



**HAL**  
open science

## Development of the ability to lexically encode novel second language phonemic contrasts

Rachel Hayes-Harb, Kyoko Masuda

► **To cite this version:**

Rachel Hayes-Harb, Kyoko Masuda. Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, 2008, 24 (1), pp.5-33. 10.1177/0267658307082980 . hal-00570737

**HAL Id: hal-00570737**

**<https://hal.science/hal-00570737>**

Submitted on 1 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Development of the ability to lexically encode novel second language phonemic contrasts

**Rachel Hayes-Harb** *University of Utah* and  
**Kyoko Masuda** *Georgia Institute of Technology*

Received July 2006; revised November 2006; accepted January 2007

Second language (L2) learners must often learn to perceive and produce novel L2 phoneme contrasts. Although both research and intuition suggest that these difficulties can be overcome to some extent with exposure to the L2, it is not known what consequences this kind of learning has for the phonological structure of the L2 lexicon. We present an experiment designed to investigate the lexical representations that learners establish for L2 words that contain novel phonemic contrasts. Specifically, we consider the acquisition of Japanese consonant length contrasts by native speakers of English: Japanese contrasts consonants such as /k/ and /kk/ while English does not. The results indicate that native English speakers do not initially encode consonant length consistently in their lexical representations of Japanese words, as reflected in both listening and production tasks. However, after one year of Japanese experience, the phonological structure of their Japanese lexicon more closely approximates that of native Japanese speakers. We conclude that significant changes to the structure of the L2 lexicon can occur even within the first year of L2 learning.

**Keywords:** L2 auditory word recognition, lexical representation, novel L2 phonemic contrasts, L2 perception and production

## I Introduction

Adult learners typically experience difficulty perceiving and producing the phonemes of a second language (L2), especially when the native

and second languages have different phoneme inventories and different phonological contrasts. This difficulty contributes to what is commonly observed as ‘foreign-accented’ listening and speaking, and there is much interest among researchers and language teachers into the sources of foreign accents and why they are often so difficult for adult second language learners to overcome.

Research on the acquisition of L2 phonological contrasts has focused primarily on explaining their relative difficulty for learners, and it is generally agreed that the influence of the native language on the acquisition of L2 contrasts depends to some extent on the relationship between the native and L2 phoneme inventories. By most accounts, the degree of difficulty of a novel contrast depends on how each of the sounds is mapped by the listener to corresponding native language categories (Bohn and Flege, 1992; Major and Kim, 1999; Best *et al.*, 2001), and this is formalized in models of second- and cross-language speech; e.g. the Perceptual Assimilation Model (Best, 1995); the Speech Learning Model (Flege, 1995); the Similarity Differential Rate Hypothesis (Major and Kim, 1999); the Native Language Magnet Model (Kuhl and Iverson, 1995); see also Boersma *et al.*, 2003; Escudero and Boersma, 2004). Learners exhibit particular difficulty acquiring a novel contrast when they initially perceive both L2 phonemes as mapping to a single native language phoneme: they must split two native language allophones into two L2 phonemes (for a discussion of allophonic split, see Eckman *et al.*, 2003). The acquisition of the English /ɪ/–/I/ contrast by native speakers of Japanese has received considerable attention in the L2 speech literature as an example of the allophonic split problem (e.g. Goto, 1971; Logan *et al.*, 1991; Bradlow *et al.*, 1997, Bradlow *et al.*, 1999; Takagi, 2002; Aoyama *et al.*, 2004). Native Japanese speakers’ difficulty perceiving and producing English /ɪ/–/I/ contrasts is believed to result from the fact that both of these sounds are sufficiently similar to the Japanese phoneme /ɾ/ (although they are not equally good members of the category; Aoyama *et al.*, 2004) that native speakers of Japanese treat them as allophones of /ɾ/. Thus they map both the /ɪ/ and /I/ sounds in their L2 input to the Japanese /ɾ/ category, effectively neutralizing the contrast to a single perceptual representation. There is evidence that L2 learners neutralize non-native contrasts in perception, even after many years of experience with an L2 (e.g. Takagi, 2002). However, with enough appropriate exposure to

the target language, learners often do improve in their ability to discriminate non-native contrasts, although they rarely reach native-like proficiency (Flege *et al.*, 1995; Takagi and Mann, 1995; Bradlow *et al.*, 1997; Bradlow *et al.*, 1999).

In light of the large literature that specifically considers the perception and production of novel phoneme contrasts by L2 learners, relatively little is known about the consequences of these new contrasts for the phonological content of learners' lexical representations. A lexical representation is the storage of a word in memory, and it contains information about the phonological, morphological, syntactic, semantic and, when available, orthographic structures of words. In the framework of generative phonology it is traditionally assumed that lexical representations contain only contrastive, or unpredictable, information about the phonological structure of words (i.e. phonemes), and that the predictable aspects of surface representations are derived by phonological rules (see Chomsky and Halle, 1968).<sup>1</sup> If lexical representations contain only contrastive information – when L2 learners have no awareness of a novel L2 contrast – they should not be able to encode information relevant to the contrast in their lexical representations, with the result that pairs of words that differ minimally with respect to the novel contrast are encoded as homophones in the learner lexicon.

There is evidence from auditory lexical decision tasks that even highly fluent L2 learners experience patterns of lexical activation that are consistent with homophonous lexical representations of L2 minimal pairs (Pallier *et al.*, 2001; Cutler and Otake, 2004). In an auditory lexical decision task, participants hear words and are asked to indicate as quickly as possible whether each word is a real word. Among the factors that influence participants' response times in lexical decision tasks is how recently a word has been encountered, with faster response times to recently-encountered words. Pallier *et al.* (2001) compared auditory repetition priming in two groups of fluent bilingual speakers of Spanish and Catalan. They found that Spanish-dominant bilinguals (but not Catalan-dominant bilinguals) showed repetition priming for Catalan minimal pairs differing only by contrasts that exist in Catalan but not in

<sup>1</sup> However, whether or not lexical representations also contain non-contrastive information is a matter of debate in the phonology literature. For a discussion of this issue and its relevance to the study of the L2 lexicon, see Pater, 2003.

Spanish (e.g. /o/–/ɔ/: *dóna–dona* ‘s/he gives’–‘woman’). For example, the Spanish-dominant bilinguals’ response to the Catalan word *dona* was faster if they had recently heard the word *dóna*, presumably because they accessed their lexical representations of both *dona* and *dóna* upon hearing either word spoken. Cutler and Otake (2004) found similar repetition priming results for English /ɪ/–/I/ minimal pairs by native Japanese speakers and for English /ɛ/–/æ/ minimal pairs by native speakers of Dutch (/ɛ/–/æ/ is a novel contrast for Dutch speakers). While repetition priming for minimal pairs may appear to indicate that the pairs are lexically encoded as homophones, this finding is also consistent with another explanation: that the learners were unable to distinguish the minimal pairs at the level of perception in these on-line listening tasks. To illustrate, Pallier *et al.* (2001) might have found the same pattern of results if the Spanish-dominant bilinguals had perceived both *dóna* and *dona* as phonologically identical and used the same phonological code to access the lexicon in both cases, whether their lexical representations of the minimal pair were homophonous or not. Therefore, this apparent homophony may instead (or also) have resulted from participants’ difficulty in perceiving the contrast.

