Journal of Memory and Language 69 (2013) 546–562

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Age effects on late bilingualism: The production development of /ɹ/ by high-proficiency Japanese learners of English

Kazuya Saito *

Waseda University, School of Commerce, 1-6-1 Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan

ARTICLE INFO

Article history:
Received 1 February 2013
revision received 16 July 2013
Available online 12 August 2013

Keywords:
Second language phonetics
Age
Late bilingualism
Ultimate attainment
English /ɹ/

ABSTRACT

The current project examined whether and to what degree age of acquisition (AOA), defined as the first intensive exposure to the target language, can be predictive of second language production attainment and nativelikeness of word-initial /ɹ/ by late English–Japanese bilinguals. Productions of /ɹ/ were elicited from 88 high-proficiency Japanese learners of English and comparison groups of 10 native English speakers and 10 low-proficiency Japanese learners of English. Tokens from word reading, sentence reading, and timed picture description tasks were assessed through listener judgements and acoustic analyses. The results demonstrated that AOA significantly predicted the attained performance of /ɹ/ at a spontaneous (picture description) but not a controlled (word and sentence reading) speech level, and with respect to third formant frequencies as determined by labial, palatal, and pharyngeal constrictions. In contrast, most Japanese learners exhibited ceiling effects regardless of AOA profiles with respect to second formant frequencies and transitional duration of first formants as determined by the degree and rate of tongue retraction. The results suggest that, whereas AOA continues to be a driving factor in the degree to which late bilinguals can benefit from additional input and interaction, such age effects may depend on different levels of phonetic processing.

Introduction

Over the past 40 years, second language acquisition (SLA) researchers have extensively examined the role of age of acquisition (AOA), defined as the age at which a learner's first period of intensive exposure to the target language begins, in the ultimate attainment and nativelikeness of second language (L2) pronunciation abilities. Whereas researchers have generally agreed on a negative correlation between AOA and the end state of L2 pronunciation proficiency, as is the case with early bilinguals who arrive in an L2 country before puberty, there is far less consensus on whether and to what degree AOA is predictive of L2 pronunciation attainment in the case of late bilinguals whose intensive exposure to the L2 begins after puberty (e.g., Birdsong, 2005 vs. DeKeyser & Larson-Hall, 2005). To examine this topic further, the current project aims to examine age effects on late bilingualism in the context of the production development of word-initial /ɹ/ by late Japanese learners of English (AOA ⩾ 16 years).[1]

Background literature

Age effects and late bilingualism

In the field of L2 speech acquisition, several theories have been proposed to explain how the first language

[1] In this paper, participants labeled as "bilinguals" are those who have already reached the upper limits of SLA (cf. rate of learning), and thus several key terminologies—plateaued and asymptotic L2 performance, end and final state of L2, ultimate attainment, and bilingualism—will be used interchangeably throughout the manuscript, following research conventions in age-related SLA literature.

* Fax: +81 03 3203 7067.
E-mail address: kazuya.saito@waseda.jp

0749-596X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2013.07.003

(L1) phonetic system interferes with adult second language acquisition (SLA) processes, especially at the *initial* and *mid* phases of L2 learning (e.g., Best, McRoberts, & Goodell, 2001 for Perceptual Assimilation Model [PAM]; Flege, 1995 for Speech Learning Model [SLM]). Scholars in this vein of L2 phonetics research have also worked to identify the factors that co-interact to mediate such L1 influence and thus *ultimately* lead certain learners to nativelike (or near-nativelike) proficiency at the *end* state of L2 production development, which occurs after several years of exposure to L2 input and interaction with native speakers. The variables that have been examined include (a) attitude and aptitude (e.g., Ioup, Boustagi, El Tigi, & Moselle, 1994), (b) motivation (e.g., Bongaerts, Van Summeren, Planken, & Schils, 1997), (c) level of education (e.g., Derwing & Munro, 2005), and (d) ethnic identity (e.g., Gatbonton & Trofimovich, 2008).

Among these variables, researchers have paid by far the most attention to learners' AOA as a relatively strong predictor of the end state of SLA: The earlier they arrive, the better their ultimate L2 performance tends to be, and the more likely it is to fall within a nativelike range. Such claims have been tested by many relevant studies which have pooled bilinguals with various AOA profiles (e.g., 0 year < AOA < 40 years), and have shown evidence for age effects, especially for early bilinguals (AOA < 16 years) in the domain of L2 phonology (Flege, Munro, & MacKay, 1995; Flege, Yeni-Komshian, & Liu, 1999; Patkowski, 1990) and L2 morphosyntax (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; Johnson & Newport, 1989). With respect to late bilinguals (AOA > 16 years) (the focus of this paper), however, previous studies have generated much disagreement, which has resulted in a great deal of theoretical discussion in regards to the nature of the SLA processes underlying early and late bilingualism. In what follows, I will review two competing theoretical accounts and their different predictions as to age effects on post-pubertal ultimate attainment and nativelikness: the Critical Period Hypothesis (CPH) and the Cognitive Aging Hypothesis (CAH).

Critical period hypothesis

Certain researchers (Abrahamsson, 2012; DeKeyser, 2000; DeKeyser & Larson-Hall, 2005; Granena & Long, 2013; Patkowski, 1990; Scovel, 2000) hold the view that any linguistic performance by late bilinguals is constrained by a loss of plasticity resulting from neural maturation after adolescence. Thus, these learners' acquisitional processes are fundamentally and qualitatively *different* from those of early bilinguals, who learn L2 automatically through mere exposure to natural input (very similarly to L1 acquisition processes). According to this theoretical position, bilinguals' access to an assumed language-specific implicit learning mechanism seems to gradually decline from early childhood (i.e., a robust AOA effect on early bilingualism), and then disappears after the mid teens (i.e., a discontinuous AOA effect on late bilingualism). In this regard, the Critical Period is defined as "the concept of an endpoint, a point beyond which learning becomes difficult or impossible" (DeKeyser & Larson-Hall, 2005, p. 97).

As a result, post-critical period SLA relies on general cognition learning processes that are intentional and explicit rather than language-specific cognition processes that are incidental and implicit (see Abrahamsson, 2012, p. 189). In the cognitive psychology literature, general skill learning (e.g., the learning of algebra, geometry, and computer programming) is characterized as the gradual proceduralization of declarative knowledge through practice and feedback, and its improvement (i.e., a decline in error rate and reaction time) follows the power law (e.g., Anderson, 1993). From a neurolinguistic perspective on SLA, mature L2 learners tend to learn an L2 in an effortful and conscious manner, drawing on metalingustic knowledge declaratively represented in the left temporal area. This is assumed to compensate for the lack of their procedural memory sustained in Broca's area and the basal ganglia, which are mainly responsible for implicit and automatic L1 acquisition (Paradis, 2009; Ullman, 2004).

Previous researchers have indeed noted a general tendency for adult L2 learners to demonstrate quick improvement over the first few months of residence (LOR), and then proceed to level off, despite additional practice and environmental input (for a review, see DeKeyser & Larson-Hall, 2005). Unlike early bilingualism, which is strongly tied to AOA, final state quality and the possibility of attaining nativelike proficiency in late bilingualism is assumed to be related to learners' individual differences, such as language learning aptitude, regardless of learners' AOA profiles. In L2 morphosyntax development, DeKeyser found that a significant predictor for near-nativelike performance of oral grammaticality judgment tests by late Hungarian learners of English was not their AOA, but their high analytical aptitude scores (see also Abrahamsson & Hyltenstam, 2008).

Cognitive aging hypothesis

In contrast to the CPH, other researchers argue that language learning capacity used in successful L1 speech acquisition remains active even after puberty and can be applied to late bilingualism (Bialystok, 1997; Flege, 1995). According to this position, the generally more salient foreign accents of older learners, compared to younger learners, could be ascribed mainly to L1 influence and the differential quality and quantity of environmental input. That is, not only does post-pubertal L2 speech learning take place in a common phonological space partially or fully organized by L1 restrictions (Best et al., 2001; Flege, 1995), but late learners also tend to have less, poorer quality interaction with native speakers than early learners (Jia & Aaronson, 2003). This suggests that adult L2 learners are able to continue to learn new sounds as long as they can intentionally or incidentally access some of the social and educational environments that bilingual children likely experience (Bialystok, 1997), such as those where the L2 can be used with native speakers on a daily basis (Flege & Liu, 2001).

Whereas late SLA is free of maturational constraints, another characteristic of this position is that AOA may continue to predict the quality of ultimate attainment and the incidence of nativelikeness over one's life span without a cutoff point. According to major L2 speech learning

theories, it is posited that the L1 and L2 systems co-exist in the same phonological space, and that thus the ongoing development of the L1 continues to affect the quality of the L2 system and vice versa. In this regard, this theoretical perspective clearly predicts age effects on adult SLA which are attributed to the mutual interaction between the L1–L2 systems: The earlier late learners arrive, the more they use the L2, and the less they hinge on the L1 (Baker, Trofimovich, Flege, Mack, & Halter, 2008; Flege, Frieda, & Nozawa, 1997).

Other researchers have claimed that the age effects on the upper limit of SLA are strongly tied to the notion of cognitive aging, such as a gradual decline in working memory, executive control, speech sound processing, or the inhibition of task-irrelevant information (Hakuta, Bialystok, & Wiley, 2003). Specifically, Birdsong (2005, 2006, 2007) linked the progressive loss of cognitive functions to the biological (but not maturational) aging process in the brain system, such as decreases in brain volume and nigrostriatal dopamine (starting at 20 years of age). For the SLA context, Birdsong (2006) maintains that the dopamine system plays a key role in "defossilization, an undoing of automatized nontargetlike linguistic performance" as well as in "suppressing and supplanting L1 routines" (p. 32).

Previous findings

The admittedly simplified theoretical constructs—CPH and CAH—provide sharply contrastive predictions in regard to age effects on ultimate attainment and nativelikeness in late bilingualism. The CPH predicts "discontinuity in the AOA-proficiency" (DeKeyser & Larson-Hall, p. 97). The CAH would suggest the notion of cognitive aging, that is, "a linear monotonic decline of learning over the [AOA] spectrum, with age effects continuing past the point at which maturation has ceased" (Birdsong, 2005, p. 115). Several studies have investigated the relationship between AOA and the ultimate attainment of late bilinguals' global pronunciation abilities; however, these have generally revealed mixed research findings.

