# MEASURING KNOWLEDGE OF ENGLISH ORTHOTACTICS IN JAPANESE LEARNERS OF ENGLISH: TOWARDS THE ESTABLISHMENT OF A TRAINING SCHEME FOR /l/-/r/ PERCEPTION

YOKO UCHIDA

## ABSTRACT

Existing regimes on training Japanese learners of English to acquire the nonnative contrast /l/ and /r/ in English focus on having the learners attend to differences in acoustics cues. Such training does show a certain amount of improvement in /l/-/r/ discrimination among Japanese learners, yet the level of achievement is not equivalent to the performance of native speakers. In light of the fact that phonological contrasts are based not only on acoustic information and that other dimensions such as phonotactics also plays a significant role, the current study explores the possibility of setting up a training scheme that supplements Japanese learners of English with an understanding of the probabilistic phonotactics of /l/ and /r/ for better discrimination of the two categories. This training would be meaningful only if it is made clear that Japanese learners of English do not possess such phonotactic knowledge while native English speakers do. As a first step, native speakers of English were tested on whether they are able to make a category decision based solely on probabilistic orthotactic information, to simulate a situation where the category is ambiguous between /l/ and /r/ in terms of acoustic cues but can be clearly defined by phonotactic constraints in English. It was confirmed that native English speakers do have knowledge on orthotactics and were overall successful in making a correct category judgment. However, although it had been predicted that Japanese native speakers lack the orthotactic knowledge of L and R and would fail in the category judgment, the Japanese speakers tested in this study performed just as well as the native speakers of English. However, having orthotactic knowledge only suggests a possibility of having phonotactic sensitivity indirectly, since processing of a written word and speech sound is not the same. Emphasis is often placed on reading and writing in the English classroom in Japan, and many students find processing spoken language more demanding. Besides, English is notorious for its unsystematic writing system and the letter-to-phoneme correspondence is very inconsistent. Further studies that directly measure the phonotactic knowledge of Japanese learners of English are necessary.

## INTRODUCTION

A large number of studies have indicated that those who begin acquiring a language later in life exhibit phonological deficits in perception, production, and processing (for a general review, see Strange, 1995). Although any infant is born with an ability to perceive any phonetic

*Second Language Studies, 29*(1), Fall 2010, pp. 97-129.

contrasts in natural languages, this ability rapidly decays around 6-12 months after birth (Werker & Tees, 1984; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), when the infant's ear comes to be tuned to one specific language, which will become his/her first language (L1; Kuhl, 2004). Language experience in L1 has such a profound effect that exposure to another language (i.e., second language, L2) after that age often leaves a trace of "foreign accent" in production, and difficulty in perception of phonetic contrasts that do not exist in the L1 persists.

Among studies on acquisition of L2 phonetic contrasts, Japanese speakers' problem in the perception of English /l/ and /r/ is perhaps one of the best documented cases. Ever since the problem was addressed in classic studies such as Goto (1971) and Miyawaki et al. (1975), the /l/-/r/ problem has been studied from various perspectives. Tsushima et al. (1994) reported that Japanese infants who are able to distinguish the contrast like any other infants in the world lose their ability by the age of 12 months as the influence of language-specific perception creeps in. Thus, English /l/ and /r/ come to be perceptually related to a Japanese category /r/, which is described as apicoalveolar tap (Vance, 2008). The difficulty in production may be overcome, but exposure to English for a long period of time even in an English speaking country is not expected to lead to improvement in terms of perception (Flege, Takag, & Mann, 1995; Takagi & Mann, 1995).

In the framework of speech perception models such as Best's Perceptual Assimilation Model (PAM; Best, 1995; Best, McRoberts, & Goodell, 2001), this perception of two L2 categories as one category in L1 is explained as "single category assimilation," which is predicted to be the most difficult L2 contrast to discern, and it is hypothesized that English /l/ and /r/ are both perceived as poor exemplars of Japanese /r/ (in some cases Japanese /w/; Best & Strange, 1992). Another model, Flege's Speech Perception Model (SLM; Flege, 1995), posits that the more distance there is between an L1 sound and a corresponding L2 sound, the easier it is for the L2 learner to master the contrast. Experimental research based on these models consistently obtained results that Japanese listeners judge English /l/ to be perceptually more similar to Japanese /r/ although great individual differences and token-to-token differences coexist (Uchida, 2001; Takagi, 2002; Hattori & Iverson, 2008). Aoyama et al. (2004), referring to studies that obtained a higher identification rate with /r/ (Flege, Takagi, & Mann, 1996; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997) pointed out that this is in accordance with the prediction of SLM. They further confirmed the validity of the hypothesis by conducting perception and production tests on the /l/-/r/ contrast with native Japanese speaking children twice, with a one-year interval, finding that greater improvement in production of /r/ than /l/ was shown in the second test.

Along with these and other theoretical implications was a question interesting both in theoretical and practical viewpoints: is it possible to modify the perception of the /l/-/r/ contrast by native speakers of Japanese through training? Attempts have been made to devise an optimal way to have the Japanese learners of English master the contrast. One of the best-known training studies was conducted by Lively and his collaborators (Logan, Lively, & Pisoni, 1991; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada & Tohkura, 1994), in which stimuli were all recorded natural tokens taken from multiple talkers (both male and female), and in five phonetic contexts, including those such as "l/r + V" (e.g. *lock-rock*), "C + l/r + V" (e.g. *glass-grass*), and "V + l/r + V". They gave the stimuli this variability in order to have learners master the distinction of the two as a category by providing examples of /l/ and /r/ that contain a full range of acoustic and phonetic cues. The Japanese participants were given two-alternative forced-choice identification tasks for training: Following a presentation of possible response alternatives contrasting in /l/ and /r/ on the computer screen, such as "light-right," the participants heard one of the items from the pair through headphones, and they responded by pressing a button on a response box. If the response was correct, the next stimulus was presented; and if incorrect, feedback was given before the next stimulus was presented.

The researchers reported that there was a significant effect in training, by which all the participants achieved a significantly better discrimination rate on the posttest than on the pretest. Moreover, the learning generalized to new tokens, i.e., to a minimal pair not presented during the training and with a different speaker. The effect of training was even retained for six months.

Regardless of the success in generalization claimed by Lively and his team, however, it must be pointed out that the learning achieved in this scheme is not comparable to the level of native English speakers. With the minimal-pair presentation method, the generalization was limited to a lower level of processing and did not take place at the level of conversation (Akahane-Yamda et al., 2004).

In addition, Takagi (2002) demonstrated native Japanese speakers' difficulty in native-level attainment of the /l/-/r/ contrast even with the identification task at a minimal-pair level. In his study focusing on individual performances, individual differences were observed among his eight Japanese participants in their baseline sensitivity at the time of pretest, which never changed even after long-term training for 15 days. Furthermore, he found every listener had a specific token or two with which misidentification persisted at a level below chance even after the training. Based on a signal detection theory hypothesis (Macmillan & Creelman, 1991) that each Japanese listener individually possesses a "dimension R-L" and that the category of each token presented is judged based on this individual dimension, Takagi analyzed additional data from each participant on the confidence-level in their category judgment for each token. Combined with the interview conducted with each participant, he observed that each listener used Japanese /r/ as a "perceptual anchor," i.e., the more similar the token presented was to Japanese /r/, the more they selected the English category /l/, and concluded that even with extensive training containing stimuli with a wide variability in acoustic information, it would be extremely difficult to have Japanese learners of English internalize a nativelike contrast of /l/-/r/.

All previous research thus clearly indicates that depending solely on the minimal-pair training does not have the effect of training the majority of Japanese speakers in the /l/-/r/ distinction to the level of native listeners. It is indeed no easy job to acquire a novel nonnative contrast when there is a large variability in the way one phoneme is realized in an utterance. With different vocal tract size and shape as well as individual differences in manners of articulation, rate of speech, and the context where a phoneme appears, all of these factors contribute to a single phonetic unit that is realized in various ways, sometimes with overlaps among categories in acoustic information (Hillenbrand et al., 1995; Miller & Liberman, 1979; Catford, 2002). This fact leads to a question about how humans acquire the internalized mental categories with their first language.

There is now sufficient evidence to believe that an infant born with a set of initial perceptual abilities makes use of pattern detection and computational abilities also innate to humans to sort out phonetic units and group them into phonemic categories based on distribution frequencies of the sounds. In contrast to the "minimal pair-based" learning, this is called "distribution-based" or "statistical" learning, and it is through learning of distributional patterns of sounds that the notion of phonemic contrast develops (Maye & Garken, 2000; Kuhl, 2004; Werker & Tees, 1984; Saffran, Aslin, & Newport, 1996).

