Title:

Variables affecting the effects of recasts on L2 pronunciation development

Running Head:

SCRUTINIZNG PHONOLOGICAL RECASTS

Author Information:

Kazuya Saito

Waseda University

School of Commerce

1-6-1 Nishi Waseda,

Shinjuku, Tokyo

169-8050, JAPAN

Email: kazuya.saito@waseda.jp

Notes on Contributors:

Kazuya Saito is an Assistant Professor of English in the School of Commerce at Waseda University (Tokyo, Japan). His research focuses on how instruction and corrective feedback can help child and adult learners develop their L2 oral skills especially in the domain of pronunciation, listening, vocabulary, and pragmatics.

Variables affecting the effects of recasts on L2 pronunciation development

Kazuya Saito (Waseda University)

The current study investigated how recasts can promote the L2 pronunciation development of word-initial /ɹ/ by Japanese learners of English in relation to two developmental stages of English /ɹ/ acquisition (i.e., change in second formant [F2] → change in third formant [F3]) as well as four affecting variables (i.e., the amount of recasts and repairs, initial pronunciation levels, explicit articulatory knowledge). Fifty-four Japanese learners of English participated in the study and received four hours of form-focused activity. While teachers gave pronunciation-focused recasts to students in the experimental group (n = 29), no recasts were directed to those in the control group (n = 25). According to the results of the rating session, which involved 20 NS listeners, their positive perception of /ɹ/ was associated with lower F3 values as the primary cue; lower F2 values were demonstrated to be secondary, and less relevant, for listeners' positive perception of /ɹ/. The results of the ANOVAs showed that whereas the F2 values of both the experimental and control groups equally decreased (i.e., due to the use of the interlanguage strategy), only the experimental group significantly lowered their F3 values (the extent of acquisition). Furthermore, a range of post-hoc analyses found that recast effectiveness was related to the amount of recasts and repairs and initial pronunciation levels.

Key words: Recasts, Second language phonetics, Pronunciation teaching, Focus on form, Form-focused Instruction, Explicit knowledge, English /ɹ/

I Introduction

Drawing the attention of Second Language (L2) learners' to form via recasts is a complex phenomenon. This is arguably because L2 learners do not always succeed in perceiving implicit feedback (such as recasts) as being language-focused, especially when the feedback is directed to morphosyntactic errors. Several observational studies (e.g., Sheen, 2006), however, have found that recasts can be quite salient to learners when their targets are L2 pronunciation errors; this suggests that recasts might be relatively facilitative of L2 pronunciation development. The current study examines in depth (a) how recasts impact L2 pronunciation development in relation to two developmental stages of the acquisition of /ɹ/ by Japanese learners of English (i.e., change in second formant [F2] → change in third formant [F3]), and (b) to what degree recast effectiveness varies according to a range of individual variables, which include the amount of recasts and immediate repairs, initial pronunciation levels, and explicit articulatory knowledge.

1 Recasts in SLA

According to Lyster and Ranta's (1997) classification of types of feedback, recasts are defined as "the teacher's reformulation of all or part of a student's utterance minus the error" (p. 46); an example of a recast from the current study is as follows[1]:

Example 1

S: It's good to start reading /ɾidɪŋ/* something.

T: Reading /ɹidɪŋ/.

S: Reading /ɹidɪŋ/ something. Even it is comic books or novels.

From a theoretical perspective, some second language acquisition (SLA) researchers assume that recasts can simultaneously provide both positive and negative evidence without compromising communicative flow. This allows for second language (L2) learners to make cognitive comparisons between their nontargetlike forms and targetlike reformulations during meaningful discourse (Long, 2007). Other researchers, however, have argued that recasts might not be the most effective type of feedback, at least for the development of L2 morphosyntax, because it has been contentious to what degree learners can succeed in perceiving the negative evidence presented in recasts (Lyster, 2007).

