# THE EARLY PHASE OF /ɹ/ PRODUCTION DEVELOPMENT IN ADULT JAPANESE LEARNERS OF ENGLISH

Kazuya Saito[1]

Waseda University

Murray J. Munro

Simon Fraser University

**Running Head**: DEVELOPMENT OF /ɹ/

[1] **Acknowledgement**: This research was funded by Waseda University Grant for Special Research Projects and the Government of Canada Post-Doctoral Research Fellowship awarded to the first author. We are grateful to Keiko Hanzawa for her help with data collection. **Address for correspondence**: Kazuya Saito, School of Commerce, Waseda University, 1-6-1 Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan; email: kazuya.saito@waseda.jp.

**Abstract**

Although previous research indicates that Japanese speakers' second-language (L2) perception and production of English /ɹ/ may improve with increased L2 experience, relatively little is known about the fine phonetic details of their /ɹ/ productions, especially during the early phase of L2 speech learning. This cross-sectional study examined acoustic properties of word-initial /ɹ/ from 60 Japanese learners with a length of residence (LOR) between one month and one year in Canada. Their performance was compared to that of 15 native speakers of English and 15 low-proficiency Japanese learners of English. Formant frequencies (F2 and F3) and F1 transition durations were evaluated under three task conditions—word reading, sentence reading, and timed picture description. Learners with as little as two to three months of residence demonstrated target-like F2 frequencies. In addition, increased LOR was predictive of more target-like transition durations. Although the learners showed some improvement in F3 as a function of LOR, they did so mainly at a controlled level of speech production. The findings suggest that during the early phase of L2 segmental development, production accuracy is task-dependent and is influenced by the availability of L1 phonetic cues for redeployment in L2.

*Key words*: L2 pronunciation, English /ɹ/, Experience effects, Task variation

**Introduction**

Much theory and research on second language (L2) segmental acquisition focuses on whether and how L2 learners expand their repertoire of phonetic categories as a result of experience with the sound system of their new language. Flege's Speech Learning Model (SLM), for instance, posits that adult L2 speech learning takes place in a common phonological space where the phonetic system of the first language (L1) is already well established. As a result, a foreign accent is a typical feature of adult L2 speech (Flege, 1995, 2003; Piske, MacKay, & Flege, 2001). During the initial stages of L2 acquisition, learners are hypothesized to perceive unfamiliar L2 segments in terms of their L1 repertoire via a process of equivalence classification. As L2 experience increases, however, learners may begin to perceive L2 sounds as distinct from their L1 counterparts, and may establish new categorical representations. Ultimately, such perception-based categories are expected to activate relevant articulatory routines such that L2 sounds are produced in an increasingly target-like manner (i.e., a perception-first view; see also Kuhl, 2000). However, the contribution of L2 experience to ultimate attainment in speech learning is mitigated by learners' age of L2 acquisition: older learners typically do not reach the same levels as younger learners, even after many years of L2 use (e.g., Flege, Munro, & MacKay, 1995).

The view that the development of new phonetic categories and restructuring of old ones is triggered and facilitated by L2 experience is widely expressed in the literature (Best & Tyler, 2007; Flege, 1995, 2003; Major, 2001; McAllister, Flege, & Piske, 2002), and a number of studies using length of residence (LOR) as a measure of experience have pinpointed specific experience benefits for both segments and prosody. For example, in a study of English vowel perception and production, Flege, Bohn, and Jang (1997) noted that speakers from a variety of

L1 backgrounds with a mean of 7 years of US residence outperformed short–term residents. With respect to prosody, Trofimovich and Baker (2006) reported better production of English rhythm by Korean speakers with 3 years of US residence than by those with only 3 months. However, the results of several other studies appear to run counter to these findings. Flege, Munro, and Skelton (1992), for instance, found no benefit of increased LOR on the intelligibility of Mandarin and Spanish speakers' productions of English word-final stops. Likewise, Cebrian (2006) observed little indication of an experience effect on Catalan listeners' perceptions of English vowels. These differences in findings may be the result of a complex array of factors that arise when LOR is used as an independent variable. Here we consider four concerns: the window of observation, the relationship between actual language experience and LOR, the differential effects of experience on particular speech phenomena and the influence of elicitation procedures.

*Window of observation.* A number of research findings suggest that experience effects are most evident in the early stages of L2 acquisition, possibly during the first year or so. Flege and Fletcher (1992) found that English L2 speakers with a mean LOR in the US of 14.3 years had weaker global foreign accents than those with only .7 years residence. However, Flege (1988) found no difference in global accent between 5-year and 1-year Chinese-speaking residents. This led to his proposal that the "amount of unaided second-language (L2) experience does not affect adults' L2 pronunciation beyond an initial rapid stage of learning" (p. 70). Evidence that this is true for global accent has been documented in Derwing and Munro's (2013) seven-year longitudinal study of Mandarin and Slavic ESL learners in Canada: Neither group showed a change in global accent after the second year of residence. With respect to more fine-grained phonetic details, however, improved performance may be evident much later on. For instance, when Baker and Trofimovich (2006) examined vowel accuracy in Korean speakers of English, they found no difference between 3-year and 1-year residents, yet 10-year residents outperformed the 1-year group. Few studies have investigated in detail the initial stages of L2 production development, and, as Piske et al. (2001) pointed out, the notion of 'early phase' has not been satisfactorily defined. However, it is clear that a full account of the L2 phonetic learning process requires a delineation of this time frame and an understanding of how it varies depending on the speech phenomena at issue. In the current study, our approach is to consider production during only the first year of residence in Canada since that time period seems most likely to allow us to observe changes.

*Experience and LOR.* Another issue that arises in connection with the study of L2 speech development is that LOR is not, in itself, a causal factor in phonetic learning. Rather, it is a proxy for "experience factors" – chiefly L2 input and interaction – which are presumed to more directly impact learning. However, as Piske et al. (2001, p. 197) observe, "LOR only provides a rough index of overall L2 experience." (p. 197). Merely residing in an L2-speaking area does not necessarily require use of the L2, particularly if a community of other L1 speakers is available for social support. Even among learners who do use the L2 regularly, the amount and quality of language experience may vary such that some learners benefit more than others despite identical LORs. This point was illustrated by Flege and Liu (2001), who found that Chinese students in the US showed LOR benefits on a variety of L2 measures, whereas non-students, who likely had less pressure to use English, did not. In addition, Derwing, Munro and Thomson (2008) observed a link between greater gains in L2 oral fluency and comprehensibility, and greater actual use of L2 for learners with very similar LORs. Such findings point to the need to be especially cautious when using LOR as a proxy for L2 experience. In the present study, we attempt to alleviate this concern by focussing on highly-motivated learners who are currently enrolled in language classes. While we cannot rule out the possibility of some individual differences in language experience, we can be reasonably certain that all participants have been exposed to their L2 on a regular basis during the time periods under consideration.

*Differential effects of experience on speech phenomena.* Bohn and Flege's (1990) research led to the conclusion that some L2 segmental distinctions are susceptible to experience effects while others are not, and Trofimovich and Baker (2006) observed the same for aspects of L2 prosody and fluency. In some cases, experience benefits may be tied to particular aspects of the L1 phonological system which can be deployed during L2 learning. Flege and Wang (1990), for example, found that speakers whose L1 permitted more word-final consonants showed greater perceptual sensitivity to the English final /t/–/d/ distinction, than did learners with more restrictive L1s. In a study of the perception of spectral and tonal aspects of Mandarin vowels, Gottfried and Suiter (1997) found that English learners had more difficulty acquiring L2 lexical tone than L2 vowel quality, perhaps because the former is not an aspect of their L1 system, while the latter is.

The possibility of exploiting L1 knowledge in L2 phonetic learning applies not only at the level of prosody and segments, but also with respect to lower-level phonetic details. In particular, the acquisition of an L2 vowel or consonant need not be an all-or-nothing process. Rather, L2 learners may perceive and produce particular cues that characterize a phonetic category, while failing to acquire others, or doing so incompletely.

