Journal of Language and Linguistics Vol. 4 No. 1 2005 ISSN 1475 - 8989

Optimal L2 Speech Perception: Native Speakers of English and Japanese Consonant Length Contrasts

Rachel Hayes-Harb

University of Utah, USA

Abstract

This paper examines the perception of Japanese singleton-geminate consonant contrasts by native speakers of English with varying degrees of experience with Japanese. In an identification experiment, it was found that while native speakers of Japanese exhibit consistent identification of consonants varying in duration as either singleton or geminate, native speakers of English do not. However, this effect was mediated by Japanese language experience—native speakers of English who have studied Japanese for up to one year exhibited more Japanese-like identification of Japanese consonants. A formal Optimality Theoretic model of second language speech perception development is proposed, building on the Boersma (1999) perception grammar and the Gradual Learning Algorithm (Boersma & Hayes 2000). The application of the model to the experimental results is demonstrated.

Introduction

Adults' perception of speech sounds is strongly influenced by the status of the sounds in the phonological system of their native language, and adult listeners typically exhibit difficulty detecting phonetic differences that do not correspond to native phonemic contrasts. However, despite initial difficulty in identifying novel sounds and discriminating non-native contrasts, adult second language learners often do develop more target-language-like perception of non-native sounds. It is the goal of this paper to present a quantitative analysis of this type of development and to propose a model that accounts for the data.

A vast literature has demonstrated the importance of the native language in determining second language speech perception patterns, especially in the initial state of second language acquisition (but see Bohn 1995 for a discussion of other relevant factors). Models of second language speech perception have been primarily concerned with predicting perception difficulties as a function of the relationship between the native and non-native languages' sound systems (e.g., the Speech Learning Model, Flege 1995 and the Perceptual Assimilation Model, Best 1995). This notion of transfer has also been demonstrated to account for a wide variety of second language phenomena outside the domain of speech perception (e.g., White 1987 for syntax; Broselow 1983, 1984 for phonology). One of the most influential hypotheses for transfer in second language acquisition research posits that the initial state in second language acquisition is a full instantiation of the native language grammar (Full Transfer), and that development beyond that state is enabled by full access to the language acquisition device (Full Access; the Full Transfer/Full Access hypothesis, Schwartz & Sprouse 1996). In addition to original work in rule-based linguistic frameworks, the Full Transfer/Full Access hypothesis has also been formalized under the constraint-based framework of Optimality Theory (OT). In these OT studies, Full Transfer is taken to mean that the learner begins with a full instantiation of the native language's hierarchy of constraints, and Full Access is taken to mean that the learner has access to all constraints and a language learning device that re-ranks constraints based on evidence in the linguistic input. Davidson (1997) and Hancin-Bhatt & Bhatt (1997) propose accounts of developmental L2 production data within the framework of OT, and more recent work has explored OT as a framework for accounting for developmental L2 perception data as well. In an OT perception grammar, proposed by Boersma (1999), the input is a set of acoustic descriptions provided by the auditory system, and the grammar uses a language-specific hierarchy of perception constraints to evaluate the "fit" of the input into the phoneme categories of the language (the candidate set). The winning candidate, then, is the phoneme category that is perceived. This theory of perception, in concert with a mechanism for OT learning, The Gradual Learning Algorithm, (Boersma 1997, Boersma & Hayes 2000), has been applied to the development of native Spanish speakers' ability to discriminate English tense-lax vowel distinctions (Escudero & Boersma 2002, 2004),

and the present study examines its ability to account for the development of native English speakers' perception of Japanese singleton and geminate consonants.

Consonant length[1] is a contrastive parameter in Japanese, as evidenced by the minimal pairs *kisaki–kissaki* ('empress'-'point of a sword'), *kika–kikka* ('your home'-'chrysanthemum'), and *oto–otto* ('sound'-'husband'). The shorter of the two consonant length classes is called singleton (or short/single), and the longer is called geminate (or long/double). Consonant duration is an important cue to the perception of consonant length contrasts in many languages (e.g., Lisker 1957, Pickett & Decker 1960, Lahiri & Hankamer 1988).[2] On the other hand, length is not a contrastive parameter for English consonants. To demonstrate this difference between Japanese and English, Hayes (2001) provided some typical intervocalic consonant durations for both Japanese and English production data (see Table 1). Note that the English consonants are similar in length to Japanese singleton consonants, and that Japanese geminate consonants are much longer than either Japanese singletons or English consonants.

TABLE 1. *Japanese and English durations for voiceless alveolar and velar stops and alveolar fricatives (reported in Hayes 2001; average durations are reported in msec with standard deviations in parentheses)*

Japanese

consonants	**t/tt**	**k/kk**	**s/ss**
singleton duration	95.7 (9.5)	81.7 (9.3)	136.1 (12.4)
geminate duration	276.1 (21.7)	223.6 (21.7)	270.1 (11.2)

English

consonant	**t**	**k**	**s**
duration	58.7 (5.6)	65 (5.9)	159.2 (13.5)

[1] The term duration refers to an acoustic property of speech segments; the term length refers to the related phonological property.

[2] It is important to note, however, that there are other possible cues to the perception of consonant length. Although most researchers agree that consonant duration is the most reliable cue in the perception of singleton-geminate contrasts, there are other factors that may play a role. Obrecht (1965) found that consonant duration had a relatively weaker role in the perception of the Arabic /s/-/ss/ contrast than either the /b/-/bb/ contrast or the /n/-/nn/ contrast. He suggested that other acoustic cues may play a role in the perception of the /s/-/ss/ contrast, including rate of formant transition, relative intensity, and time distribution of intensity. In the experiment presented in this paper, only consonant duration was manipulated.

It is predicted that these differences between the Japanese and English phoneme inventories have consequences for perception by native speakers of Japanese and English. Given that Japanese has two phoneme categories on the consonant duration continuum, it is predicted that native speakers of Japanese will identify Japanese consonants varying in duration categorically—as either singleton or geminate. On the other hand, given that English has only one phoneme category on the consonant duration continuum (which is similar in duration to Japanese singleton consonants), it is hypothesized that monolingual English speakers will identify variation in Japanese consonant length as non-phonemic, which will lead to a linear (not categorical) identification function. That is, even in a forced-choice task where they are asked to identify consonant sounds as single or double, they should be unable to assign tokens varying in duration to phonemic categories in the way that native Japanese speakers do. However, it is hypothesized that this effect of transfer will be mediated by experience with the Japanese language, and that native speakers of English who have studied Japanese will exhibit identification performance that is closer to that of native Japanese speakers. The following experiment was designed to test these hypotheses.

Identification Experiment

In the identification experiment, three groups of participants (native speakers of Japanese, monolingual English speakers, and native English speaking learners of Japanese) were asked to identify Japanese consonants as "single" or "double" (i.e., singleton or geminate). For each participant group, identification functions were plotted, and the functions were compared across groups.