With respect to L2 pronunciation errors (the focus of this paper), however, the corrective intention of recasts as negative evidence is unambiguous. Several observational studies have found that learners tend to generate more successful repairs following pronunciation-focused recasts than morphosyntax-focused recasts, and to perceive the corrective intention of these recasts. In adult EFL classrooms, Sheen (2006) noted that although pronunciation-focused recasts occurred much less frequently (21.0% of the total number of recasts) than morphosyntax-focused recasts (51.5%), students repeated the former with a higher rate of successful repair (91.8%) than the latter (70.8%) (see also Lyster, 1998). In Carpenter et al.'s (2006) experiment, when asked to watch video-taped segments of task-based interactions including recasts and repetitions, adult ESL learners identified pronunciation-focused recasts as a form of correction more accurately than morphosyntax-focused recasts (see also Mackey, Gass, & McDonough, 2000). Importantly, Carpenter et al. (2006) ascribed learners' sensitivities to pronunciation-focused recasts to the relative importance of pronunciation in successful comprehensibility: Pronunciation errors are "higher in communicative value and more likely to cause communication breakdown" than morphosyntactic errors (p. 228). As these studies indicate, recasts might be highly facilitative of L2 pronunciation development because of their perceived saliency (i.e., students likely consider pronunciation-focused recasts as such).

*a Amount of recasts and repairs***.** The amount of recasts and repairs (i.e., repetition of recasts) and their subsequent impact on acquisition is an empirical and theoretical inquiry which is open to much debate. With respect to L2 morphosyntactic development, research findings have generally shown that L2 learners benefit from recasts irrespective of opportunities for repairs (Loewen & Philps, 2006). This indicates that the effectiveness of recasts stems from repeated exposure to enhanced positive (and possibly negative) evidence; however, it might be unrelated to self-modified output following recasts.

In L2 phonology, little attention has been paid to repetitive output practice because it is viewed as reminiscent of the decontextualized practice typical of traditional pronunciation teaching (for discussion, see Trofimovich & Gatobonton, 2006). However, repetitions of recasts that occur during genuinely communicative L2 interaction can be considered as contextualized repetitive practice, which has an impact not only on accuracy but also on fluency (Segalowitz, 2003). To our knowledge, the current study is the first attempt to examine in depth the relationship between the number of recasts (*input*), the number of repetitions (*output*) and L2 pronunciation development (*acquisition*).

b Learners' initial pronunciation level. Several L2 morphosyntax studies have investigated how recast effectiveness can be influenced by learners' initial interlanguage levels, measured by a well-established sequence of development of certain grammatical features such as questions, possessive determiners, and relative clauses. For example, Mackey and Philp (1998) found that recasts positively influenced learners who were developmentally ready to acquire the target feature, but not those who lacked developmental readiness.

With respect to L2 pronunciation development, some argue that beginner learners tend to display more rapid improvement compared to advanced learners (Derwing & Munro, 2005), stressing that learners' initial pronunciation levels may be a significant variable for the effects of pronunciation instruction. In naturalistic L2 speech learning, the acquisitional value of enhanced input (including a form of instruction) according to learners' proficiency levels has continued to attract attention (see Flege, 2009). Certain researchers have argued that adult L2 learners enhance their pronunciation accuracy to a great degree *only* during the first few years of intensive exposure to the target language, and that this is followed by a plateau stage (Larson-Hall, 2006). Others have claimed that adult L2 learners are able to continue to learn new sounds in relation to the quality and quantity of L2 input even beyond the early phase of L2 learning (Flege, Takagi, & Mann, 1995).

Taken together, future research is needed in order to further examine whether enhanced input (i.e., recasts) can not only (a) help low proficiency learners to expedite the rate of initial L2 learning, but also (b) lead advanced learners to promote the continued growth of ultimate attainment of L2 learning. The current study examines this topic by closely investigating recast effectiveness on L2 pronunciation using learners' initial performance at the pre-test sessions as a covariate.

c Explicit articulatory knowledge. Although SLA theorists have debated whether recasts can help promote the acquisition of new linguistic targets or whether recasts can reinforce existing knowledge of the targets (e.g., Long, 2007 vs. Lyster, 2007), few empirical studies have directly tested the role of learners' explicit knowledge in recast effectiveness. Ellis, Loewen, and Erlam (2006) found that recasts led adult ESL learners with a great deal of explicit knowledge of regular English past tense (measured via metalinguistic tests) to manifest significant gains in their implicit knowledge (measured via timed oral imitation tests). Yet, it still remains unclear how *differential amounts* of explicit knowledge influence recast effectiveness.