On the one hand, Patkowski (1990) investigated how native speakers judged the degree of foreign accentedness of 34 highly educated ESL speakers' extemporaneous L2 production samples (AOA > 15 years). The results did not show a significant negative correlation between global accent rating scores and AOA nor any participants who could be judged as nativelike. Patkowski interpreted these findings as support for the existence of crucial discontinuities between child and adult SLA due to the passing of a critical period (see also Granena & Long, 2013).

On the other hand, Flege et al. (2006) examined foreign accentedness in the sentence production of 36 late Korean-English bilinguals (AOA > 20 years). A significant negative correlation between their accent ratings and AOA was found and interpreted as counterevidence to the CPH. Yet, they did not find any late learners who fell within the range of native speaker controls (cf. Bongaerts et al., 1997). Derwing and Munro (2013) conducted a longitudinal investigation on how 22 ESL learners improved foreign accentedness over the course of their first 7 years of residence in Canada. The results showed that their ultimate ratings were significantly correlated with AOA, indicating that the age factor continues to influence L2 pronunciation learning even during adulthood.

Notably, these conflicting results could be due to the artifacts of global foreign accentedness measures which inevitably take into account various linguistic factors, not only correct phonological skills (e.g., segmentals, suprasegmentals, syllable structures, and speech rate) but also the accurate use of lexicogrammar in L2 (see Piske, MacKay, & Flege, 2001, p. 194). For this methodological reason, therefore, these previous studies have yet to provide a detailed picture of how the age variable of late bilinguals influences the ultimate attainment of specific areas of L2 speech production learning. An exception is Birdsong (2007)'s study which examined how 22 late English learners (AOA $\geqslant$ 18 years) produced French vowel length, and also their voice onset time in stops in a word list. The results of acoustic analysis showed that two participants performed both features within the range of native speaker controls. Due to the small number of participants (n = 22), however, the author did not quantitatively investigate the relationship between their attained segmental performance and AOA. The current project is designed to extend this vein of age-related SLA research by focusing on a very different context of L2 segmental learning with a relatively large number of participants—the production development of word-initial English /ɹ/ by 88 late Japanese bilinguals—via a range of elicitation (i.e., controlled vs. spontaneous production tasks) and assessment (i.e., listener judgement and acoustic analysis) methods to scrutinize the multifaceted relationship between AOA and late bilingualism from various perspectives.

English /ɹ/

From an acoustic point of view, American English /ɹ/ has been traditionally described along multiple spectral and temporal dimensions, such as (a) third formant (F3) (1600–1900 Hz), (b) second formant (F2) (1700–2100 Hz), (c) first formant (F1) (250–550 Hz) and (d) transitional duration of F1 and F3 (50–100 ms) (Espy-Wilson, 1992; Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000; Flege, Takagi, & Mann, 1995; Hattori & Iverson, 2009). Compared to the other approximant sounds in the English phonetic system, /ɹ/ can be characterized by, in particular, a severe dip in F3 as its reliable acoustic correlate (Espy-Wilson, 1992). From an articulatory perspective, native speakers of English employ a range of tongue configurations to produce the sound, such as bunched /ɹ/ (i.e., a raised dorsum and lowered tip) and retroflexed /ɹ/ (i.e., a raised tongue tip and lowered dorsum), even within the speech of the same speaker (Delattre & Freeman, 1968). Despite variability in articulator positions, a shared articulatory parameter includes the constrictions in the labial, palatal, and pharyngeal regions of the vocal tract, which in turn create an anterior oral cavity that includes the sublingual space—a source of F3 lowering and the perceived feature of rhotocization (Espy-Wilson et al., 2000).

Because the approximant category in the Japanese phonetic system is constituted by /j/ and /w/, inexperienced Japanese listeners tend to substitute several L1

counterparts (mostly the Japanese tap) for English /ɹ/ (as well as for English /l/) both in perception (Guion, Flege, Akahane-Yamada, & Pruitt, 2000) and production (Riney, Takada, & Ota, 2000). In the analysis of speech sound perception and production of Japanese and English, however, previous cross-linguistic research has demonstrated that the Japanese tap substantially differed from English /ɹ/ in having higher F3 (2400–3000 Hz) and F2 (1700–2100 Hz) with shorter transition duration (5–20 ms) (Hattori & Iverson, 2009; Lotto, Sato, & Diehl, 2004). According to the major L2 speech theories (e.g., PAM, SLM), the acquisition of English /ɹ/ by adult Japanese learners is considered to be one of the most difficult specific instances of L2 speech learning. This in turn has generated a great deal of theoretical research on this instance as "a productive testing ground for general principles of learning and claims about adult neural plasticity" (Bradlow, 2008, p. 294). For example, Flege, Takagi, et al. (1995) showed that experienced Japanese learners (LOR > 20 years) produced the English /ɹ/-/l/ contrast in a more targetlike manner than inexperienced learners (LOR < 2 years) did. Yet, Larson-Hall (2006) failed to replicate Flege, Takagi, et al.'s (1995) findings, showing little difference between experienced and inexperienced Japanese learners.

The disagreement evident in these precursor studies calls for a more robust analysis of the topic by scrutinizing the complexities inherent in the nature of the interlanguage development of /ɹ/. First, the development of a new sound requires the mastery of more than one phonetic segment at various processing levels. That is, to move away from the substitution of the L1 counterpart (i.e., the Japanese tap) and towards the acquisition of word-initial /ɹ/, Japanese learners need to approximate both spectral and temporal domains of the L2 phonetic category by increasing their awareness towards not only primary acoustic parameters (F3) but also secondary cues (F2, transition duration). Another important aspect in measuring the interlanguage development of /ɹ/ is to take into account variation according to speaking contexts: Adult L2 learners tend to make more pronunciation errors in free speech tasks than in formal word reading tasks (Dickerson & Dickerson, 1977; Rau, Chang, & Tarone, 2009).

To scrutinize the relationship between AOA and different learning trajectories according to multiple cue weightings and task conditions, I will first elicit Japanese learners' /ɹ/ production by employing three oral tasks: (a) Word Reading (WR; i.e., reading a list of target words), (b) Sentence Reading (SR; i.e., reading sentences including target words), and (c) Timed Picture Description (TPD; i.e., using target words to describe a series of pictures). In addition, I will not only examine the perceived accentedness of Japanese learners' production of /ɹ/ via native speaker listener judgements, but also conduct a phonetic scrutiny on their cue weightings through acoustic analysis on four acoustic domains of /ɹ/—F3, F2, F1, and transition duration. In the latter analysis, our assumption is that examining the acoustic properties of participants' /ɹ/ productions could be a good index of their use of articulatory configurations: F1 for tongue height; F2 for tongue retraction; F3 for labial, palatal, and pharyngeal constrictions; and transition duration for phonemic length (Espy-Wilson et al., 2000).

Method

Participants

Participants were 88 Japanese learners of English, 20 Japanese and English controls, and 10 native English-speaking listeners.

Japanese learners of English

Data collection took place in both Montreal and Vancouver, Canada. The Japanese immigrant population is relatively low in both cities (e.g., 0.06% in Quebec and 1% in British Columbia) (Statistics Canada, 2008). The project was advertised on regional community websites and local newspapers. In order to recruit late English–Japanese bilinguals who had already reached their plateau (i.e., little room for further L2 development), two necessary conditions were specified: age of arrival in Canada greater than 16; and 6 years of LOR (for similar definitions of late bilinguals, see Birdsong, 2007; Johnson & Newport, 1989). As a result, 108 participants (31 from Montreal; 77 from Vancouver) were originally recruited.

To further narrow down the scope of the participants who used L2 on a daily basis with sufficient opportunities to practice L2, they were screened additionally based on their language background questionnaires: (a) Their self-reported use of English was above "4" (on a 6 point scale: 1. Very infrequent – 6. Very frequent) ($M = 5.4$); and (b) their primary language of communication either at home or work was English.[2] Eighty-eight Japanese learners were selected as *high-proficiency* late bilinguals for subsequent analysis in the current study (13 males and 75 females).

At the time of testing, their age ranged from 30 to 70 years ($M = 45.9$ years). They had arrived in Canada between the ages of 16 and 40 years ($M = 26.1$ years) and had lived in Canada for 6–42 years ($M = 17.8$ years). They reported 6–9 years of English learning experience (typically through grammar translation methods) in secondary school settings in Japan prior to their arrival in Canada.[3] Although most of the participants had little knowledge of French, 12 participants (8 from Montreal, 4 from Vancouver) reported limited exposure to this other official language of Canada.

Japanese and English controls

Data were collected from two control groups to establish baseline acoustic characterizations for Japanese and native English speakers' production of /ɹ/. To establish the Japanese Control, 10 native speakers of Japanese (2 males and 8 females) who had just arrived in Canada with little L2 experience (LOR < 1 month) were recruited at private language schools in downtown Montreal. They com-

[2] Eighteen Japanese participants self-rated their frequency of English use below "3" because they mainly used Japanese ($n = 10$) (e.g., their family members were Japanese and they did not work outside) and French ($n = 8$) (their business involved French-speaking customers or their partners were native speakers of French).

[3] Among the original data pool of 108 Japanese learners, two participants reported their intensive English learning experience in immersion programs in Japan. Both of them were eliminated from the final analysis because their precise AOA profile was difficult to determine.

pleted the three oral tasks (WR, SR, TPD) and their data served as a baseline for the initial state of Japanese speakers' /ɹ/ production (mean age: 27.9 years). As for the English Control, 10 native English undergraduate students at an English-speaking university in Montreal (5 males and 5 females) completed the three oral tasks. All participants were native speakers of north-eastern Canadian and American English with a mean age of 25.1 years.

Native speaking listeners

To judge the perceived intelligibility, accuracy, and goodness of English /ɹ/ production by Japanese and English speakers, 10 native speakers of English were recruited from undergraduate linguistics and psychology courses at a Canadian university. All participants were native speakers of western Canadian or American English with a mean age of 23.8 years. They reported having little frequent contact with Japanese learners of English and being unfamiliar with Japanese-accented English speech. All passed a pure-tone screening at octave frequencies between 250 and 4000 Hz.

Outcome measures

Target words

All 20 words used in the three oral tasks were Consonant–Vowel–Consonant (CVC) word-initial /ɹ/ singletons, except for the token "*Ryan*" (CVVC). These words were selected, taking into account lexical familiarity and text frequency, and phonetic contexts. First, according to the results of vocabulary profiling (Cobb, 2012), all of them fall within the first 2000 most frequent words, except "*ram*" and "*Ryan*." None of the participants reported unfamiliarity with these two words, probably because Japanese adopts many English lexical items as loan words (i.e., Katakana). Thus, the effects of lexical frequency and familiarity factors on Japanese learners' /ɹ/ production were assumed to be minimal in the current study. Another factor is co-articulation effects. Given that adult Japanese learners tend to have difficulty in producing /ɹ/ especially before front vowels (/i/, /e/) (Flege, Takagi, et al., 1995), the following vowels were evenly distributed in each task: 50% for singletons with front vowels, 50% for singletons with central and back vowels. The test tokens are summarized in Table 1.