McAllister, Flege, and Piske (2002) investigated the acquisition of Swedish vowel duration by learners from different L1 backgrounds who had more than 10 years of LOR. Perception and production accuracy depended on the overall prominence of the duration features in the L1: speakers of Estonian, a quantity language, outperformed speakers of English, which makes some use of durational cues, who, in turn, outperformed speakers of Spanish, which has no phonemic length distinctions. This outcome led to the Feature Hypothesis, that "L2 features not used to signal phonological contrast in L1 will be difficult to perceive for the L2 learner and this difficulty will be reflected in the learner's production of the contrast based on this feature" (McAllister et al, 2002, p. 230).

Baker (2010) discussed in depth how Korean learners differentially mastered various levels of articulatory features to produce English word final stop voicing contrasts (e.g., "bat" vs. "bad"). She found that Korean learners produced one relevant cue to consonant voicing (preceding vowel duration) within as little as one year of US residence, but not other cues such as stop closure duration. She attributed such differential learnability of specific articulatory features to the fact that temporal differences are used in the Korean vowel system, which may have sensitized the Korean speakers to the vowel duration differences in English.

The Feature Hypothesis was developed to account for L2 learners' tendency to draw on a range of phonetic cues to discriminate and identify L2 phonological contrasts; this featural approach can be also extended to the SLM, (Flege, 1995, 2003), which posits that L2 learners develop L2 categories in a common phonological space with their L1 counterparts. In the context of English /ɹ/ acquisition, for instance, Japanese learners may succeed in establishing an English /ɹ/ category, separate from the Japanese tap as well from English /l/, such that they can perceive and produce these three sounds without confusion and conflation (Hattori & Iverson, 2009). We propose adapting the Feature Hypothesis to explain how L2 learners improve their production of English /ɹ/, in particular, and differentiate it from its closest L1 counterpart – the Japanese tap – at a featural level: Whereas L2 learners are predicted to have minimal difficulty learning to exploit phonetic features which are used in L1, a great deal of L2 experience might be necessary for them to establish perceptual representations and articulatory routines that are entirely new. In the current investigation, we test these predictions by examining how differential L2 experience (LOR = 1 to 10 months) impacts adult Japanese learners' English /ɹ/ acquisition of existing L1 cues (i.e., second formant [F2] frequencies and transition duration for rate and degree of tongue retraction) and new cues (i.e., third formant [F3] frequencies for labial, palatal and pharyngeal constrictions) at a fine-grained phonetic level.

*Elicitation procedures.* An important concern in L2 production studies is the fact that that different elicitation procedures can yield different outcomes. Although speech perception generally entails highly automatic processing, production can occur with varying levels of attention to meaning and form (see Flege, 1993, p. 1605). As part of his Ontogeny Phylogeny Model, Major (2001) proposed that as L2 phonological learning progresses, L1 influences slowly become less strong in more formal speech production. Thus, L2 learners can be expected to make fewer pronunciation errors in formal word reading tasks than in extemporaneous speech. This expectation was borne out in Rau, Chang, and Tarone (2009), who found that Chinese learners of English mispronounced /θ/ less frequently in word and sentence reading tasks than in picture descriptions. Reduced accuracy in the latter task may have resulted from the greater demands on linguistic processing during picture descriptions, which require much more planning than do controlled reading tasks (for similar results on consonant clusters, see Lin, 2003). A cognitive perspective on second language acquisition suggests that once learners integrate certain L2 linguistic features into their long-term mental representations, they proceduralize their L2 productive knowledge through a gradual transition from effortful to automatic processing (DeKeyser, 2001, Segalowitz, 2003). A full understanding of the phonetic learning process then, requires that adult L2 performance be measured in ways that tap different levels of processing. Although some studies have used interview tasks to elicit learners' extemporaneous speech to assess global foreign accentedness (e.g., Bongaerts, Planken, & Schils, 1995; Derwing, Rossiter, Munro, & Thomson, 2004; Moyer, 1999), few have examined how learners produce specific L2 *segments* extemporaneously, likely because of the difficulties in evaluating speech elicited in such a way (Piske, Flege, MacKay, & Meador, 2011). However, research in L2 morphosyntax (Spada & Tomita, 2010) has used communicatively free productions in the form of an "activity that calls for unplanned language use directed at fulfilling some communicative purpose" (Ellis, 2002, p. 225). Relevant tasks include picture descriptions, sometimes with written or oral prompts, whereby learners are guided to use target phenomena. Our choice in the present study is to use just such an approach.

*Acquisition of English /ɹ/ by Japanese speakers.* The extensive literature on Japanese speakers' acquisition of English /ɹ/ underscores the importance of the issues raised above. Just as in other work, experience effects on the acquisition of this consonant are supported by some studies but not others. Flege, Takagi, and Mann (1995) found better /ɹ/ productions in long-term (21-year) US residents than in 2-year residents, and Flege, Takagi, and Mann (1996) observed a perceptual benefit of experience on /ɹ/ - /l/ identifications. In contrast, Larson-Hall (2006) found no LOR benefit when she compared productions of 1.1-year US residents with those of 23.2-year residents. In her review, Bradlow (2008) pointed out that examining this supposedly most difficult instance of Japanese learners' /ɹ/ acquisition serves as "a productive testing ground for general principles of learning and claims about adult neural plasticity" (p. 294).

Because the Japanese phonetic inventory lacks /ɹ/ (and /l/), Japanese speakers tend to perceptually assimilate English liquids into their native tap category (Guion, Flege, Akahane-Yamada, & Pruitt, 2000). A comparison of the English and Japanese categories is therefore useful.The articulatory properties of North American English /ɹ/ vary considerably across speakers. As Ladefoged & Maddieson (1996) note, it can be produced as a retroflex alveolar or post-alveolar approximant with lower pharyngeal constriction and lip rounding. Nonetheless, a common articulation is bunched-tongue /ɹ/, in which the tongue apex is not raised, but constrictions occur at the palate and in the lower pharynx. Irrespective of articulatory configuration, a defining acoustic property of /ɹ/ is a characteristic drop in F3 frequency (to < 2400 Hz) due to three simultaneous constrictions in the labial, palatal, and pharyngeal areas of the vocal tract (Espy-Wilson et al., 2000). In addition, F2 typically ranges between 1700 and 2100 Hz, with F1 between 250 and 550 Hz. One temporal dimension is also relevant: F1 transition duration, which is 50 to 100 ms (Espy-Wilson, 1993; Hattori & Iverson. 2009).[2]

The Japanese tap lies somewhere between English /d/, in which the tip of the tongue contacts the alveolar ridge, and English /l/, in which it creates a lateral passage for airflow along the midline (Vance, 1987). Compared to English /ɹ/, the Japanese tap has a higher F3 (2300–2600 Hz) and F2 (1600–1700 Hz), with a shorter formant transition duration (5–20 ms) (Hattori & Iverson, 2009). Based on the previous literature described in detail below, we predict that

[2] According to the extensive previous literature, native English listeners draw on various strategies to perceive Japanese learners' accented productions of English /ɹ/ (and /l/). Whereas listeners tend to use all of the relevant acoustic cues (F3, F2, transition duration) to differentiate English /ɹ/ from English /l/ (between-category perception), F3 plays a primary role in determining the extent of targetlikeness for English /ɹ/ tokens (within-category perception) (e.g., Flege et al., 1995; Saito, 2013).

Japanese learners may show a hierarchy of difficulty for acquisition of the three key phonetic properties of English /ɹ/ as follows: F3 > transition duration > F2.

It has been reported that Japanese learners have tremendous difficulty in reliably perceiving English /ɹ/, arguably because they attend little to F3 or ignore this information even after many hours of auditory training (Ingvalson, McClelland, & Holt, 2012). Although some vocalic (e.g., /ɯ/) and consonantal (e.g., /k/, /g/) sounds in the Japanese phonetic system may entail weak lip rounding with low F3 as one of their acoustic characteristics (Dohlus, 2008), F3 does not serve as a *primary* phonetic cue for any of these sounds. Rather, the five spectrally distinct long–short vowel pairs (including /ɯ/ and /ɯɯ/) are identified mainly on the basis of F1 and F2, together with phonemic duration (Nishi & Kewley-Port, 2007).