Recently, a number of studies have provided evidence in favor of a role for distributional frequencies in the input (Maye & Garken, 2001; Maye, Werker, & Gerken, 2002). Extensive training studies in the framework of this statistical learning have been conducted, in which stimuli that reflect the phonetic patterns and regularities of the language are used as input. The subjects of these studies range from infants to adults, and the results indicate a range of effectiveness, some showing learning of only the contrast used during the training and others generalized to analogous contrast (Maye, Weiss, & Aslin, 2008; Shea & Curtain, 2005; Hayes-Harb, 2007).

One possible advantage of training with statistical learning approach is that the learning is not limited to the acquisition of the new phonetic contrast but is also likely to contribute a positive effect on other dimensions of the language learning. Empirical studies have demonstrated that phonetic training and statistical segmentation prompts word learning in adults (Mirman, Magnuson, Estes, & Dixon, 2008; Perfors & Dumbar, 2010). Based on a bimodal tendency in the distribution of consonant cluster sequences in English, Christiansen, Onnis, and Hockema (2009) suggest that knowledge of phonemic distributions may be useful in speech segmentation and lexical category identification.

Every language has phonotactic patterns of its own (Vance, 2008; Catford, 2002), and it is well established that native speakers possess knowledge of distributional properties of sounds in words. Any native speaker of Japanese implicitly knows that only (C)V syllable structure is permissible in Japanese, while native speakers of English know that a sequence /tr-/ occurs in English, but not /tl-/. Considering the fact that language learning involves processing of multiple dimensions at the same time (Holt & Lotto, 2006), mastering the probabilistic phonotactics in English would be beneficial for Japanese learners of English for improving their sensitivity in /l/-/r/ categorization.

My ultimate goal is to establish a training scheme that takes probabilistic phonotactics into account. To this end, there are two premises, both of which must be confirmed to be true: one is that native speakers of English possess sound knowledge of probabilistic phonotactics of /l/ and /r/; the other is that Japanese learners of English, in contrast, do not possess this knowledge, or at most, their knowledge is not as sound as that of native speakers. If this is the case, there will be some room to train them in this aspect of language. In this paper, I conducted a set of experiments designed to clarify these two points.

## STUDY 1: ORTHOTACTIC KNOWLEDGE OF ENGLISH NATIVE SPEAKERS

As we have seen, auditory categories involve integration of multidimensional information sources, and listeners integrate these cues at different levels in making a phonetic category decision. Following this assumption, I would first like to test if probabilistic phonotactics can serve as a useful cue for category decision of /l/ or /r/ when the acoustic cues are ambiguous between the two categories.

In Study 1 (Letter Guessing Game), native speakers of American English were presented with pseudowords with one-letter missing (e.g. st*aud) on a computer screen, and were asked to guess whether the missing letter is L or R. With all the words, the possibility of occurrence of L or R was strictly controlled by having the missing letter surrounded by a sequence of sounds in such a way that the missing letter would fall into one of the following three conditions: (1) where L is most likely to appear; (2) where R is most likely to appear; and (3) where both L and R are possible with the same probability. The probability estimates were computationally assessed by Dr. Onnis based on a corpus of English. We expected that native speakers would predict L or R as a function of their phonotactic predictability. Thus, if differences in the pattern of response in the three conditions are found, this would suggest that probabilistic phonotactics contribute to the category perception of /l/ and /r/.

In the course of preparation, attempts had been made to create acoustic synthetic pseudowords in which the /l/ or /r/ part of the stimuli was masked with white noise. However, due to the poor quality of the resulting synthetic sounds and difficulty in creating synthetic sounds that do not leave an acoustic trace of /l/ or /r/ in the surrounding sounds, the current experiment took the format of asking *ortho*tactic knowledge instead of phonotactic knowledge. Pseudowords were used rather than real words in order to exclude response bias due to lexical familiarity (Ganong, 1980; Flege, Takagi, & Mann, 1996).

### Method

#### *Participants*

The 33 participants (11 males and 22 females) recruited for the Letter Guessing Game were native speakers of American English. They were all volunteers from the student population of the Psychology Department at the University of Hawai‘i at Manoa and received course credits for compensation.

### *Stimuli*

The English Lexicon Project (ELP) database (Balota et al., 2007), which contains 40,481 words and 40,481 nonwords in English with information on lexical characteristics, including phonetic transcriptions, as well as phonological and orthographic neighborhood, was employed to calculate the probability of orthographic sequences which most typically precede and follow the letters L and R, respectively. For example, the conditional probability that L occurs following a word initial sequence ST- is expected to be very low or zero, P( L | ST- ) = 0 (e.g. **stleet*, **stlange*). Conversely, P( R | ST- ) = 1 (e.g. *street*, *strange*). The distribution of probabilities of a letter sequence containing L and R thus obtained reflected the degree of ambiguity to signal the presence of either an L or R.

By writing ad-hoc Perl scripts and running them on the ELP database, possible combinations of L (or R) with two letters preceding and following it in the ELP database were exhausted, and orthotactic frames where only L appear in the database (hereafter called the “L-dominant condition”), where only R appear (“R-dominant condition”), as well as frames where both L and R appear at a same rate (“L-R possible condition”), were searched. Sequences such as *velop* (L-dominant), *strat* (R-dominant), and *ealed/eared* (L-R possible) were yielded. In searching the sequences, the word boundary (i.e., beginning and end of word) was also considered an element of the frame. Thus, sequences such as *ible_* (L-dominant) and *_tran* (R-dominant) were also yielded, with the former having an E and a word boundary at the end, and the latter preceded by a word boundary then a letter T.

Based on the results, 40 to 50 orthographic pseudowords were created by the author for each of the three conditions. After that, in order to avoid any resemblance to actual words in English, two native English speakers went through the list and selected 20 best exemplars for each condition. A total of 60 pseudowords thus created, 20 for each condition, are shown on Table 1.

Table 1

*Pseudowords for the Letter Guessing Game* (Note: An asterisk denotes the missing L/R. An underscore is a symbol that indicates a word boundary.)

| **Condition:** | **L-dominant** | | **R-dominant** | | **L-R possible** | |
|---|---|---|---|---|---|---|
| | *ib*e_* | ATIB*E | *_t*an* | T*ANTO | *_c*os* | C*OSBA |
| | pu*at | EPU*ATE | *_p*es* | P*ESOD | *ea*th* | FEA*THS |
| | ng*e_ | SUNG*E | *st*uc* | ST*UCTIN | *_p*ac* | P*ACOL |
| | mp*ic | WIMP*IC | *og*ap* | OG*APT | *ai*s_* | THAI*S |
| | *mu*at* | VAMU*AT | *_p*ec* | P*ECDO | de*in | DE*INCH |
| | *us*y_* | BAUS*Y | *_p*ov* | P*OVEG | st*y_ | VIST*Y |
| | re*at | RE*ATUL | st*at | BOST*AT | ea*ed | THEA*ED |
| | dd*e_ | BIDD*E | te*mi | STE*MIN | _f*ag | F*AGLY |
| | io*og | JIO*OG | _p*oc | P*OCOY | mp*is | AMP*ISE |
| | ve*op | VE*OPIC | pe*at | EPE*AT | ng*es | CANGLES |
| | po*it | OPPO*IT | _p*of | P*OFACK | de*iv | DELIVIC |
| | ed*y_ | BOED*Y | pa*ti | SPA*TIB | fo*d_ | SOFOLD |
| | tt*e_ | MATT*E | ve*si | VE*SIPT | hi*e_ | SEHILE |
| | al*y_ | ABNAL*Y | di*ec | DI*ECRA | nd*ed | YINDLED |
| | ul*y_ | AXFUL*Y | _p*om | P*OMPA | _p*ow | PLOWIC |
| | ca*ly | VACA*LY | te*ed | STE*EDA | _b*az | BLAZZY |
| | co*le | SCOL*EG | wa*d_ | PAWA*D | po*e_ | SUPOLE |
| | _a*lo | A*LOPT | fe*en | UFFE*EN | so*de | SOLDEIN |
| | _i*lu | I*LUGO | pe*so | PE*SOX | _p*ag | PLAGIL |
| | co*la | SCO*LAD | fo*ma | FO*MASH | te*la | STE*LAP |

*Note also:* Responses to words in grey (seven words in L-dominant and one word in LR-possible conditions) were excluded in the second analysis of Study 1. See Results section for details.