In studies designed to tease apart the contributions of perception and lexical representations in L2 word processing, Weber and Cutler (2004) and Cutler *et al.* (2006) used eye-tracking technology to observe L2 learners’ auditory word processing in greater detail. Weber and Cutler (2004) found that native Dutch speakers of English as a second language fixate their visual attention longer and more frequently on distractor pictures of words containing sounds that they are likely to confuse with sounds in an auditorily-presented target word than on less phonologically confusable distractor pictures. For example, the native speakers of Dutch in their study fixated longer and more frequently on a picture of a *pencil* when the target word was *panda* than on a less confusable distractor (e.g. *beetle* when the target word was *bottle*), suggesting that at the point where they heard the first syllable of the word *panda*, they also inappropriately activated the word *pencil*. Although this finding might at first appear to be consistent with a homophonous representation of /ɛ/ and /æ/ in native Dutch speakers’ English lexicons, Weber and Cutler (2004) also found that this pattern of inappropriate lexical activation was asymmetric, with the target *panda* causing activation of *pencil* but not

the reverse. They conclude that this asymmetry in fact provides evidence that the learners have encoded the /ɛ/–/æ/ contrast lexically but that they do not perceive the contrast in the on-line auditory word identification task. They argue that the asymmetry results from the relative phonetic similarity of the English /ɛ/ and /æ/ sounds to their closest Dutch counterpart, which is more similar to English /ɛ/ than to English /æ/. Dutch listeners may therefore perceive both English /ɛ/ and /æ/ as /ɛ/ when listening to spoken English, causing them to activate lexical entries containing /ɛ/ regardless of which of the two vowels is present in the auditory input, thus explaining why they inappropriately activate the word *pencil* when they hear the word *panda*. On the other hand, Dutch listeners' neutralization of both /ɛ/ and /æ/ to /ɛ/ in perception has the result of preventing them from searching for lexical entries containing /æ/ without the support of additional phonological information (i.e. the rest of the word), which explains why they do not inappropriately activate *pencil* when they hear *panda* but are nonetheless able to ultimately identify the picture of the panda as matching the target word *panda*. Weber and Cutler (2004) conclude that this asymmetry in lexical activation provides evidence that these Dutch learners of English have established the /ɛ/–/æ/ contrast in their lexical representations of English words but that their on-line perceptual neutralization of the contrast undermines the information stored in their lexical representations. Cutler *et al.* (2006) found a similar pattern of results for native speakers of Japanese and English /ɪ/–/I/ minimal pairs. These findings lead to the question of how learners can lexically encode a novel contrast that they cannot perceive. Cutler *et al.* (2006) speculate that explicit instruction in the language and/or knowledge of the spelled forms of words may tell learners that a pair of words is contrastive even when they cannot reliably perceive the contrast in listening tasks. This, and the issue of how these contrasts may be represented lexically, is taken up in the discussion section.

The Weber and Cutler (2004) and Cutler *et al.* (2006) studies provide evidence that highly fluent learners are able to establish lexical representations that encode novel L2 contrasts. In fact, Curtin *et al.* (1998) found that monolingual English speakers were able to contrast two novel Thai contrasts in their memory of Thai minimal triplets, even although their exposure to these contrasts was limited to the study duration. Curtin *et al.* (1998) taught monolingual English speakers minimal

triplets whose word-initial consonants reflected the Thai three-way stop voicing distinction among voiced, voiceless unaspirated and voiceless aspirated consonants (English contrasts only voiced and voiceless stop consonants word initially). During the word learning phase of their experiment, which occurred over the course of four training/review sessions spanning 11 days, participants heard the Thai words while pictures indicating their meanings were presented visually. Participants took discrimination and lexical tests on the second and eleventh days of the experiment. In the discrimination test, participants could make their judgements on the basis of auditory information alone, and did not need to access their memory of the words. In the lexical test, participants heard one of the words and were asked to choose which of two pictures matched the word, requiring them to access their memory of the newly-learned words and, in the critical test items, the two pictures were of minimal pairs reflecting either the voiceless aspirated–unaspirated or the voiced–voiceless unaspirated contrast. The English speakers' performance on both tasks was above 70% by the fourth session, indicating that they were consistently able to discriminate and encode the contrasts lexically, although their performance was less accurate than that of the native Thai speakers.

In the present study we are interested in the development of the ability to encode a novel contrast lexically: Our first objective is to investigate the abilities of learners at different stages of second language acquisition to encode a novel phonemic contrast in their memory for newly-learned L2 words. Our second objective is to investigate learners' lexical encoding of a novel phonemic contrast via two different types of tasks – listening and production – in order to inform the development of a model of L2 lexical development that is consistent with both listening and production data.

The acquisition of Japanese consonant length contrasts by native speakers of English provides an opportunity to pursue the present research objectives. Many Japanese consonants are contrastive for length: the phoneme inventory contains short (singleton) and long (geminate) consonant pairs, such as /t/–/tt/ and /s/–/ss/. For example, the words [oto] 'sound' and [otto] 'husband' form a minimal pair in Japanese.

Consonant length contrasts are realized in production primarily as differences in consonant duration in many languages, including Japanese

(Lahiri and Hankamer, 1988; Hankamer *et al.*, 1989; Cohn *et al.*, 1999; Pickett *et al.*, 1999). In Japanese, average ratios of geminate to singleton consonant durations range from 2.6 to 3.0 for voiceless stop consonants (Han, 1992; Toda, 1997; Hayes, 2002), and from 2.0 to 2.1 for geminate to singleton /s/ durations (Toda, 1997; Hayes, 2002). Other acoustic correlates of consonant length contrasts in Japanese include preceding segment duration (i.e. segments preceding geminate stop consonants are 11% longer than those preceding singleton stop consonants<sup>2</sup>; Han, 1994), voice onset time (Han, 1992), and following segment duration (e.g. word-final vowels following geminate stop consonants are 9% shorter than those following a singleton stop consonant; Han, 1994).

In English, however, consonant length is not contrastive in this way,<sup>3</sup> and native speakers of English learning Japanese must learn this novel contrast. In a perception study, Hayes-Harb (2005) asked monolingual English speakers, native English speakers with one year of Japanese experience, and native speakers of Japanese to identify Japanese consonants with manipulated durations along a continuum as 'single' or 'double'. Identification data of this type provides some information about the contrastive nature of an acoustic dimension: listeners tend to perceive tokens along contrastive continua categorically, but perceive tokens along non-contrastive continua either randomly or, depending on the task and the perceptibility of the contrast, continuously. Monolingual English speakers in the Hayes-Harb (2005) study identified the consonants in a continuous manner: the proportion of 'single' vs. 'double' responses was linearly related to the consonant's duration. This suggests that monolingual English speakers do not perceive Japanese singleton and geminate consonants (e.g. /t/ and /tt/) as contrastive, presumably because they perceive the sounds as variants of their English counterparts (e.g. /t/). However, the continuous (and not random) nature of their responses suggests that they are sensitive to differences in consonant duration even if their perception of consonant length is not categorical. Native Japanese

<sup>2</sup> In many languages, vowels preceding geminate consonants are shorter than those preceding singleton consonants (e.g. Italian: Pickett *et al.*, 1999; Icelandic: Pind, 1995). In other languages, including Japanese, vowels preceding geminate consonants are longer (e.g. Persian: Hansen, 2003).

