose native language prohibits sequences such as bz) were much less likely to distinguish the two stimuli than were French speakers (for whom ebzo is a possible structure). Thus, the Japanese speakers heard an ‘illusory vowel’ [u] (phonetically [u̯]) inside an illegal consonant sequence. This [u] is also the vowel that Japanese speakers insert in the production of foreign words, such as guurugu ‘Google’, as well as being the shortest and most variable vowel of Japanese (Keating and Huffman 1984). On this view, then, Japanese speakers insert vowels in their production of foreign words because they actually perceive these vowels in the foreign production. This conclusion is supported by the findings of Dupoux et al. (2001) that Japanese listeners perceived illusory [u] even when lexical neighborhood effects might introduce a bias toward a different vowel; for example, in a lexical decision task involving nonwords (and illegal structures) sokdo and mikdo, sokdo was classified as a real word (presumably because it was heard as the real word sokudo ‘speed’), while mikdo was not, regardless of the existence of mikado ‘emperor’. However, the question of why ebzo should be heard as ebu佐 rather than other legal structures such as ezo or enzo, is still unanswered (Peperkamp 2007). (See, e.g., Massaro and Cohen 1983, Pitt 1998 for examples of other types of misperception.)

In a followup to Dupoux et al. 1999, Dehaene-Lambertz et al. (2000) used ERP to investigate the responses of French and Japanese listeners to contrasts such as ebzo vs. ebu佐, which are potentially significant in French but not Japanese. Upon hearing the change from a standard (e.g., ebzo) to a deviant (e.g., ebu佐) while directed to make a bimanual same-different response, French listeners displayed clear mismatch negativity (MMN) responses – the typical brainwave patterns elicited by a change in an auditory stimulus. However, Japanese listeners displayed little or no MMN response to the same stimuli. Interestingly, the responses of the two NL groups were more similar at longer latencies, associated with higher-level processing. Furthermore, Jacquemot et al. (2003) found, in an fMRI imaging study of the brains of French and Japanese listeners during a task requiring discrimination of stimuli such as ebzo-ebu佐, that different brain regions were activated when processing stimuli exemplifying native language contrasts vs. non-native contrasts.

Thus there is clear evidence that even at very early levels of processing, listeners may
fail to perceive acoustic differences that do not signal potential lexical contrasts in their NL. Yet, although it has long been recognized that “adult discrimination of non-native speech contrasts is not uniformly poor” (Best, McRoberts, and Goodell 2001, 776), there is at present no model that would allow us to predict precisely which foreign structures will be accurately perceived – or, when they are inaccurately perceived, what form that misperception will take. Even if foreign language structures are perceptually assimilated to legal native structures, there is generally a range of native structures that are reasonable candidates for perceptual assimilation. We will investigate the relationship between perception and production of a foreign language by focusing on two questions that arise from the study of non-native production: (1) why are some foreign contrasts mastered more quickly than others, even though both are equally novel? and (2) what determines the direction of mispronunciation, particularly where the pronunciation of a foreign structure does not appear to be predictable from the native language grammar or the target language data?


The literature on loan adaptation and second language acquisition abounds in cases in which one particular foreign language contrast is acquired more readily than another equally novel contrast (e.g., Itô and Mester 2001). We will investigate one of these cases, though we expect that our results will be relevant to other cases of this sort.

In Japanese native (Yamato) and Sino-Japanese vocabulary, no contrast is possible between [ti] and [ʃi] or between [si] and [ʃi], because coronal obstruents are palatalized before [i], as illustrated by comparing *kat-e ‘win!’ with *kat-ʃ ‘to win’ and *hanas-e ‘talk!’ with *hanaj-ʃ ‘to talk’, where the same root takes different shapes before the imperative suffix -e and the infinitive suffix -i. (The palatal and nonpalatal consonants do however contrast in other contexts: ta ‘field’ vs. ʧa ‘tea’, saku ‘fence’ vs. ʃaku ‘wine-serving’, Itô and Mester 1995, 827.) As expected, the Japanese palatalization process is transferred in the adaptation of many loanwords (ʃiimu ‘team’; ʃi fuudo ‘sea food’). Yet English /si/ is much more likely than /ti/ or /di/ to undergo palatalization, as illustrated by *ʃiɪtuŋ ‘Citibank’ (ʃiɪʃiɪtuŋ, Itô and Mester1999), ʃiɪdi ‘CD’ (*siiʃi, Crawford 2007). As neither [ti], [di], nor [si] occurs in Japanese native vocabulary, the source of these asymmetries, which persist in the pronunciation of many English-proficient Japanese native speakers, requires explanation. (Following standard practice, we use the symbols /ʃ/, /ʃdi/, and /ʃi/ for the Japanese sounds, though these are generally pronounced as prepalatal, Vance 1987, Itô and Mester 1995).