Evidence suggests that during acquisition of English /ɹ/, Japanese speakers at least initially resort to two perceptual strategies based on their L1 knowledge. First, they give more perceptual "weight" to F2 than to F3, probably because F2 is the primary phonetic cue for their L1 approximant categories of /j/ and /w/, and is thus more salient to them (Iverson et al., 2003; Yamada, 1995). As a result, they likely articulate /ɹ/ with tongue retraction, thus generating /w/-like productions (Lotto, Sato & Diehl, 2004). A second strategy is a reliance on temporal cues (Iverson, Hazan, & Bannister, 2005; Yamada, 1995). This tendency might be due to L2 learners' general sensitivity to temporal over spectral information (Bohn, 1995; Flege, Bohn, & Jang, 1997) as well as to *some* exploitation of temporal cues in the Japanese vowel system, in which five spectrally-distinctive pairs of vowels are also temporally differentiated (Peterson & Lehiste, 1960).[3]

[3] Japanese speakers use phonemic duration as a primary phonetic cue to five short-long vowel contrasts (e.g., /i/ vs. /ii/, /u/ vs. /uu/) (Peterson & Lehiste, 1960). Yet, they do not do so for the two L1 approximant sounds (i.e., /w/ and /j/) (Bradlow, 2008). Thus, we assume that Japanese learners may have some sensitivity to the temporal aspects of English approximant /ɹ/ (Underbakke et al., 1989).

Thus, from a perceptual standpoint Japanese learners in the initial stage of acquisition of English /ɹ/ can be expected to be heavily reliant on F2, which is a familiar cue, and somewhat attuned to temporal properties, which are partially-familiar cues, at the expense of a focus on F3, which is a relatively new cue. However, the SLM predicts that as experience with English is gained, modifications to perceptual representations of word-initial /ɹ/ can be expected. At the featural level, we would expect these changes to entail a lower F2, a longer F1 transition duration, and attention to F3.

Although much investigation into L2 speech learning concerns perception, the Feature Hypothesis described earlier proposes that aspects of perception are also reflected in production. The research we will report here focuses exclusively on production. According to Flege's (1995) SLM, perceptual changes such as those described above should eventually lead to concomitant changes in articulation patterns for /ɹ/. If that is so, one might therefore expect a change in production toward narrowed labial, palatal and pharyngeal constrictions and the production of longer transitions.

*The present study.* The current study uses acoustic data to probe Japanese learners' productions of word-initial /ɹ/ during the first year of experience in an English-speaking area. Productions are elicited via three oral tasks and analyzed in terms of formant frequencies and transition durations.

*Hypotheses.* Our review of the literature leads us to propose two hypotheses to be tested. According to the Feature Hypothesis (McAlister et al., 2002), L2 learners have greater difficulty acquiring altogether new phonetic cues than adjusting existing ones. This leads to the expectation that, in terms of both perception and production, Japanese learners of English should find it relatively easy to take advantage of F2 frequency as a perceptual cue for English /ɹ/, since that cue is used across a wide range of Japanese segments, including vowels and liquids. In addition, the use of duration cues to segmental distinctions in Japanese, albeit limited, should assist Japanese speakers in producing temporal differences in segments, thus facilitating use of a lengthened transition duration as a cue for /ɹ/. However, it should be much more difficult for learners to learn to attend to F3 frequency, which is not relevant in the case of Japanese /ɾ/. On the basis of Flege's SLM (Flege, 1995, 2003), which predicts that L2 production eventually falls into line with perception, we can expect the Japanese speakers' productions of English /ɹ/ should reflect these differences in difficulty. Therefore, our first hypothesis is as follows:

> Hypothesis 1. Productions of English /ɹ/ by Japanese speakers living in Canada will show more target-like acoustic properties as a function of English experience with a predicted hierarchy of difficulty of F2 < F1 transition duration < F3.

According to the Ontogeny Phylogeny Model (Major, 2001), as L2 experience increases, learners gradually demonstrate more accurate production in formal contexts (e.g., in reading word lists) before doing so in extemporaneous productions (e.g., in explaining pictures). This leads us to our second hypothesis:

> Hypothesis 2. Productions of English /ɹ/ by Japanese speakers living in Canada will show more target-like acoustic properties as a function of English experience in tasks entailing a controlled level of production—Word Reading (WR; i.e., reading a list of target words) and Sentence Reading (SR; i.e., reading sentences including target words)—before they

do so in spontaneous speech—Timed Picture Description (TPD; i.e., using target words to describe a series of pictures).

Acoustic measurements are well-suited to an investigation such as this one, in which relatively fine-grained aspects of acquisition are to be assessed. Such details could not be adequately evaluated through listeners' judgments, for instance.

**Method**

*Japanese learners.* Sixty adult Japanese learners of English (48 females, 12 males) with a maximum of one year of residence in an English-speaking country were recruited. At the time of the project, all were enrolled in private language institutes in Montreal, Canada, to learn English abroad for academic or business reasons. They had been in Canada for a mean of 4.7 months with a range of 1–12 months. All had completed at least six years of formal English education in Japan prior to their arrival. These learners reported high levels of motivation to improve their oral proficiency skills in English for their academic and career goals (They had invested time and money to study abroad in Canada.).[4] For this reason, their time in Canada was likely to reflect the amount of English experience they had gained.

At the time of the study, their mean age was 27.8 years, with a range of 19 to 40 years. The learners were assigned to four groups ($n$ = 15 per group) on the basis of their LOR. Details are summarized in Table 1. In accordance with mean LORs, the groups are labelled as follows: *1M* (1 month), *2.5M* (2–3 months), *5M* (4–7 months), and *10M* (8–12 months). Results of a one-

[4] According to Statistics Canada, the Japanese immigrant population is relatively low (i.e., 0.06% in Quebec), which indicates a relatively limited community in the area (Statistics Canada, 2008).

way between-groups ANOVA indicated no significant differences in mean age among the four LOR groups at the time of the study, $F$ (4, 70) = .777, $p$ = .544.

*English and Japanese Comparison Groups.* For comparison purposes we collected productions from 30 native speakers of English (NE) and Japanese (NJ). For the former, 15 NE students were recruited at an English-speaking university in Montreal (mean age: 27 years, range: 20-40 years). For the latter, 15 NJ students who had never been abroad were recruited at a university in Tokyo (mean age: 18.7 years, range: 18-22 years). In contrast with the Japanese learner groups (12 males, 3 females), each of the comparison groups had eight males and seven females. Subsequent statistical tests (see below) indicated that this difference in gender balance was unlikely to have had an effect on the study's outcomes.

Table 1. *Characteristics of the speaker groups*

| Group (*n*=15/group) | Gender (F/M) | Age (years) | | LOR (months) | |
|---|---|---|---|---|---|
| | | M (SD) | Range | M (SD) | Range |
| 1M | 12/3 | 28.1 (5.24) | 22–40 | 1 (0) | 1 |
| 2.5M | 12/3 | 28.4 (5.23) | 21–37 | 2.5 (.52) | 2–3 |
| 5M | 12/3 | 25.9 (4.42) | 21–35 | 5.2 (1.21) | 4–7 |
| 10M | 12/3 | 28.9 (4.11) | 21–40 | 10.1(1.55) | 8–12 |
| NE | 8/7 | 27.3 (6.23) | 20–40 | *n/a* | *n/a* |
| NJ | 8/7 | 18.7 (1.03) | 18–22 | *n/a* | *n/a* |

*Stimuli*. Three oral tasks requiring differing kinds of processing were used to elicit productions (see below). The 20 target words, which are provided in Table 2, featured /ɹ/ in word-initial position ($n$ = 4 for TPD, $n$ = 8 for SR, $n$ = 8 for WR) and were Consonant-Vowel-Consonant (CVC) monosyllables, except for *Ryan* (CVVC). According to Cobb's (2011) vocabulary profile, all except *Ryan* and *ram* were among the 2000 most frequent English word families. Because Japanese learners have more difficulty producing word-initial /ɹ/ before front vowels than central and back vowels (Flege, Takagi, & Mann, 1995), vowel advancement was balanced across tokens in each task, with 50% of targets having front vowels and 50% having central or back vowels.