<sup>3</sup> Singleton and geminate consonants are not contrastive morpheme-internally in English. However, consonant length across a word boundary does distinguish pairs of utterances such as 'top pick' and 'topic', although pairs of this type are rare in English.

listeners' identification of the consonants indicated a sharp boundary between 'single' and 'double' responses. The native English-speaking learners of Japanese exhibited identification functions with sharper boundaries than those of the monolingual English speakers, but still differed from those of the native speakers of Japanese, indicating that their year of exposure to Japanese had impacted their ability to perceive singleton–geminate contrasts, although they were still distinguishable in their perception from native Japanese speakers.

Mah and Archibald (2003) report a case study of a native English speaker who, after four months of exposure to Japanese, differentiated singleton and geminate consonants in production by manipulating consonant duration. However, the learner's duration ratios did not match those of native Japanese speakers, and the productions were highly variable. Han (1992) found that even highly fluent native English learners of Japanese differ from native Japanese speakers in the way they differentiate singleton and geminate consonants in production, with average duration ratios ranging from 1.67 to 2.06, smaller than those typically reported for native Japanese speakers. Han (1992) also reported a great deal of variability, both between and within participants; in fact, one of the four speakers in the study 'did not make any of the desired contrasts between single and geminate stops' (Han, 1992: 119). Although it appears that learners can show evidence of differentiating singleton and geminate consonants in production very early on in their acquisition of Japanese, their productions are highly variable and do not match those of native speakers, even after many years of study.

The studies discussed above provide information about perception and production abilities only; they do not specifically shed light on whether the learners are able to establish lexical representations that contrast singleton and geminate consonants. For example, Hayes-Harb (2005) elicited listeners' judgements about the length of consonants embedded in auditorily-presented non-words, for which the learners could have no lexical representations. In another example, although Mah and Archibald (2003) limited their word stimuli to those present in the working vocabulary of the speaker, the production data was elicited with a reading task, and stimuli were presented in *hiragana*, a Japanese script that differentiates singleton and geminate consonants. For this reason, it is not clear that their speaker accessed her lexical

representations of the words, which means that her productions may have solely reflected her ability to decode the written stimuli. In the present study, we are interested not only in perception and production patterns themselves, but additionally in what they may reveal about the phonological content of L2 learners' lexicons. In the word learning phase of the present study, participants learned a set of Japanese non-words, some of which contained geminate consonants. After learning the words, participants completed a matching task (listening test phase) and a naming task (production test phase) designed to elicit information about the phonological content of their lexical representations of the new words.

## **II Listening and production experiment**

### *1 Participants*

Because we are interested in the development of native English speakers' ability to encode Japanese consonant length contrasts lexically, we tested native English speakers with no Japanese language experience ('inexperienced learners', recruited from the University of Arizona community;  $n = 12$ ) and native speakers of English with one year of Japanese language experience ('experienced learners', recruited from second-semester Japanese language classes at the University of Arizona;  $n = 12$ ). We also tested a control group of 12 native speakers of Japanese (Native Japanese), recruited at the University of Arizona and in Japan, all of whom speak the Tokyo dialect. All participants were at least 18 years old. The native English speakers were screened for L2 background: none had studied any other language containing singleton–geminate consonant contrasts (e.g. Italian or Arabic), and none reported speaking any language other than English at home or with relatives or studying any language for more than the equivalent of two college-level semesters.

### *2 Word learning phase*

Participants were taught 12 Japanese non-words words with a (C)VCV structure, where the medial C was either singleton or geminate. Four of the words had a medial geminate consonant (geminate words), four had

a medial singleton consonant (singleton words), and four were filler words.<sup>4</sup> The geminate words differed from the singleton words only in medial consonant length, to form four minimal pairs (two of which were contrasted by /t/–/tt/ and two by /s/–/ss/). Minimal pairs of this type were included in the set of words taught in order to maximize the likelihood that participants would notice that consonant length was a relevant feature of the words, and non-words were used to prevent prior familiarity with Japanese words from influencing participants' performance.

The filler words were all possible words in both English and Japanese; the medial consonants in the filler words were all singleton and did not include /t/ or /s/. All of the non-words were produced by an adult male native Japanese speaker and recorded digitally for presentation in the experiment.<sup>5</sup> The speaker had sophisticated knowledge of Japanese phonology and produced all 12 words with identical high–low pitch–accent. The geminate to singleton duration ratios in these recorded stimuli were similar to the average ratios reported in Han (1992) and Hayes (2002); durations and other acoustic information about the stimuli are provided in Appendix 1.

Each non-word was randomly assigned a meaning, which was represented by a line drawing picture. The non-words and their meanings are listed in Table 1. Participants were tested individually in a quiet room seated at a computer. They were told that they would be learning some Japanese brand names<sup>6</sup> and that their task was to learn the names and their meanings as well as possible, as they would be tested on their memory of the names in subsequent tasks. The words were presented over headphones at a comfortable listening level as the pictures indicating their meanings were displayed on a computer screen using the DMDX presentation software (Forster and Forster, 2003). Each picture appeared for two seconds, with a 500 millisecond pause between stimuli. The 12 word–picture pairs were presented five times each, for a total of 60 trials during word learning, and the 60 word learning trials were

<sup>4</sup> The filler words contain singleton medial consonants but will not be analysed as members of the singleton condition because they do not have geminate counterparts.

<sup>5</sup> The use of a single token of each word produced by a single speaker reduces the generalizability of our findings (for discussion of the importance of variability in L2 auditory input, see Pisoni and Lively, 1995).

<sup>6</sup> We referred to the non-words as brand names because the native Japanese speakers (and perhaps some of the experienced learners) were expected to notice, based on their knowledge of Japanese, that the words they were learning were not correctly matched with the items represented in the pictures.

**Table 1** The nonwords and their meanings, by word condition

Singleton words		Geminate words		Filler words	
Word	Picture	Word	Picture	Word	Picture
pete	dress	pette	stove	peme	television
keto	lemon	ketto	dresser	ehe	horn
tese	motorcycle	tesse	watch	kemo	cup
meso	piano	messo	boot	epo	onion

presented in a different random order for each participant. Presentation of the word learning trials was followed immediately by a practice task.

In the practice task, each of the 12 words was presented twice: once with a correctly matched picture and once with an incorrectly matched picture. Detecting the incorrectly matched word–picture pairs at this stage did not require participants to be sensitive to singleton–geminate contrasts. For example, in the incorrectly matched condition, the word *meso* was paired with the picture of a horn (which matches *ehe*), but would not have been paired with the picture of a boot (which matches *messo*). Participants recorded their responses by pressing a YES (the word and picture match) or NO (they do not match) key on a computer keyboard. The practice task was included to make sure that participants had achieved a minimum level of memory of the words before moving on to the listening and production test phases. Participants who scored less than 90% on the practice task began the word learning phase again; the only feedback they received was whether or not their overall performance on the practice task was accurate enough to proceed to the tests. No participant required more than three word learning–practice cycles, and number of cycles completed did not correlate with performance during the test phases.