Explicit articulatory knowledge in the current study is defined as the conscious knowledge of the articulatory configurations of L2 sounds. This concept corresponds to similar constructs in L2 phonology research such as “phonological awareness” (Venkatagiri & Levis, 2007, p. 265) and “phonological form” (Derwing & Munro, 2005, p. 388). According to the gestural theory of speech perception and production (e.g., a direct-realist position, Fowler et al., 2003), phonetic categories are gesturally-defined (e.g., manner and place of articulation). That is, explicit knowledge about relevant articulatory gestures is hypothesized to help adult L2 learners extract linguistic information from new L2 sounds with modified input (i.e., recasts) quickly and thus establish new phonetic categories effectively. Venkatagiri and Levis (2007) investigated how the differential amount of explicit knowledge of English phonological structures affects the perceived comprehensibility of adult EFL learners. The results of listener judgments showed that those with a superior understanding of L2 phonetic structures tended to show fewer accents in their extemporaneous speech production, suggesting that “comprehensibility can be promoted through promotion of awareness” (p. 276). The current study aims to investigate the relationship between learners’ explicit knowledge of English /ɹ/ (i.e., lip rounding, tongue backness and height) and the effectiveness of recasts on L2 speech pronunciation development.

2 English /ɹ/

Recasts in the current study target one of the most-well-researched cases of L2 speech learning: the acquisition of word-initial /ɹ/ by Japanese learners of English (for a review, see Bradlow, 2008). In the spectrogram, with the vertical axis representing frequency values (in Hz), the horizontal axis temporal dimensions (in ms), and a darkness of color the amount of energy, acoustic properties of English /ɹ/ can relate to F1 (250-500 Hz), F2 (900-1500 Hz), F3 (1300-1950 Hz), and the relatively long duration of the phoneme (50-100 ms) (see Figure 1) (Espy-

Wilson et al., 2000)[2]. English /ɹ/ requires, in particular, a characteristic drop in the center frequency of F3 as its reliable acoustic correlate (F3 < 2000 Hz), unlike the other approximant sounds (F3 > 2000 Hz).

FIGURE 1 AROUND HERE

With respect to the articulation of /ɹ/, although native speakers (NSs) of English are known to produce the sound with great variance (e.g., bunched vs. retroflexed /ɹ/), relevant articulatory parameters are better described in terms of vocal tract shaping: three constrictions (labial, palatal, pharynx) and a sublingual cavity that are associated with low F3 values (Espy-Wilson et al., 2000).

Because there exists no counterpart for English /ɹ/ in the Japanese phonetic system, inexperienced Japanese learners of English tend to continue to use the Japanese tap /ɾ/, which lies in "a position in a phonological space that is somewhere between English /ɹ/, /l/, and /d/" (Flege et al., 1995, p. 25). To perceive English /ɹ/, at least initially, Japanese learners tend to draw on F2 acoustic variation (e.g., low F2 values for /ɹ/ and high F2 values for /l/), regardless of F3 acoustic variation (Iverson et al., 2003).[3] Saito and Brajot (2013) found that beginner-intermediate Japanese learners (with a few years of LOR) generally produce /ɹ/ by lowering their F2 values (i.e., retracting the body of tongue for /w/-like production) while keeping their F3 values unaltered (2500-2700 Hz).[4]

Bradlow (2008) attributed these interlanguage patterns of Japanese learners (i.e., relying on F2 variation) to the cross-linguistic difference of articulatory gestures in the Japanese and English phonetic systems. That is, whereas both Japanese and English include tongue backness (related to F2 variation), a combination of constriction in the palatal and pharyngeal regions of the vocal tract and lip rounding (related to F3 variation) is used in English but not in Japanese. Taken together, Japanese learners (a) initially produce English /ɹ/, simply by drawing on L1-related articulatory characteristics (i.e., tongue retraction); and (b) gradually increase their awareness towards new articulatory configurations (i.e., lip/palatal/pharynx constrictions) in relation to their developing phonetic category for English /ɹ/ (see also Saito & Brajot, 2013). In this regard, a change in F2 value is considered to reflect the use of an interlanguage strategy (irrelevant to /ɹ/), and the reduction in F3 values is interpreted as the extent of acquisition (relevant to /ɹ/). The acquisitional sequence of /ɹ/ is visually summarized in Figure 2.

FIGURE 2 AROUND HERE

The current study examines in depth how recasts promote the shifting of learners' attention away from the interlanguage F2 strategy and towards the acquisition of the new parameter, F3, in various task and phonetic contexts.