In addition to the 60 test pseudowords, 90 foil items were created to distract the participants' attention from the test items. The foil words were also pseudowords, each of which had one missing letter, surrounded by zero to several clusters, and a forced-choice had to be made between different letter pairs, e.g., P-B, S-Z, C-K, M-N, W-V, T-D, Y-I, O-U, or I-E.

### ***Procedure***

The Letter Guessing Game (LGG) took place at the Laboratory of Second Language Acquisition and Bilingualism (SLAB). The experiment was created and administered by PsyScope X (Cohen, MacWhinney, Flatt, & Provost, 1993).

Following small amount of practice (five trials), the actual experiment took place. On each trial, one pseudoword was randomly presented to the participant on the computer screen with a question such as "Which letter is missing, 'L' or 'R'?" Once the participant made a forced-choice by guessing and pressed either key, the next word was presented on the screen. No feedback was given. Responses from 150 trials, including 60 test trials (20 pseudowords x 3 conditions) and 90 foil trials, were collected from each participant. The 60 test trials were analyzed as described below.

## Results

Table 2 gives a summary of the number and percentage of L for each stimulus in one of the three conditions, collected from the 33 English speaking participants. Another summary is given in Figure 1a, which displays the average L-responses in percentage obtained from each participant in each condition. As expected, the average L-response was the highest in the L-dominant condition (mean = 72.42%; *SD* = 14.26), followed by the LR-possible condition (mean = 46.97%; *SD* = 14.14), and the lowest in the R-dominant condition (mean = 38.18%; *SD* = 15.40). A one-way analysis of variance (ANOVA) conducted on the data revealed a significant main effect for condition ($F(2, 96) = 48.89$, $p < .001$).

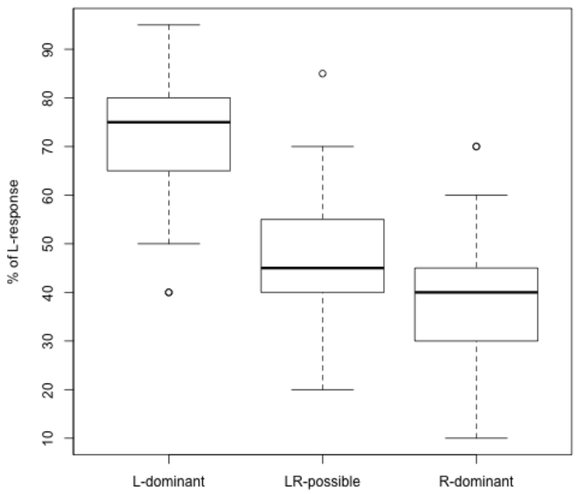
After the data collection was completed, it was observed that some of the pseudowords containing the letter L in the surroundings were more likely to receive L-response in general, which is obvious in Table 2 (i.e., the words in question are in grey.) These pseudowords were expected to have an effect on the response bias, i.e., the existence of L may have led the participants to prefer L in choosing a missing letter. The words that match this condition are: VACA*LY (ca*ly), SCOL*EG (co*le), A*LOPT (_a*lo), I*LUGO (_i*lu), SCO*LAD (co*la), AXFU*LY (fu*ly), and ABNAL*Y (al*y_) in the L-dominant condition; and STE*LAP (te*la) in the L-R possible condition. Comparing the seven words in the L-dominant condition and the remaining 13 words, a Welch Two Sample *t*-test confirmed that this was the case, with $t(18) = -2.78$, $p < .01$.

Table 2

*The Number and Percentage of L Responses for Each Pseudoword in Each Condition Collected from the 33 Native Speakers of English.*

| L-Dominant Condition | | | LR-Possible Condition | | | R-Dominant Condition | | |
|---|---|---|---|---|---|---|---|---|
| Stimulus | # of L Res | % of L Res | Stimulus | # of L Res | % of L Res | Stimulus | # of L Res | % of L Res |
| AXFU*LY | 32 | 97.0 | CANG*ES | 25 | 75.8 | PAWA*D | 20 | 60.6 |
| RE*ATUL | 30 | 90.9 | SO*DEIN | 22 | 66.7 | PE*SOX | 20 | 60.6 |
| I*LUGO | 29 | 87.9 | SOFO*D | 22 | 66.7 | STE*EDA | 20 | 60.6 |
| SCO*LAD | 28 | 84.8 | DE*INCH | 19 | 57.6 | DI*ECRA | 18 | 54.5 |
| VACA*LY | 28 | 84.8 | STE*LAP | 19 | 57.6 | P*OMPA | 18 | 54.5 |
| A*LOPT | 27 | 81.8 | SUPO*E | 17 | 51.5 | VE*SIPT | 18 | 54.5 |
| BIDD*E | 27 | 81.8 | VIST*Y | 17 | 51.5 | OG*APT | 16 | 48.5 |
| SUNG*E | 27 | 81.8 | DE*IVIC | 16 | 48.5 | P*ECDO | 15 | 45.5 |
| ABNA*LY | 26 | 78.8 | FEA*THS | 16 | 48.5 | UFFE*EN | 15 | 45.5 |
| BAUS*Y | 26 | 78.8 | P*ACOL | 16 | 48.5 | P*ESOD | 14 | 42.4 |
| BOED*Y | 24 | 72.7 | SEHI*E | 16 | 48.5 | EPE*AT | 13 | 39.4 |
| MATT*E | 23 | 69.7 | THAI*S | 16 | 48.5 | P*OCOY | 12 | 36.4 |
| VE*OPIC | 23 | 69.7 | B*AZZY | 13 | 39.4 | STE*MIN | 12 | 36.4 |
| EPU*ATE | 22 | 66.7 | F*AGLY | 13 | 39.4 | BOST*AT | 10 | 30.3 |
| JIO*OG | 21 | 63.6 | P*OWIC | 13 | 39.4 | SPA*TIB | 10 | 30.3 |
| SCO*LEG | 21 | 63.6 | YIND*ED | 13 | 39.4 | ST*UCTIN | 6 | 18.2 |
| ATIB*E | 18 | 54.5 | C*OSBA | 10 | 30.3 | P*OFACK | 5 | 15.2 |
| WIMP*IC | 18 | 54.5 | THEA*ED | 10 | 30.3 | P*OVEG | 5 | 15.2 |
| OPPO*IT | 14 | 42.4 | AMP*ISE | 9 | 27.3 | FO*MASH | 4 | 12.1 |
| VAMU*AT | 14 | 42.4 | P*AGIL | 8 | 24.2 | T*ANTO | 1 | 3.0 |

Note: Responses to words in grey (seven words in L-dominant and one word in LR-possible conditions) were excluded in the second analysis.



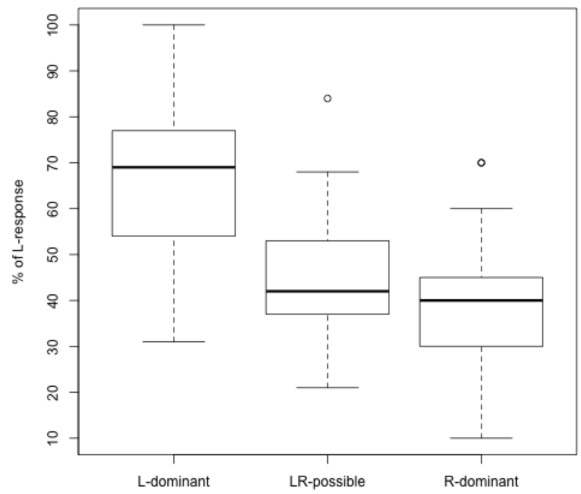




*Figure 1a.* LGG results obtained from 33 native speakers of English in the first analysis. The average percentage of L-responses per participant per condition is presented. All 20 words in each condition are included in this analysis.

*Figure 1b.* LGG results obtained from 33 native speakers of English in the second analysis. The average percentage of L-responses per participant per condition is presented. seven words in the L-dominant Condition and one word in the LR-possible condition were excluded in this analysis.