### 3 Listening test phase: auditory word–picture matching task

After completing the word learning phase, participants were tested on their ability to match the words and pictures in an auditory word–picture matching task.

*a Materials and procedure:* The method here is identical to that of the practice task described above, except that now, some test items required participants to detect mismatches that depended on

singleton–geminate contrasts. For example, in one of the incorrectly matched items, the singleton word *meso* was paired with the picture of a boot (which matches the geminate word *messo*). The set of test pairs consisted of each of the 12 non-words presented once with its correctly matched picture and once with an incorrectly matched picture, for a total of 24 pairs. The 24 test pairs were presented four times each for a total of 96 trials in the test, and the 96 test trials were presented in a different random order for each participant. The eight filler test pairs consisted of all correctly matched and incorrectly matched pairs of filler words (filler–matched and filler–unmatched). The eight target–matched test pairs consisted of each of the eight target words with their correctly matched pictures, and the eight target–unmatched test pairs consist of the eight target words with an incorrectly matched picture: these last pairs are the ones that required sensitivity only to the singleton–geminate contrast to correctly identify as ‘not matched’. The eight target–unmatched test pairs are presented in Table 2.

*b Listening test results:* Proportion correct was calculated for all participants for all pair types (Table 3). Because this matching task forced participants to choose between two options, a signal detection analysis was performed on each participant’s scores separately, where NO responses to unmatched items is the hit rate and NO responses to matched items is the false positive rate. Separate *d*-prime scores were calculated for the target pairs (target–matched and target–unmatched) and the filler pairs (filler–matched and filler–unmatched). An ANOVA revealed that there was no effect of participant group on the *d*-prime scores for filler pairs ( $F(2,33) = 2.74$ ,  $p = .08$ , partial  $\eta = .14$ ), indicating that there

**Table 2** The target-unmatched pairs in the listening test

Word	Picture	
pette	dress	( <i>should be pete</i> )
pete	stove	( <i>should be pette</i> )
tesse	motorcycle	( <i>should be tesse</i> )
tese	watch	( <i>should be tesse</i> )
messo	piano	( <i>should be meso</i> )
meso	boot	( <i>should be messo</i> )
ketto	lemon	( <i>should be keto</i> )
keto	dresser	( <i>should be ketto</i> )

**Table 3** Proportion correct data by test pair condition and subject group (standard error in parentheses)

Subject group	Filler-matched	Filler-unmatched	Target-matched	Target-unmatched
Inexperienced learner	.91 (.02)	.95 (.02)	.82 (.02)	.49 (.03)
Experienced learner	.96 (.01)	.99 (.01)	.90 (.02)	.73 (.02)
Native Japanese	.96 (.01)	.98 (.01)	.95 (.01)	.84 (.02)

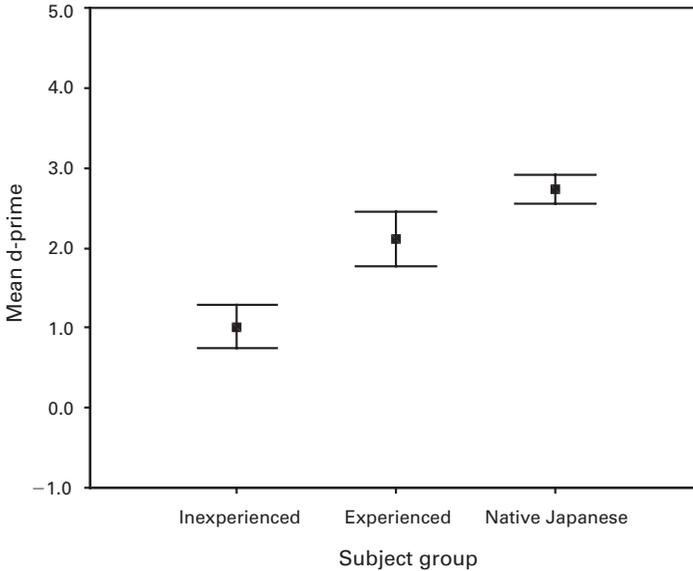
were no group differences in participants' ability to remember words in general. However, there was a significant effect of group for target pairs ( $F(2,33) = 10.25$ ;  $p < .001$ ; partial  $\eta = .38$ ). The  $d$ -prime scores for target pairs are presented in Figure 1. Planned comparisons of the groups'  $d$ -prime scores on the target pairs indicate that the inexperienced learners were less accurate than both the experienced learners ( $F(1,22) = 6.50$ ,  $p < .05$ ; partial  $\eta = .23$ ) and the native Japanese speakers ( $F(1,22) = 27.23$ ,  $p < .001$ ; partial  $\eta = .55$ ), but that performance by the experienced learners did not differ significantly from that of the native Japanese speakers ( $F(1,22) = 2.56$ ,  $p = .12$ ; partial  $\eta = .10$ ), although the relative performance of the native Japanese and experienced learner groups does follow the expected pattern.<sup>7</sup>

In order to detect that the target–unmatched items were not matched, participants must have encoded information related to the contrast (e.g. consonant length) in their memory for the words. The results of the listening task thus indicate that, on average, the experienced learners established qualitatively different lexical representations of the words they learned than did the inexperienced learners, perhaps as a result of an improved ability to encode consonant length in memory. Although as a group their average  $d$ -prime score was below that of the native Japanese speakers, the experienced learners did not differ significantly from the native speakers of Japanese on this task.

#### 4 Production test phase: naming task

Immediately following the listening test, participants took a production test.

<sup>7</sup> There was no effect of target consonant ( $F(3,102) = .82$ ,  $p = .49$ ; partial  $\eta = .02$ ) or word pair ( $F(3,99) = .81$ ,  $p = .49$ ; partial  $\eta = .02$ ).



**Figure 1** Mean d-prime scores for target test pairs by subject group; bars represent  $\pm 1$  standard error

*a Materials and procedure:* The experimenter showed participants each picture twice in random order, and participants were instructed to say the name of each picture as it was presented. Participants were given the option of saying ‘pass’ or not responding to any of the stimuli (these accounted for 2.8% of all responses), and their responses were audio-recorded.

*b Production test results:* Three independent native speakers of Japanese transcribed the participants’ responses. Where there were discrepancies among the transcriptions, the transcription provided by two of the three transcribers was used. There were very few discrepancies related to medial consonant length: most were of vowel quality or other features unrelated to the medial consonants. The transcriptions were first coded for global correctness. In this measure, a response was considered correct if it was bisyllabic and the medial consonant had the correct place and manner of articulation, ignoring at this point the consonant’s length.<sup>8</sup>

<sup>8</sup> A separate analysis with more stringent requirements for global correctness (all and only the correct phonemes were present and in the correct order) produced a similar pattern of results.

