To be published in *Second Language Research* (SAGE)

Individual Differences in Second Language Speech Learning in Classroom Settings: Roles of Awareness in the Longitudinal Development of Japanese Learners' English /ɹ/ Pronunciation

Kazuya Saito

Birkbeck, University of London

Abstract

The current study longitudinally examined a crucial individual difference variable—i.e., awareness (operationalized as explicit attention and articulatory knowledge)—in adult second language (L2) speech learning in the context of 40 Japanese learners' English /ɹ/ pronunciation development in an EFL classroom. The participants' speech, elicited from word reading, sentence reading and timed picture description tasks at the beginning and end of one academic semester, were analyzed in terms of three acoustic dimensions of English /ɹ/—third formant (F3), second formant (F2) and duration. Whereas the participants showed gains in the relatively easy aspect of the English /ɹ/ acquisition (F2 reduction) as a function of increased L2 input, their explicit awareness of accurate English /ɹ/ pronunciation played a significant role in the acquisition of the relatively difficult dimension (lengthening phonemic duration). The awareness-acquisition link was not found, however, for the most difficult dimension (F3 reduction) at least within the timeframe of the project.

Key words: Second language speech, Individual differences, Awareness, Foreign language classroom; Pronunciation, Acoustic phonetics

Individual Differences in Second Language Speech Learning: Roles of Awareness

Many child and adult learners alike practice their second language (L2) in foreign language (FL) contexts, which are typically characterized by only a few hours of target language input per week. In such contexts, the rate and ultimate attainment of learning is subject to a great deal of individual variability. To date, second language acquisition (SLA) researchers have extensively examined what kinds of learner factors relate to successful L2 lexicogrammar learning in FL classroom settings. One well-researched topic includes the role of explicit awareness (Leow, Johnson, & Zárate-Sández, 2011) and understanding (Spada & Tomita, 2010) of the target language in L2 lexicogrammar learning. To further move ahead this line of individual difference research agenda, the current study took a first step towards examining in depth how Japanese English-as-a-Foreign-Language learners differentially improved one difficult L2 segmental feature (i.e., word-initial English /ɹ/) over one academic semester according to their different degree of awareness (explicit attention and articulatory knowledge). Building on the recently-established developmental sequence framework (e.g., Saito & Brajot, 2013), their English /ɹ/ performance was analyzed for three acoustic/articulatory parameters which were assumed to represent different levels of learning difficulty: (a) lower second form (F2) for tongue retraction → (b) longer duration for prolonging phonemic length → (c) lower third formant (F3) for labial, alveolar and pharyngeal constrictions.

Background

Individual Differences in Foreign Language Classrooms

Second language learners in FL contexts access the target language primarily through FL instruction, without many opportunities to use the language outside classrooms. The nature of FL instruction is typically form- rather than meaning-oriented especially in many Asian English-as-a-Foreign-Language classrooms (the context of the current study) (Nishino & Watanabe, 2008). Despite these challenges, there is much evidence that adult L2 learners, who mainly rely on intentional and explicit learning strategies to acquire the target language, can demonstrate quicker development in classroom settings relative to young L2 learners, who tend to learn language in an incidental and implicit fashion. For example, Muñoz, (2006) showed that adult L2 learner group improved a range of L2 linguistic skills within a shorter amount of instructional time than child L2 learner group did, although the final quality of both adult and young learners' proficiency (i.e., ultimate attainment) was comparable. The adult learners' rate of learning advantage is arguably thanks to their cognitive maturity, metalinguistic awareness, literacy knowledge and experience at school (see also Muñoz, 2014).

According to previous longitudinal investigations, adult L2 learners can enhance various dimensions of L2 proficiency under FL conditions as a function of increased input and practice, such as vocabulary knowledge (Zhang & Lu, 2015; Muñoz, 2006), global speaking abilities (Baker-Smemoe & Haslam, 2013; Muñoz, 2014), and listening skills and grammar knowledge (Larson-Hall, 2008; Muñoz, 2006). When it comes to FL contexts, input is operationalized as the amount of time L2 learners spend practicing the target language inside (the number of FL classes) and outside (extra L2 practice) FL classrooms. Similar to naturalistic SLA, strong experience effects have been observed in various stages of FL learning processes. For example, Ojima, Matsuba-Kurita, Nakamura, Hoshino, and Hagiwara (2011) longitudinally tracked the L2 English vocabulary development of primary school children for three years (Grade 1-3). The results of the comprehension tests were strongly predicted by the composite input factor which

covered school lessons, home study, and private lessons outside of school/home. In Muñoz's (2014) investigation with L2 learners with extensive FL experience (10 years), their attained L2 oral abilities, measured via foreign accentedness, were strongly linked with the total number of FL classes that they had enrolled in as well as the frequency of extra practice that they engaged in outside of classrooms at the time of the project.

More importantly, what these FL studies have commonly discovered is that the extent to which L2 learners can benefit from FL instruction also greatly varies across individuals, because each learner is equipped with different processing abilities. Even if they receive the same quantity/quality of input, certain learners can show quicker and more robust gains than others do. Among a range of affecting factors which may explain such individual differences in successful L2 speech learning in classroom settings, such as aptitude (Li, 2016), motivation (Boo, Dörnyei, & Ryan, 2015) and affect (Dewaele & MacIntyre, 2014), much research attention has been given to, in particular, awareness.

Following Tomlin and Villa's (1994) definition, awareness refers to "a particular state of mind in which an individual has undergone a specific subjective experience of some cognitive content or external stimulus" (p. 193). From a theoretical perspective, awareness is believed to help L2 learners process received input more effectively and efficiently at the initial (rehearsing what they have attended in short-term memory) and later (internalizing into long-term memory) stages of interlanguage development (Schmidt, 2001). Yet, the occurrence of such successful input processing "without" awareness has remained controversial (e.g., Robinson, 1995 vs. Tomlin & Villa, 1994).

In the existing SLA literature, awareness has typically been measured via learners' self-report of underlying rules while (online) or after (offline) processing target language input, whether or not they use metalinguistic descriptions (for a methodological review, see Leow et al., 2011). For example, in Leow's (2000) oft-cited study, L2 Spanish learners engaged in a problem-solving task where they were guided to pay attention to the accurate use of the target feature (third-person agreement of irregular verbs) to successfully complete a crossword puzzle. Their awareness was measured based on the participants' verbal reports elicited by think-aloud protocols during the task and oral interviews after the completion of the task. The participants were judged to be "aware" when they referred to the target forms without mentioning the rules (e.g., anything related to irregular verb endings), or provided some form of metalinguistic explanation regarding the underlying rules (i.e., the relationship between irregular verb endings and a stem change in the preterit). The results of pre- and post-tests showed that significant L2 development occurred particularly when these learners demonstrated some degree of awareness (see also Hama & Leow, 2010; Rosa, & O'Neill, 1999).