Itô and Mester (1995,1997) account for the asymmetry between /ti/ and /si/ with a phonological model of the production grammar in which coexisting subgrammars are associated with different lexical strata. These subgrammars, formulated in the framework of Optimality Theory, consist of ranked constraints of two types: markedness (structural) constraints, which prohibit particular structures (such as /ti/ or /si/) from appearing in phonetic representations, and faithfulness constraints, which demand faithful correspondence between lexical representations and their phonetic realizations. Constraints may be violated under pressure to satisfy higher-ranked constraints, and rankings may differ from language to language. However, all subgrammars within a language are argued to maintain the same relative rankings of their markedness constraints, differing only in the position of faithfulness constraints in the hierarchy. The native vocabulary of Japanese is subject to a subgrammar in which the markedness constraints dominate the faithfulness constraints: *[ʃi] (no [si] in phonetic representation) >> *[ʃi] (no [ti] in phonetic representation) >> IDENTPAL (value for palatality cannot be changed). This ranking ensures that even when /ti/ or /si/ occurs in a lexical representation, it cannot be realized as such, and instead undergoes palatalization, the minimal
Crawford points out that the /ti-di/ asymmetry cannot reflect Japanese listeners’ interpretation of English stop aspiration as affrication; even unaspirated /t/ is subject to palatalization, as in suʃiɾu ‘steel’.

No more successful is a model based on transfer of the native language phonological distinctive features. Brown (1998) argues that foreign language learners will find foreign contrasts harder to perceive when the feature that distinguishes the foreign language phonemes plays no role in distinguishing a native language contrast. But Japanese phonology does require a feature distinguishing /ti/ from /ʃi/ and /si/ from /ʃi/, as these sounds contrast in other vocalic contexts (ta ‘field’ vs. ʃa ‘tea’, saku ‘fence’ vs. ʃaku ‘wine-serving’). Furthermore, since it is presumably the same feature that distinguishes the palatal from nonpalatal sounds, the two contrasts should pose equal difficulty, if difficulty reflects phonological specifications.

A potential solution to this problem lies in models that characterize the relative difficulty of learning new foreign language categories in terms of articulatory or perceptual similarity to native categories (e.g., the Speech Learning Model, Flege 1986, the Perceptual Assimilation Model, Best 1995, the Native Language Magnet model, Iverson and Kuhl 1996, and the L2LP model, Escudero 2005). In general, these models identify cases in which two foreign language phonemes map equally well to a single native language phoneme category as the greatest
challenge for learners. Both the English [ti]-[tʃi] and [si]-[ʃi] contrasts potentially fall into the category of the mapping of two FL sounds to a single NL category, so any explanation of their asymmetric difficulty must rest on the differences in goodness of fit between the English and the Japanese sounds. In other words, if it could be shown that English [si] and [ʃi] better approximate Japanese [ʃi] than English [ti] and [tʃi] approximate Japanese [ʃi], this would account for the asymmetry in adaptation rate. However, because the models do not provide a \textit{priori} definitions of the goodness of fit between the foreign and native sounds, each case requires comparison of the foreign and native acoustics as well as of perceived goodness of fit. In section 5 we outline a plan to compare the acoustics of the English and Japanese sequences as well as the Japanese listeners’ perception of the distance between English [ti]-[ʃi] and Japanese [tʃi] and between English [si]-[ʃi] and Japanese [ʃi].

4. Choice of Repair

4.1. Japanese Consonant Change vs. Vowel Change

As discussed above, Japanese speakers often transform English sequences [ti], [si] by palatalizing the consonant. However, changing the vowel would also repair violations of [*ti], [*si]. Although the vowel lowering strategy is significantly less common (Crawford 2007, to appear) than palatalization of the consonant (as in [ʃi]mu ‘team’) or maintenance of the original sequence (as in tiilin ‘teen’), it is attested (teʃu ‘tissue’, Itô and Mester 1995, 828). However, vowel lowering is almost never used to repair English [si] sequences (Crawford, personal communication).

The existence of forms like teʃu ‘tissue’ is surprising, as it contradicts the native language faithfulness constraint ranking $\text{IDENTPLACE(VOWEL)} \gg \text{IDENTCONSONANT}$. Furthermore, Kenstowicz (2003) has argued that precisely the opposite ranking is fixed and universal, based on the crosslinguistic preference for palatalization over vowel lowering as repair of coronal-[i] sequences. Kenstowicz’ argument is based in Steriade’s (2001) P-Map Model, in which speakers rank faithfulness constraints so as to maximize perceptual similarity between lexical and phonetic representations. The choice of [ʃi] over [te] as the phonetic realization of lexical /ti/ “implies that the stimulus pair [ti]-[ʃi] should be judged more similar than [ti]-[te]” (Kenstowicz 2003, 20), though each pair should be accurately discriminated. Alternatively, we could assume that perceptual similarity plays a much more direct role—that [ti] is more likely to be misperceived as [ʃi] than as [te], and that this misperception is in turn more likely than the perception of [si] as [ʃi]. In section 5 we outline a plan to tease apart the predictions of these two approaches.