For this reason, reanalysis was conducted after removing the eight words in question. The distribution of the average L-responses in each condition without the eight words is displayed in Figure 1b. The average L-response percentage in the L-dominant condition dropped to 66.90% (from 72.42%) compared to the first analysis, 46.39% (from 46.97%) in the LR-possible condition. There was no change (38.18%) in the R-dominant condition. In order to evaluate if the prediction still holds true without the influence of the biased data, further data analysis was carried out again (with 13 words in the L-dominant condition, 20 words in the LR-possible condition, and 19 words in the R-dominant condition). A second ANOVA still indicated a main effect for condition, $F(2, 96) = 29.27$, $p < .001$.

Tukey multiple comparisons of means conducted as a follow-up analysis showed significant differences between the L-dominant and LR-possible conditions ($p < 0.000002$), and the L-dominant and R-dominant conditions ($p < 0.000000$), and marginally significant [*sic*, editor's note] between the LR-possible and R-dominant conditions ($p < 0.09$).

The binomial test showed the there was a marginally significant [*sic*, editor's note] difference between the number of L responses and that of R responses in the LR-possible condition (L responses = 291, R responses = 336), suggesting a slight response bias toward R ($p = 0.08$). With the number of L responses in the L-dominant condition (L = 287, R = 142) and that of R responses in the R-dominant condition (L = 252, R = 408), they were both above chance ($p < .01$ in both conditions).

## STUDY 2: ORTHOTACTIC KNOWLEDGE BY JAPANESE SPEAKERS

As discussed in the introduction section, if natural languages are abundant with cues that can be helpful in distinguishing two phonetic contrasts, Japanese learners of English should attend to cues from which they could easily obtain clues for distinguishing English /l/ and /r/. The results in Study 1 suggest that probabilistic phonotactics may serve as such a cue, since native speakers of English were able to effectively identify which category fits in based solely on phonotactic information. With the data obtained in Study 1 as a baseline, the same Letter Guessing Game was administered to Japanese learners of English in Study 2. If it is shown that the performance by Japanese is significantly lower compared to English speakers, this will indicate that there may be a good chance for training Japanese English learners to improve their /l/-/r/ distinction.

Not only the lower sensitivity to phonotactics as a group, but also wide individual differences are likely to exist among participants depending on their level of English. While beginning-level learners may show low sensitivity to probabilistic phonotactics, it is likely that the more advanced the learners are, the higher their sensitivity to phonotactics will be. In order to sort out and explain possible individual differences within the Japanese participant group, two other tests and a questionnaire that will provide measures relevant to phonotactic knowledge of the Japanese learners of English are also administered in Study 2. One of the tests is the Perception Test (PT) using a synthetic /la/-/ra/ continuum. The test aims at measuring the baseline sensitivity to the English /l/-/r/ categorical difference. The other is the Vocabulary Size Test (VST) developed by Nation and Beglar (2007). This test will enable us to see how passive knowledge of English vocabulary and probabilistic phonotactics knowledge are related. The questionnaire conducted to assess the level of English language proficiency based on the participants' language backgrounds is the Language Experience and Proficiency Questionnaire (LEAP-Q) for bilingual and multilingual assessment by Marian, Blumenfeld, and Kaushanskaya (2007).

## Method

### *Participants*

All 61 participants (13 males and 48 females) were native speakers of Japanese with various levels of experience in terms of exposure to English. They included graduate students, undergraduate students, visiting researchers, students in language programs such as the University of Hawaiʻi English Learning Program and (HELP) and the English Language Institute at the Department of Second Language Studies (ELI), as well as residents of Oahu and visitors to Honolulu as tourists.

Their ages ranged from 18-47 years (mean = 30.3 years; median = 28 years), and they had lived or spent time in an English-speaking country/countries for a few weeks to 37.7 years (mean = 6.3 years; median = 2.1 years).

An additional five persons were tested but excluded from analyses for the following reasons: two reported in the LEAP-Q that they had had some hearing problem in the past; one had lived in China until age eight and had communicated in Chinese with some family members on a daily basis, so a possible influence from his Chinese knowledge could not be discounted; one was mistakenly assigned to the earlier version of the LGG test that was given to English native speakers; and one participant decided to resign after completing the LGG.

The whole procedure typically took two hours per person, and upon completion of the tasks, they were paid at a rate of $5 per 30 minutes.

### *Stimuli and Procedure*

Each participant worked on the four tasks in the following order: *Letter Guessing Game*, *Vocabulary Size Test*, *Perception Test*, and *LEAP-Q*. Some participants needed two days to complete the tasks by doing the Letter Guessing Game and the Vocabulary Size Test on Day 1 and Perception Test and LEAP-Q on Day 2, while others completed the four tasks in one day.

For those completing the tasks in one day, a minimum of a 10-minute break was inserted between the Vocabulary Size Test and Perception Test.

***Letter Guessing Game.*** The same stimuli and apparatus as the control group (Study 1) were used except for eight words (seven words in L-dominant condition and one word in LR-possible condition; shown in Table 1 in grey). These eight were replaced with the following words newly created in order to avoid response bias due to the existence of L or R in the orthotactic frame:

*L-dominant:* S*EEG (_s*ee), SHO*OG (ho*og), NA*ISH (na*is), NAB*ESK (ab*es), GUMP*ET (mp*et), WIMP*IF (mp*if), F*OOB (_f*oo)

*LR-possible:* HA*INCH (ha*in)

The structure and the number of stimuli presented to participants were also the same as the LGG conducted on native English participants in Study 1. The whole test was carried out on PsyScope, and it typically lasted for 15-20 minutes.

***Vocabulary Size Test.*** This test, developed to provide a measure of a learner's vocabulary size from the 1st 1000 to the 14th 1000 word families of English, had 14 levels, with each level containing 10 questions. By multiplying the number of correct answers (which could range from 0 to 140) by 100, the vocabulary size can be estimated.

Each question is in multiple-choice format. Following a word in a context, four definitions are listed to choose from. Here is an example:

miniature: It is a miniature.

(a) a very small thing of its kind

(b) an instrument for looking at very small objects

(c) a very small living creature

(d) a small line to join letters in handwriting

Since the original 140-question format with English definitions was found to be too demanding and time-consuming for the Japanese subjects when tested on two pilot participants, the author contacted Dr. Nation and got a permission to use a shorter version (70 questions) with Japanese definitions. In reporting the result, the author multiplied each participant's score by 2, so the score obtained ranged between 0 and 140.

Subjects were handed a pen, question sheets and a marking-sheet to fill in their answers. They were allowed to spend as much time as they wanted, but were told to answer all questions even if they had to guess. Most subjects completed the test in 20-25 minutes.

***Perception Test.*** Six sound stimuli were used to carry out the perception test, which was a discrimination task by eliciting same or different judgments to a pair of stimuli presented. The stimuli were provided by Dr. Paul Iverson, and they were equivalent to the files used in Iverson and Kuhl (1996) and Iverson et al. (2003). The sounds were synthesized (Klatt & Klatt, 1990) based on a female American English speaker's natural citation speech recordings. The six stimuli varied in the third formant (F3; acoustic information that characterizes /l/-/r/) during the consonant closure, with the frequencies equally spaced on the mel scale, to sound from good /r/ to good /l/ exemplars. Other than that, the stimuli were all identical, and had equal first (F1) and second (F2) formants, which were fixed to 327 Hz and 1074 Hz respectively. A summary of stimuli which formed an /ra/-/la/ continuum is displayed in Table 3.

The experiment was created and carried out using PsyScope. Subjects were seated on a computer with headsets in a sound-attenuated booth. Following a short practice session with 5 trials, 300 pairs of sounds were randomly presented to the subjects, one pair at a time with an

Table 3

*The Formant Frequencies for the English /la/ and /ra/ Stimuli Used in this Study.*

| Stim | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| F3 (Hz) | 1739 | 2010 | 2319 | 2670 | 3070 | 3524 |
| F2 (Hz) | 1074 | 1074 | 1074 | 1074 | 1074 | 1074 |
| F1 (Hz) | 327 | 327 | 327 | 327 | 327 | 327 |
| | ra ← | | | | | → la |

The stimuli varied in terms of the third (F3) formants during the initial consonant inter-stimulus-interval of 250 msec. The subjects pressed the "S" key if they thought the sounds were the same, or "D" if different. When either button was pressed, the next pair was presented.

***LEAP-Q.*** At the very end of the session, each participant completed a computer version of the LEAP-Q questionnaire. The information provided by them included all the languages they had learned in the past and their current use, education level, details about when they started acquiring their first and second languages (Japanese and English respectively), the age they became fluent in each language, the length of time they had used the languages, and factors that contributed to acquiring the languages. It took 15-20 minutes to complete.