These experimental findings concur with a number of classroom studies which have demonstrated that L2 learners are to show larger and more sustainable gains when they are encouraged to promote their awareness of target linguistic items via e intentional and explicit instruction than when they are simply exposed to meaningful input without any linguistic analyses as a form of incidental and implicit instruction (Spada & Tomita, 2010). To further move ahead and expand this line of awareness-acquisition research, the current study took an exploratory approach towards investigating this topic (the roles of awareness in classroom SLA) in the context of L2 pronunciation learning.

English /ɹ/

The current study focused on one of the most extensively examined topics in the field of L2 speech research—the acquisition of English /ɹ/ by adult Japanese learners (for a review, see Bradlow, 2008). Since the Japanese phonetic system has neither English /ɹ/ (nor /l/), Japanese learners tend to substitute the Japanese tap for English /ɹ/ (Riney, Takada, & Ota, 2000). Compared to the Japanese counterpart (the tap), however, the acoustic properties of English /ɹ/ are characterised as lower second formant (F2 = 1100-1600 Hz), lower third formant (F3 = 1700-2100 Hz) and longer phonemic duration (> 50 ms) (Espy-Wilson, 1992; Flege, Takagi, & Mann, 1995; Hattori & Iverson, 2009). From an articulatory standpoint, therefore, the acquisition of English /ɹ/ production requires Japanese learners to retract the tongue body (/w/-like production), make three constrictions (not only palatal/pharyngeal narrowing but also lip rounding) and prolong the duration of the sound.

From a pedagogical point of view, many researchers and teachers alike have emphasized that the acquisition of "intelligible" English /ɹ/ pronunciation needs to be prioritized in the phonological syllabus for Japanese FL learners due to its relatively high communicative value (e.g., Celce-Murcia, Brinton, Goodwin, & Griner, 2010; Riney et al., 2000). When Japanese learners continue to use the Japanese tap for English /ɹ/, it can be perceived by native speakers of English as one of four different sounds (/ɹ/, /l/, /w/ or /d/) (Sekiyama & Tohkura, 1993). This confusion could seriously hinder native (and non-native) speakers' comprehension, as the English /ɹ/-/l/, /ɹ/-/w/, /ɹ/-/d/ contrasts comprise many different minimal pairs (Catford, 1987).

Importantly, the acquisition of English /ɹ/ comprises three different developmental stages in the following order: tongue retraction (F2 reduction) → lengthening phonemic segments → labial, palatal and pharyngeal constrictions (F3 reduction). Since Japanese learners use F2 information to differentiate Japanese approximants (/w/ vs. /j/), they likely acquire nativelike F2 representations of English /ɹ/ within a short amount of residence in an English-speaking environment (e.g., < 6 months) (Saito & Munro, 2014). Although Japanese learners do not rely on temporal information in their approximant categories, the short-long distinction is marked in the Japanese vowel systems (e.g., /i/ vs. /i:/). Japanese learners with a certain amount of residence in L2 environments (e.g., 1 year) tend to become capable of producing and perceiving English /ɹ/ with sufficiently long length (50-100ms) (Flege et al., 1995; Saito & Brajot, 2013; Underbakke, Polka, Gottfried, & Strange, 1988).

In contrast, Japanese learners (including even highly experienced ones) are likely to show a lack of sensitivity to F3 information (Ingvalson, Holt, & McClelland, 2011; Saito, 2013a, 2013b), as the relevant acoustic and articulatory configurations (labial, palatal and pharyngeal constrictions) are not used as a primary cue for differentiating any phonemic contrasts in the L1 Japanese system. This could explain why many Japanese learners still demonstrate much difficulty in pronouncing nativelike English /ɹ/ even after years of learning (Ingvalson et al., 2011; cf. Flege et al., 1995). It has been suggested that the mastery of this sound may require special individual difference profiles, such as early age of acquisition (Idematsu & Holt, 2013; Saito, 2013a) and intensive training (Bradlow, 2008; Saito, 2013b).

Current Study

The current study took a first step towards exploring the extent to which a total of 40 freshman Japanese college students with varied awareness profiles differentially developed English /ɹ/ production over one academic semester. To this end, the development of English /ɹ/ was scrutinized in relation to three acoustic domains (F3, F2, duration). Subsequently, the varied

learner gains were linked to input (the amount and type of L2 use inside and outside FL classrooms) as well as awareness (explicit attention to and knowledge of English /ɹ/ pronunciation). As such, the research design would allow us to examine the relationship between learner awareness and acquisition while controlling for the other affecting factor—the quantity/quality of L2 input during the project.

Three predictions were formulated with regard to the input vs. awareness factors in different stages of English /ɹ/ development in FL classrooms (F2 → duration → F3):

1. As Japanese learners use tongue retraction (/w/-like production) as a default strategy to produce English /ɹ/, they would significantly lower their F2 values in relation to an increased amount of FL input over one academic semester.
2. Given the relative difficulty of acquiring the temporal aspect of English /ɹ/, Japanese learners who can attain more targetlike duration (> 50ms) may need not only extensive L2 input, but also a great amount of awareness of the target sound.
3. The change in the F3 dimension requires the acquisition of a new phonetic representation rather than the adjustment of existing cue weightings. Thus, it is predicted that few participating students would demonstrate robust improvement regardless of their varied amount of input and awareness within the timeframe of the project (one academic semester).

Method

The data collection process was conducted in Fall 2013. Whereas L2 speech data (English /ɹ/ production) was collected at the beginning (T1) and end (T2) of the term, the input and awareness profiles of the participating students were surveyed in a separate session within one month of the post-test. Finally, the link between acquisition and individual differences was analyzed by way of a range of statistical analyses (ANOVAs, correlations, multiple regression).

Participants

Japanese foreign language students. As a part of a larger project which aimed to survey the L2 oral proficiency of English-as-a-foreign language students with diverse backgrounds at numerous universities across the Tokyo area in Japan, a total of 40 freshman Japanese students (19 males, 21 females) were recruited from social sciences and humanities programs (business, marketing, economics, psychology) at a prestigious, large university located in downtown Tokyo. Although their pronunciation widely varied (see the Results section), they had relatively homogeneous FL backgrounds at the time of the project. First, they had used Japanese as an L1 from birth onward, and both of their parents were native speakers of Japanese. Second, they had started learning L2 English in Grade 7. Third, they had studied L2 English for only six years under FL conditions with limited exposure to the L2 in quantity (several hours of instruction per week) and quality (form- rather than meaning-oriented content). The relationship between their general oral ability scores (comprehensibility, accentedness) and motivational orientations (integrativeness, instrumentality) were reported in a different venue (Saito, Dewaele, & Hanzawa, 2017).

English baseline. To provide baseline data for English /ɹ/ production, we recruited 10 native speakers of English (5 males, 5 females). All of them were residents in Vancouver, Canada and reported little familiarity with other foreign languages (including the other official

language of Canada: French). They were undergraduate students (M_{age} = 25.2 years) at the time of the project.