4.2 Korean Stop-Nasal Sequences: Nasalization vs. Devoicing vs. Vowel Insertion

In our lab, Hwang (2006) has found additional error patterns that appear not to be explicable in terms of the NL production grammar or the foreign input. Hwang directed native speakers of Korean to repeat nonce words, recorded by a native speaker of English, which contained stop-nasal sequences. Such sequences are illegal in Korean, where a stop before a nasal obligatorily undergoes nasalization (knuk+m+ul > knʌŋ+mul ‘soup’; lap+h+man > am+m+an ‘front only’). Because Kang (2003) found evidence that Korean speakers tend to perceive release of English word-final stop as a vowel, the prenasal stops in the English stimuli were unreleased.

\[2\text{But see the findings of Cutler et al. (2000) that in changing nonwords into real words, participants were more likely to alter vowels than consonants.}\]
Participants’ accuracy in producing forms such as *tegnal*, *tebnal* was fairly low; only 52% of the responses correctly reproduced the English stop-nasal sequence (451/866). For the incorrect responses, only 18% involved nasalization, while 20% involved insertion of a vowel between the stop and the nasal (*tegnal > teginal*), and 43% involved devoicing of a voiced stop before a nasal (*tebnal > tepnal*). The latter pattern is particularly puzzling as it violates both the native language prohibition on obstruent-nasal sequences and the native language requirement that lax stops are voiced when they occur between voiced segments (e.g., /kuk+i/ > *kugi* ‘soup (nom.)’, /han+kul/ > *hangul* ‘Korean script’).

Furthermore, Hwang found two asymmetries in the choice of repair. First, while nasalization was equally likely to occur with either voiced or voiceless obstruents preceding a nasal, vowel insertion occurred almost exclusively after voiced consonants. Second, vowel insertion occurred exclusively after [g], while devoicing occurred almost exclusively with [b]. A similar [b]/[g] asymmetry was found by Major and Faudree (1996), whose Korean-speaking ESL learners devoiced 93% of final [b] in a word list reading task but only 53% of final [g]. Kabak (2003) and Kabak and Idsardi (2007) found a perceptual asymmetry related to voicing; their Korean listeners had more difficulty distinguishing [gm] and [gVm] than [km]-[kVm] pairs (as did some English subjects; see section 5 for more discussion of Kabak and Idsardi’s results.)

Once again, these patterns are problematic if viewed as an effect of the production grammar. The NL grammar of the Korean speakers has the ranking *STOP-NASAL >> DEP-VOWEL* (no vowel insertion) >> IDENT-NASAL (no change in nasality), making nasalization the preferred repair (e.g., /kuk+nul/ >*kunjmul* rather than *kukimul*). English provides no evidence that would motivate learners to rerank their constraints to favor vowel insertion. Furthermore, the production grammar would need to contain separately ranked constraints for velar and labial sounds, though neither Korean nor English provides any evidence for such rankings.

Section 5 outlines a plan to investigate possible explanations of devoicing and vowel insertion as reflections of misperception. We will also consider the hypothesis that the vowel insertion is a result of gestural mistiming, rather than actual epenthesis.

5. Proposed Project and Work Plan
5.1. Differential Rate of Acquisition: Japanese [ti]-[ʃi] vs. [si]-[ʃi]
We will examine the production and perception of these contrasts by native speakers of three different languages in the context of the following hypotheses:

- **Greater acoustic overlap impedes acquisition of new contrasts.** Japanese speakers acquire the English [ti]-[ʃi] contrast more quickly than the English [si]-[ʃi] contrast because Japanese [ʃi] overlaps with both English [si] and [ʃi], while Japanese [ti] is more distinct from English [ti].

- **Greater perceptual salience of a contrast facilitates acquisition of new contrasts.** Japanese speakers acquire the English [ti]-[ʃi] contrast more quickly than the English [si]-[ʃi] contrast because the cues signalling the [ti]-[ʃi] contrast are perceptually more salient than those signalling the [si]-[ʃi] contrast. (Of course, this hypothesis and the former are not incompatible, and both factors could well play a role.)

- **Asymmetries in perceptual salience are independent of native and foreign language structure.** If differences in perceptual salience of different contrasts are universal (as predicted by the P-Map hypothesis), even listeners whose NL has neither contrast or both contrasts might still show evidence of different degrees of sensitivity to the two contrasts.
We will compare the production and perception of the English [ti]-[ti] and [si]-[ji] contrasts by native speakers of English, Japanese and Korean (20 speakers of each), recruited from the large international population in the Stony Brook area, which includes speakers ranging from very low levels of English proficiency (newly arrived family members and/or students enrolled in the non-credit Intensive English Center) to very advanced levels (matriculated students). We will restrict participation to those who did not reside in an English-speaking country before the age of 12 and who have normal hearing, based on standard audiogram for frequencies from 250 Hz - 4KHz. From each participant we will collect detailed information concerning linguistic background, including all foreign language study, years of English study, age of exposure to English, TOEFL, SPEAK, and TSE scores. We expect, based on the results of the Pallier et al. 1997 Spanish-Catalan bilingual study discussed above as well as our own experience with such speakers, that even after extensive exposure to English, listeners will continue to experience perceptual interference from their native language sound system.

