

Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English /ɪ/ by late Japanese bilinguals*

KAZUYA SAITO

Waseda University

FRANÇOIS-XAVIER BRAJOT

McGill University

(Received: February 10, 2012; final revision received: August 30, 2012; accepted: October 16, 2012)

The current project examined whether and to what degree continued L2 input, operationalized as length of residence (LOR), and age of acquisition (AOA), defined as the first intensive exposure to the target language, can be predictive of adult Japanese learners' production of word-initial English /ɪ/. Data were collected from 65 participants, consisting of three groups of Japanese learners of English (n = 13 for Short-, Mid-, and Long-LOR groups, respectively) and two groups of baseline speakers (n = 13 for Japanese- and English-Baseline groups, respectively). Their production of /ɪ/ was elicited via three oral tasks (i.e., word reading, sentence reading, timed picture description). Acoustic analyses were carried out along four dimensions: third formant (F3), second formant (F2), first formant (F1) frequencies, and formant transition duration. The results demonstrated that (a) all learners reached native-like proficiency with respect to the use of existing cues (F2, transition duration) within approximately one year of LOR, (b) their performance was negatively related to AOA to some degree, and (c) longer LOR was predictive of the development of the new cue (F3). These results suggest that late L2 speech sound acquisition and proficiency may be characterized by different levels of phonetic processing.

Keywords: second language pronunciation, late bilingualism, experience effects, age effects, English /ɪ/

Second language (L2) speech sound production has been shown to be influenced by a number of variables, including attitude and aptitude (e.g., Ioup, Boustagi, El Tigi & Moselle, 1994; Purcell & Suter, 1980), motivation (e.g., Bongaerts, Van Summeren, Planken & Schils, 1997; Moyer, 1999), level of education (Derwing & Munro, 2005), and ethnic identity (Gatbonton & Trofimovich, 2008). Two additional extrinsic factors that have received considerable attention are (a) the amount of L2 input operationalized as length of residence (LOR) in an L2 country, and (b) age of acquisition (AOA) (defined as the first intensive exposure to the target language). Both factors are strong predictors of L2 performance for EARLY BILINGUALS who arrive in an L2 country before puberty, that is, the earlier they arrive, the longer they stay, the better their L2 performance tends to be (Flege, Munro & MacKay, 1995a; Flege, Yeni-Komshian & Liu, 1999). The continued effect of LOR and AOA on LATE BILINGUALS

who are exposed to L2 after a certain age has received somewhat less attention, however, and it is still unclear whether these variables remain influential in adult L2 speech sound acquisition in particular. The current study examines this question in the context of adult Japanese speakers' use of word-initial English /ɪ/ across different production tasks.

Length of residence, age of acquisition, and late bilingualism

Given the assumption that the development of new linguistic competencies in acquiring a second language is constrained by the amount of L2 input and the timing of a learner's first intensive exposure, second language acquisition (SLA) research has studied the role of LOR and AOA fairly extensively (Best & Tyler, 2007; Flege, 1995, 2003, 2009; Major, 2008; McAllister, Flege & Piske, 2002). The complex interrelationships among these and other important variables like motivation have resulted in mixed findings, however, disallowing simple conclusions. Understanding the role of LOR and AOA on late bilinguals' speech sound development is no exception.

* This study was funded by the Government of Canada Post-Doctoral Research Fellowship awarded to the first author. We would like to thank Murray Munro and anonymous *Bilingualism: Language and Cognition* reviewers for their useful and constructive comments on the earlier versions of the manuscript.

Address for correspondence:

Kazuya Saito, School of Commerce, Waseda University, 1-6-1, Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan
kazuya.saito@waseda.jp

With respect to LOR, some studies have shown that L2 learners with longer LOR produced L2 sounds in a more target-like fashion than those with shorter LOR with respect to consonants (e.g., Flege, Takagi & Mann, 1995b), vowels (e.g., Flege, Bohn & Jang, 1997), and prosody and stress (e.g., Trofimovich & Baker, 2006). Conversely, other studies have reported that learners' LOR was unrelated to the quality of their L2 performance (Cebrian, 2006; Flege, 1988; Flege, Munro & Fox, 1994; Larson-Hall, 2006; Munro, 1993). In their review of studies that address factors relevant to L2 accent, Piske, MacKay and Flege (2001, p. 197) conclude that differences in primary studies' methodological approaches limit LOR to "a rough index of overall L2 experience". LOR profiles vary widely across studies, for instance. Participants are compared at different points from early exposure to ultimate attainment. They also differ with respect to their daily use of L2. Such confounding variables likely contribute to the conflicting findings.

With respect to AOA, previous research has generally noted a linear decline in attained L2 pronunciation proficiency as a function of AOA when the population under consideration includes both early and late bilinguals (e.g., Flege et al., 1995a, 1999). Findings have proven controversial when analyses are limited to late bilinguals, however. For example, whereas Patkowski (1990) found that AOA was unrelated to adult L2 learners' pronunciation proficiency, Flege, Birdsong, Bialystok, Mack, Sung & Tsukada (2006) reported a significant negative correlation between global accent rating scores and AOA in the context of late bilinguals. The research in this area has generally adopted global measures of foreign accents, ranging from self-reports of proficiency level (e.g., Stevens, 1999) to native speakers' judgements of linguistic elements like segmental composition, syllable structure, and speech rate (e.g., Patkowski, 1990). As with methodological issues regarding LOR, such AOA measures are also prone to confounding effects, resulting in conflicting interpretations. Few studies have examined how late bilinguals' age variables interact to influence objective measures of L2 speech acoustic output (cf. Baker, 2010; Trofimovich & Baker, 2006).

The independent and combined effects of LOR and AOA on late L2 learner's speech development have thus proven to be complex and continue to be poorly understood, especially as regards speech sound production. Certain theoretical accounts of L2 acquisition provide a framework that may help clarify their role in the acquisition of phonetic elements. In what follows, we will briefly introduce three such accounts and their different predictions as to the interaction and consequences of experience and age on the initial and end state of post-pubertal L2 sound learning. They are the Critical Period

Hypothesis, Speech Learning Model, and Cognitive Aging Hypothesis.

Theories of late bilingualism

Critical Period Hypothesis

Some researchers argue that any linguistic performance by late bilinguals is constrained by a loss of plasticity resulting from neural maturation after adolescence (Abrahamsson, 2012; DeKeyser, 2000, 2003; DeKeyser & Larson-Hall, 2005; Johnson & Newport, 1989; Patkowski, 1990; Scovel, 2000). According to this position, young learners' ability to acquire language from mere exposure gradually declines (i.e., a robust AOA effect on early bilingualism), then ceases after puberty, such that older adolescents and adults no longer have access to an assumed language-specific incidental learning mechanism (i.e., limited AOA effects on late bilingualism). Instead, irrespective of their increasing age, post-pubertal SLA relies on some general mechanism for intentional and declarative learning. 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) due to the existence of "important discontinuities between child and adult L2 learners" (Patkowski, 1990, p. 80).

Previous research has indeed noted a general tendency for adult L2 learners to demonstrate quick improvement over the first few months of LOR, followed by a leveling-off, despite additional linguistic input (for a review, see DeKeyser & Larson-Hall, 2005). These findings are in line with suggestions that adult speech sound learning makes use of a general cognitive strategy, such as the proceduralization of declarative information or metalinguistic associations, rather than a language-specialized cognition, associated with naturalistic input and implicit learning (Abrahamson, 2012; DeKeyser, 2003; Paradis, 2009; Ullman, 2004). In this respect, continued improvement of post-pubertal SLA after the early phase of interlanguage development, if any, is not tied to LOR or AOA, but rather to individual differences, such as high language aptitude (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000).

Speech Learning Model

In contrast to the above critical-period account of late bilingualism, Flege (1995, 2003, 2009) maintains that adult L2 foreign accents can be explained by (a) the influence of the L1 sound system and (b) the limited quality and quantity of L2 input for late bilinguals. Not only must adult SLA build on a common phonological space already organized by L1 restrictions (Baker,

Trofimovich, Flege, Mack & Halter, 2008),¹ but they are typically less exposed to L2 compared to young learners (Jia & Aaronson, 2003). Flege hypothesizes that the speech learning capacity necessary for successful L1 speech acquisition remains active throughout life and may be invoked in L2 learning, provided ample exposure to L2 (see also Bialystok, 1997). If we follow this theoretical position, even post-pubertal L2 learners with sufficient L2 input are expected to continue restructuring their phonological representation and succeed in establishing new phonetic categories (see also Best & Tyler, 2007).