Participants

Participants represent three populations: native speakers of Japanese (n=8); monolingual English speakers (with no Japanese language experience; n=24); and native English-speaking learners of Japanese in their first year of Japanese language study in an undergraduate university setting (Japanese learners; n=21). The Japanese speakers were all monolingual speakers of Japanese prior to their acquisition of English at school. The

English speakers all had native English speaking parents, English was their first language, and they did not speak any other language fluently; the monolingual subset spoke exclusively English on a daily basis, and the Japanese learner subset had studied Japanese for up to two semesters at a university level.

Stimuli

A native speaker of Japanese produced Japanese nonwords of the form: CVCV (singleton frame) and CVCCV (geminate frame); the two frames were identical except for the length of the medial consonant. See Table 2 for an exhaustive list of the nonwords used in the experiment.

TABLE 2. *Nonwords from which the experimental stimuli were created*

consonant	**t – tt**	**k – kk**	**s – ss**
nonword	[ataa] – [attaa] [kuto] – [kutto]	[gakee] – [gakkee] [noka] – [nokka]	[haso] – [hasso] [hisa] – [hissa]

From each frame, a series of 13 stimuli was created. For the stop consonants /t/ and /k/, stop closure duration was manipulated; for the fricative /s/, frication duration was manipulated. Consonant duration differed by 20 millisecond (msec) intervals between 70 msec and 310 msec. The values 70 msec and 310 msec were chosen because they represent unambiguous singleton and geminate consonants durations, respectively, in Japanese. Both singleton and geminate frames were used in order to prevent other potential acoustic cues to the perception of singleton and geminate contrasts from affecting the results. By manipulating only this single acoustic parameter, all other potential acoustic cues to the contrast were held constant.

The stimulus [ataa-290] is built from the singleton frame *ataa*, with a synthesized closure duration of 290 milliseconds. The stimulus [nokka-110] is built from the geminate frame *noka*, with a synthesized closure duration of 290 msec. Notice that the stimulus [ata-290] (despite being created from a singleton frame) is more likely to be perceived as

containing a long consonant than [attaa-110] (despite being constructed from a geminate frame) because the closure duration in [ataa-290] is so much longer than in [atta-110].

There were three consonant conditions (**t–tt, k–kk**, and **s–ss**), with two nonwords in each consonant condition. From each frame a series of 13 stimuli was created, differing only in medial consonant duration. There were a total of 156 stimuli (3 consonant conditions x 2 nonwords for each consonant condition x 2 frames x 13 durations).

Method

The experiment was run using the EXPE experiment generator (Pallier, Dupoux & Jeannin 1997). Participants listened to each stimulus over headphones and were asked to judge whether they heard a word with a single or a double consonant. Subjects recorded their answers by pressing a button labeled "single" (for singleton consonants) or a button labeled "double" (for geminate consonants). The order of presentation was randomized across participants, and every participant heard every stimulus. Participants in the monolingual English group, who were expected to be unfamiliar with the concept of single and double consonants, were told that even though they were not used to hearing single and double consonants in English, their decisions should be made on the basis of the length of the consonants they heard.

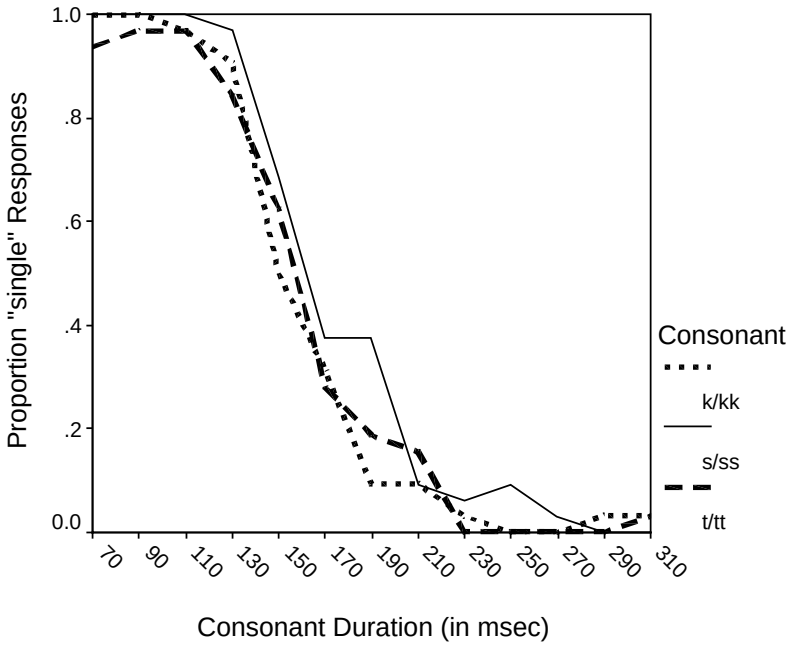
Results

Singleton responses were plotted to create identification functions for each of the three participant groups. Data points plotted in the line graphs below are averaged across participant, frame, and nonword for each participant group. The data in this format was not submitted to tests of inferential statistics; inferential analyses of an alternative coding of the data are found below. However, these identification functions are visually illustrative of the categorical versus linear nature of perception by the three participant groups.

Because singleton and geminate consonants are contrastive in Japanese, it was predicted that the native speakers of Japanese would assign the majority of stimuli to either "single"

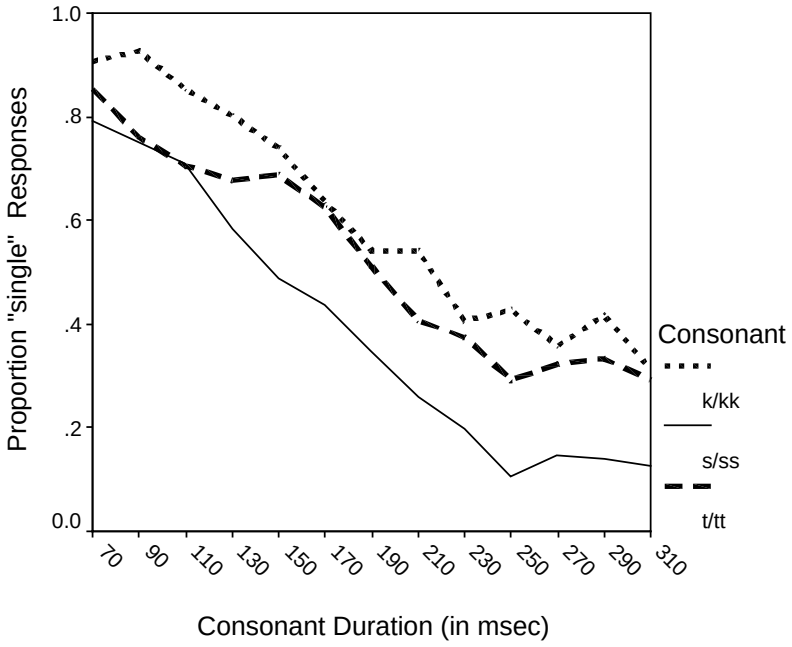
or "double" categories, and this is confirmed visually by the categorical identification function in Figure 1. Note that at the endpoints of the continuum, the native speakers of Japanese exhibited 100% and 0% "single" identification, with a sharp decline over a small portion of the duration continuum—between 130 and 230 ms. This categorical function is typical of listeners' identification of native language phonemes.