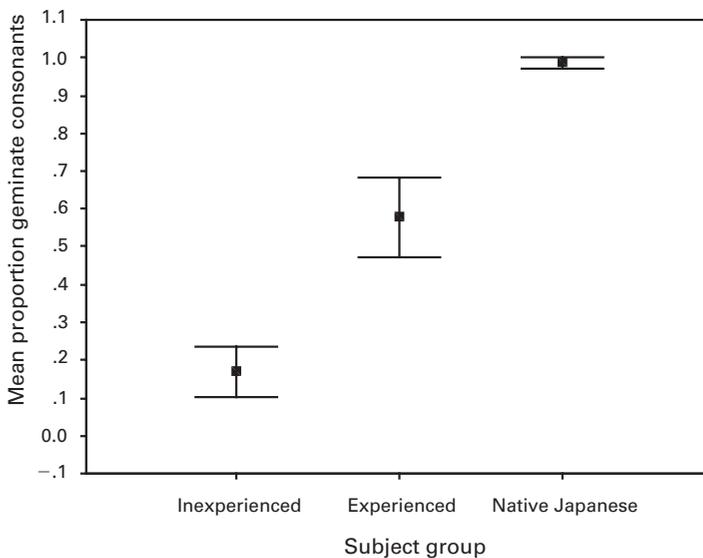
**Table 4** Proportion globally correct productions by target word condition and participant group (standard error)

Subject group	Filler	Singleton	Geminate
Inexperienced learner	0.51 (0.07)	0.52 (0.07)	0.62 (0.09)
Experienced learner	0.57 (0.09)	0.72 (0.10)	0.76 (0.09)
Native Japanese	0.75 (0.08)	0.85 (0.04)	0.92 (0.04)

The data in Table 4 was submitted to an ANOVA with target word condition as a within-participants factor and participant group as a between-participants factor. There were significant main effects of participant group ( $F(2,33) = 5.56, p < .01$ ; partial  $\eta = .25$ ) and target word condition ( $F(2,66) = 6.2, p < .005$ ; partial  $\eta = .16$ ) and the interaction of the two ( $F(4,66) < 1$ ; partial  $\eta = .01$ ) was not significant. Planned comparisons collapsing across target word condition indicate that the inexperienced learners were less accurate overall than the native Japanese speakers ( $F(1,22) = 15.72, p < .005$ ; partial  $\eta = .42$ ), but that there was no difference between the experienced learners and either the inexperienced learners ( $F(1,22) = 1.69, p = .21$ ; partial  $\eta = .07$ ) or the native Japanese speakers ( $F(1,22) = 3.52, p = .07$ ; partial  $\eta = .14$ ). That the inexperienced learners' performance was significantly less accurate on this measure than that of the native Japanese speakers was expected given that the task of remembering and producing the Japanese-sounding words should be easier in general for native speakers of Japanese than for native speakers of English who have no previous experience with Japanese-sounding words. While the comparisons between the experienced and both the inexperienced learners and the native Japanese speakers were not significant, the data follows the expected pattern: performance by the experienced learners was between that of the inexperienced learners and the native Japanese speakers.

In the next measure, which specifically considers participants' ability to distinguish singleton and geminate consonants, we included only those productions of geminate target words that were classified as globally correct above; that is, only those productions that contained the correct medial consonant in terms of place and manner of articulation. Not all participants contributed equal numbers of productions to this analysis, as participants differed in their numbers of eligible productions (participants named each of the four geminate word pictures twice, for a possible total

of eight eligible tokens). The number of eligible tokens for individual participants ranged from 2–8 for participants in the two learner groups and from 5–8 for the native Japanese speakers. These productions were coded for whether or not their medial consonants were judged to be geminates by the native Japanese listeners, and a proportion geminate consonants score (proportion of eligible productions containing geminate medial consonants) was computed for each participant. The group means are presented in Figure 2. In order to interpret the pattern of results in Figure 2, it is first necessary to determine that participants were appropriately producing singleton consonants in singleton words. An ANOVA indicated that there was no effect of group on proportion singleton consonant productions in singleton target words ( $F(2,33) = 1.50$ ,  $p = .24$ ; partial  $\eta = .08$ ). It is not surprising that participants' accuracy producing singleton consonants in singleton target words was high (inexperienced learners: 0.81, experienced learners: 0.77, native Japanese: 0.95), given that both English and Japanese have singleton consonants. On the other hand, an ANOVA revealed a significant effect of group on the proportion of geminate consonants measure ( $F(2,33) = 31.89$ ,  $p < .001$ ; partial  $\eta = .66$ ), indicating an effect of



**Figure 2** Mean proportion geminate consonants by subject group; bars represent  $\pm 1$  standard error

group on accuracy producing geminate consonants in geminate target words. Planned comparisons using the Games–Howell correction for unequal variances between groups indicate that the inexperienced learners were significantly less accurate than the native Japanese speakers ( $p < .001$ ), and that the experienced learners' performance differed from both that of the inexperienced learners ( $p < .05$ ) and the native Japanese speakers ( $p < .01$ ).<sup>9</sup>

The results of the production test indicate that the experienced learners have an improved ability relative to inexperienced learners to lexically encode and to produce geminate consonants accurately, although their performance does not yet reach native Japanese levels.<sup>10</sup>

### III Discussion

In this study we set out to investigate the development of L2 learners' ability to establish lexical representations that encode a novel phonemic contrast via learners' performance on both a listening and a production task. Because English does not have singleton–geminate consonant contrasts, we might have expected that the inexperienced Japanese learners would not establish contrastive lexical representations for the singleton–geminate minimal pairs that they learned. However, their performance on the listening task suggests that they were able to lexically contrast singleton and geminate words to some extent, which leads to the question of how they were able to do this. A follow-up auditory discrimination experiment confirmed that monolingual English listeners are in fact able to discriminate the minimal pair stimuli used in this study. In this follow-up experiment, two monolingual English speakers (neither of whom participated in the earlier study) were asked to determine whether pairs of the test stimuli (e.g. *meso* and *messo*) were the 'same' or 'different'. They discriminated the minimal pairs with 93% accuracy, indicating that the minimal pairs were discriminable even to the inexperienced learners.<sup>11</sup> Additionally, the set of words learned in the present study

<sup>9</sup> There was no effect of target consonant ( $F(1,30) = .21, p = .65$ ; partial  $\eta = .01$ ) or target word ( $F(3,51) = .89, p = .45$ , partial  $\eta = .05$ ).

<sup>10</sup> It should be noted that relatively more accurate performance by learners on the listening test relative to the production test may in part be due to the order of the tests (the listening test preceded the production test; the greater availability of episodic memory in the listening test may have contributed to more accurate performance).

contained minimal pairs, which can provide even totally inexperienced learners with information they can use to infer the importance of a novel L2 contrast over the course of a single experiment session (Curtin *et al.*, 1998; Pater, 2003; Hayes-Harb, 2006). Therefore, it seems that the inexperienced learners were able to both detect the difference between singleton and geminate consonants in the auditory input and encode pairs of newly-learned words differing only in consonant length contrastively in memory, although their performance at test was significantly less accurate than that of both the experienced learners and the native Japanese speakers. However, they were unable to accurately implement the singleton–geminate contrast in their productions, suggesting that their production abilities neutralized some contrasts that they had encoded lexically.