3 Motivation for current study

Due to the general lack of relevant studies (see Saito, 2012 for research synthesis on pronunciation teaching studies in the field of instructed SLA), Saito and Lyster (2012) took a first step towards testing how a range of form-focused instructional (FFI) techniques, including recasts, can promote the acquisition of English /ɹ/ by adult Japanese learners. The results of the acoustic analysis provided some evidence that learners receiving recasts during FFI demonstrated improvement. In the original study, accordingly with the complex nature of the interlanguage development of /ɹ/ and the multiple functions of recasts, the results raised several questions worthy of further pursuit to obtain a full-fledged understanding of recast effectiveness. To begin, although the earlier study focused exclusively on F3 as the only significant predictor of Japanese learners' performance of /ɹ/, it remains unclear how F3 as well as other acoustic properties of /ɹ/ (e.g., F2, F1, transition duration) interact to determine NS listeners' perceptions of natural /ɹ/ tokens under different task and phonetic conditions. Second, the effects of recasts on such multifaceted acoustic characteristics of /ɹ/ still need to be scrutinized. Finally, if recasts are facilitative of developing /ɹ/, we still need to answer which aspects of recasts (providing enhanced input and/or eliciting self-modified output) actually lead to such effectiveness. In this regard, drawing on the database of the original study, the current study adopted three new analyses: (a) re-examining the perception of /ɹ/ via a new rating session with 20 listeners; (b) implementing acoustic analyses not only on F3 but also on F2 for the entire data set; and (c) analyzing the individual performance of the students receiving recasts in comparison with four moderator variables (i.e., the number of recasts and repairs, initial pronunciation levels, explicit articulatory knowledge).

4 Research questions

The primary purpose of the present study is to provide a detailed description of recast effectiveness in relation to two developmental stages of acquisition of word-initial of /ɹ/ (i.e., change in F2 → change in F3), and in accordance with four moderator variables (i.e., the amount of recasts and repairs, initial pronunciation levels, and explicit knowledge). The research questions and hypotheses to be addressed are as follows:

1. How do the acoustic properties of English /ɹ/ (F1, F2, F3 and transition duration) and contextual factors (task types, ensuing vowel backness and height) relate to NS listeners' perceptions of /ɹ/?

2. How do recasts impact L2 pronunciation development of /ɹ/ in relation to F3 (the extent of acquisition) and F2 (the use of the interlanguage strategy)?
3. To what degree does recast effectiveness vary according to (a) the amount of recasts, (b) the amount of repairs, (c) initial pronunciation levels, and (d) explicit articulatory knowledge?

II Method

1 Design

The current study employed a quasi-experimental design with a pre-test and post-test in a simulated ESL classroom setting. First, after individually completing pre-tests and the initial interview, student participants received four hours of FFI treatment with and without recasts provided by two NS teachers. Two weeks after the last day of classes, the students came back to the author's office to complete post-tests as well as to have a final interview. All classes were conducted in one classroom at an English-speaking university in Montreal. Second, 20 NS listeners rated 150 speech samples produced by the participants during the pre-test sessions in order to find out which acoustic properties and contextual factors significantly affected their rating scores. Finally, based on these relevant acoustic properties and contextual factors, the students' improvement of /ɹ/ was carefully assessed.

2 Participants

a Students. Volunteer participants in the current study were 54 Japanese learners in Montreal, Canada. Their length of residence varied widely from one month to 13 years (M = 15.1 months, Mdn = 6.0 months, SD = 30.13). These participants were students either at private language institutes or English-speaking universities at the time of the project. Their mean age was 29.5 years (SD = 6.9). To avoid too much focus on pronunciation forms during instruction, the participants were told that the objective of the class was to help them gain argumentative skills in English. The pronunciation component of the lesson was not mentioned until they had completed the post-tests. After the pre-tests, the 54 students were first randomly divided into 10 classes (5-6 students per class), and grouped into either the experimental group (*n* = 29 [4 males/25 females], five classes) or the control group (*n* = 25 [5 males/20 females], five classes).

b Teachers. Two experienced ESL teachers participated in the study. The first teacher taught six classes (3 classes in the experimental group, 3 classes in the control group), and the second teacher taught four classes (2 classes in the experimental group, 2 classes in the control group).

c Listeners. Twenty NS graduate and undergraduate students in Montreal (their L1 background was a north-eastern dialect of American English) participated as listeners.