FIGURE 1. *Proportion of "single" responses by native speakers of Japanese by consonant and duration*



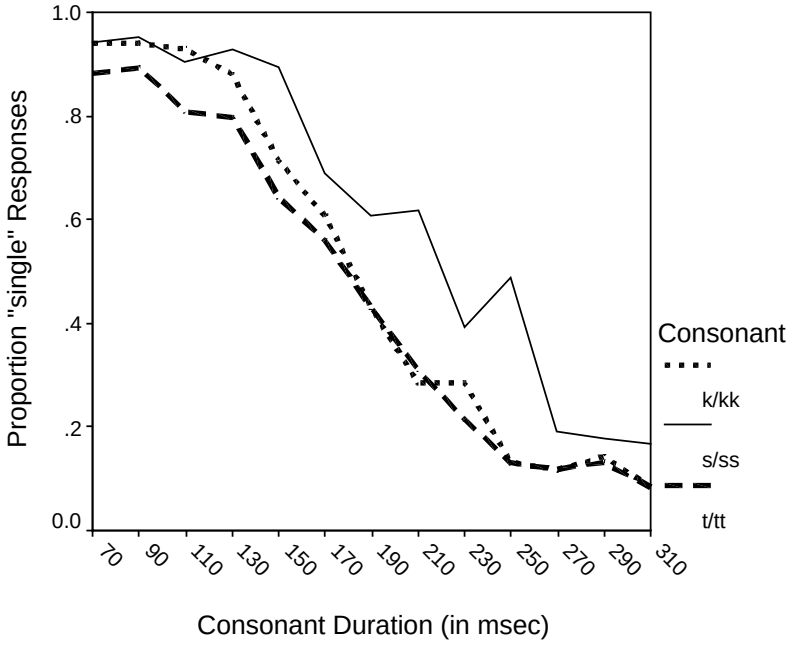
On the other hand, it was predicted that monolingual English participants would transfer their native English perceptual system to the task of identifying the Japanese stimuli and would therefore not treat the duration continuum as contrastive. The identification function in Figure 2 visually confirms this hypothesis. Note that the monolingual English participants did not reach 100% or 0% identification at the endpoints of the continuum, and that the identification function is for the most part linear.

FIGURE 2. *Proportion of "single" responses by monolingual English speakers by consonant and duration*



Finally, it was predicted that native speakers of English who have studied Japanese would exhibit a more categorical identification function of the Japanese stimuli than the English monolinguals. The data in Figure 3 visually confirms this hypothesis. Note that identification of the endpoint stimuli approaches 100% and 0% "single" relative to the monolingual English participants, and that the identification function has a sharper slope—between approximately 130 and 250 msec—than for the monolingual English participants.

FIGURE 3. *Proportion of "single" responses by learners of Japanese by consonant and duration*



The data presented above gives the visual impression that native speakers of Japanese exhibit categorical perception of consonant length, monolingual English speakers exhibit linear (continuous) perception of consonant length, and native speakers of English in their first year of Japanese study exhibit perception that is somewhere in between that of the native speakers of Japanese and the monolingual English speakers. A categorization analysis was used to quantify these visual impressions. First, each subject heard each duration (70 msec – 310 msec, at 20 msec intervals) of each consonant four times (two nonwords x two frames). An ANOVA revealed no difference between the nonwords (*ataa* versus *oto*) or the frames (singleton or geminate), so responses to all four stimuli at each duration were averaged for each participant. There are five possible averages of four responses: 100%, 75%, 50%, 25%, or 0% "single". 100% and 0% indicate that the particular participant categorized all four stimuli at the particular duration (4/4 categorized) and 75% and 25% indicate that the participant categorized three out of the four stimuli at the particular duration (3/4 categorized). 50% "single" responses were interpreted as indicating that the duration is "not categorized". Figure 4 presents a split histogram showing the percentage categorized at the 3/4 or 4/4 level, averaged across participant, for all three participant groups.

FIGURE 4. *Categorization analysis of the identification experiment results, by participant group*

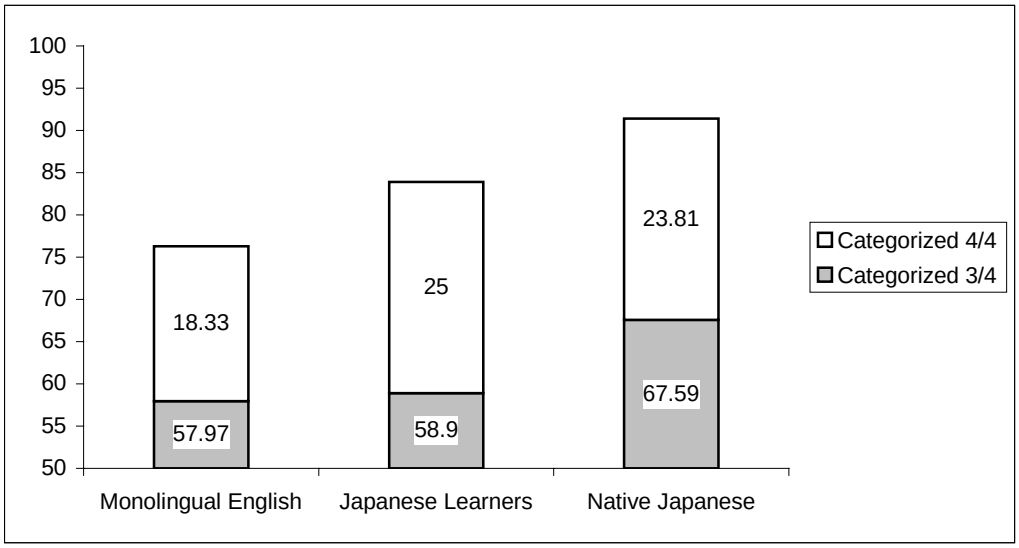


Figure 4 provides a breakdown of 3/4 and 4/4 categorization; however, the statistical analysis was performed on the sum of 3/4 and 4/4 categorized data. Monolingual English speakers categorized 76.3% of the stimuli; native speakers of Japanese categorized 91.4% of the stimuli, and Japanese learners categorized 83.9%. Native speakers of Japanese categorized more of the stimuli than the native speakers of English ($F1(1,30)=9.494$, $p<0.05$; $F2(1,12)=6.818$, $p<0.05$). Additionally, the native speakers of English who have studied Japanese for one year categorized more of the stimuli than the monolingual English speakers ($F1(1,43)=4.451$, $p<0.05$; $F2(1,12)=5.796$, $p<0.05$). These results establish that, relatively, monolingual English speakers perceive consonant length continuously, that native speakers of Japanese perceive consonant length categorically, and that learners' perception is more categorical than that of monolingual English speakers.[3]

As a result of the fact that the monolingual English participants and the Japanese learners do not have robust phonemic categories on the duration scale, their perceptual performance on this task was highly variable—they exhibited variability in performance when perceiving speech sounds that they cannot readily assign to phoneme categories.