As expected, the native speakers of Japanese exhibited high accuracy rates on both the listening and production tasks: they distinguished singleton and geminate consonants in their lexical representations of the words they learned during the experiment, and they were able to phonetically implement the contrasts stored in their lexical representations at test time.

Performance by the experienced learners differed significantly from performance by the inexperienced learners on both tests, indicating that experience with an L2 can mediate the influence of the native language on the phonological structure of the lexicon, and that one year of Japanese experience is sufficient for significant development to occur. In fact, the experienced learners did not differ from native Japanese speakers in the listening task: their one year of Japanese language experience was sufficient to make them indistinguishable from native speakers of Japanese on this particular task. However, the experienced learners' performance was significantly less accurate than that of the native Japanese speakers on the production task. In this study, target-like listening performance preceded target-like production in both the inexperienced and experienced learners.

The issue of whether performance by the inexperienced and experienced learners on these tasks unambiguously reflects the phonological

<sup>11</sup> Bohn (1995) discusses evidence that listeners tend to rely more on durational than spectral cues in cross-language perception tasks, suggesting that temporal information (as in singleton–geminate contrasts) may be particularly salient for L2 learners.

content of their lexical representations – or whether it may instead (or also) reflect their ability to phonetically implement singleton–geminate consonant contrasts in listening and production tasks – cannot be definitively resolved here. However, there are reasons to suspect that patterns of performance by the inexperienced and experienced learners in the listening task can be interpreted as reflecting participants’ lexical representations. First, as discussed earlier, monolingual English speakers’ performance on an auditory discrimination task indicated that they are indeed able to perceive the difference between the singleton–geminate minimal pair stimuli used in this study. If participants perceive the singleton–geminate distinction during the word learning phase and thus establish lexical representations that are contrastive, but their on-line perception ability during the listening task prevents them from perceiving the contrast accurately, then we might expect an asymmetry in their responses analogous to that found in Weber and Cutler (2004) and Cutler *et al.* (2006). The closest English counterpart to Japanese /t/ and /tt/ is /t/, which is more similar acoustically to Japanese /t/ than /tt/ (see Hayes-Harb, 2005). If it is the case that they are not able to discriminate between Japanese /t/ and /tt/ in the listening test, we might expect native speakers of English to perceive both as the English phoneme /t/, which should mean that they search their lexicons for only words containing /t/ regardless of which of the two consonants is present in the auditory input. For example, even if they establish contrastive lexical representations for singleton and geminate consonants, native speakers of English should perceive both *keto* and *ketto* as the lexical entry /keto/ (which they learned was the name for dresser), and respond ‘matched’ when either auditory stimulus is presented with the picture of the dresser, but they should respond ‘not matched’ when either word is presented with the picture of the lemon (which they learned was called *ketto*) simply because they do not perceive any input as /ketto/ in this on-line task. That is, if participants established contrastive lexical representations of the words, but their online perception at test had the effect of making them perceive only singleton consonants in the auditory input, we would expect a larger proportion of ‘matched’ responses when geminate auditory stimuli were paired with singleton pictures than when singleton auditory stimuli were paired with geminate pictures. However, this is not the case for either the inexperienced or the experienced learner groups

(see data presented in Table 5), as there was little difference between mean proportion ‘matched’ responses to geminate picture–singleton word and singleton picture–geminate word pairs for either learner group. On the other hand, if native English-speaking participants neutralized singleton and geminate word pairs to homophones containing only singleton consonants in the lexicon and perceived each auditory input as containing a singleton consonant (as might be expected of the inexperienced learners in particular), they should have responded ‘matched’ to every target pair. However, we did not find this, either. Although the availability of two response options (‘matched’ and ‘not matched’) in the task was likely to elicit at least some ‘not matched’ responses – unless participants detected the difference between target–matched and target–unmatched pairs – proportion ‘matched’ responses should be equal in all boxes (indicating random guessing), which we also did not find. Instead, we found weak detectability of the difference between target–matched and target–unmatched pairs for most of the inexperienced learners, indicating that even these participants are able to establish lexical representations that contrast singleton and geminate consonants, and significantly better detectability overall among the experienced learners, indicating that one year of L2 study significantly improves learners’ ability to establish a novel L2 contrast lexically.

We now turn to the question of how novel L2 contrasts are lexically encoded by learners. Under the traditional generative phonology assumption that only contrastive information is stored in lexical representations and, assuming that L2 learners initially transfer their first language (L1) phoneme inventory to L2 acquisition, learners should have relatively little difficulty lexically encoding contrasts that are

**Table 5** Mean proportion ‘matched’ responses by subject group to target test pairs in the listening task (by whether the picture and the auditory word contained singleton or geminate consonants)

	Inexperienced learner		Experienced learner		Native Japanese	
	Singleton picture	Geminate picture	Singleton picture	Geminate picture	Singleton picture	Geminate picture
Singleton word	0.82	0.52	0.89	0.28	0.94	0.15
Geminate word	0.51	0.82	0.27	0.92	0.17	0.95

common to both the L1 and the L2, and relatively more difficulty with novel contrasts. This is supported by participants' high accuracy rates on the set of filler words in the present experiment, which contained only phonemes common to English and Japanese, but their relatively less accurate performance on the set of target words, which contained novel consonant length contrasts.

The relative difficulty of lexically encoding common vs. novel contrasts has been characterized using the framework of underspecification theory (e.g. Curtin *et al.*, 1998). An assumption of underspecification theory is that phonemes are specified only by the minimal set of features necessary to distinguish them from the rest of the language's phoneme inventory. For example, English consonants must be specified for the feature [+/-voice] (voicing is contrastive in English), but not for the feature [+/-long] (consonant length is not contrastive in English). Japanese consonants, on the other hand, must be specified for both [+/-voice] and [+/-long].<sup>12</sup> Underspecification theory has been demonstrated to predict learners' difficulty perceiving and producing novel L2 contrasts (Brown, 1998; Larson-Hall, 2004), and Curtin *et al.* (1998) found that monolingual English speakers were better able to lexically encode novel Thai contrasts differing in the feature [+/-voice] (e.g. /d-t/) than contrasts differing in the feature [+/-aspirated] (e.g. /t<sup>h</sup>-t/; a feature that is not contrastive in English).<sup>13</sup> If L2 learners are initially able to lexically encode only the features specified in their native language, then learning novel L2 contrasts should involve learning to lexically encode novel features. In the case of native speakers of English learning Japanese consonant length contrasts, learners must encode the novel feature [+/-long] in order to develop native-like lexical representations of Japanese words.

What would constitute evidence that participants correctly specified the feature [+/-long] in their lexical representations? Accurate performance on the listening test may not provide sufficient evidence alone. Assuming that initially, both Japanese /t/ and /tt/ are neutralized

<sup>12</sup> The feature [+/-long] is used here to capture the singleton-geminate contrast for the purpose of illustration. However, it should be pointed out that singleton-geminate contrasts are more commonly characterized in terms of differences in number of timing slots, not distinctive features (see, e.g. Lahiri and Hankamer, 1988).