Table 2. *Target words in the production tasks*

| Task | Target words | | Sentences |
|---|---|---|---|
| | front V | central/back V | |
| TPD | read | road | n/a |
| | rain | rock | |
| SR | read | | *He will read my paper by the time I arrive there.* |
| | red | road | *She left her red bicycle on the side of the road.* |
| | race, rain | | *The race was cancelled because of the rain.* |
| | | wrong | *I can correct all wrong sentences tonight.* |
| | | Ryan, run | *Ryan does not like to run in the snow.* |
| WR | read | rough | n/a |
| | red | right | |
| | race | root | |
| | ram | room | |

*Procedure*. Individual recording sessions took place in a quiet room. Speech tokens were recorded using a Roland-05 audio recorder with a sampling rate of 44.1 kHz and 16-bit quantization, and saved as WAV files. All instructions were delivered in Japanese by the researcher (a native speaker of Japanese) to ensure the participants' clear understanding of the procedures and to avoid any possible exposure to English /ɹ/ during the instructions.

In the TPD task, participants were instructed to describe eight pictures presented randomly on cards, including four distracters, with five seconds of planning time per picture. Each picture had three key words underneath, one of which was a target item with word-initial /ɹ/. The four targets were *read, rain, road,* and *rock*. For example, a depiction of a road in the countryside with several clouds in the sky, together with the key words *blue sky, road, and cloud*) was used to elicit learners' production of /ɹ/ in *road* (for details of the other pictures, see Appendix). To familiarize speakers with the task, four distracter pictures were first presented, followed by the other four pictures covering the target words.

The TPD task was intended to elicit participants' spontaneous /ɹ/ production during language processing for message conveyance (i.e., paying simultaneous attention to not only phonological but also lexical, syntactic, and pragmatic use of L2). To prevent participants from paying too much attention to the pronunciation of English /ɹ/, distractor items were used, and the TPD task was always completed first, followed by SR and then WR. In addition, speakers were told that all of the tasks were designed to elicit their general oral production skills in English; but they were not informed about the true goal of the project (i.e., examining their pronunciation of /ɹ/) until completion of the data collection.

In the SR task, the participants read five target sentences together with three distracter sentences in a fixed order.[5] Each of the target sentences included one or two target words with word-initial /ɹ/, while the four distracter sentences did not include any (see Appendix). In contrast to the spontaneous production task (i.e., TPD), the sentence reading task allowed participants to focus on reading sentences accurately with minimal communicative pressure.

In the WR task, the participants read a list of 25 words comprising eight target words and 17 distracters in a fixed order. The distracters incorporated a number of easy and difficult English sounds (e.g., voiceless stops, interdental fricatives) (see Appendix). Due to the highly formal nature of the task, participants were expected to pay more conscious attention to

[5] Two words including /ɹ/ in word-medial position were excluded from the current analysis due to the different nature of phonetic contexts and the lack of samples. The results will be reported elsewhere.

pronouncing each of these word forms accurately, and demonstrate carefully-monitored productions of /ɹ/, possibly drawing on explicit articulatory knowledge.[6]

*Acoustic Analyses.* Acoustic analyses focused on F3, F2, and F1 frequencies, as well as transition duration. Linear predictive coding (Burg) in *Praat* (Boersma & Weenink, 2012) was used to measure the formant frequencies. Following Flege et al. (1995b), we identified word onsets by simultaneous visual inspection of the waveform and spectrogram of each token. Then a cursor was placed at the point where energy in all three formants was first clearly visible. Because the F3 of /ɹ/ is relatively low and the F3 in any preceding sound tended to continuously decline towards the beginning of each target word, the beginning of /ɹ/ in the SR and TPD productions was assumed to be the local minimum of F3. F1 transition duration was measured from the beginning point of F1 to the endpoint of the F1 transition; when the F1 peak was hard to identify, we instead used the end point of the F3 transition (Hattori & Iverson, 2009).

*Normalization.* To adjust for formant frequency variation due to individual differences in vocal tract length, raw acoustic values were submitted to the following normalization procedure (for details, see Lee, Guion, & Harada, 2006; Yang, 1996). A mean F3 value for /æ/ elicited from three monosyllabic words in WR (i.e., *man*, *map, ram*) was calculated for each talker. One female English talker was randomly selected as a reference, and her mean F3 value (3011 Hz) was divided into the mean F3 for the same words from the other talkers to provide their individual *k* factors. All formant values for each talker were then multiplied by the individual *k* factors. As noted earlier, the NE and NJ talker groups had a different gender balance than the other four

[6] Target words in the SR and WR tasks were not randomized and some items used in the TPD task (e.g., *read, rain*) were recycled. This was intentionally done so that the participants were guided to notice and increase their explicit attention to the target sound in the SR and WR tasks.

groups. In line with expectations, a series of independent-samples *t*-tests revealed significantly higher formant frequencies before normalization for the females than the males in F3 ($p < .001$), F2 ($p < .001$), and F1 ($p = .005$). After normalization, however, the gender effect was non-significant in all three cases (*ps* = .200 to .600). Consequently, we have no reason to expect the differences in gender balance across groups to be a matter of concern with respect to the between-group comparisons we report below.

To reduce the nonlinear relationship between the formant frequencies and the corresponding perceived approximant quality, all acoustic values were converted from Hz to Bark (Schroeder, Atal, & Hall, 1979)[7]:

$$Bark = 7\ln\left(\frac{Hz}{650} + \sqrt{1 + \left(\frac{Hz}{650}\right)^2}\right)$$

**Results**

In view of the typical acoustic characteristics of English /ɹ/ and the Japanese tap /ɾ/ as described in Hattori & Iverson (2009) and Lotto et al. (2004), the following benchmarks were used to interpret the data:

a) F3: 14.50–15.70 Bark for the Japanese /ɾ/ vs. 11.40–12.60 Bark for English /ɹ/

b) F2: 11.80–13.20 Bark for Japanese vs. 7.90–11.00 Bark for English

c) F1 transition duration: 5–20 ms for Japanese vs. 50–100 ms for English

---

[7] For other approaches to the conversion, see Traunmüller (1990).

Data were submitted to a series of mixed design ANOVAs with one between-group factor (Talker Group) with 6 levels (1M, 2.5M, 5M, 10M, NE, NJ) and one repeated measure (Task) with 3 levels (WR, SR, TPD). The results of all post hoc analyses (*t*-tests) reported below have been Bonferroni-adjusted (with criterion $p < .05$).[8]

*F3.* Visual inspection of Figure 1 suggests that the Japanese learners in the NJ and 1M groups (0 ≤ LOR ≤ 1 months) produced F3 values appropriate for the Japanese category (14.50-15.70 Bark) in all three tasks, while F3 frequencies in the 2.5M, 5M and 10M groups (2 ≤ LOR ≤ 12 months) were slightly lower and therefore more similar to those of English /ɹ/.

The ANOVA revealed a significant effect of Group, $F(5, 84) = 35.123$, $p < .001$, and Task $F(2, 168) = 9.067$, $p < .001$, as well as a significant Group × Task interaction, $F(10, 168) = 2.615$, $p = .006$. Post-hoc analyses indicated that under all task conditions the NE group had significantly lower F3 values than the Japanese groups, $ts = 7.256$ to $11.864$, $p < .001$ in all cases. The 2.5M and 5M groups produced significantly lower (and thus more target-like) F3 values than the NJ group only in WR ($t = 3.760$ and $4.111$, respectively, $p = .001$ and $< .001$). The 10M group outperformed the NJ group in all three task conditions ($t = 5.359$, $p < .001$ for WR; $t = 4.195$, $p = .001$ for SR; $t = 3.424$, $p = .024$ for TPD) and the 1M group in WR ($t = 3.402$, $p = .012$) and SR ($t = 3.758$, $p = .003$) but not in TPD ($t = 1.639$, $p = .208$).

Task effects were significant only for the 5M and 10M groups: the 5M group produced /ɹ/ with lower F3 in WR than SR ($t = 3.010$, $p = .011$) and TPD ($t = 2.698$, $p = .037$); the 10M group had lower F3 values in WR than in TPD ($t = 3.958$, $p = .004$).

[8] All Bonferroni-adjusted significance tests were computed via SPSS by multiplying unadjusted *p* values by the number of comparisons.