[3] The difference between performance by the Japanese learners and the native speakers of Japanese in this analysis was not significant.

This variability, as will be seen below, is an important characteristic of non-native speech perception, and is predicted in particular by the proposed model.

The Optimal L2 Speech Model

The experimental findings support a model where the development of second language phoneme categories is gradual, and that learning indicates incremental effects of the target language input. As mentioned earlier, a model of L2 speech perception must additionally account for variability non-native listeners' perception. The remainder of the paper is devoted to explaining the proposed formal model of L2 speech perception—the Optimal L2 Speech model, which takes advantage of some main insights from Boersma's (1999) OT perception grammar. Boersma's (1999) perception grammar is introduced below, followed by an account of the perception of singleton and geminate consonants in Japanese, as well as the perception of English consonant length.

According to Boersma's OT perception theory, the language-specific grammar evaluates the "fit" of the auditory input into the available phoneme categories. In this version of OT, the input is what the listener hears. The competing candidates, then, are different phoneme categories into which the input may be assigned. The role of the grammar is to assign each input to an underlying phoneme category, and the difference between languages with different sound inventories is how each grammar assigns elements of the speech signal to phoneme categories.

There are two main classes of perception constraints in this model: *CATEG(ORIZE) and *WARP (Boersma 1999). The general concept behind this model is that there is an opposition between that which is the prototypical example of a phoneme category in a particular language and that which the listener hears, which are variants of the prototype.[4] If all speech sounds exactly matched the central member of a phoneme category, then modeling the assignment of speech sounds to phoneme categories would be trivial—a

[4] For the purpose of demonstrating the model, the term 'prototype' refers to the average example in a category, though it may be the case that it should refer instead to the median value, or some other value. Distinguishing between the two (or other possibilities) is beyond the scope of the paper.

simple one-to-one mapping. However, the acoustic properties of speech are highly variable—though in systematic ways, and a more complex theory of this mapping is needed.

The *CATEG constraints are responsible for determining the inventory of phoneme categories of a language. In Figure 5, *x* refers to a phoneme category's average value along a particular acoustic dimension. As reported above, 95.7 msec is the average length of a singleton /t/ in Japanese (rounded to 96 msec). That is, a /t/ with a duration of 96 msec is the prototypical duration for a singleton /t/, and 96 msec is the representative duration for that category. Because the constraint *CATEG T/96 msec/ disallows a category with 96 msec as its prototypical duration—and this is precisely the category we need to posit for Japanese speech perception, *CATEGT/96 msec/ must be ranked lower in Japanese than *CATEGT/*y* msec/, where *y* refers to any length other than 96 msec. This is abbreviated as *CATEGT/other/.

FIGURE 5. *Speech sound categorization constraints: *CATEG (Boersma 1999)*
***CATEG/phoneme category x/**
do not categorize an input into the phoneme category *x*

The 'T' included in the constraints is a shortcut indicating that all other acoustic properties point to the input being categorized as a "voiceless coronal stop". The acoustic property of immediate interest, duration, determines whether the input is a singleton /t/ or a geminate /tt/.[5] According to the typical Japanese consonant durations reported in Hayes (2001), the lowest-ranked *CATEG constraints are those presented in Table 3.

TABLE 3. *The lowest-ranked *CATEG constraints in Japanese (all values are in msec)*

***CATEG constraint**	**definition**
*CATEGT/96/	do not assign any input to the category T/96/
*CATEGT/276/	do not assign any input to the category T/276/
*CATEGK/82/	do not assign any input to the category K/82/
*CATEGK/224/	do not assign any input to the category K/224/
*CATEGS/136/	do not assign any input to the category S/136/
*CATEGS/270/	do not assign any input to the category S/270/

[5] Similarly, 'K' and 'S' represent "voiceless velar stop" and "voiceless coronal fricative," respectively.

As mentioned above, the constraint *CATEG/other/ represents all other possible *CATEG constraints, and must be higher-ranked than the constraints listed in Table 3. For example, the constraint *CATEGT/250/ is in principle part of the grammar, but it is not relevant to the current analysis because it does not refer to an available phoneme category in Japanese. Thus, for a given analysis, /other/ should be interpreted as any value on the relevant acoustic scale that does not refer to an existent phoneme category. *CATEGT/other/ necessarily outranks the *CATEG constraints that refer to a language's existing phoneme categories; this is how a language's phoneme category inventory is delimited. For the sake of simplicity, the analyses presented here will consider only the perception of /t/ and /tt/; however, the analysis can be straightforwardly extended to perception of the /s/-/ss/ and /k/-/kk/ distinctions.

In Tableau 1, each of the four candidates represents a potential perceived phoneme category (again, specified by its prototypical duration). This ranking requires that no input be categorized into T/50/ or T/300/ (*CATEGT/50/ and *CATEGT/300/ are included in the shortcut constraint *CATEGT/other/), because T/50/ and T/300/ do not represent available phoneme categories in Japanese. Thus candidates (a) and (b) incur fatal violations of the constraint *CATEGT/other/, and candidates (c) and (d) are the potential winning candidates at this point. It remains to be explained, though, how a language chooses between the two available categories in assigning inputs—this requires a second set of constraints: the *WARP constraints.

TABLEAU 1. *Illustration of the *CATEG constraints in Japanese*

T[any duration]	*CATEGT/other/	*CATEGT/96/ *CATEGT/276/
☞ ? a. T/50/	*!	
☞ ? b. T/300/	*!	
c. T/96/		*
d. T/276/		*

*WARP constraints require similarity between the acoustic properties of the input and the prototypical acoustic properties of the categories into which they are assigned. *WARP constraints serve a similar function to faithfulness constraints in traditional OT. In Figure 6, *y* refers to an input value on a particular acoustic scale. In the case of consonant

duration, a value for *y* is a duration in msec. Some example *WARP constraints are provided in Table 4.