The rationale for including native speakers of Korean in this portion of the project is to provide a comparison group: for English speakers, both the [ti]-[ti] and [si]-[ji] distinction can serve to signal lexical contrasts, while for Korean speakers, in whose NL /s/ becomes /ʃ/ before [i], only the [ti]-[ti] distinction is contrastive. For Japanese speakers, the status of this distinction is more complex: /t/ obligatorily palatalizes before /i/ in native vocabulary, but some loanwords have been lexicalized with /ti/. (However, as Weber and Cutler (2004), Cutler, Weber, and Otake (2006) have demonstrated, lexicalization of a foreign contrast does not guarantee native-like online discrimination.) We will also include the [ka]-[ka] contrast, which is significant in Japanese and Korean but not in English. This will provide a “no contrast” baseline against which English responses to si/ji and ti/ʃi can be compared.

<table>
<thead>
<tr>
<th></th>
<th>ti-ʃi</th>
<th>si-jí</th>
<th>ka-ka</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td>contrast: teak vs. cheek</td>
<td>contrast: see vs. she</td>
<td>no contrast</td>
</tr>
<tr>
<td><strong>Korean</strong></td>
<td>contrast: ti ‘dust’ vs. ʃi ‘guy’</td>
<td>no contrast</td>
<td>contrast: karim ‘dividing’ vs. k’arim ‘small and longish’</td>
</tr>
<tr>
<td><strong>Japanese</strong></td>
<td>no contrast in native vocabulary</td>
<td>no contrast</td>
<td>contrast: kaʃa ‘freight car’ vs. k’aʃa ‘fragile, delicate’</td>
</tr>
</tbody>
</table>

**Proposed Production Study (Acoustic Analysis):** If the asymmetry in the acquisition of the two English contrasts by Japanese speakers is primarily a function of goodness of fit between the Japanese and English sounds, this should emerge in the acoustic analysis of production data. Acoustic measures of the goodness of fit will be based on a comparison of productions of native language forms in the three languages that are as similar in segmental and prosodic context as possible. Non-native production accuracy (of Korean and Japanese subjects speaking English) will be assessed using the same measures, but comparing non-native to native English forms. We will use spectral measures of the fricatives—center of gravity, skewness, amplitude ratio, and centroid frequency above F2 (e.g., Jongman et al., Edwards and Beckman 2007)—and durational and burst/fricative spectral measures for [t] versus [ʃ] (Shaw 2007).

---

[^3]: Cho (1999) demonstrates that at an earlier stage, Korean palatalized both /t/ and /s/ before /i/, but that changes in the vowel system made /t/-palatalization opaque, so that the [ti]/ʃi] contrast has now become established in the language.
Proposed Perception Study (AX, ERP): The perception experiments will involve the same sets of participants, though attrition may require us to recruit some new subjects to meet subject targets. We will use both an AX discrimination task and a study of event-related potentials carried out in Prof. Squires’ ERP lab. For the discrimination task, the syllables [si, ji, ti, tj] will be spoken by a native speaker of English and the [ka, k’a] syllables by a native speaker of Japanese. Randomized pairs of same and different syllables (e.g., [ti-tj] and [si-ji]) will be presented to each group of subjects, who will be asked to judge whether the two syllables they heard were the same or different. In the ERP task, subjects will watch a silent movie while unattended auditory stimuli play in the background. Subjects will be exposed to counterbalanced oddball paradigms (e.g., ti..ti..ti..ti) in which one member of the contrasting pair is the standard and the other is the deviant. ERPs will be recorded from 64 scalp sites referenced to linked mastoids using a NeuroScan system. The EEG will be digitized at 500 Hz, over a time base of 900 msec including a 100 msec pre-stimulus baseline. Standard (p = .8) and deviant (p = .2) stimuli will be presented in random order at a rate of one stimulus per second. Separate ERPs for standard and deviant stimuli will be averaged offline, rejecting trials with eye-movement and other artifacts. The MMN will be measured in the difference waveforms obtained by subtracting the ERP to the standard from the ERP to the deviant.

In the discrimination task we expect a strong effect of NL, so that English speakers, for example, should discriminate [ti]-[tj] and [si]-[ji] equally well, and better than [ka-k’a]. However, the ERP task might conceivably reveal finer-grained differences bearing on the question of inherent asymmetries in perceptual similarity. We expect the three groups of subjects in our experiment to exhibit significantly stronger MMN responses to changes that are linguistically significant in their native language. However, on the hypothesis that perceptual asymmetries hold independently of NL, we might see differences in response independent of NL. Some evidence for NL-independent perceptual asymmetry is presented by Hacquard and Walter (2004), who carried out an MEG study comparing responses to oddball paradigms consisting of [ba]-[da] and [ma]-[na]. They found that the [ba]-[da] change evoked a stronger MMF than the [ma]-[na] change, even though the NLs of all participants employed both contrasts. Thus, if the production asymmetry of the Japanese speakers reflects general asymmetry in perceptual salience of the [ti]-[tj] and [si]-[ji] contrasts, independent of native language, we expect that all subjects might show some difference in the strength of the MMN response to the two contrasts.