Speaking Task

From a cognitive psychology perspective (DeKeyser, 2007), any SLA phenomenon can be characterized as a transition from a "controlled" to "automatic" use of declarative knowledge, suggesting that L2 learners can enhance their processing abilities to access what they have acquired first at a controlled speech level, and later at a spontaneous speech level. Although L2 pronunciation performance has traditionally been measured via controlled speech measures, where L2 learners fully focus on the correct pronunciation of target sounds (e.g., word and sentence reading), certain researchers have adopted spontaneous speech measures (e.g., picture description, monologue, oral interview tasks: Piske, Flege, MacKay, & Meador, 2011). To tap into the participants' controlled and spontaneous processing abilities of English /ɹ/ production, three tasks elaborated in the author's precursor research (Saito, 2013a; Saito & Brajot, 2013; Saito & Munro, 2014)—word reading (WR), sentence reading (SR), and timed picture description (TPD)—were used in the current study.

Following Spada and Tomita's (2010) concept of spontaneous speech production task, TPD was assumed to reflect L2 learners' ability to produce the target sound accurately (English /ɹ/) with a primary focus on spontaneously conveying an intended message (describing each photo). Different from the spontaneous speech task, SR elicited the participants' /ɹ/ production during continuous speech with little communicative pressure. Compared to TPD and SR, WR allowed the participants to fully focus on accurate pronunciation forms. In order to avoid an excessive focus on pronunciation forms, especially in TPD (eliciting spontaneous /ɹ/ production), the three tasks were conducted in the following order (TPD → SR → WR). All participants were purposely left uninformed of the primary purpose of the tasks (accurate pronunciation of /ɹ/) until the completion of the data collection.[1]

Stimuli. All 20 target words included /ɹ/ in the word-initial position (n = 8 for WR, n = 8 for SR, n = 4 for TPD) and were Consonant-Vowel-Consonant (CVC) singletons, except for *Ryan* (CVVC). According to the results of the vocabulary profile (Cobb, 2014), the 20 words were categorized as belonging to the first 3,000 most frequent word families, except for *Ryan* and *ram*. Given the coarticulation effects of following vowel conditions (i.e., Japanese learners tend to have more difficulty in producing word-initial /ɹ/ preceding front vowels—/i/ and /e/—than central and back vowels; Flege et al., 1995), the following vowel conditions were carefully

[1] No explicit attempts were made to check the participants' degree of English /ɹ/ awareness during TPD, SR and WR performance at any data collection point; doing so would have prompted the participants to attend to English /ɹ/ and skewed the main goal of the project, i.e., the longitudinal investigation of the relationship between L2 learners' *general* awareness and L2 pronunciation development over one academic term. More importantly, the results of the study demonstrated that the participants' English /ɹ/ performance did not change across the three task modalities (TPD, SR vs. WR) (see the Results section). This in turn suggests that the potentially different amount of task-specific awareness during their completion of WR, SR and TPD did not significantly influence their English /ɹ/ performance at least in the context of the current study with N = 40 English-as-a-Foreign-Language students (cf. for clear task effects on 100+ *experienced* Japanese learners' English /ɹ/ pronunciation and the validity of the method per se, see Saito, 2013a; Saito & Brajot, 2013; Saito & Munro, 2014).

controlled in each task: 50% for singletons with front vowels and 50% for singletons with central and back vowels. The test tokens are summarized in Table 1.

Table 1
10 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

Procedure. Individual recording sessions took place in a sound proof booth at the Japanese university. To avoid any confusion in the oral test procedure, all instructions were delivered in Japanese by a research assistant (a native speaker of Japanese). All speech tokens were recorded using a Marantz PMD 660, set at 44.1 kHz sampling rate and 16-bit quantization.

1. **Timed Picture Description.** In this task, participants described seven pictures, including three distracters, with five seconds of planning time for each photo. Each picture had three key words printed underneath it, one of which was a target word (i.e., *read, rain, rock, road*). For example, one picture portrayed a straight road in the countryside with several clouds in the sky. The purpose of this picture was to elicit learners' spontaneous production of /ɹ/ in *road*. The other two key words for this picture included "*cloud*" and "*blue sky*." To familiarize speakers with the task procedure, three distracter pictures were first presented. The other four pictures featuring the target words were then presented to elicit their spontaneous production of /ɹ/. Compared to the SR and WR tasks (which included eight target words per task), four target words were used in order to prevent participants from noticing and paying too much attention to the correct pronunciation of English /ɹ/.
2. **Sentence Reading.** In this task, participants read five target sentences together with three distracter sentences. The target sentences were as follows:
    - *He will read my paper by the time I arrive there.*
    - *She left her red bicycle on the side of the road.*
    - *The race was cancelled because of the rain.*
    - *I can correct all wrong sentences tonight.*
    - *Ryan does not like to run in the snow.*
3. **Word Reading.** In this task, participants read a list of 25 words containing eight target words and 17 distracters. The target words included were *read, room, root, red, race, rough, ram*, and *right*.

Acoustic Analysis

To scrutinize the different stages of Japanese learners' English /ɹ/ development, three acoustic properties of the sound—F2 (i.e., tongue retraction), F3 (i.e., labial, palatal and pharyngeal constrictions) and transitional duration (phonemic length)—were analyzed. Whereas these cues typically explain most of the variance (70-80%) in human perception of English /ɹ/, the acoustic correlates of English /ɹ/ substantially vary depending on the criteria that human raters evaluate for. When human raters assess the accuracy of English /ɹ/ (to what degree the sound is accurately produced as English /ɹ/), they likely use F3 as primary information while paying minor attention to other cues (F2, transition duration). In the case of intelligibility judgements (whether the sound is English /ɹ/ or not), raters equally rely on all acoustic information (Flege et al., 1995).

A linear predictive coding routine in *Praat* (Boersma & Weenink, 2012) was used to measure the formant values. Drawing on the procedure for the acoustic analysis of natural speech tokens used in Flege et al. (1995), word onset was first identified based on a visual inspection of both the wave formant and spectrographic representation of each token. For WR, a cursor was placed on the point where both F2 and F3 were clearly observed. For SR and TPD, a cursor was placed on the local minimum of F3 (dip) to get the F2 and F3 values, as F3 values of English /ɹ/ are lower (1600-1900 Hz) relative to other surrounding sounds. The durational dimension of English /ɹ/ was measured via F1 transition. When the endpoint of the F1 transition was unclear, the peak of the F3 transition was used instead (Hattori & Iverson, 2009).

Because spectral information (i.e., F3, F2) significantly varies due to anatomical differences in individual vocal tract length, raw acoustic values were adjusted using the following normalization procedure (for details, see Lee, Guion, & Harada, 2006). A mean F3 value of /æ/ elicited from three monosyllabic words in the WR task (i.e., *man*, *map, ram*) was calculated for each participant (N = 40 Japanese FL students + 10 native speakers of English). One female English speaker was randomly selected as a reference, and her mean F3 value (i.e., 2950 Hz) was divided by those of the other participants to provide their own *k* factors. Then, all formant values (F3, F2) of /ɹ/ for each participant were multiplied by the individual *k* factors, respectively.