## Results

### *Letter Guessing Game (LGG)*

A summary of the LGG results in terms of the number and percentage of L responses for each pseudoword from the 61 participants is displayed in Table 4, and the average L-responses in percentage obtained from each participant in each condition are shown in Figure 2. The distribution of L responses looked different among conditions, just as in the case of native speakers of English, with the average L-response the highest in the L-dominant condition (mean = 59.59%; *SD* = 16.79), followed by the LR-possible condition (mean = 44.10%; *SD* = 14.82), and the lowest in the R-dominant condition (33.93%; SD = 14.44). A one-way ANOVA indicated a main effect of condition ($F(2, 180) = 43.027$, $p < .001$), and Tukey multiple comparisons of means showed significant differences among all three conditions, L-dominant and LR-possible conditions ($p < 0.000000$), L-dominant and R-dominant conditions ($p < 0.000000$), and LR-possible and R-dominant conditions ($p < 0.001$). The binomial test showed a significant difference between the number of L responses and that of R responses (L = 44.1% or 538 responses, R = 55.9% or 682 responses; $p < 0.01$), indicating a bias towards R responses.

Table 4

*The Number and Percentage of L Responses for Each Pseudoword in Each Condition Collected From the 61 Native Speakers of Japanese.*

| L-Dominant Condition | | | LR-Possible Condition | | | R-Dominant Condition | | |
|---|---|---|---|---|---|---|---|---|
| Stimulus | # of L Res | % of L Res | Stimulus | # of L Res | % of L Res | Stimulus | # of L Res | % of L Res |
| F*OOB | 47 | 77.0 | CANG*ES | 35 | 57.4 | P*OCOY | 35 | 57.4 |
| BAUS*Y | 43 | 70.5 | SO*DEIN | 35 | 57.4 | P*ECDO | 31 | 50.8 |
| GUMP*ET | 43 | 70.5 | DE*IVIC | 34 | 55.7 | STE*EDA | 30 | 49.2 |
| BOED*Y | 42 | 68.9 | P*OWIC | 33 | 54.1 | P*OMPA | 29 | 47.5 |
| S*EEG | 42 | 68.9 | C*OSBA | 32 | 52.5 | OG*APT | 28 | 45.9 |
| SHO*OG | 42 | 68.9 | DE*INCH | 32 | 52.5 | DI*ECRA | 24 | 39.3 |
| BIDD*E | 41 | 67.2 | THAI*S | 32 | 52.5 | P*OVEG | 24 | 39.3 |
| JIO*OG | 40 | 65.6 | P*ACOL | 30 | 49.2 | PE*SOX | 22 | 36.1 |
| SUNG*E | 40 | 65.6 | F*AGLY | 28 | 45.9 | BOST*AT | 21 | 34.4 |
| RE*ATUL | 37 | 60.7 | FEA*THS | 28 | 45.9 | P*OFACK | 21 | 34.4 |
| WIMP*IC | 37 | 60.7 | B*AZZY | 27 | 44.3 | EPE*AT | 20 | 32.8 |
| ATIB*E | 36 | 59.0 | SEHI*E | 27 | 44.3 | T*ANTO | 19 | 31.1 |
| MATT*E | 36 | 59.0 | HA*INCH | 25 | 41.0 | VE*SIPT | 19 | 31.1 |
| VE*OPIC | 35 | 57.4 | P*AGIL | 25 | 41.0 | SPA*TIB | 18 | 29.5 |
| OPPO*IT | 32 | 52.5 | VIST*Y | 24 | 39.3 | ST*UCTIN | 15 | 24.6 |
| WIMP*IF | 30 | 49.2 | SUPO*E | 22 | 36.1 | UFFE*EN | 15 | 24.6 |
| EPU*ATE | 28 | 45.9 | YIND*ED | 22 | 36.1 | PAWA*D | 14 | 23.0 |
| NA*ISH | 27 | 44.3 | THEA*ED | 18 | 29.5 | P*ESOD | 12 | 19.7 |
| VAMU*AT | 27 | 44.3 | SOFO*D | 17 | 27.9 | STE*MIN | 10 | 16.4 |
| NAB*ESK | 22 | 36.1 | AMP*ISE | 12 | 19.7 | FO*MASH | 7 | 11.5 |

*Note.* Words in grey (7 words in L-dominant and 1 word in LR-possible conditions) were newly created after words judged inappropriate in Study 1 were removed.

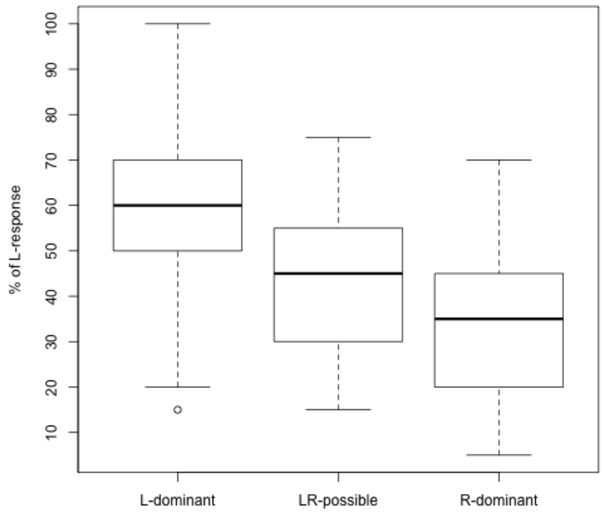
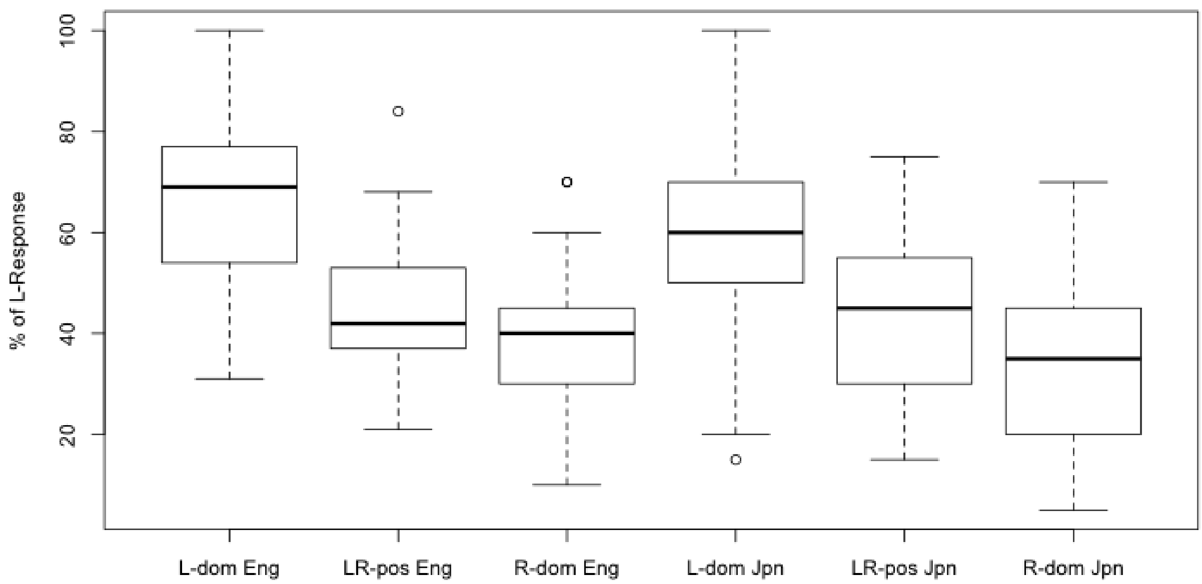



*Figure 2.* LGG results obtained from 61 Japanese learners of English. The average percentage of L-responses per participant per condition is presented.

In order to compare the performances of English speakers and Japanese speakers, LGG data for the two groups are displayed together in Figure 3. The three bars on the left are the same as the ones in Figure 1b and are based on a second analysis; the other three bars on the right are the same as the ones in Figure 2. Strictly speaking, the data of the two language groups are not comparable, since the results for the English-speaking group are based on a smaller number of stimuli (13 pseudowords in the L-dominant condition and 19 pseudowords in the LR-possible condition). Keeping this in mind, some observations and analyses will be attempted in the remainder of this section.




| | | | | | | |
|---|---|---|---|---|---|---|
| Mean (%): | 66.88 | 46.39 | 38.18 | 59.59 | 44.10 | 33.93 |
| SD: | 17.38 | 14.13 | 15.40 | 16.79 | 14.82 | 14.44 |

*Figure 3.* Comparison of LGG performance between English speakers and Japanese speakers.