<sup>13</sup> However, Pater (2003) was not able to replicate the Curtin *et al.* (1998) finding.

by native English speakers to English /t/ in their lexical representations; when they detect the difference between singleton words and geminate words, they may simply include annotative information in their lexical representations that indicates when an L2 word contains a new sound that the learner suspects is phonologically relevant (a similar suggestion is mentioned in Cutler *et al.*, 2006). For example, learners might lexically encode a geminate /tt/ consonant as /t\*/, where the '\*' might mean 'sounds different from /t/', even if they have not yet determined specifically that /t/ and /t\*/ differ specifically with respect to the feature [+/-long]. That the learners in our study in general performed better on the listening task than on the production task is consistent with this analysis: in order to perform accurately on target pairs in the listening test, learners had only to notice whether or not the auditory input or the lexical representation contained '\*'. On the other hand, performing accurately on target words in the production task required participants to have some idea about what '\*' means (e.g. to have correctly associated the feature [+long] with geminate consonants). Post-experiment feedback from the inexperienced and experienced learners provides some information about what they believed the differences between the singleton and geminate words were: some believed they were hearing vowel length contrasts, some believed the contrasts were of stress placement, and many of the experienced learners correctly identified consonant duration – some even mentioning that geminate consonants are indicated in Japanese syllabary writing with a small 'tsu' character – as the difference between the minimal pairs. Although a detailed analysis of the acoustic properties of participants' productions in the production task is beyond the scope of the present article, analyses of consonant duration, preceding and following vowel duration, and voice onset time of the production data indicate that the only cue that native English speakers in this study manipulated systematically to differentiate singleton and geminate consonants was consonant duration, and that this differentiation was highly variable both between and within learners.

Given that both the inexperienced and experienced learners in this study were able to perceive the difference between singleton and geminate consonants, one might wonder why their performance on these lexical tests, or at least the listening test, was not more accurate. Studies

of first language acquisition have provided evidence that the task of word learning may put demands on processing that obscure knowledge of newly-learned and yet-unstable contrasts (Stager and Werker, 1997; Pater *et al.*, 2004). Pater *et al.* (2004) argue that word learning incurs a greater processing load than word recognition, and that this can explain young children's ability to discriminate particular phonological contrasts in tasks that involve well-known words (Fennell and Werker, 2003), but not in tasks that involve learning new words (Stager and Werker, 1997). It is possible that adults' acquisition of L2 words – when the words contain sounds of a newly-learned contrast – incurs processing costs that make it difficult to create and/or access stable lexical representation of the new contrast. We do not know whether the experienced learners in the present study have established contrastive lexical representations of singleton and geminate consonants for familiar Japanese words: future research might examine learners' lexical representations of newly-learned vs. old words to determine whether there are differences in the phonological content of their lexical representations related to word familiarity. Additionally, an issue of particular importance in the study of the L2 lexicon is whether studies where participants are taught non-words and are subsequently tested on their knowledge of the non-words can actually provide information about learners' lexical–phonological structure. The question is whether newly-learned non-words behave like more established (i.e. actual) lexical items or whether they behave like episodic memories. While there is evidence from studies of native language non-word learning that newly-learned non-words behave differently from established lexical items in some tasks (e.g. Dumay and Gaskell, 2007), other research highlights similarities between participants' performance on newly-learned non-words and established lexical items (e.g. Shatzman and McQueen, 2006). More research is needed in the area of the time course and nature of lexicalization in second language acquisition.

As indicated by the Weber and Cutler (2004) and Cutler *et al.* (2006) studies, it is possible for learners to know that two words are contrastive even when they are not able to reliably reflect the contrast in their on-line perception and production of the words. If participants in the present study had been taught the words' spellings (in one of the Japanese syllabaries, which reliably indicate geminate consonants), they may

have been able to use the presence of this additional visual-phonological information as a cue to the consonant length contrasts. In fact, it is possible that the experienced learners' knowledge of Japanese orthography may have contributed to their relatively more accurate ability to encode the singleton–geminate contrast in their memory for the non-words in the present study. Although it is not clear exactly how knowledge of the contrastive spellings of words would translate into knowledge of the contrastive phonological content of words beyond simply drawing participants' attention to the contrast in their auditory input, recent research provides evidence that orthographic and phonological information in L2 learners' lexical representations can interact in learners' developing L2 lexicons (Komaki and Akahane-Yamada, 2004; Ota *et al.*, submitted). More research is needed to determine the role of orthographic knowledge in L2 phonological development.

Another issue that deserves further consideration is whether or not second language learners ultimately can establish native-like lexical representations with enough L2 experience. In a pilot study, we collected data from learners of Japanese with even more experience (2 or more years of Japanese experience), but found no differences between their performance and that of the experienced learners in the present study. However, the Weber and Cutler (2004) and Cutler *et al.* (2006) studies indicate that highly-fluent learners can establish lexical representations that appear to be native-like. Future research might then consider how much and what kind of exposure are necessary to attain this level of competence in a L2.

#### **IV Conclusions**

We have provided data suggesting that with increased exposure to a second language, learners can become better able to establish lexical representations of newly-learned L2 words that are differentiated by a novel contrast. We have also demonstrated that while learners' performance on the listening task provides evidence that learners have stored distinct lexical representations for single vs. geminate consonant words, performance on the production task indicates that learners may not have accurately encoded the information (e.g. the feature [+/-long]) that distinguishes singleton and geminate consonants. We attempt to account

for this difference in listening and production performance by proposing that learners may initially encode a novel L2 phoneme in the lexicon as an unfamiliar version of its closest native language counterpart (e.g. native speakers of English might encode the Japanese phoneme /tt/ as /t\*/). Thus when learners encounter the novel phoneme in the auditory input, they need only to know that it is unfamiliar to differentiate it from its native language counterpart, resulting in accurate listening performance, even if they do not yet know precisely how the new phoneme is implemented phonetically in production. However, there is evidence that many of the experienced learners in this study were able to accurately encode the feature [+/-long] in their lexical representations to accurately differentiate singleton and geminate consonants in both listening and production tests, indicating that adults can learn to encode a novel feature lexically with enough L2 experience.

Second language teachers often observe that learners can sometimes accurately perceive and produce novel sounds in controlled tasks in the classroom, but revert to foreign-accented speech in less controlled tasks, such as casual conversation. It is possible that in some classroom tasks (e.g. reading out loud), students are not required to access their lexical representations and may instead employ their 'raw' perception and production abilities. However, in casual conversation, students must access their lexical representations for words and, if those representations do not accurately encode the second language's phonological contrasts, their speech will be 'accented'. Activities that focus learners' attention on the role of novel contrasts in distinguishing between L2 words (e.g. activities involving the differentiation of minimal pairs) may help students establish contrastive lexical representations in addition to helping them develop their perception and production abilities.

### *Acknowledgements*

The authors thank audiences at the 149th Meeting of the Acoustical Society of America and the 4th International Conference on Practical Linguistics of Japanese, Suzanne Curtin, Mara Haslam, Kazutoshi Ohno, the research participants and the anonymous *Second Language Research* reviewers for their contributions to this research and manuscript.