Individual Differences

To measure the individual difference profiles of the 40 Japanese students, they participated in an individual interview in order to elicit detailed information regarding the quantity/quality of FL input, and to assess their awareness of English /ɹ/ pronunciation. The entire session took approximately one hour per participant, with all instructions delivered in Japanese by a researcher.

Measure of input. The participants' input profiles were carefully analyzed by means of an interview. In a pilot study with students at the same university, it was found that the FL classes the students were enrolled in widely varied in terms of orientation (form vs. meaning), class size (small vs. large), style (lecture vs. discussion), and peers and teachers (native vs. non-natives vs. Japanese). Instead of drawing on more fine-grained criteria that instructed SLA researchers typically use for coding different types of instruction (e.g., Spada & Tomita, 2010 for "explicit vs. implicit"; Shintani, Li, & Ellis, 2013 for "input- vs. output-based"), the following simple, straightforward and context-specific scheme was elaborated to code three categories of FL input at the university: (a) total hours of form-focused classes, (b) total hours of content-

based classes, and (c) extra L2 use outside classrooms per week. As is common at many university-level schools in Japan, the participants at this university were required to take at least 1.5 hours of form-focused lessons per week. As a way to increase the amount of FL classroom input, it was optional for the participants to take not only form-focused but also content-based classes.

- ***Total hours of form-focused classes***. Form-focused instruction referred to as any English-as-a-Foreign-Language classes with a primary focus on language. The students' performance in these classes was graded based on L2 knowledge (e.g., pronunciation, vocabulary and grammar knowledge) and/or L2 proficiency (reading, listening and writing abilities). For each class, they were also asked to report whether pronunciation instruction was featured—the pedagogic component directly relevant to the objective of the study (English /ɹ/ pronunciation development).
- ***Total hours of content-based classes***. Content-based instruction was defined as any subject matter class (e.g., business, economics, politics, history, linguistics, psychology) which delivered instruction via only English, regardless of the status of the teacher (natives, non-natives vs. Japanese) or style of the class (lecture vs. discussion). Unlike form-focused classes, all exams in these classes were designed to assess the students' content (rather than linguistic) knowledge in English.
- ***Extra L2 use per week***. Last, the participants self-reported how many minutes they spent practicing English outside of the classroom with international students (including both native and non-native speakers of English), and studying for proficiency exams (TOEIC, TOEFL). Whereas many (16 out of 40 students) did not have any opportunities to use L2 use outside of classrooms, some students spent many hours on extra L2 use activities (> 10 hours) (for details, see Results).

Measure of awareness. Following procedures in the previous literature (e.g., Leow et al., 2011), the participants' explicit awareness of target phonetic feature (English /ɹ/) was measured in terms of the amount of attention and explicit knowledge:

- ***Explicit attention to English /ɹ/ production***. This subcategory refers to the participants' awareness of pronouncing English /ɹ/ accurately vs. avoiding the substitution of the Japanese tap for English /ɹ/. During the post-hoc oral interview, they rated the following statements ("*How much are you trying to pronounce English /ɹ/ accurately?*""*How much are you trying to avoid using the Japanese tap sound for English /ɹ/*) on a 6 point-scale (*1 = I pay little attention, 6 = I always pay a lot of attention*).
- ***Explicit articulatory knowledge***. Additionally, their awareness was surveyed in terms of their explicit knowledge of two relevant articulatory configurations for English /ɹ/ production—lip rounding (for lowering F3) and tongue retraction (for lowering F2) relative to the Japanese tap ("*Do you know that you need to round lips and/or back tongue body (with tongue tip not touching the ceiling of the vocal tract) in order to pronounce English /ɹ/?*"). The participants' answers were coded "yes" or "no."

Results

The main objective of the current study was to examine the complex relationship between the participants' varied experience, input and awareness profiles, on the one hand, and three different aspects of their English /ɹ/ acquisition (F3, F2, duration) under three different conditions (Word Reading, Sentence Reading, Timed Picture Description), on the other. To this end, five different analyses were conducted, focusing on (a) the acoustic characteristics of the Japanese FL students' English /ɹ/ performance (in comparison with a native baseline); (b) their overall improvement over time as a group; (c) their individual differences in input and awareness; (d) the general relationship between their different amounts of input, awareness and English /ɹ/ acquisition; and (e) the relative weights of input and awareness in their L2 pronunciation development.

Japanese FL students vs. English Baselines

The first objective of the statistical analyses was to examine how the 40 Japanese FL students produced three different aspects of English /ɹ/ (F3, F2, duration) under three different task conditions (WR, SR, TPD) compared to 10 native baselines. The acoustic values (F1, F2, F3, transition duration) of 20 words were measured and averaged according to the task (n = 4 for TPD, n = 8 words for SR, n = 8 words for WR).

According to the descriptive results summarized in Table 2, the Japanese FL students' performance was substantially different from the native baselines both at T1 and T2. Notably, there was a substantial amount of variability in the Japanese students' /ɹ/ performance: Whereas some students produced more targetlike English /ɹ/ (F3 < 2100 Hz, F2 < 1600 Hz, duration > 50 ms), others appeared to simply substitute the Japanese tap for English /ɹ/ (F3 > 2800 Hz, F2 > 1600 Hz, duration < 50ms).

To examine the extent to which these participants' performance significantly differed between group and task contrasts, their production data (F3, F2 and duration values) at T1 were separately submitted to two-way ANOVAs with Group (FL students, Baseline) as a between-subjects factor and Task (WR, SR, TPD) as a within-subjects factor. The alpha value was set to .05. According to the results of the ANOVAs, the Japanese FL students produced English /ɹ/ with significantly higher F3 values, $F(1, 48) = 40.080$, $p < .001$, $\eta^2 = 0.493$, higher F2 values, $F(1, 48) = 15.421$, $p < .001$, $\eta^2 = 0.246$, and longer duration, $F(1, 48) = 90.233$, $p < .001$, $\eta^2 = 0.689$. To summarize, whereas all acoustic properties of English /ɹ/ substantially differed between the Japanese FL students and native baselines, their production performance was comparable across the different task conditions (showing little influence of speaking tasks).