Task description

Three tasks were designed to elicit participants' production of /ɹ/ at a spontaneous (TPD) and a controlled (SR, WR) level.

TPD

With the view of elaborating and validating outcome measures to assess learners' spontaneous use of certain linguistic structures, the SLA literature has emphasized the importance of eliciting learners' performance in a *communicative context* (i.e., they are required to pay simultaneous attention to grammatical, phonological, lexical, and pragmatic aspects of language to convey their intended message) (Spada & Tomita, 2010) and within a realistic *time limit* (i.e., they are not given much planning time to access explicit knowledge stored in general memory) (Ellis, 2005). One example of such spontaneous measures is a picture description task (e.g., Rau et al., 2009).

Table 1
20 Tokens in the controlled and spontaneous production tests in relation to following vowel conditions.

A. Timed picture description	
Following vowels	
[front]	**read**, **rain**
[central/back]	**road**, **rock**
B. Sentence reading	
Following vowels	
[front]	read, rain, red, race
[central/back]	run, Ryan, road, wrong
C. Word reading	
Following vowels	
[front]	**read**, red, **race**, ram
[central/back]	rough, **right**, root, **room**

Note: **Words in bold** were used for both acoustic analysis and listener judgment.

In the current study, participants described a picture by using three key words (one of which was always a target word) after 5 s of planning time. For example, they were given three word cues—"road" "blue sky" and "cloud"—to describe a picture of a road going off into the distance under blue sky with a cloud. There were four target pictures with four distracter pictures in total. The four target words were *read, rain, road,* and*rock.* To familiarize learners with the task procedure, four distracter pictures were first randomly presented; the other four pictures including target words were then randomly presented to elicit their spontaneous production of /ɹ/.

In order to prevent participants from paying too much attention to the target sound, efforts were made as follows. First, participants remained uninformed about the true intention of the entire project until their completion of all tasks (the project was advertised to investigate general oral skills of Japanese learners of English). Second, the participants took the TPD before the SR and WR in which target forms became rather salient (see below). Last, four (instead of eight) words were used in this task to minimize the number of learners' encounters with /ɹ/-words.

SR

In this task, participants read five target sentences, together with three distracter sentences[4]:

1. He will read my paper by the time I arrive there.
2. She left her red bicycle on the side of the road.
3. The race was cancelled because of the rain.
4. I can correct all wrong sentences tonight.
5. Ryan does not like to run in the snow.

In contrast to the spontaneous production task (i.e., TPD), the sentence reading task allowed participants to focus only on reading sentences accurately without much communicative pressure.

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

WR

In this task, participants read a list of 25 words. consisting of eight target words (i.e., *read, red, race, ram, rough, right, root,* and *room*) and 17 distracters including a number of easy and difficult English sounds (e.g., voiceless stops, interdental fricatives). Due to the highly formal nature of the task, participants were assumed to pay more conscious attention to pronouncing each of these word forms in a correct way, and demonstrate their carefully-monitored production of /ɹ/, possibly drawing on their explicit articulatory knowledge, if any.

Procedure

All participants (88 Japanese learners, 20 Japanese and English controls) first completed all three tasks in the following order: (a) TPD, (b) SR, and (c) WR. After the recordings, they filled in a language background questionnaire and then moved onto a personal interview with the researcher. Individual recordings were conducted in a quiet room at university labs, community centers, and participants' homes. Speech tokens were recorded with a Roland-05 audio recorder, set at 44.1 kHz sampling rate and 16-bit quantization, and a unidirectional condenser microphone. In total, 2160 tokens were recorded (108 participants × 20 tokens [$n = 4$ in TPD, 8 in SR, and 8 in WR]).

Acoustic analysis

In accordance with the acoustic analysis performed for natural speech tokens of word-initial /ɹ/ established by Flege, Takagi, et al. (1995), formants were measured through linear predictive coding spectra by means of *Praat* (Boersma & Weenik, 2012). The beginning of word-initial English /ɹ/ was first identified using both the spectrographic representations and wave forms of the speech tokens. The cursor was put at the point where the energy for all three formants became visible in order to measure F3, F2 and F1 in Hz. As for /ɹ/ tokens in continuous speech (TPD, SR), the severe dip in F3 (local peak) was identified, given that English /ɹ/ exhibits relatively low F3 values compared to other vowel and consonant sounds in the English phonetic system. Because spectral information (i.e., F3, F2, F1 values) varies considerably due to anatomical differences in individual vocal tract length (especially in connection with gender), all raw acoustic values were adjusted by the talker as follows (for a more detailed account, see Lee, Guion, & Harada, 2006; Yang, 1996).

Given that F3 of open vowels (F1 > 600 Hz) is a reliable indicator of one's vocal tract length (e.g., the lower F3 indicate the longer vocal tract) (Yang, 1996), a mean F3 value of /æ/ elicited from 10 monosyllabic words in the word reading task (i.e., *man*, *map, ram*) was calculated for each speaker. Next, one female native speaker of English was randomly selected as a reference, and her mean F3 value (3011 Hz) was divided by those of the other participants to derive their own *k* factors. Then, all formant values (F3, F2, F1) of /ɹ/ for each participant were multiplied by the individual *k* factor. Finally, all acoustic values in Hertz were converted into Bark (Boersma & Weenik, 2012; Schroeder, Atal, & Hall, 1979) in order to reduce the nonlinear relationship between the formant frequencies and the corresponding perceived semivowel quality.[5]

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

Temporal aspects of English /ɹ/ were analyzed based on two types of transition duration. Raw transition duration was measured by dragging a cursor from the beginning point of F1 transition to the endpoint of the F1 or F3 transition (Hattori & Iverson, 2009). To take into account the possibility of a significant difference in the speech rate among participants, normalized transition duration—the ratio of transition duration per syllable—was also calculated by dividing raw transition duration by the syllable length. "*Ryan*" (CVVC) in SR was excluded from the normalized transition duration analysis, due to its relatively long syllable length compared to the other CVC singletons.

Listener judgements

Stimulus preparation

Given the potential for listener fatigue due to the large number of original /ɹ/ tokens ($N = 2160$), listener judgements were collected for only a subset of the recorded words: *read, race, right, room* from the WR and TPD conditions only. These target words were representative of the entire controlled and spontaneous production samples of /ɹ/ according to following vowel conditions (50% for singletons with front vowels, 50% for singletons with central and back vowels). The test tokens are reproduced in bold in Table 1.

The 864 /ɹ/ tokens (8 tokens × 108 participants) were normalized for peak amplitude, and saved as Wave files. In terms of the words embedded in continuous speech streams (i.e., TPD), the researcher carefully listened to the speech samples multiple times in order to put a cursor on the onset of the word (where any component of /ɹ/ could be heard), and move towards its offset by 5 ms steps. To avoid significant distortion of extracted words, inflected endings (rain<u>ed</u>, read<u>ing</u>) were included if they sounded more natural.

Procedure

The listening sessions were conducted individually in a quiet room at the Canadian university in Vancouver. Stimuli were randomized and presented to the listeners via a laptop computer. Listeners were told that all tokens were produced as /ɹ/ either by Japanese learners of English with various LOR profiles or by native speakers of English. They were instructed to base their judgments as much as they could on only the /ɹ/ component—instead of the entire words—because their ratings would have otherwise been

[5] The validity of the proposed normalization procedure was tested and confirmed with the English baseline dataset which noted pre-existing significant differences in formant frequency values between genders (five females vs. five males). Originally, a set of independent-samples *t*-tests revealed significantly higher acoustic values for females than males in F3 ($p = .004$), F2 ($p = .012$), and F1 ($p = .004$). This gender difference became non-significant after the normalization ($p = .200$–$.600$).

influenced by other pronunciation errors typical of Japanese learners, such as vowels (e.g., /ɹId / for /ɹid/), and suprasegmentals (e.g., slow speech rate). A "repeat" button was available to allow the listeners to hear an item up to three times before making a judgment.

Nine-point descriptors were adopted and modified from Flege, Takagi, et al.'s (1995) 6-point scale, and the listeners judged the quality of word-initial /ɹ/ by choosing one of the response alternatives: 1 (*Nativelike /ɹ/*) → 2 (*good /ɹ/*) → 3 (*probably /ɹ/*) → 4 (*possibly /ɹ/*) → 5 (*neutral exemplars, neither /ɹ/ nor /l/*) → 6 (*possibly /l/*) → 7 (*probably /l/*) → 8 (*good /l/*) → 9 (*Nativelike /l/*). The 9-point descriptors were assumed to correspond to three stages of interlanguage pronunciation development of /ɹ/ by Japanese learners as follows:

1. Japanese learners without much awareness of English /ɹ/ tend to substitute the Japanese tap (Riney et al., 2000) whose acoustic properties are substantially similar to English /l/ (Hattori & Iverson, 2009) and thus most likely judged as English /l/ by native speakers of English (i.e., Descriptors: "6"-"9") (Sekiyama & Tohkura, 1993).
2. When Japanese learners make some efforts to produce English /ɹ/ instead of the Japanese tap, they are reported to use various interlanguage strategies (e.g., retracting tongue body, lengthening phonemic length), and the acoustic properties of their /ɹ/ production likely stretch between English /ɹ/ and the Japanese tap (Lotto et al., 2004).[6] In this case, exemplars could be perceived as neither /ɹ/ nor /l/ (i.e., Descriptor "5" *neutral exemplars*).
3. Whereas many Japanese learners acquire English /ɹ/ with different degree of foreign accentedness (i.e., Descriptors: "2"-"4"), very few Japanese learners could fall into a range of native speakers of English (i.e., Descriptor "1" *Nativelike /ɹ/*).

After rating five familiarization stimuli not included in the subsequent listening materials, the listeners judged the stimuli in three equal blocks with 5-min breaks between blocks. The entire listening session took approximately 90 min.

Results: acoustic analysis

Baseline

To establish baseline /ɹ/ performance in acoustic terms, data from the Japanese beginners and native English speakers was compared (see Fig. 1).