FIGURE 6. *Speech sound categorization constraints: *WARP (Boersma 1999)*
***WARP[acoustic value *y*]→/phoneme category *x*/**
do not categorize an input with the value *y* into the phoneme category *x*

TABLE 4. *Some example *WARP constraints (all values are in msec)*

***WARP constraint**	**Definition**
*WARPT[100]→T/96/	do not assign any input of T[100]) to the category T/96/
*WARPT[100]→T/276/	do not assign any input of T[100]) to the category T/276/
*WARPT[300]→T/96/	do not assign any input of T[300]) to the category T/96/
*WARPT[300]→T/276/	do not assign any input of T[300]) to the category T/276/

The *WARP constraints include specification of an input length (e.g. T[300]) and the name of a phoneme category (e.g. T/96/) to which it must not be mapped. An important feature of the *WARP constraints is their universal ranking with respect to each other, based on the acoustic distance between the input specification and the prototypical value for the relevant phoneme category. The larger the distance, the higher the ranking—this captures the generalization that if a listener categorizes a consonant *x* with a length of T[150] as a singleton, consonant *x* with a length of T[90] will not be categorized as belonging to a category with a longer prototypical duration (the geminate category).

TABLE 5. *Acoustic distances represented by example *WARP constraints*

***WARP constraint**	**acoustic distance**
*WARPT[100]→T/96/	4 msec
*WARPT[100]→T/276/	176 msec
*WARPT[300]→T/96/	204 msec
*WARPT[300]→T/276/	24 msec

These distances like those presented in Table 5, determine the universal ranking of the *WARP constraints with respect to each other. Only *WARP constraints referring to the same input value have rankings relative to each other—there is no need to rank constraints like *WARPT[300]→T/96/ and *WARPT[100]→T/96/ because they are never in conflict with each other. The constraint *WARPT[300]→T/96/ is ranked higher than the constraint *WARPT[300]→T/276/ because the former represents an acoustic distance

of 204 msec, while the latter represents a distance of only 24 msec. These universal rankings ensure that inputs are categorized into the most similar available phoneme category. However, because languages have limited inventories of phoneme categories, the grammar must determine the most similar phoneme category that is available. This is accomplished via the interaction of *CATEG and *WARP constraints.

The relative ranking of particular *CATEG constraints versus particular *WARP constraints determines, among other things, whether there are separate "singleton" and "geminate" categories, and what the prototypical duration is for each category. According to the experimental results discussed above, an input of T[110] is perceived as belonging to the length class T/96/ by native speakers of Japanese, as formalized in Tableau 2.

TABLEAU 2. *Japanese, input of T[110]*

T[110]	*CATEGT/110/	*WARP T[110]→T/276/	*WARP T[110]→T/96/	*CATEGT/96/ *CATEGT/276/
☞ a. T/96/			*	*
b. T/276/		*!		*
c. T/110/	*!			

In Tableau 2, *WARPT[110]→T/276/ is universally ranked above *WARPT[110]→T/96/ because the former represents a distance of 166 msec, while the latter represents a distance of only 14 msec. In Japanese, an input of T[250] is perceived as belonging to the category with T/276/ as its prototype. It cannot be categorized into the category T/250/ because that category is not available in Japanese (due to the highly-ranked *CATEGT/250/ constraint). Given the available phoneme categories T/96/ and T/276/, the universally-ranked *WARP constraints determine that the input of T[250] must be categorized as T/276/, the best available option (see Tableau 3).

TABLEAU 3. *Japanese, input of T[250]*

T[250]	*CATEGT/250/	*WARP T[250]→T/96/	*WARP T[250]→T/276/	*CATEG T/96/ *CATEG T/276/
a. T/96/		*!		*
☞ b. T/276/			*	*
c. T/250/	*!			

In Tableau 3, *WARPT[250]→T/96/ must be ranked higher than *WARPT[250]→T/276/ because the former represents a distance of 154 msec, while the latter represents a distance of only 26 msec. This ranking produces the correct result: an input of T[250] is categorized by native speakers of Japanese as a "geminate" (belonging to the category T/276/).

In this model, cross-linguistic variation with respect to consonant duration categories is formalized as variant rankings of the relevant constraints. Based on the Hayes (2001) acoustic study discussed above, it is assumed that the prototypical durations of the English /t/, /k/, and / s/ consonants are 59, 65, and 159 msec, respectively. The phoneme categories corresponding to these prototypical durations are captured in the ranking in Figure 7.

FIGURE 7. *Ranking of the *CATEG constraints in English*
***CATEGT,K,S/other/ >> *CATEGT/59/, * CATEGK/65/, * CATEGS/159/**

In English, an input of T[150] is perceived as belonging to the length class T/59/ because there is no other available phoneme category in English (*CATEGT/59/ is the only low-ranked *CATEG constraint in English). Under this analysis, in English, an input of T[250] should also perceived as belonging to the length class T/59/ (see Tableaux 4 and 5).

TABLEAU 4. *English, input of T[150]*

T[150]	*CATEG T/150/ *CATEG T/other/	*WARPT[150]→T/59/	*CATEGT/59/
☞ a. T/59/		*	*
b. T/250/	*!		
c. T/other/	*!		

TABLEAU 5. *English, input of T[250]*

T[250]	*CATEG T/250/ *CATEG T/other/	*WARP T[250]→T/59/	*CATEG T/59/
☞ a. T/59/		*	*
b. T/250/	*!		
c. T/other/	*!		

Because there is no length category in English other than T/59/, any variant of T along the duration continuum should be perceived as belonging to this category. This analysis

predicts that native speakers of English will perceive any stimulus along a duration continuum as belonging to the category "single", and that the monolingual English participants in the experimental study would have identified each stimulus as belonging to the category T/59/. The experimental results for monolingual English speakers indicate that this is not, in fact, the case—instead, the monolingual English speakers perceived consonant duration continuously. This result may be, in part, an artifact of the experimental method—given that participants were instructed to identify each stimulus as containing either a "single" or a "double" consonant, monolingual English speakers were biased to hear at least some of the consonants as being "double"—belonging to a non-native category. Their tendency to assign Japanese consonants to the category "double" was linearly related to the duration of the consonant—a continuous identification function is consistent with not having distinct categories on the duration continuum. In order to understand why the identification function of the monolingual English speakers was continuous in this way, it is necessary to understand how variability is formalized in this model.

It was shown above that participants in the monolingual English and the Japanese learner groups exhibited variable perceptual performance. That is, relative to the native speakers of Japanese, the native speakers of English were less likely to categorize the experimental stimuli.[6] Here it is argued that the continuous nature of perception by the monolingual English participants is a direct result of the variable (stochastic) ranking of the relevant *CATEG and *WARP constraints in their grammars. In addition, the more categorical nature of the perceptual performance of the Japanese learners results from a gradual re-ranking of the relevant constraints over the course of their Japanese language experience. An additional theoretical mechanism, the Gradual Learning Algorithm (GLA; Boersma 1997, Boersma & Hayes 2000), predicts these kinds of variability in grammar, and is able to specifically account for: (i) the variability of the performance by the monolingual English speakers; (ii) the decreased variability in the Japanese learners' performance; and (iii) the relative lack of variability in the performance by the native speakers of Japanese.