Flege and his colleagues have reported evidence that continued L2 input, through increased LOR, is significantly correlated with L2 speech learning improvements, provided the main language of communication is L2 (e.g., university-level international students), and not L1 (e.g., migrant workers) (Flege & Liu, 2001; see also Flege et al., 1997; Flege & Fletcher, 1992; Flege & MacKay, 2004; Piske et al., 2001).² Such findings are interpreted as evidence against a Critical Period Hypothesis which rejects the mediating role of environmental input in late bilingualism.

Although the Speech Learning Model explains how adult L2 learners may develop new phonetic categories depending on the amount and quality of L2 input, it is unclear how it may be applied to the question of separable effects of experiential (LOR) and age (AOA) variables in the context of late bilingualism.³ One possibility is that LOR predicts L2 proficiency regardless of AOA (i.e., late bilinguals with sufficient L2 exposure demonstrate less accented pronunciation, despite varying AOA profiles; Flege et al., 1999; Piske et al., 2001). On the other hand, AOA may be a driving factor as to how much L2 learners can benefit from the additional input and interaction that may be associated with LOR (i.e., late bilinguals' accentedness tends to be correlated with their AOA; Flege et al., 2006). In other words, while proponents

of this position would predict improvements in adult L2 pronunciation as a function of LOR, the role of AOA remains ambiguous.

Cognitive Aging Hypothesis

A third theoretical position rests on the notion of “cognitive aging” (i.e., a decline in working memory, executive control, speech sound processing, or inhibition of task-irrelevant information) as a mediating factor in L2 production, processing and learning (for a review, see Hakuta, Bialystok & Wiley 2003). Birdsong (2005, 2006) links this progressive loss of cognitive functions to the biological (but not maturational) aging process in the brain, such as decreases in brain volume and nigrostriatal dopamine (starting at age 20 years). In the SLA context, for example, Birdsong (2006, p. 32) proposes that the dopamine system plays a key role in “defossilization, an undoing of automatized nontargetlike linguistic performance” as well as “suppressing and supplanting L1 routines”. The expected extent of adult L2 acquisition, according to this position, declines with age, such that AOA may best predict the degree of ultimate attainment. LOR, on the other hand, would not be a useful measure to predict post-pubertal L2 performance.

Three major theoretical positions thus provide fairly straightforward predictions as to the role of LOR and AOA in L2 speech sound acquisition. A strong interpretation of the Critical Period Hypothesis precludes any role for LOR or AOA. A Speech Learning Model would suggest a possible relation between LOR and adult L2 proficiency, without saying much about AOA. A Cognitive Aging Hypothesis predicts declining potential in relation to advancing AOA, regardless of possible LOR effects. Building on this admittedly simplified array of hypothetical constructs, the current study examines the relationship between LOR and AOA in the specific case of late (adult) native Japanese speakers' acquisition of North American English /ɪ/.

English /ɪ/

Within the broad category of sonorants, the phoneme /ɪ/ is further described as a semivowel, approximant, and liquid, when in consonantal position. Acoustically, /ɪ/ can be characterized along various dimensions, such as intensity, duration and spectral bandwidth. A clear delimitation of the phoneme requires, in particular, a description of its formant frequencies. The first two formants of /ɪ/ generally have values within the range of a central rounded vowel /ə/ (Espy-Wilson, 1992). Unlike such vowels, however, the third formant (F3) shows a characteristic drop in center frequency, approaching and sometimes merging with the lower, second formant (F2).

¹ According to Baker et al.'s (2008, p. 319) Interaction Hypothesis, “L1 indeed exerts a powerful influence on adults' ability to learn L2 sounds, and . . . variation from native L2 speaker perception and production is often traceable to their L1” (see also Flege, 1995, 2003).

² Given that international university-level students tend to have a high level of education and are very motivated to become very proficient in English relative to migrant workers, it is unclear which variables (including the amount of exposure to native speakers' input) led to their superior L2 speech performance (see DeKeyser & Larson-Hall, 2005; Flege, 2009).

³ To explain the relatively strong AOA effects in the context of EARLY bilingualism, Flege (2009, p. 184) claims that AOA is a “macrovariable” related to a wide range of affecting variables including frequency of L1/L2 use, state of neurological and cognitive development, and degree of L1 development. He did not specifically explain how L2 performance of LATE bilinguals with fully developed L1 systems who receive consistent native-speaker input can be related to AOA (see Flege et al., 2006).

This can also be described as a narrowing of the distance between F2 and F3. The perceptual effect of the low F3 ($x = 1300\text{--}1950$ Hz) is an ‘r-coloring’ or rhotacization (Espy-Wilson, 1992), a key feature in the perception of the English phoneme (Espy-Wilson, Boyce, Jackson, Narayanan & Alwan, 2000).

With respect to articulatory configuration, English /ɹ/ is produced along a range of possible configurations by native speakers, and that range can even be observed within the same speaker (Delattre & Freeman, 1968). The acoustic effect appears to be a shift in F4/F5 values (Zhou, Espy-Wilson, Boyce, Tiede, Holland & Choe, 2008) and is not directly relevant to F3 lowering nor, by extension, to the perceived feature of rhotacization. Relevant articulatory parameters are better described in terms of vocal tract shaping: three constrictions (labial, palatal, pharyngeal) and a sublingual cavity that is associated with the low third or “r-formant” (Espy-Wilson et al., 2000).

Since the Japanese phonetic system does not include the phoneme /ɹ/, it is posited that Japanese speakers perceptually assimilate English liquids into their native tap category (Guion, Flege, Ahahane-Yamada & Pruitt, 2000). In an analysis on speech sound production of Japanese and English, Lotto, Sato and Diehl (2004) found an F2/F3 distribution of the Japanese tap that overlaps with English /ɹ/ and /l/ distributions, an F2 range that does not reach lower /ɹ/ and /l/ F2 values (presumably because this extends into the formant space of Japanese /w/), and a strong F2/F3 correlation suggesting a consistent F3–F2 relative distance across frequencies. They report F2 values falling between about 1000 and 3000 Hz, and F3 values between 1500 and 3500 Hz (for similar results, see Hattori & Iverson, 2009). Although the transition duration of English /ɹ/ ($x = 50\text{--}100$ ms) is generally longer than that of the Japanese tap ($x = 5\text{--}20$ ms) (Hattori & Iverson, 2009), it has a wide range of natural variation (e.g., certain instances of /ɹ/ can be as short as a Japanese tap) and is not considered a significant acoustic correlate of /ɹ/ (Flege et al., 1995b).

At least two crosslinguistic influences have been identified in Japanese speakers’ acquisition of English /ɹ/ both in perception and production: (a) a preference for F2 variance (vs. F3; effects of tongue retraction), tied to the importance of F2 variance in Japanese (e.g., Japanese tap vs. /w/) (Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann & Siebert, 2003; Lotto et al., 2004), and (b) an overreliance on temporal rather than spectral cues (Iverson, Hazan & Bannister, 2005; Yamada, 1995). Thus, if Japanese learners are in fact overly reliant on F2 distributions and temporal dimensions at the expense of variation in F3 (Lotto et al., 2004), the following framework for Japanese speakers’ acquisition of English /ɹ/ (the Japanese tap → English /ɹ/) may be proposed:

1. A move away from (but not a complete abandonment of) a “Japanese speakers’ default strategy” of F2 (1700–2100 Hz → 900–1500 Hz) and duration (5–20 ms → 50–100 ms) dependency.
2. Attention to new, unfamiliar parameters such as F3 variance (2400–3000 Hz → 1600–1900 Hz).
3. Associated shift in orolingual articulation that includes narrowed labial constriction (for word-initial tokens) and an adequate sublingual cavity (for F3 resonance).

The components of this framework are not expected to occur simultaneously, resulting in important differences in observed productions by Japanese learners of English.

Motivation for current study

As mentioned earlier, some disagreement exists in the literature as to the degree of influence LOR and AOA have on late L2 speech sound acquisition. Flege et al. (1995b), for instance, examined the relationship between LOR and adult Japanese speakers’ production of /ɹ/. Results showed that Japanese learners with 20 years LOR not only outperformed Japanese learners with two years experience, but also managed to fall within the range of native speakers of English. Larson-Hall (2006), by contrast, failed to replicate Flege et al.’s (1995b) study. Whereas Japanese learners with short LOR showed some gain in word-initial /ɹ/ production, experienced Japanese learners with extensive LOR actually showed a decline in L2 speech performance. Based on these results, Larson-Hall suggested that, although the first few months of LOR could facilitate L2 pronunciation development, increasing AOA and chronological age may negatively influence the quality of adult L2 speech production in the long term.