The five acoustic components of /ɹ/ were separately submitted to two-way ANOVAs with one between-group factor (Japanese, English) and one repeated measure (Task: WR, SR, TPD). Main effects of Group were found for F3, $F(1,18) = 133.58$, $p < .001$, $d = 4.68$; F2, $F(1,18) = 64.854$, $p < .001$, $d = 3.26$; transition duration, $F(1,18) = 261.25$, $p < .001$, $d = 4.87$; and normalized transition duration, $F(1,18) = 409.88$, $p < .001$, $d = 5.03$. No significant main or interaction effects of Task were found for any context ($p > .05$). Both Japanese and English controls produced similar F1 values (3.00–6.00 Bark) in all task conditions ($p > .05$).

Thus, compared to nativelike performance, the initial state of Japanese /ɹ/ production was characterized by significantly higher F3 ($M = 15.40$ Bark) and F2 ($M = 11.97$ Bark) with shorter raw ($M = 19.86$ ms) and normalized ($M = 5\%$) transition duration with substantially large effects ($d = 3.00$–5.00). This suggests that the beginner Japanese learners tended to substitute the Japanese tap for /ɹ/ under all task conditions.

Given the pattern of distinct formant distributions between Japanese and English controls, the remainder of the data set was analyzed with respect to the following benchmark (the initial state of Japanese /ɹ/ production → nativelike /ɹ/ production) focusing only on F3, F2, and raw and normalized transition duration:

- F3: 15.30–15.50 Bark (2800–3000 Hz) → 12.30–12.50 Bark (1800–1900 Hz).
- F2: 11.90–12.10 Bark (1700–1800 Hz) → 9.70–9.90 Bark (1200–1300 Hz).
- Raw transition duration: 15–20 ms → 80–100 ms.
- Normalized transition duration: 4–5% → 20–25%.

Japanese learners

Descriptive results

A visual inspection of Fig. 1 indicates that experienced Japanese learners (LOR ⩾ 6 years) generally produced /ɹ/ with lower F3 and F2 values and longer transition duration (i.e., more targetlike exemplars compared to Japanese controls). Moreover, some of their /ɹ/ productions substantially overlapped with a range of English controls, especially in the domain of the secondary cues (F2, raw and normalized transition).

Task effects

To investigate the role of task conditions in the experienced Japanese learners' /ɹ/ performance, a set of one-way ANOVA were conducted on each acoustic domain—F3, F2, raw and normalized transition duration—with one within-subject factor (Task: WR, SR, TPD). Significant main effects of Task were found in F3, $F(1,87) = 31.36$, $p < .001$ and F2, $F(1,87) = 19.81$, $p < .001$. Bonferroni post hoc comparisons showed significantly different contrasts as follows:

- F3: WR ($M = 13.32$ Bark) < SR ($M = 13.51$ Bark) < TPD ($M = 13.70$ Bark) ($p < .01$).
- F2: WR ($M = 9.84$ Bark) < TPD ($M = 10.20$ Bark) ($p < .001$), SR ($M = 9.94$ Bark) < TPD ($p = .001$).

In short, the results suggest that the experienced Japanese learners produced more targetlike exemplars of /ɹ/ (i.e., lower F3 and F2) at a controlled rather than spontaneous speech level.

[6] In Polka and Strange's (1985) listening experiment with synthesized tokens on a rock-lock continuum, listeners perceived stimuli with intermediate spectral (F3, F2) and temporal (F1 transition duration) values as neither /ɹ/nor /1/, but rather as /w/or /d/.

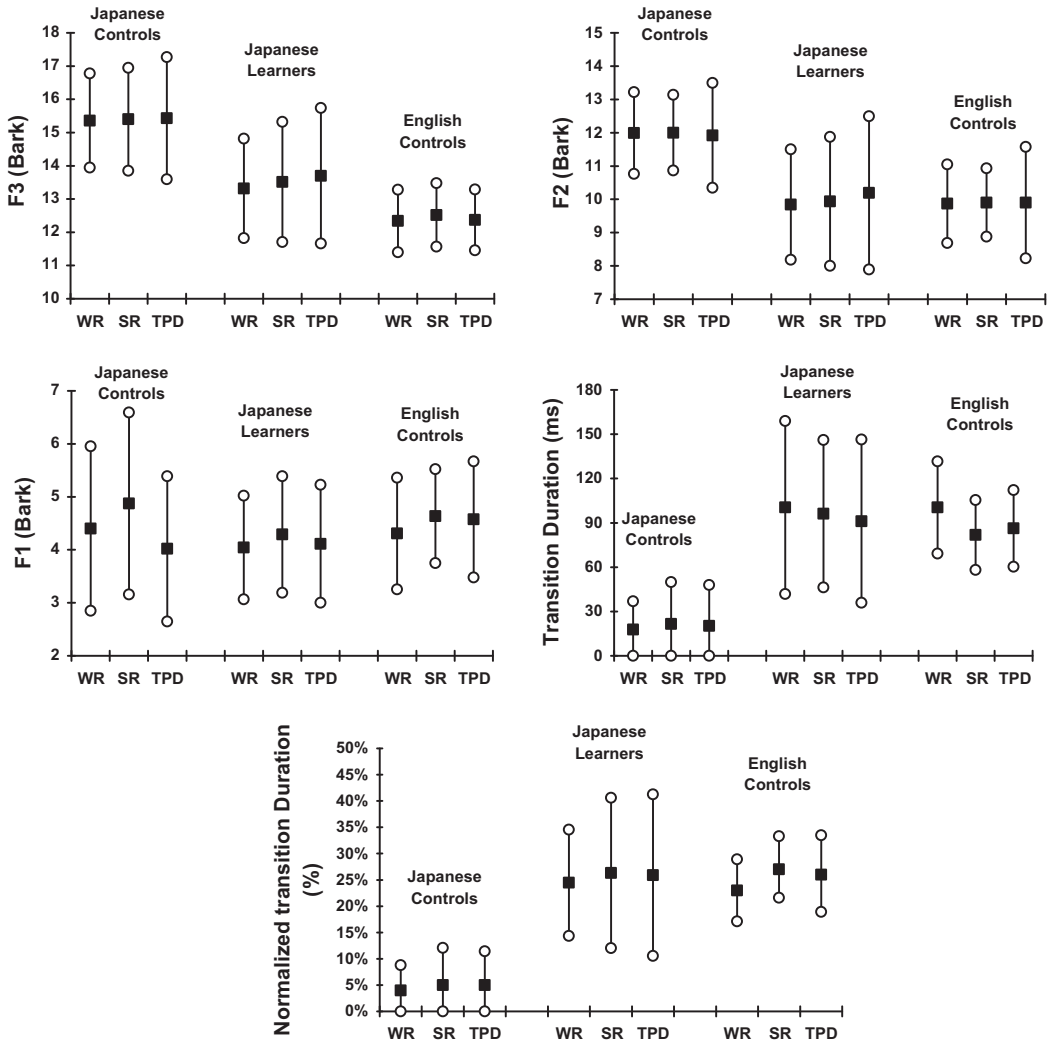

Fig. 1. Two SDs and mean values of F3, F2, raw and normalized transition duration in Group (English, Japanese controls) × Task (WR, SR, TPD) context.

AOA effects

For a robust correlation analysis, a series of Grubbs' tests were first conducted to find any outliers according to five acoustic domains (F3, F2, F1, raw and normalized transition duration) under three task conditions (WR, SR, TPD). The results identified one participant as a significant outlier in terms of F3 under WR ($z = 3.67$) and SR ($z = 3.79$) at a $p < .05$ level ($z < 3.34$). This participant was eliminated from subsequent statistical analyses of F3 under WR and SR. Simple correlation analyses revealed a significant relationship between AOA and F3 (the earlier they arrived in Canada, the lower their F3 tended to be) in the context of WR, r (86) = .266, $p = .013$, SR, r (86) = .251, $p = .019$, and TPD, r (87) = .356, $p < .001$ (see Fig. 2).

Next, to examine the actual shape of the significant AOA-F3 function, I conducted a series of trend (i.e., linear vs. quadratic vs. cubic connections) and piecewise regression analyses (i.e., the existence of any discontinuities in the regression line). Whereas one-way ANOVAs did not support the quadratic and cubic trends in any of the contexts ($p > .05$), they identified significant linear trends in the AOA and their /ɹ/ production in WR, $F(1,66) = 5.607$, $p = .021$, SR, $F(1,67) = 5.130$, $p = .027$, and TPD, $F(1,67) = 12.185$, $p = .001$. According to the results of the piecewise regression analyses, a single sloping line without a break point was identified as the best fitting function for WR, $F(1,85) = 6.470$, $p = .001$, SR, $F(1,85) = 5.667$, $p = .002$, and TPD, $F(1,86) = 12.551$, $p < .001$.

Third, I checked whether F3 did or did not correlate more strongly with AOA than the other acoustic dimensions (F2, raw and normalized transition duration) in WR, SR, and TPD, respectively. It was found that the F3-AOA correlation was not significantly stronger than the F2-, F1-, and duration-AOA correlations in any task context ($p > .05$). This in turn indicates that, whereas AOA significantly predicts only F3 attainment, AOA may be somewhat

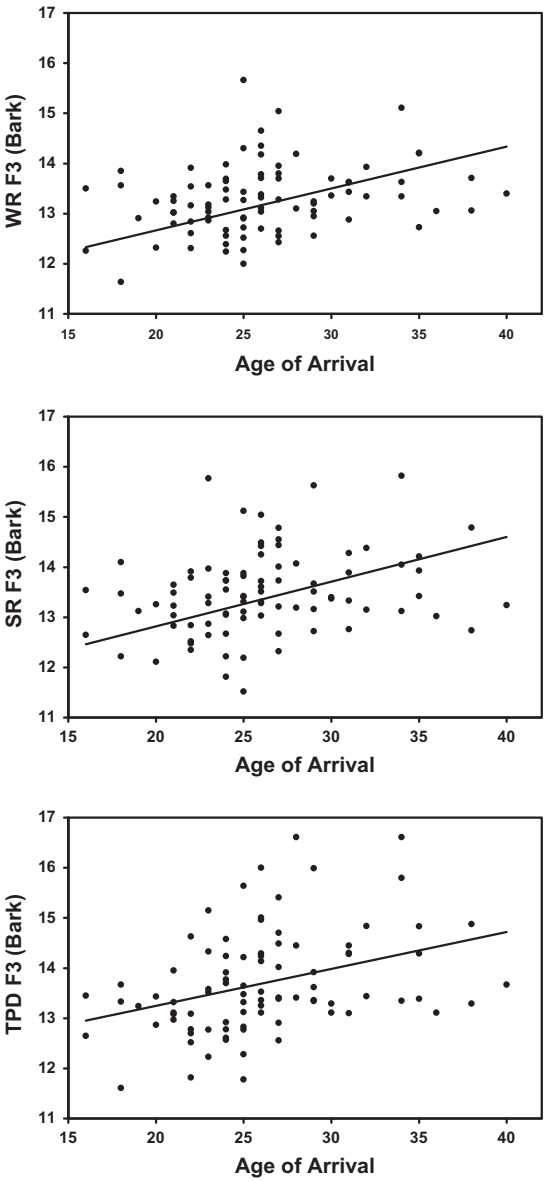

Fig. 2. F3 values in word reading, sentence reading and picture description plotted as a function to AOA.

related to the development of the F2 and transition duration.[7]

Last, because many previous age-related SLA studies noted that AOA effects are likely confounded with learners' LOR (e.g., Flege, Munro, et al., 1995) (i.e., the earlier they arrive in an L2 country, the longer they stay) and the two variables were indeed significantly correlated in the current study, $r(87) = -.315$, $p = .003$, partial correlation analyses were also conducted to examine the effects of AOA on F3 variation with LOR effects factored out. With variances in LOR controlled, the correlation between AOA and F3 values remained statistically significant in WR, $r(84) = .263$, $p = .014$, SR, $r(84) = .248$, $p = .021$, and TPD, $r(85) = .340$, $p < .001$. With respect to the use of the secondary (F2, raw and normalized transition duration), however, no significant correlation was found in any contexts ($p > .05$). The results of the correlation analyses are summarized in Table 2.