-----------------------------------------

INSERT FIGURE 1 HERE

-----------------------------------------

*F2*. Figure 2 suggests that F2 in the 1M group was appropriate for the Japanese tap category (11.50 Bark), while the Japanese learners with greater LOR appeared to produce more English-like F2. While the effect of Group proved significant, $F$ (5, 84) = 13.520, $p < .001$, no significant Task or interaction effects were observed, $F$ (2, 168) = 1.855, $p = .160$, and $F$ (10, 168) = 1.091, $p = .372$, respectively. According to post hoc analyses, (a) there was no significant difference between the NJ group and the 1M group ($t = 2.531$, $p = .199$), (b) the other Japanese groups (2.5M, 5M, 10M) produced significantly lower (and thus more target-like) F2 than the NJ group (*ts* = 3.268 to 5.041, $p < .05$); and (c) the NE group differed significantly from all of the Japanese groups (NJ, 1M, 2.5M, 5M) (*ts*= 3.477 to 7.761, $p < .05$), except the 10M group ($t = 2.510$, $p = .125$).

-----------------------------------------

INSERT FIGURE 2 HERE

-----------------------------------------

*F1*. Mean F1 data are shown in Figure 3. In the analysis, Group, Task and interaction effects proved non-significant for F1 ($p > .05$).

-----------------------------------------

INSERT FIGURE 3 HERE

----------------------------------------

*F1 Transition Duration.* Figure 4 appears to show longer, and therefore more English-like, transitions in the longer LOR groups than in the NJ group. A significant effect of Group, $F$ (5, 84) = 20.387, $p < .001$, was observed, but no significant Task or interaction effects emerged, $F$ (2, 168) = 0.097, $p = .907$, and $F$ (10, 168) = 1.417, $p = .177$. Post hoc comparisons showed that the NE group exhibited significantly longer durations than all Japanese groups (*ts* = 4.029 to 9.180, $p < .001$). In addition, whereas the 2.5M and 5M groups produced more native-English-like transition durations than the NJ group (*ts* = 3.940 and 4.122, $p = .003$ and .001, respectively), the 10M group outperformed not only the NJ group ($t = 5.151$, $p < .001$) but also the 1M group ($t = 3.759$, $p = .005$).

----------------------------------------

INSERT FIGURE 4 HERE

----------------------------------------

**Discussion**

This cross-sectional study of /ɹ/ productions by Japanese learners of English indicates differential success in the production of the relevant acoustic cues – F2, F3, and transition durations – during the first year of Canadian residence. Significant effects of both LOR and task type were observed, but these varied according to the acoustic cues. To interpret the results, we turn our attention to the two specific hypotheses related to L1 influence and task demands (see Table 3).

Table 3. *Summary of predictions and findings*

| Predictions | Findings |
|---|---|
| **L1 Effects**:<br>• New (F3) > Partially-familiar (transition duration) > shared L1/L2 cues (F2) | **F3**:<br>• Significant improvement limited to controlled processing task<br>• Non-native-like attainment<br>• Task effect after LOR of 5 months<br><br>**Transition Duration**:<br>• Significant improvement<br>• Non-native-like attainment<br>• No task effect |
| **Task Effects**:<br>• Spontaneous performance > controlled performance | **F2**:<br>• Significant improvement<br>• Native-like attainment<br>• No task effect |

*Predicted hierarchy of difficulty*. On the basis of McAlister et al.'s (2002) Feature Hypothesis we predicted that among the cues needed for accurate /ɹ/ perception and production, F3 should prove the most difficult to acquire because of its limited relevance in Japanese, particularly with respect to /ɾ/, which is the closest Japanese phonetic category to English /ɹ/. On the other hand, the development of perceptual representations and corresponding production routines that accurately reflect the role of F2 should be comparatively easy because of F2's importance in a wide range of Japanese segments. Finally, the /ɹ/ transition duration could be expected to show an intermediate degree of difficulty, given that some duration contrasts exist at the segmental level in Japanese that would presumably have sensitized the learners to the durational phenomena. Since Flege's (1995, 2003) SLM predicts that production comes to reflect perceptual representations, we therefore predicted a hierarchy of difficulty for production as follows: F2 < F1 transition duration < F3.

Our results for F2 indicate that, irrespective of task, Japanese learners with even very short LOR (2.5M) had begun to produce more target-like F2 frequencies for /ɹ/ compared to native Japanese talkers without any experience abroad. Moreover, those with approximately 10 months of residence produced native-like F2 values. This outcome supports the prediction that the development and mastery of satisfactory articulatory configurations for F2 was straightforward, occurring very early during Canadian residence.

The results for F3 indicate that the Japanese talkers with more than a few months of LOR showed some improvement especially in the controlled tasks (WR, SR). While all three Japanese learner groups produced significantly lower F3 values than the NJ group in WR, those with long-term LOR (10M) outperformed those with short-term LOR in WR and SR, though no group showed fully native-like performance. Taken together, these results support the prediction that F3 should pose considerable difficulty for Japanese learners in their acquisition of English /ɹ/ in all task conditions over only one year of LOR.

With respect to a key temporal characteristic of /ɹ/, all talker groups produced transition durations that were significantly shorter than those of the native English talkers. Thus, none of the groups can be said to have developed target-like mastery of the transition duration cue. Nonetheless, while the intermediate LOR groups (2.5M, 5M) showed significantly greater transition durations than the NJ group, the group with the longest LOR (10M) outperformed both the NJ and the shortest LOR group (1M), apparently reflecting partial acquisition. Such improvement was robust across the task conditions. Therefore, in conjunction with the findings for F2 and F3, the results are consistent with our predication of an intermediate level of difficulty for the acquisition of transition duration.

These findings generally support the Feature Hypothesis (McAlister et al., 2002) which proposes that the degree of success in L2 speech learning depends on the extent to which relevant phonetic cues are used in the L1 phonetic system. Our production data echo previous findings on Japanese learners' interlanguage strategies in the perception of English /ɹ/, such as their sensitivity to F2 variation (Iverson et al., 2003) and to temporal properties (Underbakke et al., 1988). While the results may also be consistent with a perception-first view of L2 phonetic learning, which assumes that difficulties in L2 production arise as a result of underlying perceptual limitations, we have not evaluated that possibility since this study includes no perceptual tests. In addition, it should be noted that our findings do not allow us to rule out the possibility that some of the difficulties experienced by the learners were due mainly or partially to their inability to establish native-like production routines, irrespective of the status of their perceptual representations. In particular, it is still conceivable that some Japanese learners of English might develop native-like perception of English /ɹ/, yet exhibit poor production, though ample evidence indicates that, even with extensive training, fully native-like perception of /ɹ/ is unlikely.

*Task effects.* On the basis of Major's (2001) Ontogeny Phylogeny Model, we predicted that task effects would be evident among Japanese learners of English with longer, but not shorter, LOR, and that such effects would be reflected in more native-like performance on tasks requiring fewer processing demands. While no task effects proved significant for the F1, F2, or transition data, both aspects of this hypotheses received some degree of support from the results for F3. First, the only groups to show task effects were those with the greatest LOR (5M and 10M). Second, assuming that processing demands were the least in the word reading task, the greatest in the picture description task, and intermediate in the sentence reading task, performance on F3 tended to be better when a task was less demanding. In general, all of the Japanese groups produced more native-like F3 than the NJ group in word reading, while the 10M group had more accurate F3 values than the 1M group in word reading and sentence reading.

This differential difficulty in producing L2 sounds across task conditions might be explained in terms of the processing resources available to speakers that would allow accurate articulatory configurations to be realized. In particular, the demands of the more processing-intensive tasks may have prevented the learners from exploiting all their knowledge about English /ɹ/. To account for such phenomena, several theoretical constructs have distinguished two broad types of L2 knowledge: explicit vs. implicit knowledge (DeKeyser, 2003), learned vs. acquired systems (Krashen, 1981), conscious vs. automatic processing (Segalowitz, 2003), and unintegrated and integrated knowledge (Jian, 2007). In the word reading task, the learners could probably focus more attention on monitoring their pronunciation of /ɹ/, perhaps drawing on explicit articulatory knowledge about tongue configurations and lip rounding, which may not yet have been fully automatized. In the picture description task, however, the talkers were called upon to produce English /ɹ/ while activating many different aspects of their linguistic knowledge, including phonology, morphosyntax, lexis, and pragmatics, while under time pressure. If some articulatory routines associated with accurate /ɹ/ had been learned explicitly, but not fully automatized, it might have been possible for the talkers to deploy them only when processing requirements were relatively light.