It is possible that such conflicting results stem from complexities inherent in the nature of interlanguage development of /ɹ/. Both previous studies used only perceptual ratings (accentedness and phoneme identification scales) as measures. Perceptual rating methods have been promoted as a gold standard in L2 pronunciation research (Piske et al., 2001), and are by definition the only standard with respect to the relevance for speakers and listeners. Given that the acquisition of /ɹ/ requires the mastery of more than one phonetic segment (using various strategies such as tone retraction, lengthening phonemic duration, and labial/palatal/pharyngeal constrictions), exclusive use of perceptual judgement might not be the most appropriate method to examine changes in the relative weighting of existing cues, such as F1 vs. F2 or duration, or increased awareness and use of new cues – namely lowered F3. For instance, although native speaking listeners mainly rely on F3 variation to perceive English /ɹ/ (e.g., Iverson

et al., 2003), they can also identify /ɪ/ by paying attention to temporal aspects of the sound (i.e., transition duration > 50 ms) even when F3 information is unaltered (i.e., a phonetic trading relation: Underbakke, Polka, Gottfried & Strange, 1988). A more precise assessment of possible interlanguage developmental effects would therefore involve an examination of how Japanese learners approach English /ɪ/ along each acoustic dimension as a function of LOR and AOA.

Another important aspect in measuring the interlanguage development of /ɪ/ is to take into account variation according to task: L2 learners tend to make more pronunciation errors in free speech tasks than formal word reading tasks (Lin, 2003; Rau, Chang & Tarone, 2009). In a review of L2 pronunciation research, Major (2008) points out that L2 learners initially show L1-related errors (e.g., L1 to L2 substitutions), but later begin to show characteristics of an “interlanguage phonology” (i.e., universal errors), involving greater variation in speech sound production across styles and linguistic contexts. In earlier work, Dickerson and Dickerson (1977) made a casual observation that Japanese speakers’ production ability was strongly influenced by styles (i.e., speaking contexts). Although Japanese learners could intelligibly produce /ɪ/ in word lists, the accuracy of their pronunciation performance fell below chance level in free conversation. Larson-Hall (2006) had Japanese learners produce target sounds embedded within a 600-word story in an attempt to control for possible conscious monitoring of those sounds. At the beginning of the study, however, participants were still told that the purpose of the task was to test overall pronunciation. The paragraph reading task also did not rule out conscious attention to pronunciation accuracy and hyper-articulation.⁴

To assess learners’ abilities to spontaneously use certain linguistic structures, L2 research has emphasized the importance of eliciting language within a COMMUNICATIVE context (i.e., they are required to pay simultaneous attention to grammatical, phonological, lexical, and pragmatic aspects of language: 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). As one example, target-word-prompted picture description tasks provide a means of tapping into an individual’s ability to formulate context-appropriate phrases and simultaneously incorporate target productions. With respect to L2 phonology, Rau et al. (2009) made use of such a technique. The authors found that Chinese learners of English mispronounced /θ/ more frequently in a picture description task than in word and sentence reading tasks.

⁴ Larson-Hall (2006) did not find any difference in Japanese speakers’ production of /ɪ/ between word and paragraph reading tasks, ostensibly due to similar task demands across tasks.

This discrepancy in performance was argued to arise from increased demands on linguistic processing without the benefit of increased planning time available during the reading tasks (see also Lin, 2003, for a similar argument regarding consonant cluster production).

Following this line of thought, the current study elicited Japanese speakers’ production of word-initial /ɪ/ along different levels of processing via 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). Subsequently, we examined the effects of LOR and AOA on the development of four acoustic dimensions of /ɪ/ (i.e., F1, F2, F3 and transition duration) across task contexts.

Method

Participants

Japanese learners of English

The data were collected at an English-speaking university in Montreal, Canada. Based on the results of an initial interview, 39 Japanese learners of English who met the following criteria were selected: (a) LOR above 8 months (including intensive exposure to English during that period of time) and (b) daily use of English (they mainly used English at school or work). According to a preliminary language background questionnaire, they had studied English for several years through grammar translation methods in Japan and arrived in Canada at a mean age of 27.3 years (range: 19–39 years, *SD* = 5.4). Mean LOR was 4.1 years (range: 8 months – 13 years, *SD* = 3.7) and mean age was 31.4 years (range 21–43 years, *SD* = 5.8).

Length of residence profile

The participants were categorized into three groups – Short LOR (8 months $\leq x \leq 1$ year), Mid LOR (1 year < $x < 5$ year), Long LOR (5 years $\leq x \leq 13$ years) (for details, see Table 1). The basis for the categorization into Short-, Mid-, and Long-LOR was motivated by prior knowledge from relevant literature: Short- LOR (8 months $\leq x \leq 1$ year) for those who partially or fully complete the initial quick improvement over the first year of LOR (Munro & Derwing, 2008); (b) Mid-LOR (1 < $x < 5$ years) for those who do not have much room for the rate of learning advantage (Munro, 1993); and (c) Long-LOR (> 5 years) for those who are assumed to have reached their ultimate attainment (Johnson & Newport, 1989; Thompson, 1991). For a similar categorization, see Baker (2010) and Trofimovich and Baker (2006). Since the group analysis (i.e., ANOVA) is subject to the influence of the categorization described here, L2 performance was

Table 1. Characteristics of four Japanese groups and one English group.

	Group									
	Japanese-Baseline (<i>n</i> = 13)		Short-LOR (<i>n</i> = 13)		Mid-LOR (<i>n</i> = 13)		Long-LOR (<i>n</i> = 13)		English-Baseline (<i>n</i> = 13)	
	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range	<i>M</i> (<i>SD</i>)	Range
Testing age (years)	28.2 (5.7)	21–40	28.9 (7.3)	21–40	30.5 (3.6)	25–36	34.9 (4.1)	28–43	21.3	20–28
AOA (years)			28.1 (0.1)	19–35	27.6 (1.2)	21–34	26.3 (2.5)	21–38		
LOR in Canada (years)			0.9 (0.1)	0.7–1.0	2.9 (1.2)	1.2–4.7	8.6 (2.5)	5–13		
Gender	11 females, 2 males		10 females, 3 males		10 females, 3 males		12 females, 1 male		6 females, 7 males	

AOA = age of acquisition; LOR = length of residence

also investigated using correlation analysis (see the results section).

Whereas most of the participants with LOR around one year were studying abroad to improve their English skills for future business and academic careers, those with LOR more than one year were either graduate students at English-speaking universities or full-time workers who dealt mainly with English-speaking customers.

Age of acquisition profile

The AOA profile of the 39 learners was distributed as follows: (a) Early adulthood ($19 \leq x \leq 24$ years, $n = 15$), (b) Mid adulthood ($25 \leq x \leq 30$ years, $n = 15$), and (c) Late adulthood ($31 \leq x \leq 39$ years, $n = 9$). Due to the unequal number of participants for each category, the predictive role of AOA was examined via correlation analysis.

Baseline speakers

Data were collected from two control groups to establish baseline acoustic characterizations for Japanese and native English speakers' production of /ɹ/. With respect to the Japanese Baseline, 13 native speakers of Japanese who had just arrived in Canada with little L2 experience (LOR < 1 month) were recruited at private language schools in downtown Montreal. They completed 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: 28.5 years). To establish the English Baseline, 13 native English undergraduate students at an English-speaking university in Montreal completed the three oral tasks (mean age: 21.3 years).

Type of target words

All twenty words used in the three oral tasks were Consonant–Vowel–Consonant (CVC) word-initial /ɹ/

singletons, except for the token *Ryan* (CVVC). Efforts were made to control lexical factors (i.e., text frequency and familiarity of target words) which are known to exert an influence on L2 segmental production (Bundgaard-Nilsen, Best & Tyler, 2011). According to the results of the vocabulary profiling (Cobb, 2011), 20 words fell into the first 2000 most frequent spoken words, except *ram* and *Ryan*. None of the participants reported unfamiliarity with the two words; we therefore assumed that the effects of lexical factors on Japanese speakers' production of /ɹ/ were minimal.