3 Interview

The author individually interviewed all participants in his office right after the post-test sessions (the final interview). The purpose of the interviews was to ascertain whether the learners had had any pre-existing explicit knowledge of /ɹ/ in a retrospective manner after the project was completed. All communication was in Japanese and audio-recorded.

4 Instructional treatment

The main theme of the four one-hour meaning-oriented lessons was to acquire English argumentative skills entailing logical thinking, negotiation and debating skills, as well as public speaking abilities. A variety of form-focused practice activities were also embedded in these contexts so that learners were encouraged to notice and practice the target feature during meaningful discourse. Each activity included at least one target word with /ɹ/ in various phonetic contexts (see Table 1). These target words were highlighted in red in the power point presentations used to deliver the lessons (i.e., typographically enhanced input). Among the activities included were (a) Argument Critique, (b) English debating, (c) Argument Creation, and (d) Public Speaking:

1. *Argument critique*: The students were given a set of logically problematic passages such as "***Ryan*** was able to drive his car well in the ***rain*** yesterday. So, he can drive his car in the snow without any problems," and were taught how to find and critique these problems (i.e., analogy problem).
2. *Debating*: With the critique techniques that they had learned, the students debated several topics such as "Is ***run***ning inside better than ***run***ning outside?" and "Is it good to have a ***rain***y day?"
3. *Argument creation and public speaking*: After learning how to use a logical argumentative sequence (i.e., Introduction → Evidence → Objection → Defense → Conclusion), the students were paired up to develop their own arguments and make a public speech in front of the class. They were encouraged to maintain good control over their eye contact and speak in a loud, clear voice. Topics used in this activity included: "Is ***read***ing comic books good for children?" and "Is a sense of '***rat race***' among students harmful (e.g., tests, entrance examinations)?"

In addition to the main activities, three other activities were used as warm-up games, all of which required students to distinguish /ɹ/ in perception and production at least from Japanese tap /ɾ/ in order to win.

1. *English Karuta*: Thirty-six cards are placed on a table. Each card represents one lexical item, and has a relevant picture with the first letter of the word. When the teachers call out a word from a list, students try to find and pick up the called card as soon as possible.

In order to get many cards, the students have to pay attention to perceptual difference between /ɹ/ and /l/.

2. *English card game*: Each card has two identical sentences except for one minimally paired word (e.g., "I found a beautiful ***leaf***" vs. "I found a beautiful ***reef***"), and the students, in pairs, take turns reading one of the sentences; the other student has to guess which sentence was read. In order to attain many cards, the students have to differentiate their production of the English /ɹ/-/l/ contrast (there were 36 cards in total); the purpose of this activity was to promote their awareness of their production of the English /ɹ/-/l/ contrast, but not their perceptual abilities (note that all students were Japanese learners of English).
3. *Guessing game*: Each card concerns one vocabulary item which is orthographically written on the right-hand corner of the card. Learners, in pairs, take turns explaining what the word is without saying the name of it, and guessing the vocabulary item his or her partner is trying to describe.

The instructional treatment highlighted 38 target words which included /ɹ/ in various phonetic contexts (25 words for word-initial positions, 3 words for word-medial position, and 10 words for consonant clusters). All 38 words were minimally-paired with /l/, which required the students to make a clear distinction between the /ɹ/ and /l/ contrast. These words are presented in Table 1.

TABLE 1 AROUND HERE

----

5 Pronunciation-focused recasts

In the experimental group, the teachers were asked to reformulate students' mispronunciations and unclear pronunciations of /ɹ/ without altering the original meaning of the utterances; this is known as pronunciation-focused recasts. In order to control the linguistic characteristics of recasts (e.g., length [recasting only errors or entire sentences including errors], number of changes [recasting single or multiple errors]) which greatly influence the saliency and efficacy of recasts (Sheen, 2006), the teachers were instructed to consistently recast only one word using a falling intonation, without adding any additional meaning, see the following example.[5]

Example 2 in Argument Critique

S: Analogies are not similar. Driving in the heavy rain /ɾeɪn/* is…

T: Rain /ɹeɪn/

S: Rain. Driving in the heavy rain /ɹeɪn/ is…

6 Control group

The twenty-five learners in the control group also received comparable lessons in English argumentative skills with FFI activities. However, they received no feedback on their pronunciation errors, only on ungrammaticality and inappropriate lexical choices (e.g., "you should say, 'she is manipulative' instead of 'she is tricky'") or an error related to the content of the various lessons (e.g., "you need more evidence to support your idea").