**Conclusions and Implications**

While the findings reported here provide useful insights into the production of specific phonetic cues during the first year of residence in an L2-speaking environment, a full picture of segmental acquisition requires probing beyond the time frame selected for this study. An obvious direction for future work is an examination of improvement on production of F3 and F1 transition durations beyond one year of residence. For any future L2 research on segmental learning, however, we note that the relationship between L1 and L2 phonetic cues (i.e., new vs. familiar cues) and elicitation methods (controlled vs. spontaneous tasks) must be carefully considered. Previous longitudinal work indicates that some aspects of adult segmental learning may plateau before the end of 1 year of L2 experience (Munro & Derwing, 2008), such that LOR may be a good predictor of acquisition phenomena only during the first several months of residence. However, this may be true for some, but not all, segments and some, but not all, phonetic cues that distinguish segments. Given Flege et al.'s (1995) finding that Japanese learners with 15 years of US residence produced target-like /ɹ/ more often than those with only 2 years of residence, a detailed examination of the acquisition of individual cues over extended time frames using both cross-sectional and longitudinal designs may be worthwhile. Such work may help resolve contradictions in research findings like that posed by Larson-Hall (2006), who was not able to replicate Flege et al.'s (1995) outcome.

Another potential research direction is further evaluation of how classroom instruction affects the rate of learning of English /ɹ/. Given that exposure to L2 for one year seems to be insufficient to trigger significant change in the development of F3, instruction might be effective, or even necessary, for the earlier development of robust perceptual representations of /ɹ/ that lead to target-like articulatory configurations across a wide range of tasks requiring both controlled and automatized processing (cf. Saito, 2014).

Finally, we acknowledge that, although we adopted an acoustic approach for the current study out of necessity, it would be interesting to collect perceptual data for comparison reasons. Specifically, identification and discrimination responses to native and non-native productions could also provide us with useful insights. However, such studies need to be carefully designed in terms of the selection of tokens and listeners. Although native speakers are attuned to the acoustic properties of English /ɹ/ (e.g., F3, F2, F1, transition duration), their sound and word recognition patterns are also subject to the influence of lexical factors, such as text frequency and neighbourhood density (Bradlow & Pisoni, 1999) and familiarity with accented speech (Kennedy & Trofimovich, 2008). Furthermore, it might be relatively difficult for human listeners to accurately evaluate the quality of particular L2 sounds and words embedded in spontaneous speech samples because their judgments would likely be affected by not only phonological but also lexical, morphosyntactic, and pragmatic use of language (Piske et al., 2001).

**References**


Baker, W. (2010). Effects of age and experience on the production of English word-final stops by Korean speakers. *Bilingualism: Language and Cognition, 13*, 263-278.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*, 231-250.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception. In O. Bohn, & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer*. Retrieved from http://www.praat.org.

Bohn, O.-S. (1995). Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379–410). Timonium, MD: York Press.

Bohn, O.-S. & Flege, J. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303-328.

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In: D. Singleton , & Z. Lengyel (eds.), *The age factor in second language acquisition.* Clevedon: Multilingual Matters.

Bradlow, A. R. (2008). Training non-native language sound patterns. In J. Hansen & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287-308). Philadelphia, PA: John Benjamins.

Bradlow, A., & Pisoni, D. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener- and item-related factors. *Journal of the Acoustical Society of America, 106*, 2074-2085.

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

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34*, 372-387.

Cobb, T. (2012). *The compleat lexical tutor*. Retrieved from http://www.lextutor.ca/vp/

Dalston, R. M. (1975). Acoustical characteristics of English/w, r, 1/ spoken correctly by young children and adults, *Journal of the Acoustical Society of America*, 57, 462-469.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-51). Cambridge, UK: Cambridge University Press.

DeKeyser, R. (2003). Implicit and explicit learning. In M. Long & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 313-348). Malden, MA: Blackwell.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J.F. Kroll & A.M.B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Derwing, T. M., Munro, M. J. (2013) The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning*, *63*, 163-185.

Derwing, T. M., Rossiter, M.J., Munro, M. J. & Thomson, R. I. (2004). L2 fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

Derwing T. M., Munro, M. J., & Thomson, R. I. (2008) A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics, 29,* 359-380.

Dohlus, K. (2008). *The role of phonology and phonetics in loanword adaptation: German and French front rounded vowels in Japanese.* Berlin: Peter Lang.

Espy-Wilson, C.Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *Journal of the Acoustical Society of America, 92,* 736-757.

Espy-Wilson, C.Y., Boyce, S.E., Jackson, M., Narayanan, S. & Alwan, A. (2000). Acoustic modeling of American English /ɹ/. *Journal of the Acoustical Society of America, 108,* 343-356.

Flege, J. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*, 70-79.

Flege, J. (1993). Production and perception of a novel second-language phonetic contrast. *Journal of the Acoustical Society of America*, 93, 1589-1608.

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

Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

Flege, J., Bohn, O-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.

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

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

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

Flege, J., Munro, M., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*, 3125–3134.

Flege, J., Munro, M., & Skelton, L. (1992). Production of the word-final English /t/-/d/ contrast by native speakers of English, Mandarin and Spanish. *Journal of the Acoustical Society of America*, *92*, 128-143.

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

Flege, J. & Wang, C. (1990). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, *17*, 299-315.

Gottfried, T. L. & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin tones, *Journal of Phonetics, 25,* 207–231.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /ɹ/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*, 3267-3278.

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate/r–l/on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15,* 255-274.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language acquisition. *Language Learning, 57*, 1-33.

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness*, *19*, 171-185.

Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science, 97,* 11850-11857.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Larson-Hall, J. (2006). What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /ɹ/ and /l/ by Japanese speakers. *Language and Speech, 49*, 521-548.

Lee, B., Guion, S.G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition 28*, 487-513.

Lin, Y. (2003). Interphonology variability: sociolinguistic factors affecting L2 simplification strategies. *Applied Linguistics, 24*, 439–464.

Lotto, A. J., Sato, M., & Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /ɹ/ and /l/. In J. Slifka, S. Manuel, & M. Matthies

(Eds.), *From sound to sense: 50+ years of discoveries in speech communication* (pp. C381-C386). Cambridge, MA: Research Laboratory of Electronics at MIT.

MacKain, K.S., Best, C.T. and Strange, W. (1981) Categorical perception of English /ɹ/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics, 2*, 369-390.

Major, R. (2001) *Foreign accent: The ontogeny and phylogeny of second language phonology*. Mahwah, NJ: Lawrence Erlbaum Associates.

McAllister, R., Flege, J., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics, 30*, 229-258.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition*, *21*, 81-108.

Munro, M. J. (1993). Production of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech, 36,* 39–66.

Munro, M. & Derwing, T. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning, 58,* 479-502.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. Journal of Speech, Language, and Hearing Research, 50, 1496-1509.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English, *Journal of the Acoustical Society of America, 97,* 1286-1296.

Piske, T., Flege, J., MacKay, & Meador, D. (2011). *I*nvestigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul & Dziubalska-Kołaczyk, K.

(Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 195-205). Switzerland: Peter Lang.

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*, 191–215.

Rau, D., Chang, A., & Tarone, E. (2009). Think or sink: Chinese learners' acquisition of the voiceless interdental fricative. *Language Learning*, *59*, 581-621.

Saito, K. (2013). Age effects on late bilingualism: The production development of /ɹ/ by high-proficiency Japanese learners of English. *Journal of Memory and Language*.

Saito, K. (2014). Communicative focus on L2 phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics, 34*.

Schroeder, M.R., Atal, B.S. & Hall, J.L. (1979). Optimizing digital speech coders by exploiting masking properties of the human ear. *Journal of the Acoustical Society of America*, 66, 1647-1652.