A glance at Figure 3 indicates that, overall, the distributions of L responses in the three conditions are similar between the English speakers and the Japanese speakers, although there appears to be a tendency in the Japanese group to choose R responses somewhat more often, and the range of variability is larger. A 2 (Language: English vs. Japanese) x 3 (Condition: L-dominant, LR-possible, R-dominant) analysis of variance (ANOVA) on the L-response percentage showed a main effect of condition, $F(2, 92) = 125.68$, $p < 0.001$, and a marginally significant [*sic*, editor's note] main effect of language, $F(1, 92) = 3.07$, $p < 0.08$. There was no significant interaction between language and condition, $F(1, 92) = 0.99$, $p = 0.37$ (Figure 4). Thus, the findings indicate that all participants, both English and Japanese speakers, changed the frequency of L responses depending on the condition in a significant way: L-response was preferred most in the L-dominant condition, and L-response was preferred least in the R-dominant condition. Also, English speakers tended to give more L-responses regardless of the condition, but there was no significant difference in the performance between native speakers of English and Japanese learners of English. Moreover, no significant difference was observed between the two language groups in the way they responded in a specific condition(s).

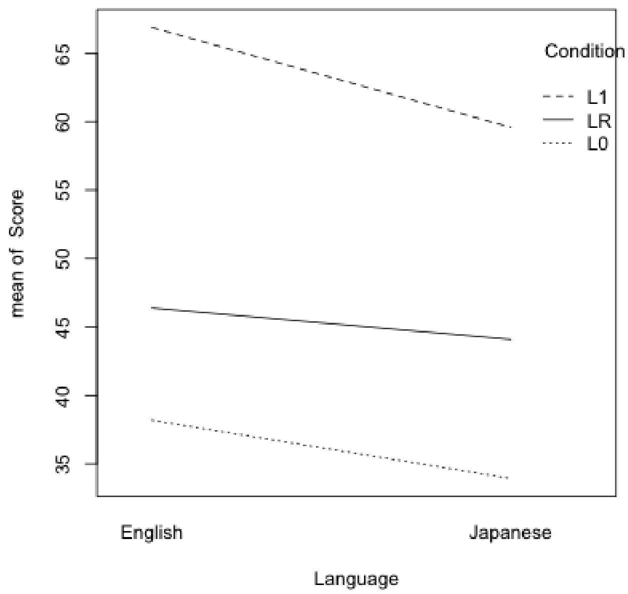



*Figure 4.* The effect of language and condition on L-responses.

As an index of how well each participant performed in the LGG task, percent correct rates for the L-dominant condition and the R-dominant condition were added and divided by 2 for every participant. For example, if a participant gave a response “L” 14 times (out of 20 trials) in words from the L-dominant context and 8 times (out of 20 trials) in words from the R-dominant context, the percent correct rates for L and R are L(c) = 14/20*100 = 70(%) and R(c) = (20-8)*100 = 60(%); thus the participant’s LGG performance index (LGG-PI) = (70+60)/2 = 65.5. The index obtained from each participant this way was used in examining the relationship between the phonotactic knowledge and factors such as sensitivity to English /l/-/r/, vocabulary size, and the language background, which is described in the following sections.

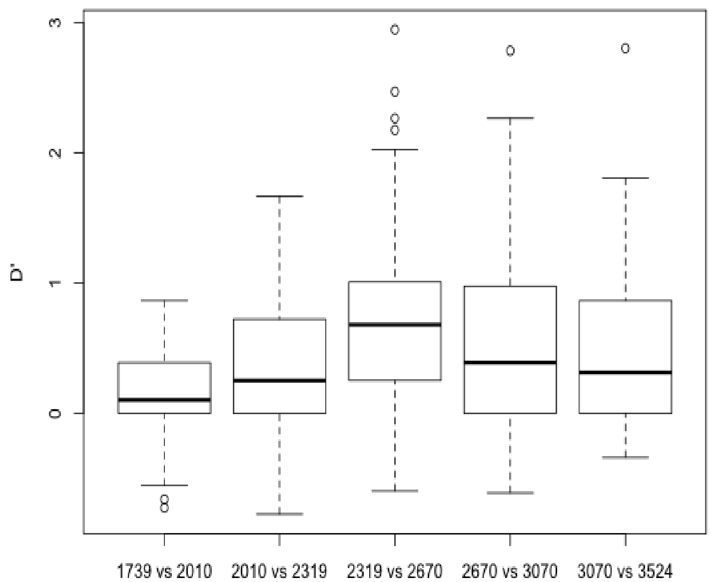



*Figure 5.* Distribution of D’s obtained per participant per stimulus pair.

***Perception Test (PT)***

Figure 5 shows the distribution of D’s obtained from the participants from five combinations. The mean D’s and the standard deviations were:

1739Hz vs 2010Hz: mean = 0.13; *SD* = 0.35

2010Hz vs 2319Hz: mean = 0.30; *SD* = 0.52

2319Hz vs 2670Hz: mean = 0.76; *SD* = 0.72

3070Hz vs 3524Hz: mean = 0.46; *SD* = 0.67

3070Hz vs 3524Hz: mean = 0.56; *SD* = 0.68

The scatterplot in Figure 6 intends to compare the participant’s sensitivity to English /l/ and /r/ with the LGG score. The D’ of 2319Hz vs 2670Hz from each participant is used to represent sensitivity. As can be seen from the scatterplot, the D’ cannot be used as a measure to explain their performance in the LGG. There was no correlation between the LGG score and sensitivity to /l/-/r/, r = -0.08.

Without the sensitivity measure from native speakers of English, it is not easy to tell what exactly the relatively high D’ of 0.76 obtained with the stimulus pair 2319Hz and 2670Hz from Japanese participants indicates, but it suggests that some participants perceived some differences between the two sounds. Yet, according to Iverson (personal communication), the boundary between /l/ and /r/ was supposed to come somewhere around 3070Hz, and in this sense, the sensitivity demonstrated by the participants is different from the one that native speakers of English possess. After the data collection was completed, the author presented the stimulus pairs to two native speakers of English, and confirmed that the perceptual boundary of English /l/ and /r/ was between 2670Hz and 3070Hz. The stimulus pairs were also presented to two Japanese speakers (who did not participate in the experiment), and their perceptual expression was that the stimulus with 2319Hz sounds Japanese /ra/ and the stimulus with 3070Hz to Japanese /wa/.

Thus, the data collected are not sufficient to claim whether any of the participants in the current study have sensitivity to the difference between English /l/ and /r/. In order to use the D' data as a sensitivity measure, sensitivity data should be collected from native speakers of English as well. Further, more thorough data should be collected, i.e., not only same/different judgment but also what category the respective stimulus falls on should be collected from Japanese speakers.

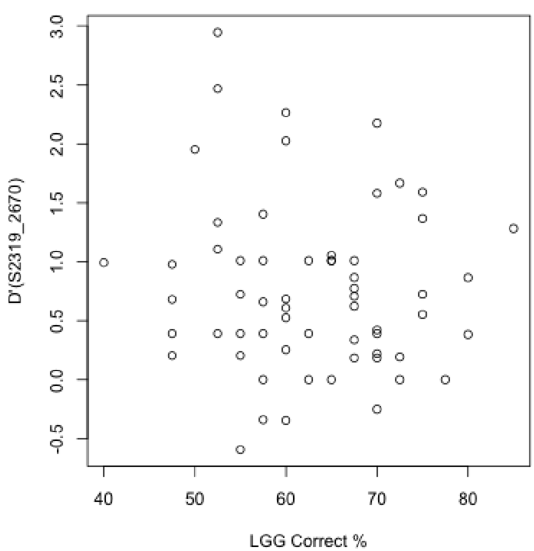



*Figure 6.* Scatterplots of the LGG score in relation to the D' of 2319 vs 2670.

### *Vocabulary Size Test (VST)*

Not much was found on the relationship between the LGG task score and the vocabulary size. The graph (Figure 7) with the LGG score on the x-axis and the VST score on the y-axis does not show any linear trend. No correlation was indicated between the LGG score and the VST score, *r* = -0.02.