While a small number of previous studies (e.g., Best et al. 1988) compare adult listeners’ responses to two novel contrasts, the proposed study would to our knowledge be the first to use ERP methodology to explicitly examine the relationship between second language learners’ problems in pronunciation and their ability to perceive the relevant foreign language contrasts. Having production data for some of the same participants will allow an even closer examination of the relationship between perception and production accuracy. We hypothesize that a stronger MMN response to a foreign language contrast may correlate with more accurate or more consistent native-like production in the same individuals.

Proposed Lexical Stratum Perception Study (ERP): As the [ti]-[tj] change is potentially significant for Japanese listeners only in foreign vocabulary, an additional ERP study will be carried out to investigate whether Japanese listeners’ perception of the [ti]-[tj] and [si]-[ji] contrasts is affected by cues biasing them to interpret tokens as associated with a native or foreign lexical stratum. Some evidence for such an effect comes from a study by Moreton and Amano (1999), who investigated whether Japanese listeners’ classification of ambiguous stimuli was affected by the presence of phonological cues biasing listeners toward hearing a word as native or foreign. In the Japanese lexicon, long vowels occur only in native words. Moreton
and Amano’s Japanese listeners tended to identify the same vowel as long in bisyllabic stimuli containing structures (such as [hj]) which occur only in Sino-Japanese vocabulary but as short in a bisyllabic stimulus containing structures (such as [fa], phonetically [ϕa]) which occur only in foreign vocabulary.

We will use ERP to investigate Japanese listeners’ responses to the oddball nonce word paradigms [hja(canvas)-[hjaf[i] and [ϕati]-[ϕaf[i], both recorded by a native speaker of Japanese. If the Japanese lexical strata are in fact associated with subgrammars, each assigning a different status to the [ti]-[t[i] distinction, we expect a stronger MMN response to the [hjati]-[hjaft[i] change (which invokes the foreign stratum, where the [ti]-[t[i] change could potentially signal a lexical contrast), than to the [ϕati]-[ϕaf[i] change (where this distinction is not contrastive). We will include pairs [hjati]-[hjaft[i] and [ϕati]-[ϕaf[i], along with [hjas[i]-[hjas[i] and [ϕas[i]-[ϕas[i] (for comparison) in the ERP experiments targeting Japanese listeners. English listeners will serve as controls. (See Darcy et al. (in press) for evidence from a word detection task that advanced L2 learners were able to compensate for the patterns of contextual variation appropriate to each language. See also Mitterer and Blomert (2003), who found that Dutch listeners exhibited different MMN strengths to standard-deviant pairs with contextually appropriate variation (such as tuin banc-tuim banc ‘garden bench’ where lexical /n/ optionally assimilates to following /b/) than to pairs with contextually inappropriate variation (tuin stoel-tuim stoel ‘garden stool’).

5.2. Differential Repair
5.2.1. Japanese Palatalization vs. Vowel Lowering in English /ti/ and /si/
In this component we will investigate why modifications of English [ti] to Japanese [t[i] are considerably more frequent than modifications of [ti] to [te], which are in turn more frequent than modifications of [si] to [se], in the context of the following hypotheses:

- **A particular repair strategy is favored because it maximizes the perceptual similarity between the foreign and the adapted form.** Listeners may be able to accurately perceive a particular FL structure (e.g., [ti]) but will produce it as the most similar NL structure (e.g., [t[i]). The relative perceptual similarity of different structures is independent of NL.

- **A particular repair strategy is more common because it represents the most likely misperception of the foreign form.** Listeners may not consistently perceive the FL structures accurately, and the more frequent repair strategies will reflect more frequent confusions.

**Proposed Relative Similarity Study (AX, ERP):** Based on frequency of use in repairs, we predict the following similarity relationships: \( \Delta[s[i]-[se] > \Delta[s[i]-[t[i], \Delta[s[i]-[se] > \Delta[t[i]-[te], \Delta[t[i]-[te] > \Delta[t[i]-[t[i]. \) Following up on the preceding set of discrimination/ERP experiments, we will test Japanese, Korean, and English speakers’ ability to discriminate these pairs using both an AX behavioral discrimination task, and ERP. If the Japanese preference for specific repairs results not from inaccurate perception but rather from NL-independent perceptual similarity relationships (used by learners to project constraint rankings, Steriade 2001a,b, Kenstowicz 2003), we expect that speakers of all 3 languages should successfully discriminate [ti]-[te], [ti]-[t[i], [si]-[t[i], and [si]-[se]. However, for all 3 groups, MMN responses to the perceptually more similar distinctions should be weaker than responses to the less similar distinctions, even for those whose NL makes use of both contrasts (following Hacquard and Walter 2004). On the other hand, if inaccurate production is at least in part due to inaccurate perception, we should
expect Japanese speakers to find perceptual similarity reflected even in the discrimination task.