In order to control for coarticulation effects on prevocalic position (Japanese speakers tend to have difficulty in producing /ɹ/ preceding front vowels such as /i/ and /e/: see Flege et al., 1995b), the following vowels were evenly distributed in each task: 50% for singletons with front vowels, (/i, eɪ, ε, æ/), 50% for singletons with central and back vowels (/ʌ, u, oʊ, ɔ, aɪ/).⁵ The test tokens are summarized in Table 2.

Task description

Timed picture description

In this task, participants were first given a picture with three word prompts, one of which was a target word. For example, they were given three word cues – “rain”, “table”, “drive way” – to describe a picture where rain was falling on the top of a table in a driveway. After five seconds of planning time, they described the picture.

There were eight pictures in total (including four distracters). In order to elicit spontaneous productions without drawing attention to the target sound, four (instead of eight) words were used in this task. The four target

⁵ Vowels and diphthongs have been categorized according to Canadian English normative descriptions (Boersma & Pater, 2007; Labov, Ash & Boberg, 2006).

Table 2. *Twenty tokens in the controlled and spontaneous production tests in relation to Following Vowel conditions.*

A. Timed picture description		
Phonemic category	Following vowel	Example token
[front]	/i, eɪ/	<i>read, rain</i>
[central, back]	/oʊ, ɔ/	<i>road, rock</i>
B. Sentence reading		
Phonemic category	Following vowel	Example token
[front]	/i, eɪ, ε, æ/	<i>read, rain, red, race</i>
[central, back]	/ʌ, oʊ, ɔ, aɪ/	<i>run, Ryan, road, wrong</i>
C. Word reading		
Phonemic category	Following vowel	Example token
[front]	/i, eɪ, ε, æ/	<i>read, race, red, ram</i>
[central, back]	/ʌ, u, aɪ/	<i>rough, right, root, room</i>

words were *read, rain, road, and rock*. To familiarize speakers 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/.

Sentence reading

In this task, participants read five target sentences, together with three distracter sentences:⁶

- (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.

Word reading

In this task, participants read a list of 25 words which consisted of eight target words and 17 distracters. These target words were *read, red, race, ram, rough, right, root, and room*. The distracters included a number of easy and difficult sounds (e.g., voiceless stops, interdental fricatives).

Procedures

All 52 Japanese and 13 English speakers completed the three oral tasks, a language-background questionnaire and

interview. 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 order to avoid conscious focus on pronunciation in the context of TPD, data were collected as follows: (a) all participants were told that the tests were to measure general oral English skills including grammar, lexical, and pragmatic use of language (they were not told that the focus of the project was their pronunciation until they finished all of the recordings); and (b) the tests took place in the following order: TPD → SR → WR. Due to the nature of the task, performance on the controlled production tests (WR, SR) could still have been influenced by conscious monitoring of /ɹ/ pronunciation (reading a list of words and sentences enables – and possibly primes – participants to focus on the correct production of words).

Acoustic analyses

Procedure

Following Flege et al.'s (1995b) procedure, formant values were determined using the linear predictive coding routine available in Praat (Boersma & Weenink, 2011). Word onset was visually identified from the spectrographic display of each token. A cursor was placed at the point where all three formant bands were clearly observed. Target phonemes embedded in continuous speech were identified by local peaks in F3 (F3 of the preceding sounds tends to continue to decline toward the beginning of the word because F3 of /ɹ/ is relatively low). Transition duration was measured from the beginning point of the F1 transition to the endpoint of the F1 or F3 transitions (Hattori & Iverson, 2009).

⁶ 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.

Normalization

Since spectral information (i.e., F3, F2, F1 values) varies considerably due to anatomical differences in individual vocal tract length, raw acoustic values were adjusted according to the following normalization procedure (for details, see Lee, Guion & Harada, 2006; Yang, 1996).

Since F3 of open vowels ($F1 > 600$ Hz) is a reliable indicator of vocal tract length (Yang, 1996), a mean F3 value of /æ/ elicited from three monosyllabic words in WR (i.e., *man*, *map*, *ram*) was calculated for each participant. One female native speaker of English was randomly selected as a reference, and her mean F3 value (i.e., 3011 Hz) was divided by those of the other participant to provide their own k factors. Then, we multiplied all formant values (F3, F2, F1) of /ɪ/ for each participant by the individual k factor, respectively. Furthermore, all acoustic values in Hertz were converted into Bark using the formula described in the Praat manual (Boersma & Weenink, 2011; Schroeder, Atal & Hall, 1979) in order to reduce the nonlinear relationship between the formant frequencies and the corresponding perceived semivowel quality.⁷

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

Acoustic values (F1, F2, F3, transition duration) of 20 words were measured and averaged according to task ($n = 8$ words for WR, $n = 8$ words for SR, $n = 4$ for TPD).

Results

Given the pattern of distinct formant distributions between /ɪ/ and the Japanese tap discussed earlier (Hattori & Iverson, 2009; Lotto et al., 2004), the following benchmark (Japanese tap → English /ɪ/) was used to interpret the acoustic results:

- F3: 14.10–15.70 Bark (2400–3000 Hz) for the Japanese tap → 11.40–12.60 Bark (1600–1900 Hz) for English /ɪ/
- F2: 11.80–13.20 Bark (1700–2100 Hz) for the Japanese tap → 7.90–11.00 Bark (900–1500 Hz) for English /ɪ/
- Transition duration: 5–20 ms → 50–100 ms

⁷ We tested and confirmed the validity of the proposed normalization procedure with the English baseline dataset which noted pre-existing significant differences in formant frequency values between genders (six females vs. seven males). Originally, a set of independent-samples t -tests revealed significantly higher acoustic values for females than males in F3 ($p < .001$), F2 ($p = .008$), and F1 ($p = .007$). This gender difference became non-significant after the normalization ($p = .500$ – $.800$).

In the following analyses, any reduction in F3 is interpreted as an indication of the activation of the new cue, and any change in F1, F2 and transition duration an index of the resetting of the existing cues.

ANOVA

Baseline

First, we examine whether and to what degree Japanese learners of English need to make efforts to approximate English /ɪ/ in the four acoustic domains (F3, F2, F1, transition duration) under various task conditions by comparing the two baseline groups. The four acoustic components of /ɪ/ 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 (a) F3, $F(1,24) = 162.060$, $p < .001$; (b) F2, $F(1,24) = 36.810$, $p < .001$; and (c) transition duration, $F(1,24) = 177.119$, $p < .001$. Main effects of Task were not significant in any acoustic domains ($p > .05$). The results of Bonferroni post-hoc comparisons further revealed that the Japanese-Baseline /ɪ/ production was significantly different from the English-Baseline, regardless of task conditions, with respect to higher F3 ($M = 15.22$ vs. 12.47 Bark), higher F2 ($M = 11.63$ vs. 9.94 Bark) and a shorter transition duration ($M = 28.33$ vs. 88.19 ms) ($p < .001$). That is, the Japanese learners with LOR less than 1 month tended to substitute the Japanese tap for English /ɪ/ regardless of the task condition (see Figures 1–4).⁸

New cue (F3)

Next, we turn our attention to the interlanguage development of /ɪ/ for the three groups of Japanese learners (Short-, Mid-, and Long-LOR) with respect to the baseline groups (Japanese, English) in terms of the use of the new cue (F3). A two-way ANOVA with one between-group factor (five Groups: Short-, Mid-, and Long-LOR: Japanese- and English-Baseline) and one repeated measure (three Tasks: WR, SR, TPD) was computed on the learners' F3 values as the dependent variable. The results show a significant Group × Task interaction effect, $F(4,60) = 3.241$, $p = .018$.

Bonferroni multiple comparisons of task effects revealed significantly higher F3 values for the Short-LOR group in TPD ($M = 14.59$ Bark) versus WR and SR conditions ($M = 14.01$ Bark, $p < .001$; $M = 14.19$ Bark,

⁸ Their substitution of the Japanese tap for English /ɪ/ at the beginning of LOR (< 1 month) indicates the lack of adequate instruction and awareness towards accurate pronunciation in the EFL education system in Japan. This finding is not surprising at all, because pronunciation teaching has received little research attention in the field of second language education, and practitioners tend to be “heavily influenced by commonsense intuitive notions” (Derwing & Munro, 2005, p. 380).