Nativelikeness

To determine how many participants produced /ɹ/ in a nativelike manner as a function of AOA, I employed the following procedure as used in other similar age-related SLA studies with late bilinguals (e.g., Birdsong, 2007; Flege, Munro, et al., 1995, 1999, 2006). First, 88 Japanese learners were categorized into four groups based on their AOA profiles at 5-year intervals: adolescent arrivals ($n = 8$) (AOA: 16–20 years); early adulthood arrivals ($n = 36$) (AOA: 21–25 years); mid adulthood arrivals ($n = 28$) (AOA: 26–30 years); and late adulthood arrivals ($n = 16$) (AOA: 31–40 years). The four groups of the Japanese learners significantly differed in their AOA, $F(3,84) = 206.04$, $p < .001$. Second, the mean and standard deviation (SD) of the 10 English controls was calculated for each acoustic domain. Last, I counted how many Japanese learners' /ɹ/ performance fell within two SDs of the English controls' mean values.

According to the descriptive results summarized in Table 3, the percentage of the Japanese learners who met the 2-SD criterion seemed to linearly decline from 50–70% to 10–20% across the four groups in terms of the primary cue (i.e., F3). In contrast, most of the Japanese learners (80–90%) reached nativelike proficiency in terms of the secondary cues (i.e., F2, raw and normalized transition duration) in spite of their various AOA profiles.

Results: listener judgment

Evaluation criteria

To assess the listeners' judgements of /ɹ/ productions the following rubric was established:

- Intelligibility refers to "whether participants' productions matched the English /ɹ/ category." Percentage scores were calculated for each token based on how many of the 10 listeners judged the sound as English /ɹ/—1 (*Nativelike* /ɹ/) to 4 (*possibly* /ɹ/).
- Accuracy is defined as "which category participants' production fell into in the context of the English /ɹ/-/l/ continuum" (9-point scores were used without any modification). As few empirical studies have closely examined the relationship between the quality of possibly-to-nativelike English /l/ and the Japanese tap, it also needs to be acknowledged that the accuracy criteria above (1. *Nativelike* /ɹ/ – 9. *Nativelike* /l/)

[7] According to the nativelike analysis mentioned below, a small portion of Japanese learners (10–20%) indeed noted non-nativelike F2 and duration representation, and there is a great possibility that their production quality was influenced by their AOA profiles. In this regard, AOA appears to play a crucial role in determining the extent to which L2 learners can ultimately master all acoustic properties of new sounds, if at all there exists any room for further improvement.

Table 2
Summary of simple and partial correlation analyses between L2 proficiency and AOA.

Dependent variable	Task condition	Simple correlation	Partial correlation (LOR controlled)
F3	WR	$r = .266$ ($p = .013$)*	$r = .263$. ($p = .014$)*
	SR	$r = .251$ ($p = .019$)*	$r = .248$ ($p = .021$)*
	TPD	$r = .356$ ($p < .001$)**	$r = .340$ ($p < .001$)**
F2	WR	$r = .094$ ($p = .384$)	$r = .050$ ($p = .643$)
	SR	$r = .179$ ($p = .095$)	$r = .170$ ($p = .115$)
	TPD	$r = .199$ ($p = .063$)	$r = .189$ ($p = .079$)
Transition duration	WR	$r = -.173$ ($p = .188$)	$r = -.093$ ($p = .389$)
	SR	$r = -.064$ ($p = .555$)	$r = -039$ ($p = .717$)
	TPD	$r = -.129$ ($p = .233$)	$r = -.090$. ($p = .407$)
Normalized duration	WR	$r = -.160$ ($p = .138$)	$r = -.128$ ($p = .237$)
	SR	$r = -.138$ ($p = .200$)	$r = -.156$ ($p = .150$)
	TPD	$r = -.047$ ($p = .663$)	$r = -.028$ ($p = .797$)
Accuracy (9-point)	WR	$r = .143$ ($p = .187$)	$r = .103$ ($p = .346$)
	TPD	$r = .256$ ($p = .017$)*	$r = .206$ ($p = .056$)
Intelligibility (2-point)	WR	$r = -.077$ ($p = .477$)	$r = -.028$ ($p = .798$)
	TPD	$r = -.215$ ($p = .045$)*	$r = -.157$ ($p = .147$)
Goodness (5-point)	WR	$r = .200$ ($p = .063$)	$r = .159$ ($p = .144$)
	TPD	$r = .275$ ($p = .009$)**	$r = .224$ ($p = .037$)*

Note: AOA, age of acquisition; LOR, length of residence; WR, word reading; SR, sentence reading; TPD, timed picture description.
* Statistical significance at $p < .05$ level.
** Statistical significance at $p < .01$ level.

Table 3
The ratio and raw number of Nativelike Japanese learners at a 5-year AOA intervals.

	F3			F2			Transition duration			Normalized duration		
A. Acoustic analysis												
AOA profiles	WR	SR	TPD	WR	SR	TPD	WR	SR	TPD	WR	SR	TPD
Adolescent arrivals (%)	75	50	50	88	88	100	88	100	100	88	88	100
($n = 8$)	(6)	(4)	(4)	(7)	(7)	(8)	(7)	(8)	(8)	(7)	(7)	(8)
Early adulthood arrivals (%)	69	42	53	97	86	83	83	92	86	86	78	69
($n = 36$)	(25)	(15)	(19)	(35)	(31)	(30)	(30)	(33)	(31)	(31)	(28)	(25)
Mid adulthood arrivals (%)	43	32	18	89	93	82	82	86	82	89	86	75
($n = 28$)	(12)	(9)	(18)	(25)	(26)	(23)	(23)	(24)	(23)	(25)	(24)	(21)
Late adulthood arrivals (%)	25	38	13	94	75	75	81	88	81	81	88	81
($n = 16$)	(4)	(6)	(13)	(15)	(12)	(12)	(13)	(14)	(13)	(13)	(14)	(13)

	Intelligibility		Accuracy		Goodness	
B. Listener judgment						
AOA profiles	WR	TPD	WR	TPD	WR	TPD
Adolescent arrivals (%)	100	100	25	38	25	38
($n = 8$)	(8)	(8)	(2)	(3)	(2)	(3)
Early adulthood arrivals (%)	53	67	11	25	11	25
($n = 36$)	(19)	(24)	(4)	(3)	(4)	(3)
Mid adulthood arrivals (%)	57	57	14	21	14	21
($n = 28$)	(16)	(16)	(4)	(6)	(4)	(6)
Late adulthood arrivals (%)	38	50	0	13	0	13
($n = 16$)	(6)	(8)	(0)	(2)	(0)	(2)

Note: Adolescent arrivals ($16 \leqslant$ age of acquisition [AOA] $\leqslant 20$ years); early adulthood ($21 \leqslant$ AOA $\leqslant 25$ years); mid adulthood ($26 \leqslant$ AOA $\leqslant 30$); late adulthood ($31 \leqslant$ AOA $\leqslant 40$); WR, word reading; SR, sentence reading; TPD, timed picture description.

could conflate how to evaluate the learners who noted little awareness towards English /ɹ/; their substitution of the tap without any effort to produce English /ɹ/ (supposedly assigned to "9") could not be necessarily equivalent to nativelike English /l/ (labeled as "9").

- Due to the primary goal of the project (examining in-depth the differences and similarities between experienced Japanese learners' and native English speakers' /ɹ/ production) and the potential ambiguity of the /l/ category under the 9-point scale,[8] a third

[8] Whereas the significant acoustic and articulatory differences between the Japanese tap and English /ɹ/ are well-documented (e.g., Flege et al., 1995b), few empirical studies have examined to what degree the former sound is actually dissimilar or similar to English /l/ (cf. Hattori & Iverson, 2009). In other words, we do not yet know whether the participants' substitution of the Japanese tap (i.e., no effort to produce English /ɹ/) should be evaluated as nativelike, good, probably, or possible /l/.

criteria, goodness, was included in order to analyze the degree to which participants' production was targetlike within (rather than between) a category of English /ɹ/. While the original scores were used from "1" (*Nativelike* /ɹ/) to "4" (*possibly* /ɹ/) without any adjustment, the non-/ɹ/ category—5 (*neutral exemplars*) to 9 (*Nativelike good* /l/)—was equally assigned "5."

Inter-rater reliability

To measure the intraclass correlation between the 10 native speaking listeners, Cronbach's alpha was computed at .943 for the intelligibility dataset, .961 for the accuracy dataset, and .944 for the goodness dataset (n = 864, respectively). The results indicate an acceptable level of inter-rater agreement compared to previous L2 pronunciation research (e.g., Derwing & Munro, 2013). By pooling over listeners, I assigned mean rating scores to each /ɹ/ token produced by the participants.

Baseline

To establish a baseline for the initial state of Japanese /ɹ/ production relative to nativelike /ɹ/ production, I will first examine the intelligibility (%), accuracy (9-point), and goodness (5-point) of the Japanese and English controls. The descriptive results of the listener judgment according to the three domains are summarized in Fig. 3.