5.3.2. Korean Nasalization vs. Devoicing vs. Vowel Insertion

Participants in Hwang’s (2006) study employed devoicing and vowel insertion in stop-nasal sequences, in addition to the NL strategy of nasalization. Furthermore, devoicing generally applied only to labials (e.g., tebnal > tepnal) and vowel insertion only in a velar context (e.g., tegnal > teginal). We will investigate these patterns in light of the following hypotheses:

**Perception Hypothesis— for Korean listeners (Hwang 2006):**

- **Bilabial closure induces a bias toward perception as voiceless; velar closure induces a bias toward perception as voiced.** This would account for the greater likelihood for [b] than [g] to be devoiced. Prediction: [bN]-[pN] harder to distinguish than [gN-kN] (Korean and English speakers).

- **Voicing induces a bias toward perception of an illusory vowel.** This would account for the greater likelihood of vowel insertion after voiced than voiceless stops. Prediction: [gN]-[giN] harder to distinguish than [kN]-[kiN] (Korean).

- **Therefore, among voiced stops, velar closure induces a bias toward perception of an illusory vowel.** This would account for the greater likelihood of vowel insertion in [gN] than in [bN]. Prediction: [gN]-[giN] harder to distinguish than [bN]-[biN] (Korean).

Voicing in Korean is fully predictable, with voiced stops occurring only as an allophonic variant of lenis stops surrounded by voiced segments. Because Korean syllable contact restrictions do not allow stops to surface before sonorant consonants, any voiced stop in Korean will be followed by a vowel. Therefore, if Korean listeners interpret stop voicing as a cue for the presence of a following vowel, and if they hear the stop as voiced, they are likely to hear a vowel following it. The vowel [i], the epenthetic vowel in loanwords such as *tʰeŋi*i ‘tennis’ (Kang 2003), is the likely candidate for this illusory vowel.

Whether Korean listeners can in fact perceive English voicing contrasts is an important question. In an AX discrimination task, Kabak (2003) and Kabak and Idsardi (2007) found that Korean speakers residing in the US discriminated [km] from [gm] nearly as well as English speakers. However, they argue that voicing contrasts do not affect Korean speakers’ perceptual mapping, based on participants’ discrimination of stimulus pairs such as [kC]-[koC], [gC]-[goC], and [tʃC]-[tʃiC]. Participants performed about equally well with respect to the first two pairs, successfully discriminating the CC from the CVC stimuli (with the exception of the only stop-nasal pairs, [gm]-[gvm]). However, participants were less successful in discriminating CC-CVC pairs in which the first consonant was palatal. Kabak and Idsardi argue that the discrimination judgements are determined by the syllable structure restrictions of Korean. Because palatal consonants may not appear in Korean codas, they argue, listeners supply an illusory vowel following the palatal (in this case [i], the vowel that is normally epenthized after a palatal consonant). In contrast, [k], though banned before a nasal consonant in Korean, may occur in syllable codas. From this Kabak and Idsardi conclude that only restrictions on possible syllables, not restrictions on possible sequences, induce perception of an illusory vowel.

However, because [g] is also not possible in Korean codas (being limited to prevocalic position), the fact that listeners discriminated gC-gVC pairs about as well as kC-kVC pairs (except for gm-gVM) is problematic for the claim that only restrictions on possible syllables induce illusory vowels. Kabak and Idsardi (2007, 44) therefore argue that “the place and manner information
retrieved from the stimuli was mapped onto the underlying phoneme, /k/, of which [g] is a predictable variant in Korean.” An alternative analysis of this data is proposed by Steriade (2007), who suggests that the successful discrimination of the kC-kVC and gC-gVC pairs may have resulted from the fact that Kabak and Idsardi used English [o] as the vowel in the kVC and gVC stimuli, rather than the Korean epenthetic [i]. Steriade provides evidence from loanwords suggesting that Korean listeners map English [o] to Korean [u] rather than the epenthetic [i]. Thus, she argues that Korean listeners’ success in discriminating the CC and CVC pairs results from the fact that the CVC stimuli included a clearly non-Korean element.