Segalowitz, N. (2003) Automaticity and second languages. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382-408). Oxford: Blackwell Publishers.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60,* 263-308.

Statistics Canada. (2008). *2006 Census of Canada topic based tabulations, ethnic origin and visible minorities tables: Ethnic origin, for population, for Canada, provinces and territories, 2006 census.* (Catalogue number 97-562-XWE2006002). Retrieved June 3, 2012 from Statistics Canada: http://www12.statcan.ca/census-recensement/2006/dp-pd/hlt/97-562/index.cfm?Lang=E

Trofimovich, P., & Baker, W. (2006). Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, *88*, 97-100.

Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of /ɹ/ and /l/ by Japanese learners of English. *Journal of the Acoustical Society of America, 84*, 90–100.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds: Perception of American English /p/ and /l/ by native speakers of Japanese. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305–320). Timonium, MD: York Press.

Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24*, 245-261.

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

## Appendix

### Pictures used in Timed Picture Description Task (with Target Items Italicized)

- Participants described a dog lying on the floor which is reading a comic book by using three key words (dog, *read,* lying).
- Participants described a table left on the drive way in the heavy rain" by using three key words (*rain,* table, drive way).
- Participants described "three guys who were playing rock music with one guy singing a song and the other two guys playing guitars" by using three key words (three guys, guitar, *rock music*).
- Participants described "a long stretched road under the blue sky with a lot of clouds" (blue sky, *road,* cloud).

### Sentences used in Sentence Reading Task (with Target Items Italicized)

1. My dogs eat two times each day.
2. He will *read* my proposal by the time I arrive there.
3. She left her *red* bicycle on the side of the *road*.
4. He takes a bus to go to a hospital.
5. The *race* was cancelled because of the *rain*.
6. I can correct all *wrong* sentences tonight.
7. I like these apples so much.
8. *Ryan* does not like to *run* in the snow.

### Words used in Word Reading Task (with Target Items Italicized)

1. man
2. book
3. desk
4. tall
5. *room*
6. bus
7. *red*
8. music
9. *root*
10. tom
11. japan
12. *rough*
13. ship
14. chair
15. sea
16. map
17. man
18. *ram*
19. sip
20. *race*
21. *read*
22. subway
23. *right*
24. she
25. yellow

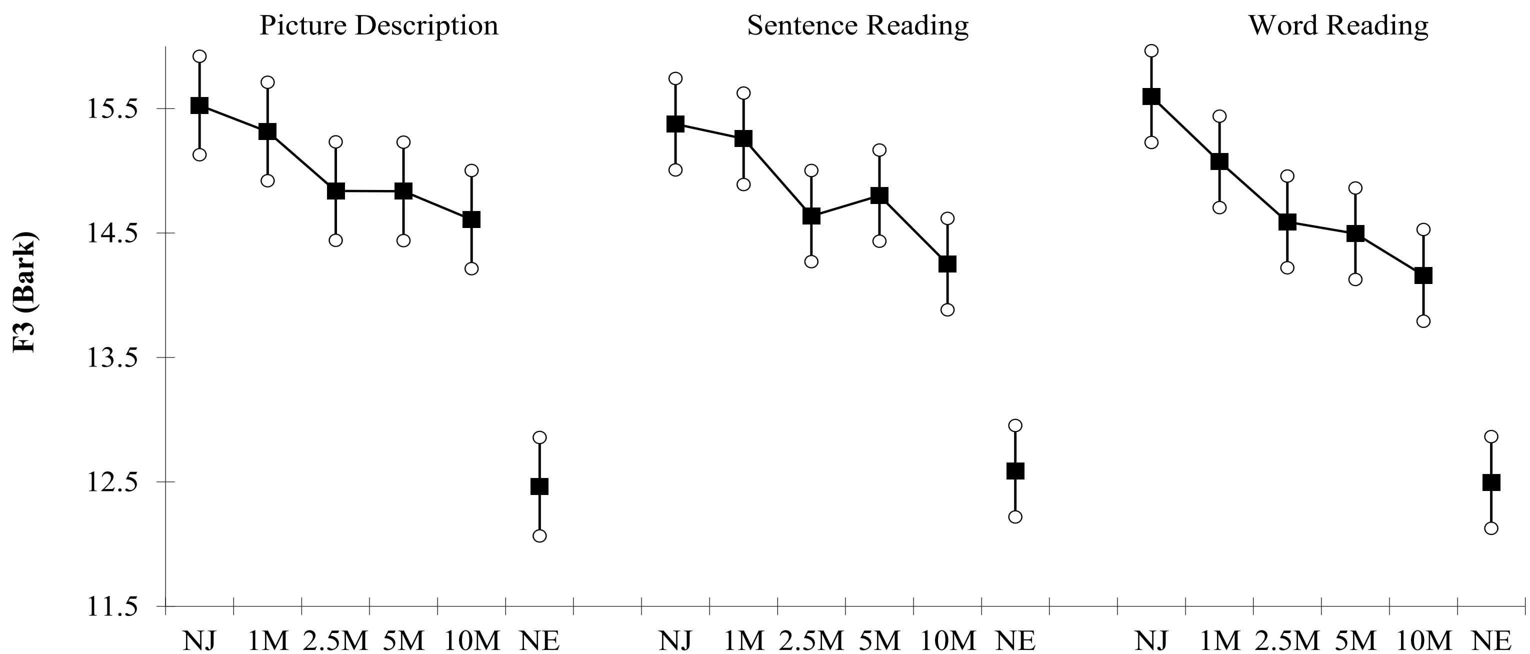


*Figure 1*. Mean normalized F3 (Bark) for the 6 talker groups on the 3 tasks. 95% confidence intervals are shown.

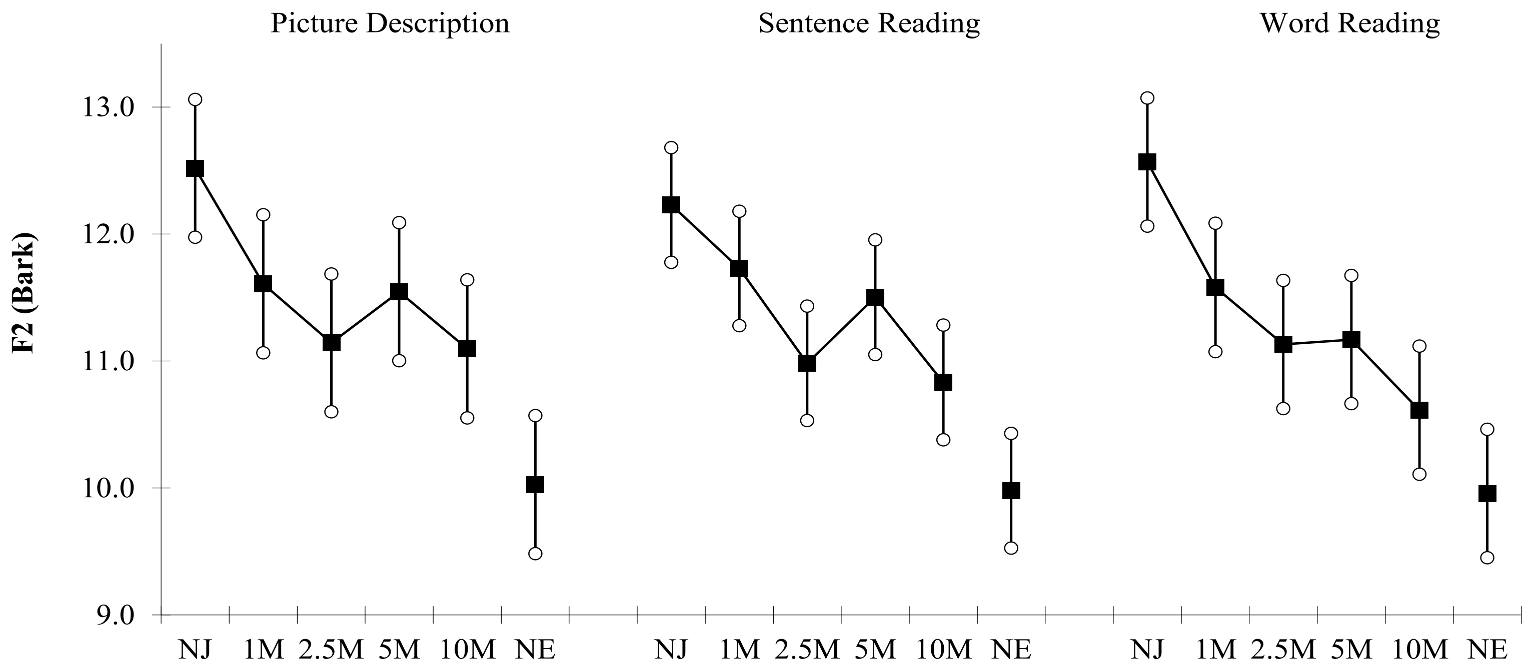


*Figure 2*. Mean normalized F2 (Bark) for the 6 talker groups on the 3 tasks. 95% confidence intervals are shown.