## V References

- Aoyama, K., Flege, J.E., Guion, S.G., Akahane-Yamada, R. and Yamada, T.** 2004: Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32, 233–50.
- Best, C.T.** 1995: A direct realist view of cross-language speech perception. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press.
- Best, C.T., McRoberts, G.W. and Goodell, E.** 2001: Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners native phonological system. *Journal of the Acoustical Society of America* 109, 775–94.
- Boersma, P., Escudero, P. and Hayes, R.** 2003: Learning abstract phonological from auditory phonetic categories: an integrated model of the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*.
- Bohn, O.-S.** 1995: Cross-language speech perception in adults: first language transfer doesn't tell it all. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press.
- Bohn, O.-S. and Flege, J.E.** 1992: The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14, 131–58.
- Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R. and Tohkura, Y.** 1997: Training Japanese listeners to identify English /r/ and /l/; volume IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101, 2299–2310.
- Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B. and Tohkura, Y.** 1999: Training Japanese listeners to identify English /r/ and /l/: Long term retention of learning perception and production. *Perception and Psychophysics* 61, 977–85.
- Brown, C.A.** 1998: The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research* 14, 136–93.
- Chomsky, N. and Halle, M.** 1968: *The sound pattern of English*. New York: Harper and Row.
- Cohn, A.C., Ham, W.H. and Podesva, R.J.** 1999: The phonetic realization of singleton–geminate contrasts in three languages of Indonesia. *Proceedings of the 14th International Congress of Phonetic Sciences*.
- Curtin, S., Goad, H. and Pater, J.V.** 1998: Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research* 14, 389–405.

- Cutler, A., Weber, A. and Otake, T.** 2006: Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34, 269–84.
- Cutler, A. and Otake, T.** 2004: Pseudo-homophony in non-native listening. Unpublished paper presented at the Acoustical Society of America, New York.
- Dumay, N. and Gaskell, M.G.** 2007: Sleep-associated changes in the mental representation of spoken words. *Psychological Science* 18, 35–39.
- Eckman, F.R., Elreyes, A. and Iverson, G.K.** 2003: Some principles of second language phonology. *Second Language Research* 19, 169–208.
- Escudero, P. and Boersma, P.** 2004: Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26, 551–85.
- Fennell, C. and Werker, J.** 2003: Early word learners' ability to access phonetic detail in well-known words. *Language and Speech* 46, 245–64.
- Flege, J.E.** 1995. Second language speech learning: theory, findings, and problems. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press.
- Flege, J., Munro, M. and MacKay, I.** 1995: Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16, 1–26.
- Forster, K.I. and Forster, J.C.** 2003: DMDX: a Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers* 35, 116–24.
- Goto, H.** 1971: Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. *Neuropsychologia* 9, 317–23.
- Han, M.S.** 1992: The timing control of geminate and single stop consonants in Japanese: a challenge for nonnative speakers. *Phonetica* 49, 102–27.
- 1994: Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96, 73–81.
- Hankamer, J., Lahiri, A. and Koreman, J.** 1989: Perception of consonant length: voiceless stops in Turkish and Bengali. *Journal of Phonetics* 17, 283–298.
- Hansen, B.B.** 2003: Production of Persian geminate stops: Speaking rate effects. Unpublished MS, Department of Linguistics, University of Texas at Austin.
- Hayes, R.** 2002: The perception of novel phoneme contrasts in a second language: a developmental study of native speakers of English learning Japanese singleton and geminate consonant contrasts. *Coyote Papers* 12, 28–41.
- Hayes-Harb, R.** 2005: Optimal L2 speech perception: native speakers of English and Japanese consonant length contrasts. *Journal of Language and Linguistics* 4, 1–29.

- 2006: Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research* 23, 1–31.
- Komaki, R.** and **Akahane-Yamada, R.** 2004: Japanese speakers' confusion of phonemically contrasting English words: a link between phoneme perception and lexical processing. *Proceedings of the 18th International Congress on Acoustics*.
- Kuhl, P.K.** and **Iverson, P.** 1995: Linguistic experience and the 'perceptual magnet effect'. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press.
- Lahiri, A.** and **Hankamer, J.** 1988: The timing of geminate consonants. *Journal of Phonetics* 16, 327–38.
- Larson-Hall, J.** 2004: Predicting perceptual success with segments: a test of Japanese speakers of Russian. *Second Language Research* 20, 33–76.
- Logan, J.D., Lively, S.E.** and **Pisoni, D.B.** 1991: Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America* 89, 874–76.
- Mah, J.** and **Archibald, J.** 2003: Acquisition of L2 length contrasts. Liceras, J.M., Zobl, H. and Goodluck, H., editors, *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*. Somerville, MA: Cascadilla.
- Major, R.C.** and **Kim, E.** 1999: The similarity differential rate hypothesis. *Language Learning* 49, Supplement 1, 151–83.
- Ota, M., Hartsuiker, R.J.** and **Haywood, S.** submitted: When a FAN is FUN: near-homophone effects in bilingual visual word recognition.
- Pallier, C., Colomé, A.** and **Sebastián-Gallés, N.** 2001: The influence of native-language phonology on lexical access: exemplar-based vs. abstract lexical entries. *Psychological Science* 12, 445–49.
- Pater, J.** 2003: The perceptual acquisition of Thai phonology by English speakers: task and stimulus effects. *Second Language Research* 19, 209–23.
- Pater, J., Stager, C.** and **Werker, J.** 2004: The perceptual acquisition of phonological contrasts. *Language* 80, 384–402.
- Pickett, E.R., Blumstein, S.E.** and **Burton, M.W.** 1999: Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica* 56, 135–57.
- Pind, J.** 1995: Speaking rate, voice-onset time, and quantity: the search for higher-order invariants for two Icelandic speech cues. *Perception and Psychophysics* 57, 291–304.
- Pisoni, D.B.** and **Lively, S.E.** 1995: Variability and invariance in speech perception: a new look at some old problems in perceptual learning. In Strange, W., editor, *Speech perception and linguistic experience*. Timonium, MD: York Press. 433–59.
- Shatzman, K.B.** and **McQueen, J.M.** 2006: Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science* 17, 372–77.

- Stager, C.L. and Werker, J.F.** 1997: Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388, 381–82.
- Takagi, N.** 2002: The limits of training Japanese listeners to identify English /r/ and /l/: eight case studies. *Journal of the Acoustical Society of America* 111, 2887–96.
- Takagi, N. and Mann, V.** 1995: The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics* 16, 379–405.
- Toda, T.** 1997: Strategies for producing mora timing by non-native speakers of Japanese. *Acquisition of Japanese as a Second Language* 1, 157–94.
- Weber, A. and Cutler, A.** 2004: Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50, 1–25.

**Appendix 1** Acoustic properties of nonword stimuli

Nonword stimulus	Consonant	Preceding vowel	VOT	Following vowel
<i>Singleton consonants:</i>				
meso	86.0	100.6	n/a	111.5
tese	90.3	62.7	n/a	72.5
pete	61.4	64.5	15.0	104.1
ket <sub>o</sub>	60.9	49.5	13.5	77.3
<i>Geminate consonants:</i>				
messo	185.4	119.5	n/a	110.9
tesse	185.0	67.1	n/a	109.2
pette	181.4	77.2	9.3	156.3
ketto	177.0	64.9	17.7	95.0

*Notes:* All values are consonant durations in milliseconds. The consonant duration ratio (geminate to singleton) is 2.2, 2.0, 3.0 and 2.9 respectively.