In any case, the fact that nearly all instances of vowel insertion in Hwang’s data occurred after [g] remains to be explained. (This fact is congruent with Kabak and Idsardi’s (2007) finding that listeners, both Korean and (surprisingly) English, had somewhat more difficulty in discriminating [gm]-[gVm] than [km]-[kVm]. Kabak and Idsardi’s stimuli did not contain labial-nasal sequences, so no comparison of g-m and b-m is recoverable from their data). The tendency to hear a vowel after [g] but not [b] would follow if Korean listeners were more likely to hear [g] as voiced than to hear [b] as voiced. Interactions between place and voicing in production are crosslinguistically common (and perhaps universal; see Cho and Ladefoged 1999). Furthermore, both English speakers and chinchillas have been shown to have different category boundaries for voicing in labial and velar stops in CV syllables (Lisker and Abramson 1964, Kuhl and Miller 1978). Reporting on an experiment with English listeners, Benki (2001) notes that “bilabial and alveolar stimuli were more likely than velar stimuli to be classified as voiceless, largely independent of F1 transition pattern” (Benki 2001,17). If the Korean participants were less likely to perceive [b] as voiced than [g], then the greater likelihood of devoicing of [b] but vowel insertion after [g] is expected.

Pilot Perception Study: In order to test the hypothesis that the vowel inserted in [gN] sequences reflects misperception, Hwang, Broselow, de Leon, Squires, and Sandiford (2007) recorded stimuli CVCNVC and CVCiNVC produced by a Korean-English bilingual speaker who was instructed to produce the words as English words but containing the [i] vowel of Korean. They then synthesized 2 continua, ranging from, e.g., teginal (80 msec vowel) to tegnal (no vowel), in increments of 20 msec. 8 Korean and 7 English speakers in an AX discrimination task were presented with pairs of stimuli and asked to identify the members as the same or different. Both English and Korean listeners showed evidence of a similar perceptual boundary for the bN-biN continuum. However, for the gN-giN continuum, the Korean participants exhibited much weaker evidence of a boundary than the English speakers.

The results of the discrimination study were used to construct a comparison of Korean and English listeners’ responses to the change from gN to giN and from bN to biN. Stimuli were constructed using the pair on either side of the point where English listeners had a categorical boundary for gN vs. giN (20 vs. 40 msec). These were also the pairs that the Korean listeners were most likely to identify as different, though their percentage of ‘different’ judgements was much lower than that of the English listeners. Using a passive listening oddball paradigm, it was found that while both English and Korean listeners showed evidence of an MMN to the bN-biN contrast, Korean listeners showed a much weaker response to the gN-giN contrast than did the English listeners. The black line indicates the response to a standard (e.g., [gn] following [gn]) and the lighter line indicates response to a deviant (e.g., [gin] following [gn]). The point at which the difference between the deviant and the preceding standard was audible was approximately 250 msec, so the MMN peaks fall within the window of 160-220 msec following the mismatch (Luck 2005):
Proposed Production Study (Acoustic Analysis, Ultrasound): To complement the repetition task employed in the pilot study, we will conduct a production task in which 20 native speakers of Korean read English sentences containing target stimuli in phrase-initial position. These productions will be transcribed by phonetically trained native speakers of English. The production data will be analyzed to compare the Korean and English productions in terms of the cues used to signal voicing contrasts (following methodology in Hwang and Tserdanelis 2006). Based on Hwang and Tserdanelis’ (2006) study of Korean productions of coda voicing in nonsense words, we expect the acoustic analysis of the production data to show that Koreans have mastered some of the acoustic cues to the English voicing contrast better than others. Comparing our perception and production data will yield insight into whether and how the production cues used are determined by perceptual salience.

We will also investigate the possibility that vowel insertion actually represents gestural mistiming rather than true insertion of a vocalic target. Davidson (2003, 2006, 2007) and Davidson and Stone (2003) have shown, using ultrasound imaging, that vowels perceived as inserted into non-native consonant clusters may lack the characteristics associated with true vowels, and are better analyzed as an effect of insufficient overlap between the consonantal gestures (see also Hall 2003). In order to distinguish true vowel insertion from gestural mistiming, we will analyze both acoustic and articulatory data. We will assess vowel formants and duration in non-native nonce forms (e.g., tegnal) and forms presented in Korean orthography with underlying vowels (e.g., teg). In addition, we will analyze ultrasound data on tongue shape and position, collected using the ultrasound facilities at the Cornell University Phonetics Laboratory. The ultrasound data will add an additional perspective on the usefulness of articulatory analysis to evaluate language learners’ abilities and progress in learning, and will provide the investigators with the opportunity to develop expertise with this new methodology, which is being applied more frequently to a wide range of questions concerning the mental representation of sound structure.

Proposed Perception Study (AX, ERP): In addition to investigating listeners’ responses to (i) the gn-gVN and bN-bVN distinctions, we will also investigate (ii) gN-kN vs. bN-pN, to determine whether [b] is more likely to be perceived as voiceless than is [g] (hypothesized to hold for both Korean and English speakers), and (iii) gN-gVN vs. kN-kVN, to determine whether an illusion
vowel is more likely to be perceived after the voiced than the voiceless stop (hypothesized to be a language-specific effect of Korean grammar). For subjects who participated in both the production and perception studies, we will also investigate the correlation between their performance in the perception experiments and the effectiveness with which they signal these distinctions.