Table 2

Descriptive Results of English /ɹ/ Production by Japanese FL Students (n = 40) at T1 and T2, and native baselines (n = 10)

		Japanese students (T1)			Japanese students (T2)			Native Baseline		
		M	*SD*	*Range*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
F3	WR	2664	385	1944-3444	2642	440	1909-3132	1928	113	1693-2088
	SR	2640	349	2052-3294	2627	364	2002-3397	1931	106	1714-2060
	TPD	2699	371	1919-3397	2739	289	2061-3367	1923	84.4	1766-2060
F2	WR	1654	259	1164-2130	1643	274	1210-2098	1319	177	1084-1647
	SR	1635	265	1088-2121	1598	145	1177-2099	1310	132	1140-1517
	TPD	1675	262	1175-2213	1685	258	1180-2231	1369	184	1077-1610
Duration	WR	30.5	21.5	10.0-93.1	32.8	22.7	10.0-78.3	91.9	10.5	68.9-121.5
	SR	30.0	17.9	10.0-66.1	34.2	17.5	10.0-65.7	80.9	11.0	69.1-115.3
	TPD	33.1	19.4	10.0-74.5	34.1	16.2	10.0-61.0	89.2	12.7	59.7-117.8

Notes. T1, T2 = the beginning and end of the project; WR for word reading; SR for sentence reading; TPD for timed picture description

Group Analyses

The second objective of the statistical analyses was to explore whether the FL students significantly improved their English /ɹ/ production as a "group" over one academic semester. To this end, their acoustic data (F3, F2, duration) at T1 and T2 were submitted to two-way General Linear Model ANOVAs with Time (T1, T2) and Task (WR, SR, TPD) as within-subjects factors. The ANOVAs did not find significant Time effects in any acoustic dimensions of English /ɹ/—F3, $F(1, 78) = .006$, $p = .939$, $\eta^2 = .026$, F2, $F(1, 78) = .377$, $p = .543$, $\eta^2 = .010$, duration, $F(1, 78) = .072$, $p = .790$, $\eta^2 = .002$. The results here showed that the participants did not significantly enhance their English /ɹ/ performance at a $p < .05$ level, suggesting that the amount of the participants' improvement could be substantially different between individuals.

Input and Awareness Data

The third objective of the statistical analyses was to survey the individual difference profiles of the 40 FL students in terms of quantity and quality input (the amount/type of input inside and outside classrooms) and awareness (explicit attention to and knowledge of English /ɹ/ pronunciation) (see Table 3).

- **Input**. The participants' FL experience (i.e., what type of input they processed) widely varied, indicating that they used different strategies to increase their exposure to FL input by taking form-focused classes (16.5-79.5 hours per semester) and content-based classes (0-99 hours per semester), and/or engaging in conversation activities with native and non-native speakers outside FL classrooms (0-140 min per week). In terms of the content of the form-focused classes, the endpoint interview confirmed that none of the participants had received any pronunciation training throughout the project.
- **Awareness.** A majority of the FL students demonstrated a great amount of attention towards avoiding the substitution of the Japanese tap for English /ɹ/ ($M = 5.4$ on a 6-point scale), and to a lesser degree, towards pronouncing English /ɹ/ accurately ($M = 2.5$). However, only half of the participants actually had explicit articulatory knowledge of English /ɹ/ production (lip rounding, tongue retraction).

Table 3
Descriptive Statistics of Individual Differences among 40 Japanese Foreign Language Students

A. Quantity and quality of input	*Mean*	*SD*	*Range*
(1) Total hours of form-focused instruction	58.0	22.2	16.5-79.5
(2) Total hours of content-based instruction	31.3	34.0	0-99
(3) Extra L2 use outside classroom per week (min)	27.5	48.8	0-140
B. Awareness	***Mean***	***SD***	***Range***
(4) Awareness of pronouncing English /ɹ/ accurately	2.5	7.7	1-5
(5) Awareness of avoiding using the Japanese tap for English /ɹ/	5.4	2.2	3-5
(6) Explicit articulatory knowledge (lip rounding)	Yes = 21, No = 19		
(7) Explicit articulatory knowledge (tongue retraction)	Yes = 17, No = 23		

The inter-relationship among the seven individual difference factors was also investigated via a set of Pearson correlation analyses. According to the results (summarized in Table 4), several intriguing patterns were found: (a) those who took more content-based classes were

likely registered in more form-based classes, but did not necessarily spend more time in conversational activities with native and non-native speakers outside FL classrooms; (b) those enrolled in more content-based classes likely exhibited greater awareness of accurate English /ɹ/ pronunciation; and (c) those with more explicit attention to English /ɹ/ pronunciation tended to have more explicit articulatory knowledge of English /ɹ/ production.

Table 4
Pearson Correlations between the Input and Awareness Factors

	Content-based classes	Extra L2 use	Awareness (accurate /ɹ/)	Awareness (the L1 avoidance)	Lip rounding	Tongue retraction
	r	*r*	*r*	*r*	*r*	*r*
A. Input						
Form-focused classes	.52*	.31	.20	.23	.29	-.06
Content-based classes		.30	.33*	.16	.19	.04
Extra L2 use			.11	-.10	-.01	-.02
B. Awareness						
Awareness of accurate /ɹ/				.36*	.37*	.50*
Awareness of the L1 avoidance					.23	.12
Lip rounding						.02

Note. * indicated $p < .05$.

Individual Differences and L2 Pronunciation Development

The fourth objective of the statistical analyses was to identify whether the participants' improvement patterns were related to their individual difference profiles—(a) quantity and quality of input and (b) explicit attention and articulatory knowledge of English /ɹ/ pronunciation. For this analysis, their gain scores for each acoustic dimension (F3, F2, duration) and task (WR, SR, TPD) were used as dependent variables. For example, when one FL student reduced their F3 values from 2775 Hz (T1) to 2350 Hz (T2) at the word reading task, his gain score in F3 × WR was 425 Hz. The participants' performance over time is summarized in Table 5.

Table 5
Descriptive Statistics of 40 FL Students' Gain Scores

		M	*SR*	*Range*
F3	WR	-21.6	213	-415 to 584
	SR	-12.2	214	-508 to 455
	TPD	40.2	290	-644 to 616
F2	WR	-10.5	134	-326 to 264
	SR	-36.5	135	-390 to 220
	TPD	10.4	198	-580 to 491
Duration	WR	2.3	12.9	-41.9 to 31.2
	SR	4.1	13.1	-30.1 to 40.3
	TPD	-5.2	21.6	-52.5 to 35.5

To provide a general picture of the relationship between input (form-focused/content-based classes, extra L2 use), awareness (explicit attention, articulatory knowledge) and L2 speech learning (F3, F2, duration) under different task conditions (WR, SR, TPD), the participants' gain scores were submitted to a set of partial correlation analyses. In this statistical analysis, the participants' pre-test scores (T1) were used as a covariate and statistically controlled for. As such, the influence of the participants' previous FL experience prior to the project on the amount of their gain scores was factored out from the following analyses. This decision was crucial, as the participants had studied English for six years in various classroom settings before entering the university, and their varied previous FL backgrounds may have influenced L2 learning patterns during the semester when the data collection took place (see Aguilar & Muñoz, 2014 for the role of initial L2 proficiency in SLA in FL classrooms).