A central assumption of the GLA is that variability in linguistic performance results directly from the probabilistic nature of relative constraint rankings. Unlike in traditional OT, where constraints have fixed rankings relative to each other, in this stochastic version of OT, constraints have variable rankings. Main tenets of the GLA include: (i) constraints have absolute—not relative—ranking values; (ii) at evaluation time, an element of random noise disturbs these absolute ranking values such that constraints with sufficiently close ranking values may switch their order; and (iii) the likelihood that two constraints will switch their order at evaluation time is a function of the overlap of their normal distributions. When two constraints have the same absolute ranking value, their normal distributions overlap completely (because they are identical), and there is a 50/50 chance that either will dominate at evaluation time.

When two constraints have different absolute ranking values,[7] the one with the higher ranking value will typically outrank the other, and the probability of this happening decreases as the constraints have ranking values that are further apart. There is, in principle, the possibility that any two constraints will switch their relative ranking at evaluation time. According to the GLA, then, constraints appear to be strictly ranked relative to each other because of the tendency for them to be ranked in a particular order at evaluation time. Variability in this model is not the result of optionality—it is principled in this stochastic grammar.

Modeling the Monolingual English data

Before demonstrating how the model accounts for this data, one additional note is necessary: Boersma (1999) points out that even when listeners have only one category on a particular acoustic continuum (as do the native speakers of English on the duration continuum), it is possible to encode a cutoff, where a listener perceives an input as being too different from the available phoneme category. In the present model, the consequence in this case is for a null candidate to be considered optimal by the grammar. This "cutoff"

[6] According to the stochastic model of grammar assumed in this paper, even native language grammars are variable. It is the *degree* of variability that distinguishes native from non-native grammars.

[7] At this point, I avoid assigning actual ranking values—the assignment of actual ranking values necessitates additional discussion that is outside the scope of this paper. The basics of the model are nonetheless demonstrable.

is encoded in the constraint hierarchy by ranking *WARP constraints related to the excessive acoustic distance above a new constraint, *NULL, described in Figure 8.[8]

FIGURE 8. *NULL
***NULL(PARSE)**
all inputs must be categorized

For the purpose of illustration, we will assume that although English has the category T/59/, an input of T[250] is too different to be perceived as a member of the category T/59/.

TABLEAU 6.[9] *Monolingual English, input of T[250]*

T[250]	*CATEG T/250/ *CATEG T/other/	*WARPT[250]→T/59/	*NULL	*CATEGT/59/
☞ a. null			*	
b. T/59/		*!		*
c. T/250/	*!			
d. T/other/	*!			

In Tableau 6, the null candidate is optimal because the only available category, T/59/, is prevented from being optimal due to the size of the difference between it and the input duration. The forced choice task in the experiment caused monolingual English speakers (who presumably have no "double" consonant category) responded "double" when null parses were optimal and no other response was possible. Now we will see why perception of consonant length by this participant group was continuous.

The rankings represented in Tableau 6 are not strict rankings (the dotted lines in the tableau indicate this). Instead, each constraint has a ranking value on an absolute ranking scale, and depending on the closeness of the ranking values for the constraints, their relative rankings can switch at evaluation time. Constraints that are ranked closer to each other are more likely to have a switched ranking at evaluation time—that is, the more closely-ranked the relevant constraints are, the more variable performance is. Thus

[8] The constraint *NULL is a departure from the Boersma (1999) proposal.

[9] Given that a primary tenet of the GLA is that there are no *absolute* rankings, the lines dividing constraints in tableaux from this point in the paper on will be dotted lines. The order of constraints does, however, represent the *most likely* dominance relation for the particular evaluation time.

performance that is most variable results from the clustering of *CATEG and *WARP constraints near each other on the absolute ranking scale.

Notice, however, that performance by monolingual English speakers was not entirely random—the continuousness of the identification function is the result of universal ranking of the *WARP constraints. A prediction of the universal ranking of *WARP constraints is that listeners will perceive non-native contrasts continuously—the more acoustic distance there is between the input value and the available phoneme categories, the more likely it is that the null candidate will win. Thus the rankings of interest in the identification task performance by the monolingual English participants are the rankings among the constraints that force a null versus a T/59/ choice. The relevant constraints are *NULL and the relevant *WARP constraints, which are presented in Table 6).

TABLE 6. **WARP constraints for the category T/59/*

*WARPT[70]→T/59/	*WARPT[170]→T/59/	*WARPT[250]→T/59/
*WARPT[90]→T/59/	*WARPT[190]→T/59/	*WARPT[270]→T/59/
*WARPT[110]→T/59/	*WARPT[210]→T/59/	*WARPT[290]→T/59/
*WARPT[130]→T/59/	*WARPT[230]→T/59/	*WARPT[310]→T/59/
*WARPT[150]→T/59/		

Each of the constraints in Table 6 frequently switches ranking with *NULL because each of these constraints has an absolute constraint ranking that is close enough to that of *NULL. How frequently they switch ranking with *NULL is a function of their proximity to *NULL on the ranking scale. Recall that the *WARP constraints are universally ranked: *WARPT[70]→T/59/ is the lowest-ranked and *WARPT[310]→T/59/ is the highest-ranked of the relevant constraints. For the purpose of illustration, we will assume that *WARPT[190]→T/59/ and *NULL have the same ranking value (*x*). This is because monolingual English speakers identified T[190] stimuli as "single" 50% of the time. Because these two have the same ranking value, the chance for either to dominate at evaluation time is 50% (see Tableau 7).

TABLEAU 7. *Monolingual English, input of T[190]*

T[190]	*CATEG T/190/ *CATEG T/other/	*WARPT[190]→T/59/ >> *NULL *NULL >> *WARPT[190]→T/59/	*CATEGT/59/
(50%) ☞ a. null		*	
(50%) ☞ b. /59/		*	*
c. /190/	*!		
d. /other/	*!		

In Tableau 7, candidates (a) and (b) each have a 50% chance of being optimal (candidate (a) resulted in a "double" response in the experiment, and candidate (b) resulted in a "single" response). for an input of T[190]. Now we will turn to input durations that are shorter and longer than 190 msec. Consider the absolute ranking values presented in Table 7.