5.4. Work Plan

<table>
<thead>
<tr>
<th>Summary of Proposed Studies</th>
<th>English</th>
<th>Japanese</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative perceptual salience: ti-[ʃi], si-[ʃi], ka-[kʰa] Production, AX, ERP</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Perceptibility relative to lexical strata: hjati-hja-[ʃi], φati-φa-[ʃi]</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Perceptibility of different repairs: ti-[ʃi]-te and si-[ʃi]-se Production, AX, ERP</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Perceptibility of different repairs: Korean stop-nasal Production, Ultrasound, AX, ERP</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

For all of the studies planned, production data will be collected before subjects take part in the perception experiments. For the perception studies, half will participate first in the AX component and half first in the ERP study. Stimuli will be produced using speakers who are not participants in any of the studies. The studies of the [si]-[ʃi], [ti]-[ʃi], [ti]-[te] and [si]-[se] contrasts are independent of the Korean stop+nasal study, and thus can be run in parallel.

The first step in the investigation of the [si]-[ʃi], [ti]-[ʃi], [ti]-[te] and [si]-[se] contrasts is establishing target English values to compare with similar Korean and Japanese native forms, and to use to evaluate Korean and Japanese speakers’ English production accuracy. Therefore, the English production component will be done prior to running the non-native speakers, so that analysis of the English data is well underway, or completed, when the non-native speakers are run. Beyond this critical ordering, Japanese and Korean subjects will be run in the production, and then in the perception studies, as they become available. Results of our analyses will be presented at professional conferences, and written up for publication, as they are completed.

**Year 1:** Run English and Japanese production studies for [si]-[ʃi], [ti]-[ʃi], [ti]-[te] and [si]-[se] and do data analysis in preparation for perception studies; perform ultrasound study of Korean stop+Nasal epenthetic vowels and English controls; develop stimuli for Korean stop+Nasal study and begin running perception experiments.

**Year 2:** Complete production data analysis and develop monosyllabic stimuli for [si]-[ʃi], [ti]-[ʃi], [ti]-[te] and [si]/[se] AX and ERP studies and bisyllabic stimuli for lexical context study ([hjat]-[hja-[ʃi]], [φat]-[φa-[ʃi]], [hjas]-[hjasi], [φas]-[φa-[ʃi]]); begin running AX and ERP studies; analyze ultrasound data.

**Year 3:** Continue running perception studies and continue analysis of data; where possible, bring back subjects from previous year to investigate possible changes due to a year in an English environment.

6. Prior support and relationship to PIs’ research and teaching

The project described above is closely connected to the research and teaching of the investigators. The Department of Linguistics houses the programs in ESL teaching and teacher preparation, and both Broselow and Huffman regularly teach courses for current and
prospective foreign language and ESL teachers, as well as courses related to the theory of second language acquisition. Squires teaches in the Biopsychology Program in the Department of Psychology, and all three participate in the interdisciplinary Language, Mind, Brain seminar. A consistent focus of Broselow’s research has been the implications of language contact phonology for theories of phonology and acquisition. Most recently, she has received support through NSF’s Dissertation Improvement Program (award #0418670) which provided funding for an investigation of how the articulatory and perceptual properties of glottal stop contribute to its crosslinguistic patterning (Borroff 2005, 2006). Huffman’s recent research, supported by NSF grant #0325188 “Adaptive Spoken Dialog with Human and Computer Partners”, has focused on adaptation between speakers of different languages and dialects. Huffman (in press) emphasizes the need for attention to multiple forms of language use, while Stent, Huffman, and Brennan (to appear) describes speakers’ clear speech adaptations to errors by a simulated speech recognition system, finding evidence for both global articulatory adjustments and local, specific changes in segmental targets. Squires’ research focuses on the sensory and cognitive processes of the human brain, and has included work on ERP and fMRI in stroke patients with aphasia, and neurological correlates of reading disability. Dr. Susana De Leon, who will serve as postdoctoral associate on the proposed project, used ERP methodology in Squires’ lab as part of her Ph.D. dissertation investigating whether phoneme-discrimination difficulties play a part in reading disability in college-age students. Funding of the project will allow Jiwon Hwang, who will be a Research Assistant on the project, to continue her dissertation research on Korean speakers’ pronunciation and perception of English.

7. Broader impacts

This is an interdisciplinary project involving researchers in formal phonology, phonetics, second language acquisition, and the neurophysiology of language processing. It will be one of the few studies to investigate the relative perceptibility of foreign language sequences using both behavioral tasks and ERP, and to relate the findings of the perception studies to formal models of grammar and grammar acquisition. The findings of the project have implications for theories of language acquisition and for hypotheses concerning the plasticity of the neural structures underlying speech processing, as well as for the field of foreign language instruction. Students at both undergraduate and graduate levels will be encouraged to develop related projects in second language acquisition, and the graduate students supported on this project will receive valuable multidisciplinary training which will help prepare them to make future contributions to cognitive science. Findings will be disseminated to audiences in phonology, phonetics, second language acquisition, and the neurophysiology of language processing.