According to the results of the partial correlation analyses (summarized in Table 6), a significant relationship was identified between the FL students' gain scores in F2 dimensions and the number of content-based classes that they had enrolled during the semester under all task conditions (r = -.38 to -.41). The FL students' gain scores in the F2 and duration dimensions in the spontaneous speech task (TPD) were significantly correlated with the amount of awareness/attention towards accurate English /ɹ/ pronunciation (r = .36). The partial correlation analyses did not find any significant effects of individual differences on the participants' F3 performance at a $p < .05$ level.

Relative Weights of Input and Awareness in L2 Pronunciation Development

To further examine the extent to which the input and awareness factors *differentially* related to the participants' English /ɹ/ acquisition, a series of stepwise multiple regression analyses were performed, using the participants' gain scores in each acoustic (F3, F2, duration) and task (WR, SR, TPD) domain as dependent variables, and their input and awareness profiles as independent variables. Similar to the partial correlation analyses mentioned earlier, the participants' pre-test scores were also entered into the regression model as another independent variable in order to control for the influence of background FL experience (i.e., prior to the project).

Table 6
Partial Correlations between the Input/Awareness Factors and Gain Scores

		Input						Awareness							
		Form-focused classes		Content-based classes		Extra L2 use		Awareness (accurate /ɹ/)		Awareness (the L1 avoidance)		Lip rounding		Tongue retraction	
		r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
F3	WR	.06	.71	-.01	.94	-.01	.98	.04	.79	.13	.42	.01	.98	.01	.92
	SR	-.03	.81	-.07	.63	.04	.68	-.06	.68	-.17	.29	-.03	.83	-.22	.16
	TPD	.11	.47	-.11	.50	.02	.88	-.26	.100	.13	.40	-.27	.08	-.19	.23
F2	WR	-.20	.21	-.41**	.01	-.07	.65	-.23	.14	-.07	.64	.06	.69	-.24	.13
	SR	-.33*	.04	-.38*	.016	-.10	.54	-.10	.54	-.13	.40	.13	.40	-.09	.57
	TPD	-.25	.12	-.41**	.01	-.20	.21	-.35*	.02	-.06	.71	-.27	.08	-.16	.31
Duration	WR	.03	.84	-.04	.77	.20	.21	.03	.84	-.25	.11	.09	.57	-.03	.83
	SR	-.11	.49	-.07	.67	-.23	.15	.20	.21	.15	.35	-.15	.36	.23	.14
	TPD	-.01	.92	.03	.84	.12	.44	.36*	.02	.10	.54	.13	.40	.25	.11

Note. * indicates $p < .05$, ** indicates $p < .01$

Several necessary steps were carefully taken ensure the appropriateness of conducting the multiple regression analyses with the relatively small sample size of the study ($N = 40$). First, the decision was made to reduce the number of independent variables by focusing only on those predictors revealed by the partial correlation analyses to be significant (i.e., form-focused/content-based classes, explicit attention to English /ɹ/). Second, the normality of each dependent and independent variable was confirmed by Kolmogorov-Smirnov tests ($p > .05$). Finally, the results of the power analysis showed that the observed power for the multiple regression with 40 participants and four predictors to generate a medium-to-large effect size ($R^2 = .50$-$.70$) was .98, which could be considered as relatively large in the field of SLA research (Larson-Hall, 2010).

As summarized in Table 7, the results of the multiple regression analyses showed that the variance in the participants' gain scores (T1 → T2) was primarily predicted by their initial proficiency, and secondarily by either the number of content-classes (for F2 acquisition in WR, SR and TPD) or the awareness of accurate English /ɹ/ pronunciation (for duration acquisition in TPD). In short, the results here suggest that even after statistically controlling for previous FL experience, the input and awareness factors were still uniquely associated with the relatively easy (F2) and difficult (duration) aspects of the participants' English /ɹ/ acquisition, respectively. Yet, the regression models did not identify any significant links between such learner factors (input/awareness) and the participants' F3 reduction—the most difficult aspect of English /ɹ/ acquisition.

Table 7
Results of Multiple Regression Analysis Using Input and Awareness Variables as Predictors of L2 Speech Learning under Different Acoustic (F3, F2 Duration) and Task (WR, SR, TPD) Conditions

Predicted variable	Predictor variables	*Adjusted* R^2	R^2 *change*	*F*	*p*
F3 (WR)	Pre-test scores	.76	.76	124.612	$p < .001$
F3 (SR)	Pre-test scores	.66	.66	78.459	$p < .001$
F3 (TPD)	Pre-test scores	.40	.40	26.225	$p < .001$
F2 (WR)	Pre-test scores	.76	.76	122.844	$p < .001$
	Content classes	.80	.04	75.544	$p < .001$
F2 (SR)	Pre-test scores	.74	.73	110.708	$p < .001$
	Content classes	.78	.04	66.380	$p < .001$
F2 (TPD)	Pre-test scores	.50	.50	38.335	$p < .001$
	Content classes	.58	.08	26.265	$p < .001$
Duration (WR)	Pre-test scores	.68	.68	83.842	$p < .001$
Duration (SR)	Pre-test scores	.53	.53	43.075	$p < .001$
Duration (TPD)	Awareness	.15	.15	6.507	$p = .015$

Note. The variables entered into the regression equation included the number of form and content classes, and the participants' awareness of accurate English /ɹ/ pronunciation.

Discussion and Conclusion

The current study aimed to examine how a total of 40 Japanese FL students with various awareness profiles differentially developed their English /ɹ/ performance when their input factor (their L2 use inside/outside classrooms) was statistically controlled for one academic semester. Specifically, the FL students' interlanguage development was scrutinized according to three different aspects of English /ɹ/ acquisition—F3 reduction (bilabial, palatal and pharyngeal constrictions), F2 reduction (tongue retraction) and duration (prolonged phonemic length)—under three task conditions (Word Reading, Sentence Reading, Timed Picture Description). The acoustic and task dimensions were assumed to represent different levels of difficulty in learning (F2 < duration < F3) and processing (WR < SR < TPD).

The results of the ANOVAs showed that the participants failed to significantly enhance any acoustic dimension (F3, F2, duration) of their English /ɹ/ pronunciation as a "group" over one academic semester. The lack of significant group improvement could be due to several reasons. For example, the FL students in the study did not receive any focused pronunciation instruction in class during the project. Indeed, half of them did not have adequate L2 articulatory knowledge for producing English /ɹ/ (lip rounding, tongue retraction). The lack of pronunciation training and phonological awareness among the students participating in the study may not be surprising; the same situation has been reported in many foreign language (Saito & van Poeteren, 2012) and second language (Derwing & Munro, 2005) classrooms all over the world. Another possible interpretation could be that the acquisition of English /ɹ/ may still require a tremendous amount of meaningful input and interaction that L2 learners can normally access in naturalistic settings where the L2 is used on a daily basis (e.g., Flege et al., 1995). This in turn suggests that L2 learners need special individual difference profiles in order to make the most of the acquisitionally-limited L2 learning experience in FL settings (Muñoz, 2014).