The three criteria of /ɹ/ (intelligibility, accuracy, goodness) were separately submitted to two-way ANOVAs with one between-group factor (Japanese, English) and one repeated measure (Task: WR, TPD). Main effects of Group were found for intelligibility, $F(1,18) = 94.166$, $p < .001$, $d = 3.28$; accuracy, $F(1,18) = 155.526$, $p < .001$, $d = 3.84$; and goodness, $F(1,18) = 292.296$, $p < .001$, $d = 5.28$. No main and interaction effects of Task were found at any contexts ($p > .05$).

According to the results, beginner Japanese learners' /ɹ/ production was highly unintelligible (M = 31.0% in "intelligibility") and categorized as neutral exemplars or possibly English /l/ rather than English /ɹ/ (M = 5.75 in "accuracy" [9-point] and M = 4.38 in "goodness" [5-point]). This indicates that the initial state of English /ɹ/ production could be synonymous to the substitution of the Japanese tap in line with the results of the acoustic analysis mentioned above as well as with previous findings (Sekiyama & Tohkura, 1993).

Japanese learners of English

Descriptive results

The listener judgment of the production of /ɹ/ by 88 experienced Japanese learners is plotted in Fig. 3. A visual inspection indicates that their /ɹ/ production was not only quite intelligible (M = 88.4% in "intelligibility") but also perceived as good exemplars of /ɹ/ (M = 2.76 in "accuracy" and 2.60 in "goodness"). Despite a great deal of variance among the learners, some seemed to fall into the nativelike category.

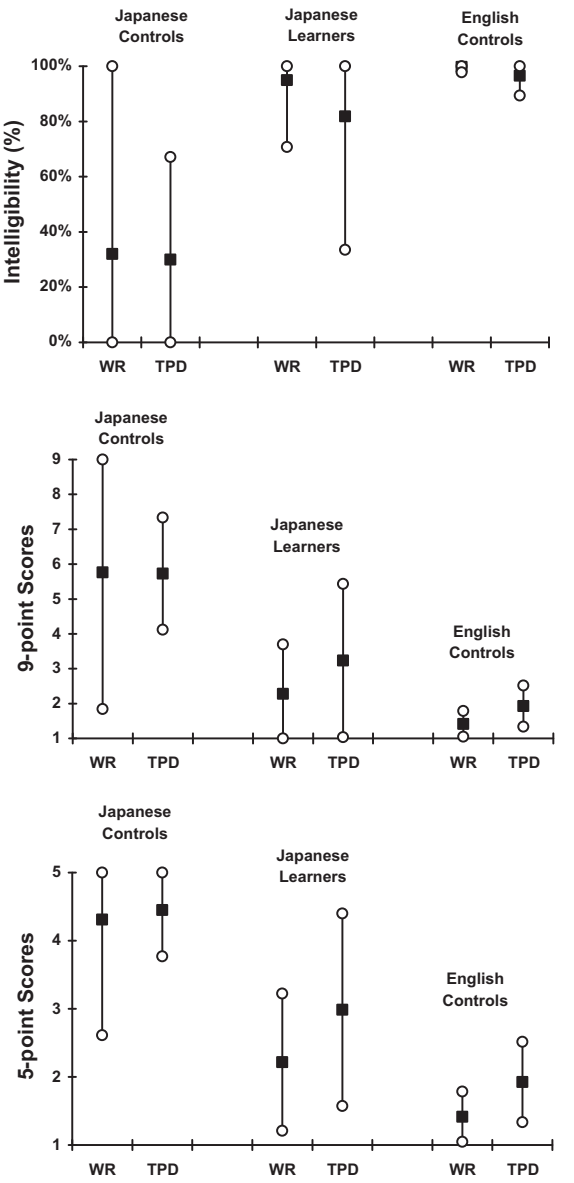

Fig. 3. Two SDs and mean values of intelligibility, accuracy and goodness in Group (Japanese learners, English, Japanese controls) × Task (WR, TPD) context.

Task effects

Effects of task were assessed through a set of paired-samples t-tests on each listening criterion—intelligibility (%), accuracy (9-point), and goodness (5-point)—with one within-subject factor (Task: WR, TPD). Significant main effects of Task were found in (a) intelligibility, $t(87) = 6.759$, $p < .001$, (b) accuracy, $t(87) = -11.72$, $p < .001$, and (c) $t(87) = -14.098$, $p < .001$. The results indicate that, although Japanese learners tended to produce good exemplars of /ɹ/ at a controlled speech level, their spontaneous performance of /ɹ/ was slightly less intelligible ("probably /ɹ/").

AOA effects

According to the results of the Grubbs' tests, one participant's /ɹ/ performance on the WR task (invariably perceived as English /l/) was considered as an outlier at a $p < .05$ level for intelligibility ($z = 5.78$), accuracy ($z = 5.47$) and goodness ($z = 4.18$), and was thus eliminated from the subsequent statistical analyses. A set of simple correlation analyses revealed that for the controlled production task (WR), the correlation between listener judgment and AOA was not significant in the domains of (a) intelligibility, r (86) = −.077, $p = .477$, (b) accuracy$r$ (86) = .143, $p = .187$, or (c) goodnesss, r (86) = .200, $p = .063$. The distribution of the listening scores in Fig. 4 showed that the lack of statistical significance could be due to ceiling effects (i.e., most of the experienced Japanese learners produced good exemplars of /ɹ/ without much variance).

For the spontaneous production task (TPD), however, there is a significant weak-to-moderate AOA effect in the domain of (a) intelligibility, r (87) = −.215, $p = .045$, (b) accuracy$r$ (87) = .255, $p = .017$, and (c) goodnesss, r (87) = .275, $p = .009$ (see Table 2 and Fig. 5).

In terms of the actual shape of the significant regression line between AOA and spontaneous /ɹ/ production, linear trends were found in the domains of intelligibility, $F(1,67) = 3.990$, $p = .050$, accuracy, $F(1,67) = 5.656$, $p = .020$, and goodness, $F(1,67) = 6.747$, $p = .012$. The piecewise regression analyses again identified a single sloping line without a break point as the best fitting function for intelligibility, $F(1,86) = 4.257$, $p = .04$ accuracy, $F(1,86) = 6.049$, $p = .02$ and goodness, $F(1,86) = 7.060$, $p < .001$.

Last, to control the variable of the learners' LOR profiles, partial correlation analyses were also conducted. When the LOR effect was factored out, the correlation between listeners' judgment scores and AOA in the picture description task became non-significant in intelligibility, r (85) = −.157, $p = .147$, and marginal in accuracy, r (85) = .206, $p = .056$, but remained significant in goodness, r (85) = .224, $p = .037$.

Nativelikeness

To estimate how many Japanese learners produced /ɹ/ within range of the 10 English controls, the 2-SD criteria approach was employed again for the listener judgment scores. As summarized in Table 3, the incidence of nativelikeness linearly decreased according to their AOA profiles (16–40 years): 100% to 40–50% (intelligibility) and 30–40% to 0–10% (accuracy and goodness).

Acoustic correlates of /ɹ/

To investigate the relationship between the relevant acoustic properties of /ɹ/ and human perception of /ɹ/, a set of multiple regression analyses was computed with each of the three listener judgment criteria—intelligibility (%), accuracy (9-point), goodness (5-point)—as dependent variables ($N = 864$ tokens: 8 tokens × 108 participants), and with the corresponding four acoustic domains of /ɹ/ (F3, F2, raw and normalized transition duration) as independent variables.

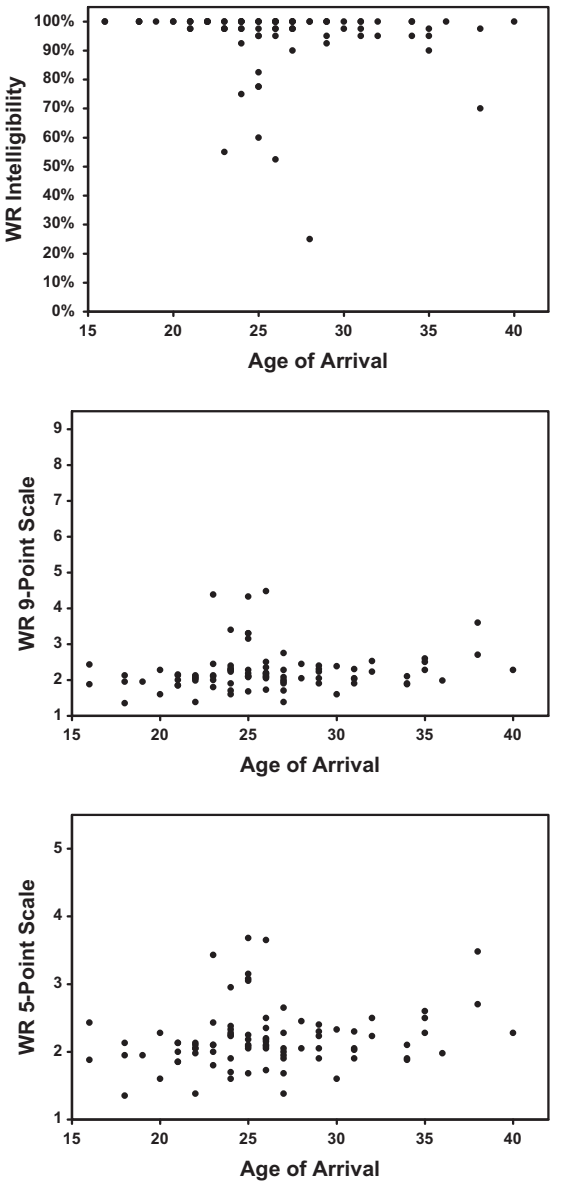

Fig. 4. Listener judgment scores in the controlled production task (word reading) plotted as a function of AOA.

Intelligibility

The model was significant, $F(4,859) = 179.140$, $p < .001$, accounting for 45.5% of the variance in the listeners' intelligibility scores. The model identified all acoustic variables as significant predictors: F3 ($\beta = -.343$, $t = -9.702$, $p < .001$), F2 ($\beta = -.078$, $t = -2.234$, $p = .026$), transition duration ($\beta = .234$, $t = 6.373$, $p < .001$), and normalized transition duration ($\beta = .164$, $t = 4.834$, $p < .001$).