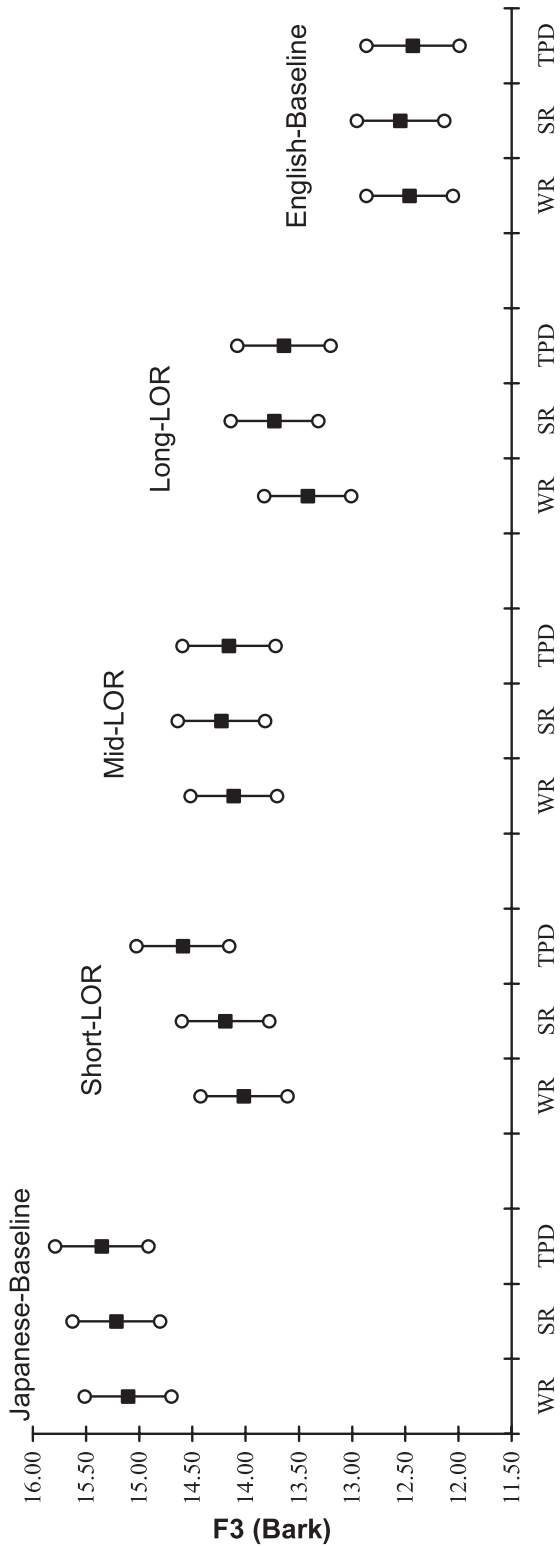


Figure 1. 95% confidence intervals and mean values of F3 (Bark) in Group \times Task. The Short-LOR group noted significant simple Task effects (lower F3 for WR, SR than TPD). English speakers generally produced lower F3 than Japanese learners. The Mid- and Long-LOR groups outperformed the Japanese-Baseline group at all task contexts. The Long-LOR group produced lower F3 than the Short-LOR group in TPD.

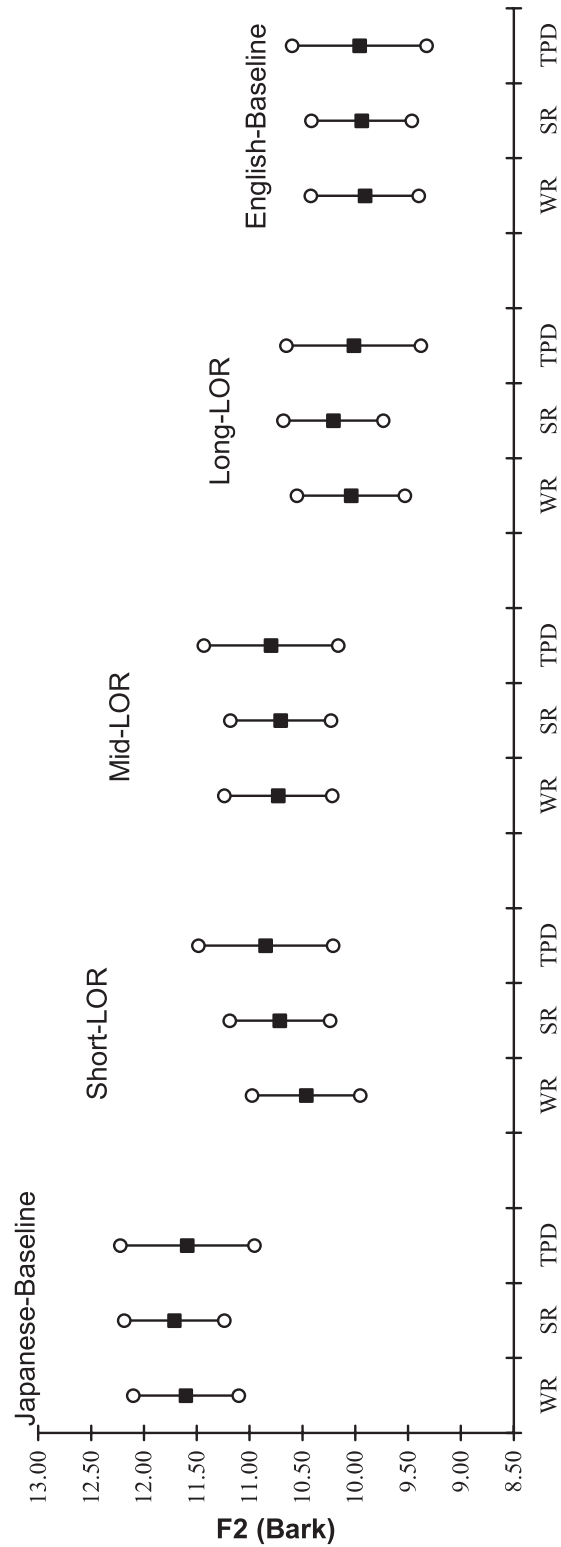


Figure 2. 95% confidence intervals and mean values of F2 (Bark) in Group \times Task. Whereas the Japanese-Baseline produced higher F2 values than the English-Baseline group, no significant group difference was found between the Japanese learners (Short-, Mid-, and Long-LOR) and the English-Baseline at a $p < .05$ level.

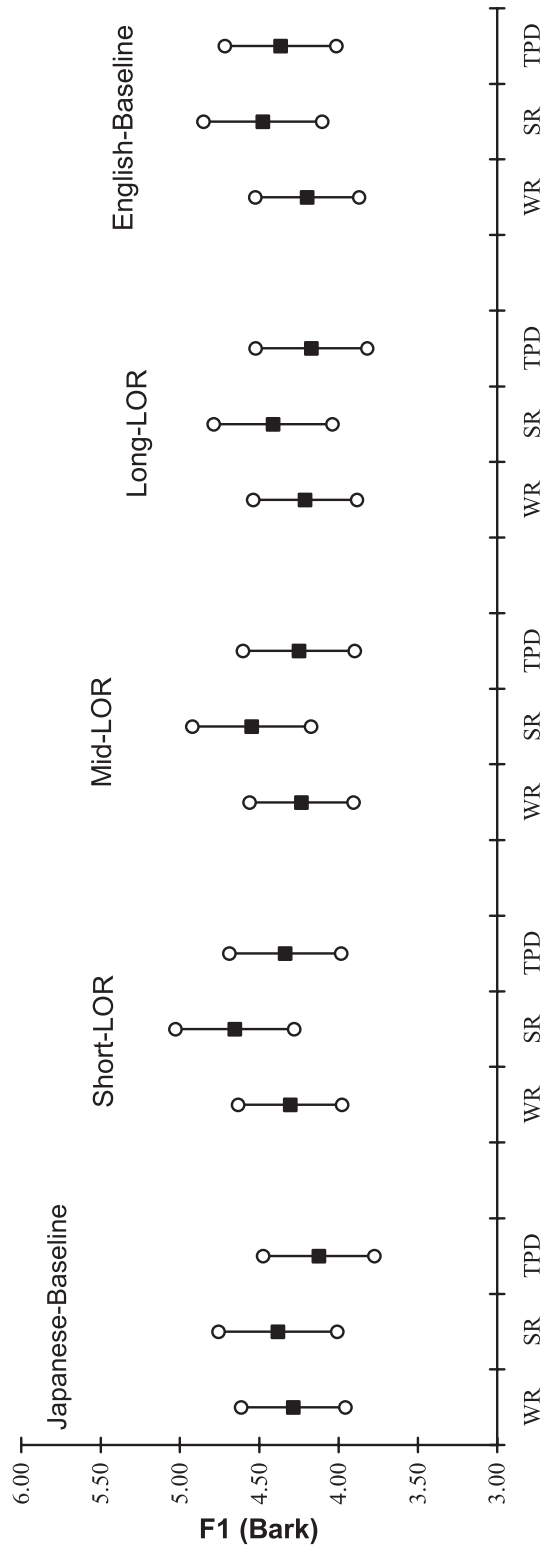


Figure 3. 95% confidence intervals and mean values of F1 (Bark) in Group \times Task. Neither significant Task nor Group effects were observed in any contexts at a $p < .05$ level.