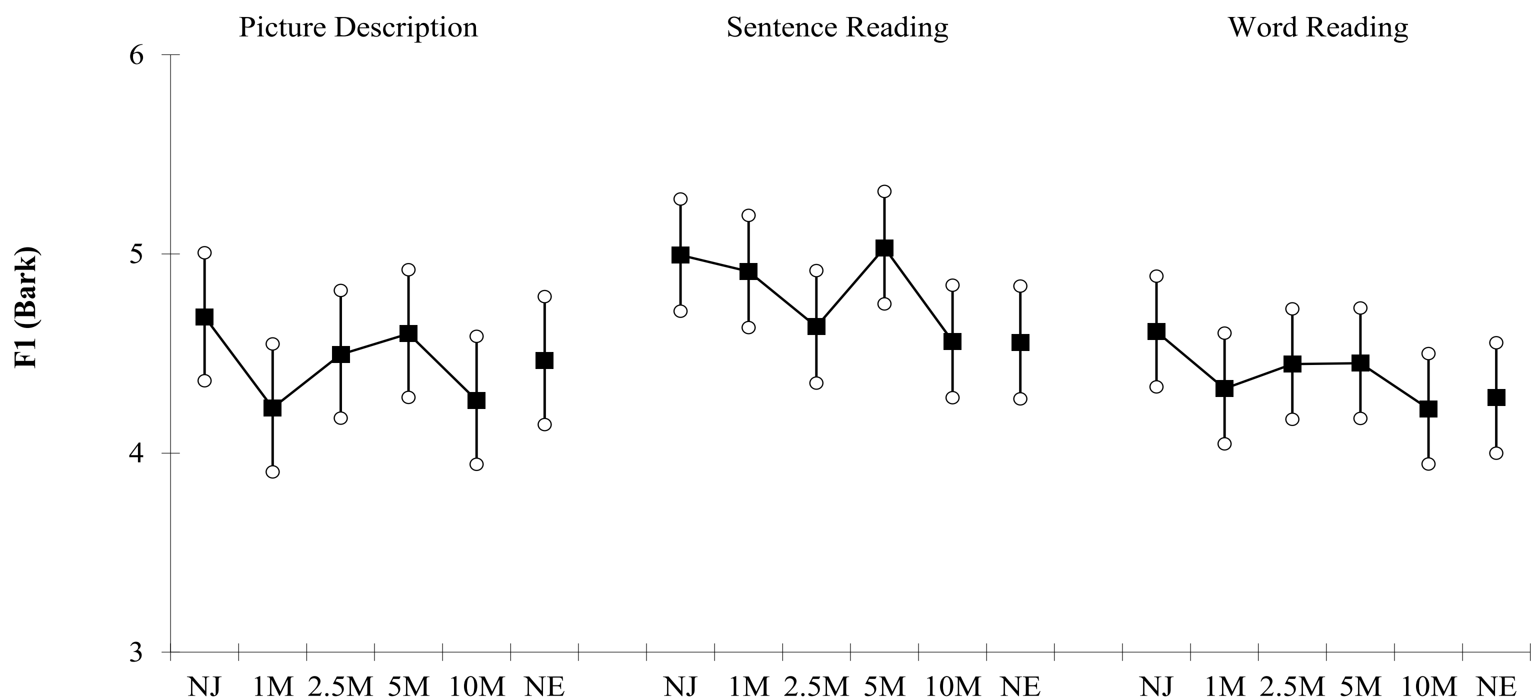


*Figure 3*. Mean normalized F1 (Bark) for the 6 talker groups on the 3 tasks. 95% confidence intervals are shown.

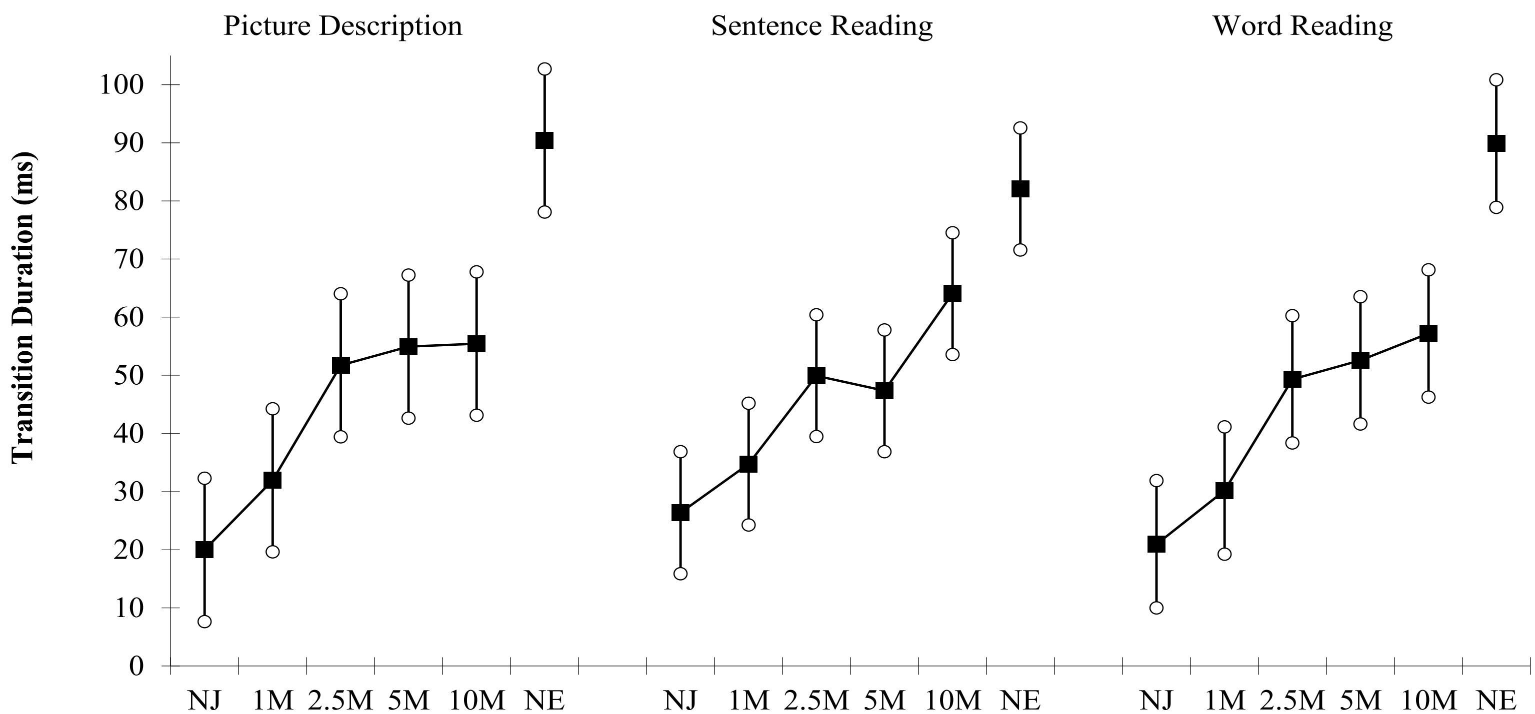


*Figure 4*. Mean transition durations (ms) for the 6 talker groups on the 3 tasks. 95% confidence intervals are shown.