In fact, the survey data showed that the FL students greatly differed in terms of not only the quantity and quality of input they were exposed to inside and outside FL classrooms during the project, but also the amount of their explicit awareness towards accurate English /ɹ/ pronunciation. Importantly, the results of partial correlation and multiple regression analyses equally demonstrated that such individual difference factors—input and awareness—were uniquely tied to three different stages of English /ɹ/ production development—F3, F2 and duration.

First and foremost, one of the input factors (i.e., the number of content-based classes) was significantly related to the participants' improvement in F2 (retracting tongue body for lowering F2 values < 1600 Hz). Whereas all the participants were required to take, at least, 1.5 hours of form-focused instruction per week, it was only certain students that decided to take content-based classes beyond their regular FL syllabus. It could be said that based on this these FL students enrolled in content-based classes may have made additional efforts to expand and enrich their FL experience, as they were able to benefit from both form- *and* meaning-oriented input. The findings here are in line with the extensive discussion in the L2 speech literature, which has stated that input is instrumental to successful L2 speech learning, as even adult L2 learners maintain the capacity to continuously and adaptively change their phonetic representations throughout their life span as long as they use the target language regularly, constantly and frequently (e.g., Flege, 2009; Saito, 2015).

What the current study adds, however, is that input under FL conditions (with limited input in quantity and quality) may facilitate only the development of relatively easy aspects of L2 speech learning—F2 reduction in the English /ɹ/ acquisition. Comparatively, none of the

input factors (number of form-/content-based classes; extra L2 use) yielded any significant associations with the other domains of English /ɹ/ production which entail more learning difficulty—F3 reduction and duration. Rather, a significant link was found between one of the awareness factors (i.e., explicit attention towards pronouncing English /ɹ/ accurately) and the relatively difficult aspect of English /ɹ/ acquisition—the FL students' durational change (prolonging the phonemic length > 50ms).

Taken together, I would like to provide a tentative hypothesis as to the intricate relationship between input, awareness and L2 pronunciation development in FL classrooms. In the case of the acquisition of English /ɹ/, especially in acquisitionally-limited contexts (i.e., foreign language classrooms), input alone may impact only on the relatively easy dimension of English /ɹ/ pronunciation (F2 reduction). Adult Japanese students may need not only sufficient input, but also direct explicit attention to acquiring various aspects of accurate English /ɹ/ pronunciation. This is because such awareness is assumed to help adult L2 learners become sensitized to any received input so as to notice, understand and learn the cross-linguistic differences of the target sound—i.e., lowering F2 *and* prolonging phonemic length. Even if ample input is provided, unaware learners may have difficulty achieving such noticing and learning without any explicit phonetic instruction, and instead may continue to apply the Japanese counterpart (the Japanese tap) to the L2 English approximant categories.

Similarly, some L2 speech researchers have addressed the role of attention (i.e., explicit awareness of L2 phonetic structures) in adult L2 speech learning and provided cross-sectional findings. For example, Venkatagiri and Levis (2007) demonstrated that learners with more phonological awareness tend to produce more comprehensible L2 speech. To further this, our study has provided the first "longitudinal" support for the acquisitional value of L2 learners' explicit phonological awareness in a FL classroom setting. The results here suggest that enhancing L2 learners' awareness via explicit phonetic instruction may be necessary if L2 learners are to acquire a phonetic cue which is not fully used in their L1 system (the durational cue in L2 English /ɹ/). The suggestion here also concurs with positive findings resulting from classroom intervention studies (e.g., Derwing & Munro, 2005).

Notably, the results of the current study did not find explicit articulatory knowledge (lip rounding, tongue retraction) to be a significant predictor of any aspects of English /ɹ/ development. This indicates that the explicit phonetic knowledge of learners relevant to acquisition should be conceptualized on a perceptual basis (whether they can hear the perceptual properties of English /ɹ/) rather than based on articulatory configuration (whether they know how to produce English /ɹ/). This concurs with the claim of many researchers that L2 pronunciation learning is mainly driven by L2 learners' change in perceptual ability (for detailed accounts of the perception-first view in L2 speech learning, see Flege, 2009). From a pedagogical perspective, the findings indicated that explicit phonetic instruction should not only provide explicit articulatory explanation on how to pronounce English /ɹ/, but also help learners perceive its key acoustic features. For example, it has been shown that intensive exposure to acoustically-enhanced input promotes the concurrent development of L2 perception and production abilities to some degree, suggesting that perception and production improved similarly or in related ways (Bradlow, 2008; Saito, 2013b).

Of course, however, the lack of evidence on the predictive capacity of articulatory knowledge for /ɹ/ development (the null result) does not necessarily imply that it is perception awareness that is important. To further examine the relationship between explicit perception- and production-based awareness and L2 speech learning, we must wait for future studies to

conceptualize what constitutes explicit perceptual and articulatory L2 knowledge, and then elaborate a range of instruments by which to assess the various dimensions of target-sound awareness from multiple angles.

For example, as operationalized in Hama and Leow (2010), it would be interesting to include not only offline (e.g., post interviews) but also online (e.g., think-aloud protocols) measures, especially when L2 learners engage tasks which guide them to accurately use L2 pronunciation with a primary focus on meaning conveyance. In my precursor research (Saito, 2013b; Saito & Lyster, 2012), Japanese learners participated in discussion tasks where they were asked to debate a range of topics which all included a number of target words with word-initial English /ɹ/ (e.g., "***Is running outside better than running inside?***"), highlighted in red as a form of input enhancement (to promote students' noticing of English /ɹ/). Future studies of this kind can examine whether and to what degree their awareness of English /ɹ/, measured during and after the task, would actually impact their improvement of the target sound (measured via pre-/post-tests).

Last, it is important to remember that no such individual difference-proficiency associations were found for the participants' improvement in the F3 dimension (creating bilabial, palatal and pharyngeal constrictions simultaneously for lowering F3 < 2100 Hz). As noted earlier, there is much evidence that Japanese learners demonstrate a tremendous amount of difficulty in establishing the entirely new phonetic representation of F3 variance to reliably perceive and produce nativelike English /ɹ/ forms. To improve the intelligibility and accuracy of English /ɹ/ pronunciation, Japanese learners are reported to rather focus on enhancing the F2 and temporal dimensions of English /ɹ/ (Ingvalson et al., 2011; Saito & Brajot, 2013). For example, Ingvalson, Holt and McClelland's (2012) training study established that only a subset of Japanese learners can develop sensitivity to F3 variation when exposed to synthetic /ɹ/-/l/ tokens varying only in the F3 domain. In line with our predictions and previous relevant findings, the results of the current study confirmed that: (a) F3 development is unlikely to occur within one semester of FL learning; and (b) none of the input/awareness factors included in the study marked those exceptional students who significantly reduced their F3 values in English /ɹ/ production (-415 Hz to -614 Hz).