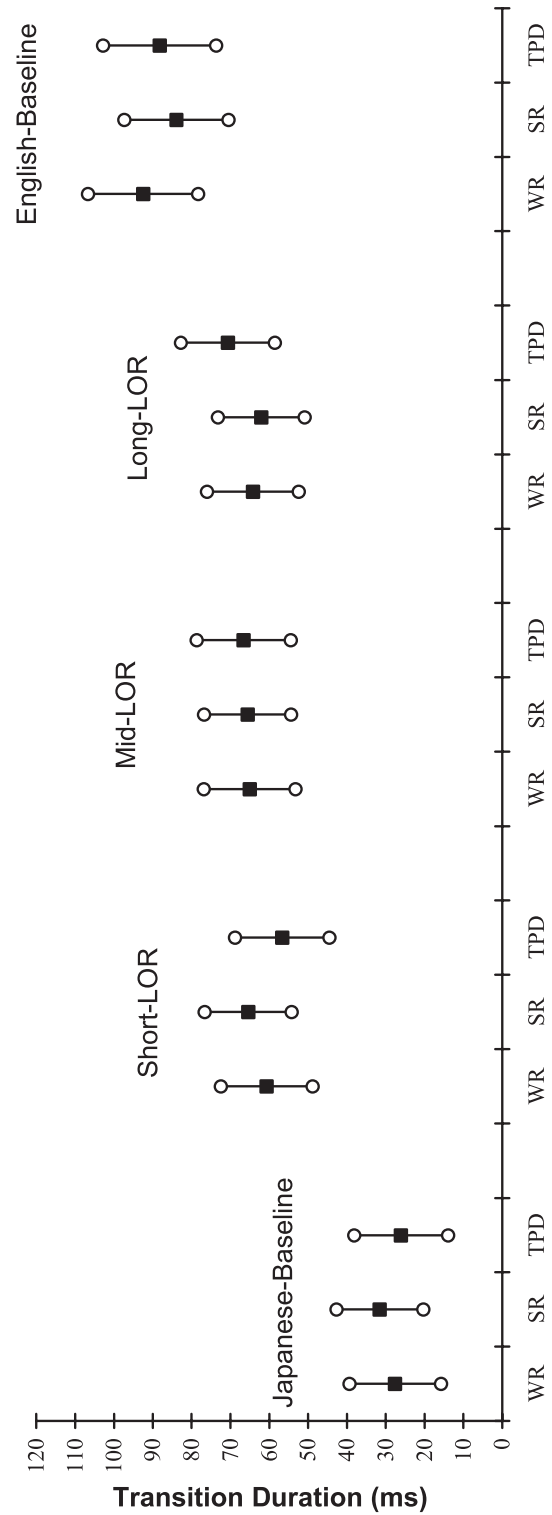


Figure 4. 95% confidence intervals and mean values of transition duration (Bark) in Group \times Task. Neither significant Task nor Group effects were observed in any contexts at a $p < .05$ level. Whereas the Japanese-Baseline produced longer transition duration than the English-Baseline group, no significant group difference was found between the Japanese learners (Short-, Mid-, and Long-LOR) and the English-Baseline at a $p < .05$ level.

$p = .005$, respectively). This task-specific difference was not observed for the other groups (i.e., Mid-, Long-LOR; Japanese-, English-Baseline) ($p > .05$).

Bonferroni multiple comparisons also showed significant group differences according to task condition. All of the Japanese groups produced significantly higher F3 values (> 12.60 Bark) than the English-Baseline (< 12.60). Whereas the Mid- and Long-LOR groups differed significantly in their F3 values from the Japanese-Baseline in all task contexts ($p < .001$), the Short-LOR group outperformed the Japanese-Baseline only in WR and SR.

Turning to within-Japanese comparisons (8 months $<$ LOR $<$ 13 years), only the Long-LOR group produced significantly lower F3 values ($M = 13.64$ Bark) when compared to the Short-LOR group ($M = 14.59$ Bark) in the TPD condition ($p = .013$). No differences were identified between Groups in the WR or SR conditions.

In short, (a) F3 frequencies of Japanese learners did not reach native-like values in any context, (b) a drop in F3 peak frequency was observed in Japanese learners with at least one year LOR in the controlled reading conditions only, and (c) F3 values reached their lowest values in the group of subjects with greater than five years of LOR, in both controlled and more spontaneous speech-elicitation conditions. The descriptive results of F3 values are plotted in Figure 1.

Minor cues (F2, F1, transition duration)

To investigate how LOR was associated with a change of existing cues (change in F2, F1, transition duration), the following subsections present the results of inferential statistics for each acoustic dimension.

F2

A two-way ANOVA (Group \times Task) indicated a significant main effect only of Group, $F(4,60) = 7.303$, $p < .001$. Bonferroni multiple comparisons identified a significant difference in F2 between baseline Japanese and English groups, but failed to show a difference between Japanese speakers with some LOR and the English group ($M = 10.09$ – 10.74 Bark, $M = 9.94$ Bark respectively; $p > .05$). The descriptive results of F2 values are plotted in Figure 2.

F1

A two-way ANOVA (Group \times Task) did not show any significant contrasts in the domain of F1 ($p > .05$). The descriptive results of F1 values are plotted in Figure 3.

Transition duration

A two-way ANOVA (Group \times Task) yielded a significant main effect only of Group, $F(4,60) = 15.264$, $p < .001$. According to Bonferroni multiple comparisons, whereas both the Japanese-Baseline and Short-LOR groups produced /ɪ/ with significantly shorter transition ($M = 28$ ms and 61 ms) than the English-Baseline group

($M = 88$ ms) ($p = .013$, $p < .001$, respectively), no significant difference was observed between the Mid- and Long-LOR groups ($M = 65.66$ ms and 65.48 ms) and the English-Baseline ($p > .05$). The descriptive results of transition duration are plotted in Figure 4.

Taken together, the results imply that, with respect to the acoustic properties under consideration, all Japanese learners (even those with approximately one year of LOR) had already (a) moved away from the Japanese tap category (with its higher F2 and short duration) and (b) reached native-like values for English /ɪ/ F2 frequency and transition duration.

Correlation analyses

Since grouping the Japanese learners into Short-, Mid-, and Long-LOR could have affected ANOVA results, we completed a set of correlation analyses without these predetermined categories (39 speakers, LOR $>$ 8 months). LOR and AOA were not significantly correlated among the Japanese speakers in the current study, $r(37) = -.229$, $p = .160$. Reported correlations may thus be expected to provide insight into specific effects LOR or AOA may have on measured acoustics individually.

Primary cue (F3)

LOR was found to be negatively correlated with F3 ($n = 39$) (the longer participants stayed in Canada, the lower their F3 tended to be) in WR at a $p < .05$ level, $r(37) = -.356$ ($p = .026$), and in TPD at a $p < .01$ level, $r(37) = -.476$, $p = .002$. Although AOA was positively correlated with F3 (the older they became, the higher their F3 values tended to be), this relationship was not statistically significant, $r(37) = .160$ – $.260$ ($p > .05$).

Minor cues (F2, F1, transition duration)

In the context of TPD, F2 was moderately related to LOR, $r(37) = -.350$, $p = .029$ (the longer Japanese learners stayed in Canada, the lower their F2 values tended to be), as well as strongly connected with AOA, $r(37) = -.413$, $p = .009$ (the earlier Japanese learners arrived in Canada, the lower their F2 values tended to be). In addition, the correlation between transition duration and AOA approached significance in SR, $r(37) = -.289$ ($p = .074$) (the earlier Japanese learners arrived in Canada, the longer their phoneme transition length tended to be). Yet, F1 was not significantly related to LOR or AOA in any context ($p > .05$).

Taken together, the results suggest that (a) LOR is strongly related to the primary cue (F3), such that Japanese speakers may show continued development toward native-like ranges, whereas (b) AOA is associated with the minor cues (F2, transition duration) that appear to be more relevant at early stages of L2 speech sound acquisition. The results are summarized in Table 3.