TABLE 7. *Monolingual English absolute ranking values*

constraint	**ranking value**	**constraint**	**ranking value**
***NULL**	***x***	***WARPT[190]→T/59/**	***x***
*WARPT[70]→T/59/	*x-6*	*WARPT[210]→T/59/	*x+1*
*WARPT[90]→T/59/	*x-5*	*WARPT[230]→T/59/	*x+2*
*WARPT[110]→T/59/	*x-4*	*WARPT[250]→T/59/	*x+3*
*WARPT[130]→T/59/	*x-3*	*WARPT[270]→T/59/	*x+4*
*WARPT[150]→T/59/	*x-2*	*WARPT[290]→T/59/	*x+5*
*WARPT[170]→T/59/	*x-1*	*WARPT[310]→T/59/	*x+6*

An input of T[70] is the most likely input to be categorized by monolingual English speakers as "single". The ranking value of x-6 ensures that *WARPT[70]→T/59/, out of all of the relevant *WARP constraints, has the lowest probability of being ranked above *NULL. And an input of T[310] is the input that is most likely to be categorized as "double", because it is the input that has the greatest likelihood of resulting in a null parse—*WARPT[310]→T/59/ is the highest-ranked of all of the relevant *WARP constraints. That each increase in duration had a linear effect on the likelihood that monolingual English speakers would perceive an input as "single" or "double" is captured in this model.

The next section demonstrates how the model can be applied to the data from the Japanese learner group. The native speakers of English learning Japanese modify this

initial grammar in a gradual way to accommodate L2 Japanese input and to form new phoneme categories.

Modeling the Japanese learner data

A model of the Japanese learners' performance needs to account for the increased likelihood that inputs will be categorized either as "single" or "double"—it must account for the increasingly categorical identification by Japanese learners. Assuming that learners begin the process of second language acquisition with a full instantiation of their first language constraint hierarchy (Schwartz & Sprouse 1996), their learning task is to develop two new phoneme categories: T/96/ and T/276/. Thus the ranking values presented in Table 7 above are part of the learners' initial state. In principle, however, the learners also have access to the constraints listed in Figure 9, in addition to a similar set that makes reference to the category T/276/.

FIGURE 9. *Constraints made available by Full Access: *WARP and *CATEG constraints relating to the category T/96/*

*WARPT[70]→T/96/	*WARPT[170]→T/96/	*WARPT[270]→T/96/
*WARPT[90]→T/96/	*WARPT[190]→T/96/	*WARPT[290]→T/96/
*WARPT[110]→T/96/	*WARPT[210]→T/96/	*WARPT[310]→T/96/
*WARPT[130]→T/96/	*WARPT[230]→T/96/	*CATEGT/96/
*WARPT[150]→T/96/	*WARPT[250]→T/96/	

Because neither T/96/ nor T/276/ represents an available category in the learners' native English, the *CATEG constraints associated with each of these durations initially outrank *NULL. The initial state ranking, then, for these constraints is provided in Figure 10.

FIGURE 10. *Initial state ranking of the relevant constraints*
***CATEGT/96/, *CATEGT/276/ >> the *WARP constraints >> *NULL >> *CATEGT/59/**

Boersma (1999) and Boersma & Hayes (2000) propose that development under the GLA is error-driven.[10] That is, the interlanguage grammar re-organizes (adjusts constraint ranking values) when learners detect errors in their perceptual performance (Escudero &

[10] While error detection must play a role in the development of novel phoneme categories, it may not be the primary driving force in L2 perceptual development. Instead, it is possible that the distribution of input values over a learner's experience with the L2 causes restructuring. This argument is based on findings reported in Maye & Gerken (2000), but is not explored here.

Boersma 2002). Take the following example: a Japanese learner perceives (in error) an input of T[270], intended by the speaker to mean *otto* 'husband', as *oto* 'sound'. We will assume that the Japanese learner detects this error because the semantic context indicates that the intended meaning was 'husband' and not 'sound'. The ensuing interlanguage development consists of raising the ranking values for all of the constraints violated in the learner's incorrect winner (*oto* 'sound') and lowering the ranking values for all of the constraints violated by the intended winner (*otto* 'husband'), as demonstrated in Tableau 8.

TABLEAU 8. *Japanese learners, input of T[276]*

T[276]	*CATEG T/276/	*NULL	*WARPT[276]→T/59/	*CATEGT/59/
a. null		*! →		
b. T/276/	*! →			
☞ c. T/59/			← *!	← *

Because either candidate (a) or candidate (b) will give the desired result—the perception of T[276] as a "double" consonant—the constraints violated by either (a) or (b) are lowered. And the constraints violated by the incorrect winner, candidate (c), are raised.[11] In Tableau 8, the arrows indicate the direction of the error-driven reassignment of ranking values. After enough perception errors, this grammar will restructure sufficiently to perform in the manner indicated in the Japanese learners' performance. Crucially, this restructuring occurs gradually, as each perception error has only incremental influence on the reassignment of constraint ranking values.

The demotion of *CATEGT/276/ accounts for the increasingly categorical perception of consonant length by Japanese learners—the closer the ranking value for *CATEGT/276/ gets to that of *NULL, the greater the likelihood that it will be outranked by *NULL at evaluation time, creating a true "double" response on the part of the learners. The word true is intended to contrast this "double" response to the "double" responses by the monolingual English group. In the latter case, a "double" response was an artifact of the

[11] The amount by which constraints are lowered and raised in during learning of this type depends on the assumed plasticity of the grammar at the particular point in acquisition. Determining plasticity values is beyond the scope of this paper.

experimental method, and was the response chosen by the monolingual English speakers when their grammars chose a null candidate as optimal. In the Japanese learners, although null candidates are still often chosen as optimal, the grammar is beginning to develop a "double" category on the duration continuum—due to the gradual demotion of *CATEGT/276/. It is the combination of "double" responses from the demotion of *CATEGT/276/ and "double" responses from an optimal null parse that account for the increasingly categorical performance by Japanese learners. In theory, with enough exposure to Japanese, *CATEGT/276/ would be demoted sufficiently to account for all "double" responses.

What is responsible for "single" responses by the Japanese learners? There are two approaches to this question. The first approach depends on the grammar's acceptance of Ts within the Japanese range of durations (centered around 96 msec) into the English category with a prototypical duration of 59 msec. Undoubtedly, 96 msec is sufficiently close to 59 msec that "single" responses can be due to the absorption of T[96] into the English category T/59/. As long as learners correctly perceive inputs like T[96] as singleton, no error-driven restructuring occurs, as illustrated in Tableau 9.

TABLEAU 9. *Japanese learners, input of T[96]*

T[96]	*CATEG T/96/ *CATEG T/other/	*WARPT[96]→T/96/	*NULL	*CATEGT/59/
a. null			*!	
b. T/96/	*!	*!		
☞ c. T/59/				*
d. T/other/	*!			