Accuracy

Whereas the model explained 47% of the variance in the listeners' accuracy scores, $F(4,859) = 190.452$, $p < .001$, it found three out of the four acoustic variables as significant predictors: F3 ($\beta = .438$, $t = 12.533$, $p < .001$), transition

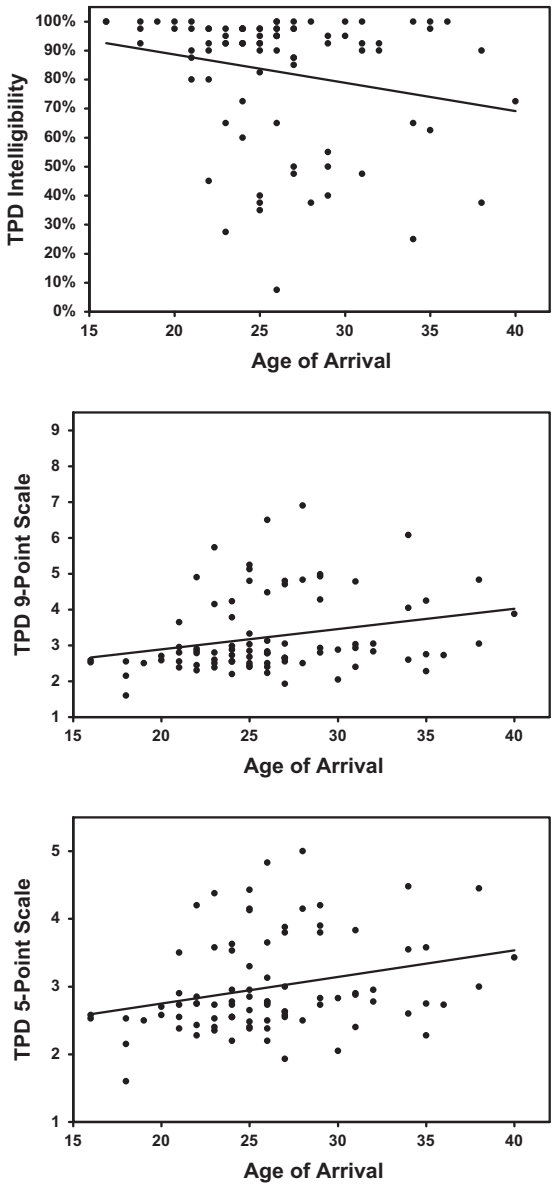

Fig. 5. Listener judgment scores in the spontaneous production task (picture description) plotted as a function of AOA.

duration ($\beta = -.295$, $t = -8.130$, $p < .001$), and normalized transition duration ($\beta = -.067$, $t = -1.986$, $p < .001$).

Goodness

According to the model, 45.4% of the variance in the listeners' goodness scores was explained by the acoustic properties, $F(4,859) = 178.785$, $p < .001$, and two out of the four variables were found to be significant predictors: F3 ($\beta = .445$, $t = 12.559$, $p < .001$), and transition duration ($\beta = -.334$, $t = -9.077$, $p < .001$).

To summarize, the results indicate that (a) native speaking listeners use all of the relevant acoustic cues (F3, F2, raw and normalized transition duration) to differentiate English /ɹ/ from English /l/ (between-category perception); and (b) they tended to draw on low F3 as a primary cue and longer transition duration as a secondary cue to determine the extent of targetlikeness for English /ɹ/ tokens (within-category perception).

Discussion

Situated in the production development of word-initial /ɹ/ by high-proficiency late Japanese bilinguals (AOA $\geqslant$ 16 years, LOR $\geqslant$ 6 years), the current study examined whether and to what degree the correlation between AOA and L2 proficiency levels remains significant in late SLA. Our findings allowed us to evaluate the two competing theoretical positions on age-related SLA in late bilingualism—discontinuous (Critical Period Hypothesis) or continuous (Cognitive Aging Hypothesis) AOA effects on the quality of attained L2 production ability and the incidence of nativelikeness.

First, the listener judgments revealed clear age effects, especially when L2 production was elicited at a spontaneous speech level. As shown in Fig. 4, although many Japanese learners showed ceiling effects under the controlled production task (word reading), the ultimate attainment of their spontaneous /ɹ/ production (timed picture description) varied widely and correlated moderately with their age of arrival in Canada. Importantly, the AOA-proficiency regression line was best identified as a singly linear one, and including a breakpoint in the function did not improve fit to data. Second, the results of the acoustic analysis provided a detailed picture of which domains of their /ɹ/ performance were subject to the influence of age effects. Specifically, AOA was a weak predictor for the primary cue (F3) acquisition at a controlled speech level (word and sentence reading) and a moderate-to-strong predictor at a spontaneous speech level (timed picture description) without any non-linearities throughout the data. The incidence of nativelikeness in this domain was probable, but nativelike results were less likely in proportion to their increasing AOA profiles. However, most Japanese learners attained nativelike proficiency in the secondary cues (F2, transition duration) despite their varied AOA profiles, demonstrating no predictive role of AOA in this domain.

These two findings fail to support the predictions of the strong version of the CPH that AOA effects are absent in late bilingualism (due to a loss of plasticity),[9] and that reaching nativelike production is not related to AOA profiles (but rather to other socio-psychology factors including aptitude). The findings can, however, be well accounted for by the Cognitive Aging Hypothesis, which holds that AOA effects continue to be observed even after puberty as predictors for the degree of success in SLA and the likelihood of nativelikeness. On the one hand, the experienced Japanese learners (LOR $\geqslant$ 6 years) produced /ɹ/ significantly better than the Japanese controls (LOR < 1 month), indicating that late bilinguals maintained the ability to learn new sounds

[9] As Birdsong and Molis (2001) and Hakuta et al. (2003) emphasized, a wide variety of statistical analyses, such as trend and piecewise regression analyses, need to be included in order to investigate and confirm significant discontinuities (i.e., CP) in the AOA/attainment function, if any.

through extensive length of residence in an L2 speaking country (Bialystok, 1997; Flege, 1995). On the other hand, the experienced Japanese learners' asymptotic /ɹ/ performance was negatively constrained by how late they arrived in Canada. This in turn suggests that AOA may be a driving factor in the degree to which late bilinguals can benefit from additional input and interaction (Derwing & Munro, 2013; Flege et al., 2006). By taking into account task type (controlled vs. spontaneous) and multiple cue weightings (primary vs. secondary) in L2 speech learning processes, however, the current study further revealed the complex mechanisms underlying age-related adult L2 production learning. That is, the ultimate attainment and nativelikeness of late bilingualism are multidimensional phenomena characterized by multiple learning curves and plateaus as a result of the interactions of AOA, processing abilities, and L1 influence.

Task effects

One variable that is potentially relevant to age effects in late bilingualism is task condition. In the study, three different speaking tasks were used to measure two types of processing abilities for producing English /ɹ/, both at a controlled and spontaneous speech level. Whereas conscious attention to pronunciation accuracy and hyper-articulation of /ɹ/ was allowed in the controlled processing mode, participants were required to attend to various domains of language—phonological, morphological, syntactical, lexical, pragmatic accuracy in L2—under time pressure in the spontaneous processing mode, which deprived them of the attentional resources required to monitor their correct pronunciation of /ɹ/. According to the results, the AOA-proficiency relationship appeared to be especially strong when their performance was assessed by way of spontaneous (but not controlled) production tasks.

Several possible interpretations of such task-specific findings can be made in line with relevant theories on controlled versus spontaneous speech production. With respect to the integrated state feedback control architecture of general speech production (for details, see Hickok, Houde, & Rong, 2011), a lexical conceptualization of auditory or written input initiates phonological encoding via the internal model which predicts the articulatory features of the targets (Levelt, Roelofs, & Meyer, 1999). This semantic-to-phonological transcoding process subsequently generates neuromascular signals which activate and control the movements of the articulators in actual output (Ventura, Nagarajan, & Houde, 2009). Notably, this model also assumes that speech production can occur in other processing modes, where the external input directly connects with the internal articulatory representation in order to monitor and adjust their output quickly, but bypasses any semantic processing (McCarthy & Warrington, 1984).

Based on the findings in the study, the theoretical perspectives indicate that (a) most of the late Japanese learners produced good English /ɹ/ exemplars under controlled speech conditions, probably by matching orthographic /ɹ/ forms to relevant articulatory gestures of /ɹ/ (using language as an object); and (b) their AOA profiles were significantly predictive of the ultimate attainment of /ɹ/ when they produced the sound for both form and meaning (using language for message-conveyance).

The results here also concur with Major's (2008) proposal for task variation in adult L2 production development characterized as a transition from L1-related to universal errors: Although learners without much L2 experience (Japanese controls) likely transfer L1 counterparts (the Japanese tap) to produce L2 sounds in the early acquisition stages (i.e., L1-related errors), their performance demonstrates several universal principles, such as differential performance at a controlled and spontaneous speech level (i.e., universal errors). This vein of interlanguage phonology research has demonstrated that, when tested via formal tasks (e.g., word reading), adult L2 learners can attain relatively advanced L2 proficiency by deliberately drawing on their explicit articulatory knowledge stored in their general memory; they are, however, unable to do so in the context of less constrained outcome measures (e.g., picture description) (Dickerson & Dickerson, 1977; Rau et al., 2009). In this regard, many SLA (e.g., Spada & Tomita, 2010) and L2 phonology (e.g., Major, 2008) researchers have emphasized the importance of adopting a range of tasks, including more cognitively demanding ones, to better reveal the present state of learners' linguistic abilities, excluding monitoring effects otherwise present.

In sum, it is suggested that the effect of age on late bilingualism may be evident in task modalities which simulate the semantic–phonological processing operations involved in their naturalistic learning and speaking contexts (spontaneous measure), but not in modalities which reflect the amount of explicit articulatory knowledge (controlled measure). This is arguably because AOA may be influential in determining to what degree late bilinguals can attain a more robust representational system of English /ɹ/ (i.e., qualitative change), as well as various levels of processing ability (i.e., quantitative change) to produce the sound in a more *nativelike* and *automatic* manner for communicative purposes.

First language Effects

Another important variable is the influence of the L1 system on the phonetic features the learners are trying to acquire. Although almost all Japanese learners equally produced /ɹ/ with low F2 (tongue retraction) and long transition duration (prolonging phonemic segment) in a nativelike fashion, they exhibited somewhat limited performance in the F3 domain, with evident AOA effects on their ultimate attainment and nativelikeness. The difference in Japanese bilinguals' acquisition of /ɹ/ could be ascribed to the degree to which the relevant phonetic cues are used in the L1 phonetic system, i.e. new (F3) vs. existing (F2, transition duration) cues.