Table 3. Correlation analyses (F3, F2, F1, duration vs. LOR, AOA).

Acoustic dimensions		LOR	AOA
F3	WR	$r = -.356 (p = .026)^*$	$r = .160 (p = .331)$
	SR	$r = -.315 (p = .051)$	$r = .264 (p = .104)$
	TPD	$r = -.476 (p = .001)^{**}$	$r = .201 (p = .321)$
F2	WR	$r = -.270 (p = .097)$	$r = .123 (p = .456)$
	SR	$r = -.254 (p = .119)$	$r = .246 (p = .131)$
	TPD	$r = -.350 (p = .029)^*$	$r = .413 (p = .009)^{**}$
F1	WR	$r = -.103 (p = .534)$	$r = .075 (p = .648)$
	SR	$r = -.122 (p = .458)$	$r = .014 (p = .932)$
	TPD	$r = -.088 (p = .594)$	$r = .120 (p = .429)$
Transition duration	WR	$r = .038 (p = .814)$	$r = -.253 (p = .120)$
	SR	$r = .063 (p = .702)$	$r = -.289 (p = .074)$
	TPD	$r = .199 (p = .224)$	$r = -.236 (p = .148)$

LOR = length of residence; AOA = age of acquisition; WR = word reading; SR = sentence reading;

TPD = timed picture description

* = statistical significance at a $p < .05$ level; ** = statistical significance at a $p < .01$ level

Discussion

To support ongoing investigations into the role of LOR and AOA in the acquisition of L2 phonology, the current study measured acoustic indices of word-initial /ɪ/ production accuracy, under different task conditions, in post-pubertal Japanese learners of American English. Third formant frequency measures of participants' productions on a picture-description task were associated with LOR, but not AOA. This distinction was not observed on more controlled speech-elicitation tasks. Other acoustic dimensions (F2 and transition duration) approached native-like values for all groups, in which case AOA, and not LOR, proved to be the relevant variable.

Task effects

Secondary acoustic cues for word-initial /ɪ/ production by Japanese speakers of English did not differ significantly across task conditions. The primary cue of lowered F3 frequency did, however, and this in particular for inexperienced Japanese participants (Short-LOR: 8 months $\leq x \leq$ 1 year). Following Major's (2008) proposal for task variation in interlanguage phonology, the results suggest three distinct developmental patterns: (a) Japanese learners initially use the Japanese tap regardless of task contexts (i.e., L1 substitution errors), (b) they begin to produce more target-like exemplars of /ɪ/ at a controlled speech level, but have not yet generalized to spontaneous speech contexts (i.e., universal errors), and (c) they produce intelligible exemplars of /ɪ/ with little variance across all task conditions after the first few years of LOR (i.e., acquisition).

Subjects' better performance on reading tasks may have resulted from having sufficient time to access explicit knowledge of relevant articulatory gestures (Lin, 2003; Rau et al., 2009). Conversely, spontaneous speech tasks require speakers to attend to various domains of language under time pressure, limiting the amount of attentional resources they can use to produce /ɪ/. In this respect, L2 performance under more cognitively demanding tasks may better reveal the present state of learners' representational system and processing abilities, excluding effects of conscious monitoring otherwise present (i.e., AUTOMATIZED KNOWLEDGE: Segalowitz, 2003). This in turn suggests that task variation should be taken into account and analyzed separately when addressing the effects of factors like LOR on adult SLA.

Length of residence effects

Japanese learners' production of /ɪ/ shows significant improvement only during the first year of LOR, especially under controlled production tests (DeKeyser & Larson-Hall, 2005; Larson-Hall, 2006). By taking into account task variance (controlled vs. spontaneous) and multiple cue weightings (the resetting of existing cues and activation of new cues), however, the results of the current study further revealed two interesting findings. First, Japanese participants could be differentiated according to LOR by measuring F3, especially within the context of a picture description task. An analysis of variance indicated that LOR was a significant factor of primary cue (F3) acquisition. An additional correlation analysis further identified a relationship between LOR and F3

and F2. In both cases, this was only when performance was assessed in the picture description task. Second, both inexperienced and experienced Japanese participants demonstrated native-like acoustic values for minor cues of /ɪ/ (F2, transition duration). An analysis of variance could not distinguish participants from the English-Baseline as a function of LOR.

The plateau in performance despite increased LOR, observed from a gross analysis of the data, appears to be in line with predictions of the Critical Period Hypothesis. Controlling for effects such as task elicitation method, however, it is possible to draw a distinction in performance as a function of LOR. The results from the correlation analyses therefore fail to support a Critical Period Hypothesis, which rejects both continuous LOR effects beyond the early phase of SLA and the attainment of native-like proficiency. The change in performance is likewise not in line with the Cognitive Aging Hypothesis which, as stated in the introduction, assumes LOR would not be a useful measure for predicting late L2 acquisition. These findings are, however, in line with Flege's (1995, 2003, 2009) view that late bilinguals may continue to show improved L2 phonological skills later in life as L2 experience increases. The Japanese learners in the current study who had sufficient opportunities to interact with native speakers showed some evidence of continued flexibility in phonological acquisition (Flege & Liu, 2001). Developing upon the hypothesis of the Speech Learning Model, the results of the current study suggest that acquisition of L2 phonetic cues in late learners is not unidimensional: In order to provide more detailed suggestions of how such experience effects may take place among adult L2 speech learning, any measure of production (or perception) must be able to take into account changes in psycholinguistic processing and various acoustic domains in segmental learning as the learner becomes more proficient (Major, 2008).

The findings presented here suggest that Japanese speakers of English quickly master acoustic aspects of /ɪ/ that partially overlap with their L1, even without much L2 experience. Namely, a small amount of L2 experience appears to be enough to change the relative weights of the already-existing acoustic dimensions for differentiating English /ɪ/ from the Japanese tap, by reducing F2 dimensions (i.e., /w/-like production) and lengthening the duration of the segment (i.e., 5–20 ms → 50–100 ms). Similar conclusions have been forwarded for native English speakers learning length distinctions in Swedish vowels (McAllister et al., 2002) and spectral information (but not lexical tone) in Mandarin vowels (Gottfried & Suiter, 1997). In both cases, these successfully-acquired L2 features are used to signal phonological contrasts in their L1 (English). To summarize, experience effects on the resetting of the existing cues can be observed within the first few years of LOR.

In contrast, a larger amount of L2 experience might be needed to establish a new phonetic representation of spectral information (in this case, the primary F3 cue), as well as activate relevant articulatory configurations (i.e., labial/palatal/pharyngeal constrictions). Importantly, such LOR effects can be discerned by comparing these measures across tasks where a distinction can be drawn between 'conscious hyper-articulation' and generalization to more natural speaking contexts. The former is possible during 'easy' tasks that allow the speaker to directly monitor his or her speech sound production. The latter may be inferred from performance in tasks that tax working memory. This in turn suggests that additional L2 experience (i.e., communicative use of target linguistic features in real-life speaking conditions beyond the first few years of LOR) plays an appreciable role in allowing late bilinguals to attain a more robust representation of the new phonetic cue (i.e., qualitative change), as well as various levels of processing abilities (i.e., quantitative change) to produce /ɪ/ in a more native-like and AUTOMATIC manner.

Age of acquisition effects

Unlike the LOR effects, age of acquisition was identified as a marginally significant predictor of the acquisition of secondary cues (F2, transitional duration) and a poor predictor of the primary cue (F3) among the adult Japanese learners in this study. Changes in secondary cues in approximating English /ɪ/ apparently plateaued for individuals with one or more years LOR. This was expected as these cues are already in place as L1 contrasts. The non-native, primary cue did not show a clear levelling-off effect, however, even for individuals with greater than five years LOR (proposed as an upper limit for late bilingual ultimate attainment by Johnson & Newport, 1989, and Thompson, 1991).