This makes the prediction that when L2 input "fits" into a native language phoneme category, the interlanguage grammar will be unlikely to encounter perceptual errors. This lack of perceptual errors and resultant lack of grammatical restructuring explains the absorption of second language phoneme categories into existing first language categories. In perceptual performance, this means that when L1 and L2 categories are similar enough, perceptual assimilation occurs, and the L2 category is likely to retain qualities of the native language category at least temporarily. However, when an L2 sound is sufficiently distinct from exiting native language categories to warrant the creation of a

novel L2 category, there should be less interference from existing native language categories to hinder the development of a target-like category. This phenomenon is well-attested in the literature on L2 speech perception (see Flege 1987, 1995).[12] In other words, it should be more difficult to establish a precisely target-like category where the L2 category competes with an existing, close L1 category (in this case, the English /t/ absorbs the Japanese /t/ category) than where the L2 has a category that does not exist at all in the native language (e.g., Japanese /tt/). In the proposed model, since English does not have a /tt/ category or anything similar to it, learners "start from scratch" in the development of the novel /tt/ category.

In the proposed model, L2 speech perception development results from the gradual reassignment of absolute ranking values associated with the relevant speech perception constraints; the L2 learner makes perceptual errors and restructures the interlanguage grammar accordingly. In this way, both variability and gradual development fall out from the properties of the learning mechanism. An additional prediction of this model is that L2 learners are less likely to make perceptual errors where first and second language phoneme categories are similar enough; this results in greater difficulty in reaching precisely target-like categories when the L2 category is not perceived as distinct from the native language category. However, in the case of a novel L2 category, where the input is perceived as sufficiently distinct from existing L1 sounds to warrant the formation of a novel category, development results in more target-like perception. This has the result of similar but not quite target-like performance in the cases where the first and second language categories are sufficiently similar, but predicts that entirely novel L2 categories should be more target-like. Much research has been devoted to this phenomenon (e.g. Flege 1987, 1995); it is a natural result of the learning model presented here.

[12] The nature of the potential native language interference may depend on whether or not the novel phoneme category requires sensitivity to an acoustic dimension that is not used in the native language.

Conclusions

The experimental study presented in this paper provided evidence that monolingual English speakers perceive Japanese consonant length continuously; that is, they perceive consonant duration as having an linear effect on the likelihood that a stimulus is a singleton or a geminate consonant. Their perception differs markedly from perception by the native speakers of Japanese, who perceived consonant length categorically. This means that they have two distinct categories for consonant length on the duration continuum, and that they perceive stimuli as belonging to either the singleton category or the geminate category, with little in between. Of interest in the present work is how native speakers of English develop the ability to perceive this novel consonant length contrast. It was found that native speakers of English in their first year of Japanese language study perceive Japanese consonant length more categorically than monolingual English speakers, but that they still differed from native speakers of Japanese, indicating that second language learners can develop the ability to perceive a novel phonetic contrast, but that this development is gradual. The proposed model of second language speech perception takes advantage of and extends an OT model of speech perception (Boersma 1999) and an OT learning algorithm (the GLA; Boersma & Hayes 2000), and allows for explicit, testable hypotheses about the nature of L2 perceptual development.

Acknowledgements

I would like to thank Mike Hammond, Mary Zampini, Janet Nicol, Kazutoshi Ohno, Adam Ussishkin, Lisa Shannon, Paul Boersma, and Paola Escudero for their input in this project. Earlier versions of this work were presented at *Laboratory Phonology 8* and the 2002 *Conference on Contrast in Phonology* (Hayes 2002a,b). All errors are my own.

About the Author

Dr Hayes-Harb is Assistant Professor of Linguistics at the University of Utah, Salt Lake City, Utah, USA. Her research interests include: bilingual speech perception, and second language phonology.

Email: hayes-harb@linguistics.utah.edu

References

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange, ed. *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*. Timonium, MD: York Press. Pp 171-204.

Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21, 43-58.

Boersma, P. (1999). On the need for a separate perception grammar. Ms, University of Amsterdam.

Boersma, P. & B. Hayes. (2000). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86.

Boersma, P. & D. Weenink (1992-2001). Praat, a system for doing phonetics by computer. http://www.praat.org.

Bohn, O.-S. (1995) Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange, ed. *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*. Timonium, MD: York Press. Pp. 275-300.

Broselow, E. (1983). Non-obvious transfer: On predicting epenthesis errors. In S. Gass & L. Selinker, eds. *Language Transfer in Language Learning*. Pp. 269-280.

Broselow, E. (1984). An investigation of transfer in second language phonology. *International Review of Applied Linguistics* 22: 253-269.

Davidson, L. (1997). An Optimality Theoretic approach to second language acquisition. Honors Thesis, Brown University.

Escudero, P. & P. Boersma. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In B. Skarabela, S. Fish, and A.H.-J. Do, eds. *Proceedings of the 26th Annual Boston University Conference on Language Development*. Pp. 208-219.

Escudero, P. & P. Boersma. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26, 4, 551-585.

Flege. J.E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47-65.

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 Speech Research*. Timonium, MD: York Press. Pp. 229-273.

Grieser, D. & P. Kuhl. (1989). Categorization of speech by infants: Support for speech sound prototypes. *Developmental Psychology* 25, 4, 577-88.

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

Hayes, R.L. (2001). Singleton-geminate contrasts and second language acquisition: Native speakers of English learning Japanese. *Linguistics Society of America Annual Meeting* (LSA) Poster Session, Washington, DC, January 2001.

Hayes, R.L. (2002a). The perception of Japanese consonant length by non-native listeners. *Laboratory Phonology 8.* Poster, New Haven, CT, June 2002.

Hayes, R.L. (2002b). An OT model of L2 speech perception. *Conference on Contrast in Phonology*. Poster, Toronto, CA, May 2002.

Lahiri, A. & J. Hankamer. (1988). The timing of geminate consonants. *Journal of Phonetics* 16, 327-338.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33, 42-49.

Maye, J. & L.A. Gerken. (2000). Learning phonemes without minimal pairs. In S.C. Howell, S.A. Fish, and T. Keith-Lucas, eds. Proceedings of the 24th Annual

Boston University Conference on Language Development. Somerville, MA: Cascadilla Press. Pp. 522-533.

Obrecht, D.H. (1965). Three experiments in the perception of geminate consonants in Arabic. *Language and Speech* 8, 31-41.

Pallier, C., Dupoux, E. & Jeannin, X. (1997). EXPE: An expandable programming language for on-line psychological experiments. *Behavior Research Methods, Instruments and Computers*, 29, 3, 322-327.

Pickett, J.M. & L.R. Decker. (1960). Time factors in perception of a double consonant. *Language and Speech* 3, 11-17.

Schwartz, B. & R. Sprouse. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research* 12, 1, 40-72.

White, L. (1987). Markedness and second language acquisition: The question of transfer. *Studies in Second Language Acquisition* 9, 261-80.