To close, two primary limitations need to be addressed with an eye towards future investigations. First, it needs to be acknowledged that the findings of the current study were based on a relatively small number of participants (N = 40 Japanese FL students) and were focused only on the acquisition of word-initial English /ɹ/. Thus, these findings need to be replicated with a larger number of L2 learners with varied age (child vs. adult), proficiency (beginner vs. advanced) and cognitive (analytical vs. functional) profiles in various learning contexts (foreign vs. second language settings). It is noteworthy that the target sound of the study—English /ɹ/—has relatively high communicative value, as its mispronunciation tends to result in a negative impact on native speakers' global assessments of L2 speech (e.g., Riney et al., 2000 for foreign accentedness). Therefore, it would also be intriguing to revisit the role of awareness in the acquisition of L2 sounds with less communicative value (with its mispronunciation generating little impact on L2 assessment), such as segmentals with low functional load (e.g., English /θ, ð/) (Catford, 1987). In these cases, awareness factors may have stronger predictive power for successful L2 speech learning. This is arguably because the mispronunciation of these features does not hinder the flow of L2 commutation, and adult L2 learners would otherwise have a great deal of difficulty in noticing these sounds without awareness.

Second, although three different tasks were adopted to elicit English /ɹ/ production at controlled and spontaneous speech levels, the statistical analyses did not find any significant role for task modality. The results hinted that the cognitive demand of these tasks was not sufficiently different from each other. More research attention should thus be given towards the conceptualization, elaboration and refinement of L2 tasks that can tap into the multifaceted nature of adult L2 learners' speech abilities. In the domain of L2 grammar, previous SLA studies have convincingly shown that L2 learners differ in their attentional orientations to the accurate, fluent and complex use of language, and that their performance greatly changes according to different task conditions on these dimensions, such as the presence/absence of preplanning time, online planning time, task repetition and background/foreground information (Ahmadian & Tavakoli, 2014). Recently, some scholars have begun to investigate the generalizability of the topic to L2 speech learning, indicating that such task manipulations can also impact accuracy of phonetic forms (Solon, Long, & Gurzynski-Weiss, 2016). Additionally, it needs to be acknowledged that the findings of the study were exclusively limited to the three production measures. Given that much learning initially takes place at the perception (rather than production) level (Bradlow, 2008; Flege, 2009), future studies should include a range of perception *and* production measures to explore the impact of awareness on different stages of L2 speech learning (perception → production).

References

Aguilar, M., & Muñoz, C. (2014). The effect of proficiency on CLIL benefits in Engineering students in Spain. *International Journal of Applied Linguistics, 24*, 1-18.

Baker-Smemoe, W., & Haslam, N. (2013). The effect of language learning aptitude, strategy use and learning context on L2 pronunciation learning. *Applied linguistics*, *34*, 435-456.

Boersma, D., & Weenink, P. (2012). *Praat: Doing phonetics by computer version 5.3.14.* Retrieved from http://www.praat.org

Boo, Z., Dörnyei. Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System, 55*, 147-157.

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.

Catford, J. C. (1987). Phonetics and the teaching of pronunciation: a systemic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory* (pp. 83-100). Washington, D.C.: TESOL.

Celce-Murcia, M., Brinton, D., Goodwin, J., & Griner, B. (2010). *Teaching pronunciation: A course book and reference guide*. Cambridge: Cambridge University Press.

Cobb, T. (2014). *The compleat lexical tutor*. Retrieved from http://www.lextutor.ca/vp/

DeKeyser, R. (2007). Skill acquisition theory. In J. Williams., & B. VanPatten (Eds.). *Theories in Second Language Acquisition: An introduction* (pp. 97–113). Mahwah, NJ: Erlbaum.

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

Dewaele, J.-M. & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and Enjoyment in the Foreign Language Classroom. *Studies in Second Language Learning and Teaching,* 4, 237-274.

Espy-Wilson, C. (1992). Acoustic measures for linguistic features distinguishing the semivowels /wjrl/ in American English. *Journal of Acoustical Society of America, 92*, 736–751.

Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp.175–190). Clevedon: Multilingual Matters.

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

Hama, M., & Leow, R. P. (2010). Learning without awareness revisited. *Studies in Second Language Acquisition*, *32*, 465-491.

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

Idemaru, K., & Holt, L. L. (2013). The developmental trajectory of children's perception and production of English/ɹ/-/l/a). *The Journal of the Acoustical Society of America*, *133*, 4232-4246.

Ingvalson, E. M., McClelland, J. L., & Holt, L. L. (2011). Predicting native English-like performance by native Japanese speakers. *Journal of Phonetics, 39*, 571-584.

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

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*, 35–63.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

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.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior. *Studies in second language acquisition*, *22*, 557-584.

Leow, R. P., Johnson, E., & Zárate-Sández, G. (2011). Getting a grip on the slippery construct of awareness: Toward a finer grained methodological perspective. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 61-72). Washington, D.C.: Georgetown University Press.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition.* doi: 10.1017/S027226311500042X

Muñoz, C. (2006). The effects of age on foreign language learning: The BAF Project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning (pp. 1–40)*. Clevedon: Multilingual Matters.

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics, 35*, 463–482.

Nishino, T., & Watanabe, M. (2008). Communication-oriented policies versus classroom realities in Japan. *TESOL Quarterly*, *42*, 133-138.

Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., & Hagiwara, H. (2011). Age and amount of exposure to a foreign language during childhood: Behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neuroscience research*, *70*, 197-205.

Piske, T., Flege, J., MacKay, & Meador, D. (2011). Investigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 195–205). Frankfurt am Main: Peter Lang.

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

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language learning*, *45*, 283-331.

Saito, K. (2013a). Age effects on late bilingualism: The production development of /ɹ/ by high-proficiency Japanese learners of English. *Journal of Memory and Language, 69*, 546-562.

Saito, K. (2013b). Re-examining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition, 35*, 1-29.

Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning, 65*, 563-595.

Saito, K., & Brajot, F. (2013). Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English /ɹ/ by late Japanese bilinguals. *Bilingualism: Language and Cognition, 16*, 847-863.

Saito, K., Dewaele, J.-M., & Hanzawa, K. (2017). A longitudinal investigation of the relationship between motivation and late second language speech learning in classroom settings. *Language and Speech*, *60*, 614-632.

Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning, 62*, 595-633.

Saito, K., & Munro, M. (2014). The early phase of /ɹ/ production development in adult Japanese learners of English. *Language and Speech, 57*, 451-469.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

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

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-based versus production-based grammar instruction: A meta-analysis of comparative studies. *Language Learning, 63*, 296–329.

Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2016). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*. Advanced online publication. doi: 10.1017/S0272263116000425

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

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in second language acquisition*, *16*, 183-203.

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.

Venkatagiri, H., & Levis, J. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness, 16*, 263–277.