When considering theoretical implications, it is tempting to interpret these results in light of the Cognitive Aging Hypothesis, since it allows for specific predictions with respect to age, but the results are not clear enough to make any strong claims about the hypothesis. Since age-related effects may be most evident once an L2 learner has reached a plateau (Birdsong 2005, 2006), we could speculate that AOA effects are anticipated largely for those variables that show little change with LOR – which is in fact the case for F2 in the picture description task. Given that we do not see a similar pattern across all measures (or at least all secondary cues) and that our subjects' LOR ranged from 8 months to 13 years, we are limited to the conclusion that the data provide, at best, mild support for the Cognitive Aging Hypothesis. As concerns the Speech Learning Model, these results are not directly relevant since it makes no clear predictions as to the role of AOA in late bilingualism. They do, however, argue against a

strong interpretation of the Critical Period Hypothesis which, as mentioned earlier, rejects the possible role of LOR or AOA in adult L2 learners. That is, although the theoretical account assumes a fundamental and qualitative difference between adult (limited AOA effects) and child SLA (robust AOA effects), it cannot explain why the current study found AOA to be predictive of late Japanese learners' production of F2 and transition duration despite having passed an assumed critical period.

Another variable to be addressed is the type of linguistic target which, many researchers argue, influences the extent of AOA effects on late bilingualism (Baker, 2010; Birdsong, 2005, 2006; Flege et al., 1999; Trofimovich & Baker, 2006). In a series of experiments with late Korean–English bilinguals, Baker and Trofimovich (Baker, 2010; Trofimovich & Baker, 2006) proposed that AOA can be a determinant of learning when target phonetic features are related to the acquisition of non-salient, redundant, or/and infrequent phonetic cues with less communicative value. For example, examining the acquisition of final stop consonants, Baker (2010) found that the mastery of not only a primary cue (i.e., preceding vowel duration), but also a minor cue (i.e., closure duration) was highly related to AOA ranging from 21 to 29 years. Similarly, Derwing and Munro (in press) illustrated that AOA was predictive of accentedness (i.e., phonological native-likeness of utterances) but not comprehensibility (i.e., how easy it is to understand what they say) of experienced ESL bilinguals' extemporaneous production. LOR, on the other hand, may facilitate the acquisition and use of other phonetic features which are perceptually salient and of high communicative value (e.g., Trofimovich & Baker, 2006, for the acquisition of stress-timing).

Given that developing F3 representation is a necessary (but not sufficient) condition for producing /ɪ/ (e.g., Japanese speakers' /ɪ/ production without any F3 reduction would be highly unintelligible to native speakers of English), it seems reasonable to assume that the acquisition of the primary acoustic cue for /ɪ/ is better understood in relation to LOR than to AOA. Compared to F3, however, the resetting of F2 and transition duration might have less of an impact on native speakers' perception of /ɪ/ (Flege et al., 1995b; Iverson et al., 2003; see also Underbakke et al., 1988). These secondary acoustic domains should be investigated in relation to AOA.

Conclusion

The current study examined whether and to what degree L2 input, in terms of length of residence (LOR) and the time of first intensive exposure to L2 (AOA), can be advantageous for, or detrimental to, Japanese adult L2 pronunciation of English /ɪ/. Separate analyses were

conducted on four acoustic dimensions of Japanese speakers' production of /ɪ/ (F3, F2, F1, transition duration) under three task conditions (WR, SR, TPD). The results led to three broad conclusions. First, whereas Japanese learners demonstrated native-like values for dimensions also relevant in L1 (F2 and transition duration) within approximately one year of LOR, more accurate production of the primary acoustic cue (F3) was evident only in individuals with longer LOR. Second, AOA effects were related to the secondary cues to some degree, in line with possible effects of cognitive aging on attained L2 proficiency, described as "postmaturational decline in sensitivity" (Birdsong, 2006, p. 119). Third, LOR effects were particularly evident for the primary cue when performance was measured at a spontaneous-speech level (i.e., timed picture description), indicating that LOR may be associated with the development of more robust L2 speech sound mechanisms that allow for more accurate productions in less constrained contexts. This in turn supports "the predictive power of L2 input" even beyond a first few years of LOR (Flege, 2009, p. 188).

To close, we would like to emphasize that the correlation between LOR and F3 values was moderate ($r = .300-.500$), underlining the fact that LOR remains one of many factors relevant to adult L2 speech production development. In addition, the fact that the role of LOR was found to be significant under LIMITED conditions suggests that future studies of this kind need to carefully control for such variables as different acoustic cues and speech elicitation method. Assessing L2 phonology within a continuum from controlled to spontaneous production provides a means to gain insight into changing speech sound representation and production. Lastly, we should acknowledge that the timed picture description task, while more ecologically valid than word reading, is not an indication of truly spontaneous speech, nor, for that matter, of implicit knowledge. Few studies have examined how learners produce specific segmental sounds in a spontaneous manner, due in part to "the inherent difficulty in analyzing conversational speech under controlled conditions" (Piske, Flege, MacKay & Meador, 2011, p. 197). In this respect, we hope to see more research elaborate on valid and reliable measures of the complex phenomenon that is L2 speakers' changing performance across controlled and automatic communicative contexts.

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.

- 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., & 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.
- Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second Language Research*, 13, 116–137.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In Kroll & de Groot (eds.), pp. 109–127.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56, 9–49.
- Boersma, P., & Pater, J. (2007). Constructing constraints from language data: The case of Canadian English diphthongs. Handout, 38th Annual Meeting of the North East Linguistic Society (NELS 38), Ottawa. <http://www.fon.hum.uva.nl/paul/presentations/CanadianRaisingNELS2007.pdf> (retrieved December 2, 2012).
- Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer. <http://www.praat.org> (retrieved September 1, 2011).
- 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.
- Bundgaard-Nilsen, R., Best, C., & Tyler, M. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, 33, 433–461.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34, 372–387.
- Cobb, T. (2011). The complete lexical tutor. <http://www.lex tutor.ca/vp/> (retrieved September 1, 2011).
- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499–533.
- 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 Kroll & de Groot (eds.), pp. 88–108.
- 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., & Munro, M. (in press). The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning*.
- 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. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84, 70–79.
- Flege, J. (1995). Second language speech learning. In Strange (ed.), pp. 233–277.
- 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. (2009). Give input a chance! In T. Piske & M. Young-Scholten (eds.), *Input matters in SLA*, pp. 175–190. Clevedon: Multilingual Matters.
- 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., 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., & Fletcher, K. (1992). Talker and listener effects on the perception of degree of 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., & MacKay, I. [R. A.] (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1–34.
- Flege, J., Munro, M., & Fox, A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95, 3623–3641.
- Flege, J., Munro, M., & MacKay, I. R. A. (1995a). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125–3134.
- Flege, J., Takagi, N., & Mann, V. (1995b). 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.
- Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin tones. *Journal of Phonetics, 25*, 207–231.
- 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.
- 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., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, 847–857.
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*, 3267–3278.
- 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.
- Kroll, J. F., & de Groot, A. M. B. (eds.) (2005). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford & New York: Oxford University Press.
- Labov, W., Ash, S., & Boberg, C. (2006). *The Atlas of North American English: Phonetics, phonology, and sound change*. New York: Mouton de Gruyter.
- Larson-Hall, J. (2006). What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /r/ 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 /r/ and /l/. In J. Slifka, S. Manuel & M. Matthies (eds.), *From sound to sense: 50+ years of discoveries in speech communication*, pp. C381–386. <http://www.rle.mit.edu/soundtosense/conference/pdfs/fulltext/Saturday%20Posters/SB-Lotto-STS.pdf> [Online conference proceedings]
- 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.
- 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. (1993). Productions 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.
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.
- Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics, 11*, 73–89.
- Piske, T., Flege, J., MacKay, I. [R. A.], & Meador, D. (2011). Investigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul & K. Dziubalska-Kolaczyk (eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010*, pp. 195–205. Switzerland: Peter Lang.
- Piske, T., MacKay, I. [R. A.], & Flege, J. (2001). Factors affecting degree of foreign accents in an L2: A review. *Journal of Phonetics, 29*, 191–215.
- Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning, 30*, 271–287.
- Rau, D., Chang, A., & Tarone, E. (2009). Think or sink: Chinese learners' acquisition of the voiceless interdental fricative. *Language Learning, 59*, 581–621.
- 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.
- Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (eds.), *The handbook of second language acquisition*, pp. 382–408. Oxford: Blackwell.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263–308.
- Stevens, G. (1999). Age at immigration and second language proficiency among foreign-born adults. *Language in Society, 28*, 555–578.
- Strange, W. (ed.) (1988). *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press.
- Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning, 41*, 177–204.

- 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 /ɪ/ and /i/ 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 Strange (ed.), pp. 305–320.
- Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24, 245–261.
- Zhou, X., Espy-Wilson, C., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /ɪ/. *Journal of the Acoustical Society of America*, 123, 4466–4481.