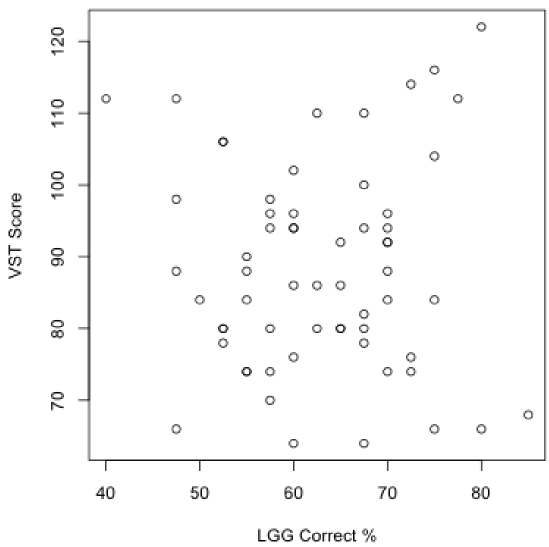



*Figure 7.* Scatterplots of the LGG score in relation to the VST score.

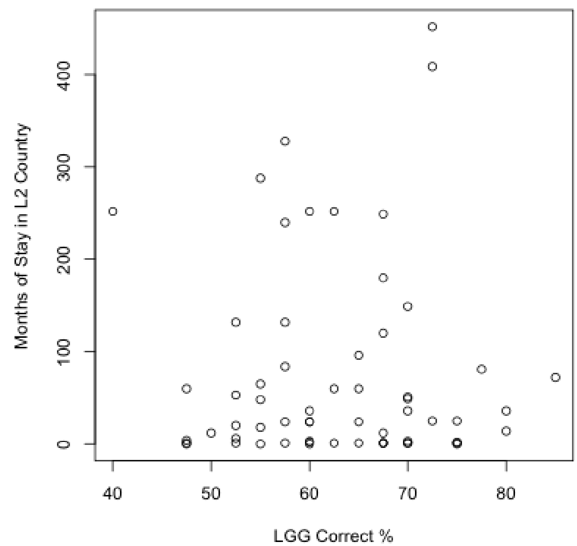



*Figure 8.* Scatterplots of the LGG score in relation to months of stay in the L2 country.

***LEAP-Q***

As can be seen from the scatterplot in Figure 8, the participant's length of stay in an English-speaking country cannot be used as a measure to explain their performance in the LGG. There was no correlation between the LGG score and the months of stay in an L2 country, *r* = -0.001.

Other than the length of stay, over 30 measures on their language backgrounds were collected through LEAP-Q. Among the measures obtained, two sets of multiple regressions were conducted focusing on "English learning experience" and "current use and proficiency of English."

Factors considered relevant to English learning experience included the "years of formal education," "length of stay in an English-speaking country," "length of study in an English-speaking school," "contribution of friends to learning English (on a scale from 0 to 10)," "contribution of TV to learning English (scale: 0-10)," and "contribution of reading to learning English (scale: 0-10)." None of the independent variables entered predicted sensitivity to orthotactics as measured by the Language Guessing Game.

Factors considered relevant to current use and proficiency of English included “current exposure time to English (in %),” “preference of reading in English over other languages (in %)[1],” “length of study in an English-speaking school,” “contribution of friends to learning English (on a scale from 0 to 10),” “contribution of television exposure to learning English (scale: 0-10),” and “contribution of reading to learning English (scale: 0-10).” A multiple regression conducted on these five factors indicated that “preference of reading in English over other languages” predicted the LGG score (beta = 0.137, $t = 2.748$, $p < .01$). No other entered variable predicted the LGG score significantly. The R-squared for the overall equation was significant but modest ($R^2 = 0.179$, $F(5,55) = 2.4$, $p < .05$).

The multiple regression results thus suggest some modest relationship between the performance in the LGG task and language preference in reading, i.e., Japanese speakers who prefer to read texts in English tended to perform better in LGG.

## DISCUSSION

The aim of the current study was to measure Japanese speakers’ sensitivity to probabilistic orthotactics of English /l/ and /r/, which was expected to reflect their phonotactic knowledge of the phonemic contrast. The results obtained were intended to provide a grounds for creating a training scheme that enables Japanese learners of English to raise their sensitivity to the /l/-/r/ contrast through learning phonotactic distribution of the phonemes.

To this end, four tasks were assigned to a group of native speakers of Japanese. The Letter Guessing Game (LGG) was conducted with a prediction that Japanese learners of English would have less sensitivity to English orthotactic rules than native speakers of English and that there would be a wide variability in the level of sensitivity. The variability was predicted to stem from other linguistic factors, and several indexes were collected through additional tasks. The Perception Test (PT) was designed to measure the sensitivity to the English /l/ and /r/ contrast in terms of acoustic cues, and it was expected that those who have sensitivity to the phonetic contrast would also be sensitive to the phonotactic distribution of /l/ and /r/. The Vocabulary

[1] Responses are based on the question and had to be answered in percentages: When choosing to read a text available in all your languages, in what percentage of cases would you choose to read it in each of your languages? Assume that the original was written in another language, which is unknown to you.

Size Test (VST) was administered as a measure of proficiency (refs Nation) and vocabulary size was expected to correlate with knowledge of the /l/-/r/ distribution in English. LEAP-Q asking detailed language background of each participant was meant to provide variables that would explain the variability in the LGG performance among the Japanese participants.

Contrary to these predictions, however, the Japanese participants in the current study did not show any significant difference in their sensitivity to English phonotactics compared to native speakers of English. The indexes collected from the additional three tasks did not yield meaningful outcomes to explain the variability within the Japanese participants. Thus, in what follows, I will shift my attention away from the PT, VST, and LEAP-Q, and discuss the results yielded by LGG and its implications. I would like to conclude this paper by proposing future directions for data collection and analysis.

First, what can we conclude about the Japanese participants' knowledge on probabilistic phonotactics from the result of LGG? There are a few possibilities. It may be that the Japanese speakers' ability to make use of phonotactic distribution is comparable to that of native speakers, and the problem of /l/-/r/ distinction stems solely from the lack of sensitivity to the acoustic cues. This could be the case, since there was no significant difference between the English speaking and the Japanese speaking groups in the LGG. Yet, this cannot be confirmed unless a new set of experiments (with the pseudowords used for testing the Japanese group) is given to a new group of native English speakers.

Also, it is too soon to conclude that the Japanese speakers' knowledge on phonotactics is equivalent to that of native listeners until we conduct a phonotactic version of the LGG, or the Phoneme Guessing Game (PGG), and get evidence that the Japanese speakers' performance is no different from that of native speakers. However, it is very unlikely that Japanese speakers will perform in the PGG as well as in the LGG. This is because emphasis is often placed more on learning orthographic words than the phonetic words in second language classrooms, and even with good knowledge of English orthography, this does not guarantee possession of sufficient knowledge of English phonotactics. The inconsistent writing system in English may add to the difficulty of connecting letters to sounds in a systematic way. Furthermore, even if a native speaker of Japanese possesses knowledge of probabilistic phonotactics equivalent to that of native speakers, sound is temporary compared to written language whose image is stable, and therefore, it is hard to believe that a nonnative speakers would perform exactly the same in the two tasks that require distinct skills.

It goes without saying that the LGG using the new set of pseudowords (those used with the Japanese participants) must be carried out with a new group of native speakers of English in order to make more accurate comparisons with the Japanese participants possible. It may also be reasonable to test another group of Japanese speakers, preferably a more homogenous group or two. In the case of the Japanese participants in the current study, there was wide variability in a number of factors, such as age, length of stay in an English speaking country, level of education, the type of exposure they had had, and the way they acquired English, all of which seem to have interacted to yield a very complicated picture of lexical knowledge. In order to partial out effects of lexical knowledge, it seems better to control other factors. For example, it may be possible to select two groups of college students, all of whom have been brought up in a monolingual environment without any substantial exposure to authentic English, but one group consisting of English majors and the other of students majoring in a field other than language.

Some consideration will be necessary in creating the PGG in order to directly measure the phonotactic knowledge of the Japanese speakers. In the PGG, a pseudoword has to be pronounced with its /l/ or /r/ area made ambiguous and replaced by non-speech sound, such as a tone. However, this is certainly a challenge, since it is extremely difficult to create stimuli that do not leave any trace of /l/ or /r/ in the surrounding sounds. For example, if it is a vowel that is surrounding /l/ or /r/, there will be a trace in the way formants glide. Consonants will change quality, too. For example, the /t/ in “tree” and “try,” followed by an /r/, tends to affricate and becomes like [tʃriː] and [tʃraɪ] (Catford, 2002). Although leaving this as it is may mean leaving some acoustic cues and may be considered a failure to control the cues, it is also arguable that knowing this phenomenon is part of phonotactic knowledge. After all, different levels go hand in hand and serve together as cues in language processing (Holt & Lotto, 2006).