In the Japanese phonetic system, F3 is not used as essential information to differentiate any consonantal or vocalic sounds. Furthermore, since the L1 counterpart of the Japanese tap has a wide variation of F3 (2400–

3000 Hz), native speakers of Japanese are apt to even ignore this acoustic information for speech perception (Lotto et al., 2004). Conversely, Japanese learners likely give more weight to F2, rather than F3, in perceiving English /ɹ/ (Iverson et al., 2003), and produce the sound with lower F2 (i.e., /w/-like production) (Lotto et al., 2004). This is arguably because F2 is already used to separate L1 approximant categories (i.e., /j/ vs. /w/) and is thus salient to them. In addition, Japanese learners demonstrate high sensitivity to temporal cues rather than spectral cues (Underbakke, Polka, Gottfried, & Strange, 1988). This tendency could be attributed to L2 learners' general preference for temporal cues over spectral cues (Bohn, 1995) as well as to some exploitation of temporal cues in the L1 system (i.e., five spectrally distinctive pairs of vowels are temporally differentiated) (Peterson & Lehiste, 1960).

Japanese learners indeed demonstrate a different amount of difficulty acquiring English /ɹ/ according to different acoustic domains of the sound. Because Japanese learners quickly adjust and attend to F2 and transition duration of English /ɹ/, they typically show *moderate* improvement on their perception and production abilities after the first few years of immersion in L2 countries (e.g., Larson-Hall, 2006) and several hours of intensive auditory training (e.g., Bradlow, 2008). Notably, certain researchers have pointed out that their improved performance rarely reaches a nativelike level, arguably due to their insufficient awareness towards F3 variation (Ingvalson, Holt, & McCLELLAND, 2012). It has been shown that learners' ability to reliably and robustly identify English /ɹ/ is highly related to their sensitivity to the F3 cue (Hattori & Iverson, 2009), and auditory training can lead only a small portion of (but not all) Japanese learners to enhance their F3 representational system and attain nativelike performance (Ingvalson et al., 2012).

Our findings on the multifaceted cue weightings in L2 segmental learning relative to the L1 system coincide with the major theoretical views that L2 learners have more difficulty establishing new cues than changing the relative weights of cues already present in the L1 system (e.g., McAllister, Flege, & Piske, 2002 for the Feature Hypothesis). Recent research has indeed demonstrated a differential amount of difficulty in learning new versus existing phonetic features at the consonantal (Baker, 2010), vocalic (McAllister et al., 2002) and prosodic (Trofimovich & Baker, 2006) levels of L2 production development. Following this line of thought, the findings of the current study yielded evidence that AOA could be predictive of the development of a new cue (e.g., F3), but that an increasing age might not exert any negative influence on the resetting of existing cues (e.g., F2, transition duration), which is typically characterized by quick improvement without any age-related disadvantage. This is important because maturational accounts in the Critical Period Hypothesis would predict more difficulty with learning in general; the fact that the predictive power of AOA for learning difficulty was observed only along specific dimensions (the new F3 cue) suggests that there is a degree of plasticity—the persistent influence of age of acquisition and L1 entrenchment on L2 ultimate attainment—even after puberty, as suggested by the Cognitive Aging Hypothesis (e.g., Bialystok, 1997).

Limitations

Due to the exploratory nature of the project, however, I address here several topics worthy of future research attention. First, our findings on AOA-proficiency decline were discussed exclusively in the context of English /ɹ/. This segment has a relatively high functional load on successful communication, and L2 learners likely notice and make efforts to acquire the sound despite its tremendous learning difficulty (Riney et al., 2000). Thus, it would be intriguing to examine how AOA can be predictive of late bilingualism in relation to other various phonetic features (especially non-salient ones with less communicative value) both in perception and production for late learners with diverse L1 profiles (cf. Abrahamsson, 2012).

Second, although the age effects on late learners were interpreted as support for the Cognitive Aging Hypothesis, the methodological limitations of the study did not necessarily allow the complete rejection of the CPH. That is, there could be a critical period for language learning, but late learners who use more domain general learning processes could be also disadvantaged by aging-related cognitive decline. To further examine the robustness of the critical period in bilingualism, it would be intriguing to test whether and to what degree the same test materials could generate AOA effects with early bilinguals. Investigating the differences or similarities in AOA functions between early and late bilinguals will in turn shed light on whether any dramatic change in language acquisition takes place around puberty.

The third limitation of the current study is the lack of analysis and discussion as to the underlying causes of the negative AOA-proficiency correlation in adult SLA. Given that AOA constitutes a wide range of affecting variables including L1–L2 interaction (e.g., Baker et al., 2008) and a biologically driven general decline in learning abilities (e.g., Birdsong, 2005), future research needs to tease apart and examine which factors relatively account for the age effects. In this respect, one promising research direction is to adopt similar test procedures—WR, SR, TPD—to elicit not only their L2 (English /ɹ/) but also L1 (the Japanese tap) production, which will consequently serve as a good proxy for their bilingual phonological representation relative to monolinguals of English and Japanese (e.g., Flege et al., 1997). In addition, it would be also intriguing to assess participants' states of neurological and cognitive development via valid instruments previously used in the cognitive psychology literature (e.g., the Simon task: Bialystok, Craik, Klein, & Viswanathan, 2004). Such future studies will allow us to examine whether and to what degree not only participants' AOA but also their levels of previous phonetic development and cognitive aging are differentially predictive of their ultimate attainment of /ɹ/.

Finally, the Timed Picture Description task (whereby participants were given target words to explain a set of pictures), though more ecologically valid than word reading, is not an index of truly spontaneous speech, let alone

implicit knowledge. Few studies have examined how learners produce specific segmental features in a spontaneous and automatic manner due to the difficulties in evaluating conversational speech elicited in controlled settings (Piske et al., 2001). In this respect, I hope to see more research on this topic that will elaborate and validate reliable measures to capture various aspects of learners' L2 phonetic knowledge.

Conclusion

To test the existence or absence of the negative relationship between increasing AOA and L2 proficiency in late bilingualism, the current study examined the role of AOA in ultimate attainment and nativelikeness of /ɹ/ production by high-proficiency Japanese learners. The results of the acoustic analysis and listener judgment identified significant AOA effects on late bilingualism precisely where much L2 experience is required to acquire the sound—the development of the new articulatory parameter (low F3 for labial, palatal, pharyngeal constrictions) (Bialystok, 1997; Flege, 1995; McAllister et al., 2002) entailing various levels of processing abilities (i.e., spontaneous /ɹ/ production) (Major, 2008). In contrast, most Japanese learners attained nativelike performance in terms of resetting existing articulatory parameters (low F2 for tongue retraction, long transition duration for lengthening phonemic segment) under single task condition (i.e., controlled /ɹ/ production). The results lend empirical support to the CAH in that "the decline in attained L2 proficiency is not linked to maturational milestones, but persists over the age spectrum" (Birdsong, 2005, p. 125), which may in turn constitute "a serious challenge to the CPH function" (DeKeyser & Larson-Hall, 2005, p. 97).

To close, it needs to be emphasized that the findings should be interpreted with caution due to several methodological limitations. More L2 phonology research of this kind needs to examine precisely what characterizes age effects on late bilingualism (e.g., L1–L2 interaction vs. cognitive aging) by including participants with various AOA profiles ($0 < x < 40$) as well as assessing their production performance in various L1 and L2 phonetic contexts. For such future studies, however, it is strongly recommended that variation in phonetic segments (new vs. existing cues) and elicitation methods (controlled vs. spontaneous tasks) should be taken into account and analyzed separately when addressing possible effects on adult SLA, such as AOA.

Acknowledgments

This study was funded by the Government of Canada Post-Doctoral Research Fellowship. I would like to thank Murray Munro, Pavel Trofimovich and the three anonymous reviewers from the *Journal of Memory and Language* for their useful and constructive comments on the earlier versions of the this manuscript. I am also grateful to Midori Adachi, Yuki Matsumura, Noriko Yamane, Keiko Onishi, Yukiko Simon, and Tonarigumi for their efforts to organize the data collection for this the project.

References

Abrahamsson, N. (2012). Age of onset and nativelike L2 ultimate attainment of morphosyntactic and phonetic intuition. *Studies in Second Language Acquisition, 34*, 187–214.

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition, 30*, 481–509.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

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., Flege, J. E., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech, 51*, 316–341.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*, 775–794.

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research, 13*, 116–137.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290–303.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*, 9–49.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. Munro (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 99–116). Amsterdam: Benjamins.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109127). New York: Oxford University Press.

Boersma, P., & Weenik, D. (2012). *Praat: Doing phonetics by computer*. <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.

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*, 447–465.

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.

Cobb, T. (2012). *The complete lexical tutor*. <http://www.lextutor.ca/vp/>.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499–533.

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.

Delattre, P., & Freeman, D. (1968). A dialect study of American r's by X-ray motion picture. *Linguistics, 44*, 29–68.

Derwing, T., & Munro, M. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*, 379–397.

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.

Dickerson, L., & Dickerson, W. (1977). Interlanguage phonology: Current research and future directions. The notions of simplification, interlanguages, and pidgins, and their relation to second language pedagogy, *Actes du 5ième Colloque de Linguistique Appliquée de Neuchatel* (pp. 18–30). Geneva: Droz.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a Second Language: A psychometric study. *Studies in Second Language Acquisition, 27*, 141–172.

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., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics, 34*, 153–175.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*, 169–186.

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

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

Flege, J., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory & Language, 41*, 78–104.

Gatbonton, E., & Trofimovich, P. (2008). The ethnic group affiliation and L2 proficiency link: Empirical evidence. *Language Awareness, 17*, 229–248.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*, 311–343.

Guion, S., Flege, J., Ahahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*, 2711–2725.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science, 14*, 31–38.

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.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron, 69*, 407–422.

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.

Ioup, G., Boustagi, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition, 16*, 73–98.

Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, 847–857.

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. *Applied Psycholinguistics, 24*, 131–161.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of ESL. *Cognitive Psychology, 21*, 60–99.

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.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75.

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.

Major, R. (2008). Transfer in second language phonology: A review. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 63–94). Amsterdam: John Benjamins.

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.

McCarthy, R. A., & Warrington, E. K. (1984). A two-route model of speech production: Evidence from aphasia. *Brain, 107*, 463–485.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: Benjamins.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 11*, 73–89.

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

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

Polka, L., & Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /ɹ/ and /l/. *Journal of the Acoustical Society of America, 78*, 1187–1197.

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

Riney, T., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly, 34*, 711–737.

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.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics, 20*, 213–223.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics, 21*, 427–444.

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). Statistics Canada. <http://www12.statcan.ca/census-recensement/2006/dp-pd/hlt/97-562/index.cfm?Lang=E>. Retrieved 03.06.12.

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.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.

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.

Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience, 10*, 58.

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