In order to make the PGG more feasible, and at the same time make it possible to sort out what actually the Japanese learners’ problem is regarding the phonotactics of /l/ and /r/, I would like to propose the following procedure. First, conduct the LLG that takes exactly the same format as the current version. The stimuli will be presented on the screen, and participants will respond by choosing L or R as the missing letter. There will be the L-dominant condition, the LR-possible condition, and the R-possible condition. This will measure the participant’s orthotactic knowledge. After that, conduct the PGG. Stimuli will be created in the following way. First, calculate the phonotactic probabilities by making use of the ELP corpus, and make pseudowords based on the results. Then have the words pronounced by a native speaker of English. The L-dominant words will be pronounced with an /l/, and the R-dominant words will be pronounced with an /r/. As for the LR-possible words, half of them will be pronounced with /l/, and half of them with /r/. Finally, cover the /l/ or /r/ area with a tone, and present them randomly through headsets. It will be impossible to remove all the traces of /l/ or /r/ in the surrounding sounds, so with the L-dominant and R-dominant words, listeners will be tested on their sensitivity to both phonotactic and acoustic cues (i.e., the two cannot be completely separated). In the case of words in the LR-possible condition, the listeners will be tested only on the sensitivity to the acoustic cues surrounding the /l/ or /r/ that have been made ambiguous. The procedure is summarized in Table 5.

I predict four patterns of performance among the Japanese listeners:

| | | | |
|---|---|---|---|
| | LGG | PGG (L-dominant & R-dominant) | PGG (LR-possible) |
| (1) | X orthotactic | X phonotactic + acoustic | X acoustic |
| (2) | ○ orthotactic | X phonotactic + acoustic | X acoustic |
| (3) | ○ orthotactic | ○ phonotactic + acoustic | X acoustic |
| (4) | ○ orthotactic | ○ phonotactic + acoustic | ○ acoustic |

Based on what has been discussed above, I predict the phonotactic task (PGG) would be more difficult than the orthotactic task (LGG). Those who perform very well on both the LGG and PGG (those who perform as (4)) as well as those in (3) who fail only in the LR-possible context of PGG (i.e., who lack in acoustic sensitivity only) will not be the target of the /l/-/r/ phonotactic training that will follow the two tests. The /l/-/r/ phonotactic training will await Japanese learners of English who show either pattern (1), who lack both orthotactic and phonotactic sensitivity, or (2), who possess orthotactic knowledge but lack in phonotactic sensitivity.

Table 5

*Letter Guessing Game + Phoneme Guessing Game*

| | L-dominant | LR-possible | | R-dominant |
|---|---|---|---|---|
| LGG: | F*OOB, BAUS*Y | CANG*ES, DE*IVIC | | P*OCOY, T*ANTO |
| | << Present the pseudowords with an asterisk on the computer screen >> | | | |
| | ↓ | ↓ | | ↓ |
| | With all three conditions testing the orthotactic knowledge | | | |
| PGG: | xLxxx, xxxxLx | xxxxLxx | xxRxxxx | xRxxxx, xRxxxx |
| | << With all four conditions, record pseudowords as natural tokens, then cover the /l/ or /r/ area with white noise and present through headsets>> | | | |
| | ↓ | ↓ | ↓ | ↓ |
| | phonotactic + acoustic | acoustic | acoustic | phonotactic + acoustic |

## CONCLUSION

"Learn with a context!" is what every language learner has been repeatedly told. When you learn a new word "suffocate," for example, you know it is more effective to learn it in a sentence "He was suffocated by the smoke," rather than memorizing the word by itself. When it comes to learning of L2 phonological contrasts, however, the instructors' attention tends to be focused too much on how the sounds contrast in the same contrast, too little information on what kind of distribution each category has is left out, although the phonotactics of a language is a significant element in speech processing. For many Japanese learners of English, English /l/ and /r/ are assimilated to Japanese /r/, and English words "play" and "pray" are both assimilated into Japanese /purei/. Likewise, English "tray" sounds like Japanese /torei/, and without knowledge of English phonotactics, they will have no idea if it is "tlay" or "tray" in English. Incorporating the element of phonotactic knowledge into the /l/-/r/ training will surely be beneficial to such learners.

## ACKNOWLEDGMENTS

This work was supported by a Language Learning Research Grant awarded to L. Onnis at the Center for Second Language Research. The author would like to thank Dr. Luca Onnis for providing her the opportunity to participate in this phonotactics project and for his guidance throughout the research period. She is also grateful to Daniel Jackson for his assistance in conducting the experiment sessions, and Lawrence Davis for his advice in calculation of D-primes. She is indebted to Dr. Richard Schmidt for his suggestions in writing this paper.

**REFERENCES**

Akahane-Yamada, R., Kato, H., Adachi, T., Watanabe, H., Komaki, R., Kubo, R., Takada, T, & Ikuma, Y. (2004). ATR CALL: A speech perception/production training system utilizing speech technology. *The 18th International Congress on Acoustics, III*, 2319-2320.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*, 233-250.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). *The English Lexicon Project. Behavior Research Methods, 39*, 445-459.

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

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305-330.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America, 125*, 469-479.

Bradlow, A., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299-2310.

Catford, J. C. (2002). *A practical introduction to phonetics*. Oxford: Oxford University.

Christiansen, M., Onnis, L., & Hockema, S. (2009). The secret is in the sound: From unsegmented speech to lexical categories. *Developmental Science, 12*(3), 388-395.

Cohen J. D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257-271.

Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). Timonium, MD: York.

Flege, J. E., Takagi, N. & Mann, V. (1995). Japanese adults can learn to produce English /I/ and /l/ accurately. *Language and Speech*, *38*, 25-55.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1161-1173.

Ganong, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110-125.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia, 9*, 317-323.

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*, *125*, 469-479.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*, 65-94.

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *Journal of the Acoustical Society of America*, *108*, 3013-3022.

Holt, L., & Lotto, A. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*(5), 3059-3071.

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99*, 1130-1140.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tokura, Y., & Kettermann, A. (2003). A perceptual interference account of acquisition difficulties with non-native phonemes. *Cognition, 87*, B47-B57.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820-857.

Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*, 831-843.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606-608.

Lively, S., Logan, J., & Pisoni, D. (1993). Training Japanese listeners to identify English /r/ and /l/: II. the role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustic Society of America, 94*, 1242-1255.

Lively, S. E., Pisoni, D. B., & Yamada, R. A. (1994). Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96,* 2076-2087.

Logan, J., Lively, S. E, & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustic Society of America*, *89*, 874-885.

Macmillan, N. A., & Creelman, D. C. (1991). *Detection theory: A user's guide*. New York: Cambridge University.

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 1-28.

Maye, J., & Gerken, L. (2000). Learning phonemes without minimal pairs. *BUCLD 24 Proceedings*, 522-533.

Maye, J., & Gerken, L. (2001). Learning phonemes: How far can the input take us? *In 25th annual conference of the BUCLD.*

Maye, J., & Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122-134.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*(6), 457-465.

Mirman, D., Magnuson, J., Graf Estes, K., & Dixon, J. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108,* 271-280.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of r and l by native speakers of Japanese and English. *Perception & Psychophysics, 18*, 331-340.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.

Shea, C., & Curtin, S. (2005). Learning allophones from the input. *In 29th annual conference of the BUCLD.*

Strange, W. (1995). Cross-language studies of speech perception: An historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York.

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*, 2887-2896.

Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of Englis. *Applied Psycholinguistics 16* (4), 379-405

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishiki, K., Kohno, M., Menyuk, P., & Best, C. T. (1994). Discrimination of English /r-1/and /w-j/ by Japanese infants at 6-12 months: Language-specific developmental changes in speech perception abilities. Paper presented at the International Conference on Spoken Language Processing.

Uchida, Y. (2001). *Perception of English syllable-initial consonants by Japanese listeners.* Unpublished Ph.D dissertation, Tokyo Univ. of Foreign Studies.

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge: Cambridge University.

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

Werker, J., & Yeung, H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences, 9*(11), 519-527.