7 Teacher training

Four 1-hour teacher training sessions were held over a two day period prior to the experiment. The researcher explained the purpose and content of the FFI materials and demonstrated when and how recasts should be directed to students.

8 Measures

All of the testing sessions took place individually in a quiet room. The author explained and led the testing procedure in Japanese. Recordings were made via a speech analysis software, *Praat* (Boersma & Weenik, 2009)[6], as well as a unidirectional microphone (DM-20SL).

a Materials. The learners completed three oral tasks; the same materials were used during both the pre- and post-test sessions (*n* = 2,376 tokens). Out of the 38 words featured in the form-focused activities, 14 words including /ɹ/ in word-initial positions were used in the pre-/post-test sessions. All of these words were consonant-vowel-consonant (CVC) singletons except for one (*Ryan* as CVVC; see the words with asterisks in Table 1). Each participant generated 22 word-initial singleton tokens in the pre- and post-test sessions respectively (out of 14 words, eight were tested twice but in different tasks).

1. *Word reading*: In this task, the learners read a list of ten target words (i.e., *read, room, root, rule, red, race, rough, row, ram*, and *right*) together with 15 distracters. All of these words were CVC singletons. In total, 1080 tokens were produced (10 words × 54 students × 2 test sessions = 1,080 tokens).
2. *Sentence reading*. In this task, the learners read five sentences where eight target words were included together with three distracter sentences:

   He will <u>read</u> my paper by the time I arrive there.

   She left her <u>red</u> bicycle on the side of the <u>road</u>.

   The <u>race</u> was cancelled because of the <u>rain</u>.

   I can correct all <u>wrong</u> sentences tonight.

   <u>Ryan</u> does not like to <u>run</u> in the snow.

   In total, 864 tokens were created (8 words × 54 students × 2 test sessions = 864 tokens).

3. *Timed picture description*. In this task, the learners described eight pictures, each of which was accompanied by three words that learners were prompted to use in their descriptions. Whereas four pictures served as distracters and contained no target words, the other four pictures included four target words (i.e., *read, rain, rock, road*). Note that this task was timed such that the learners had only five seconds to prepare before describing a given picture. In total, 432 tokens were created (4 words × 54 students × 2 test sessions = 432 tokens).

b NS baseline. For comparison purposes, six native speakers of English (three males, three females) who were undergraduate students at an English-speaking university in Montreal completed the same tasks (Word Reading, Sentence Reading, and Timed Picture Description).

9 Listener ratings and acoustic analyses

Twenty NS listeners rated 150 tokens produced during the pre-test sessions by 30 of the 54 Japanese learners of English on a 9-point scale. Subsequently, four speech properties (F1, F2, F3 and duration of F3 transition) and three contextual factors (task types, following vowel backness and height) of these speech tokens were carefully analyzed by the author in order to see which variables significantly influence the NS listeners' rating scores of English /ɹ/.

a Token preparation. Drawing on the set of 1,188 tokens produced by learners in the pre-test sessions, 150 speech samples were created. First, 30 learners were randomly selected (15 from the experimental group and 15 from the control), and then each learner contributed five words (two from Word Reading, two from Sentence Reading, and one word from the Timed Picture Description).[7] Second, after listening to the original speech samples several times, the target words within those speech samples were identified and isolated by the author using *Praat*.

With *Praat* (a speech analysis application), any component of /ɹ/ can be heard by moving the cursor onto the onset of the target word and moving it towards its offset by 5 ms.

b Acoustic properties and contextual factors. For the 150 speech tokens, the author measured the F1, F2, and F3 values in Hz and transitional duration of F3 in milliseconds by means of *Praat* (Boersma & Weenik, 2009), and categorized them based on (a) task type (word reading, sentence reading, picture description), and (b) according to the backness (front, central, back) and height (high, mid, low) of subsequent vowels, respectively. Given that Flege et al.'s (1995) study is one of very few studies that acoustically analyzed natural tokens of /ɹ/ produced by Japanese learners, the current study followed their procedure in order to measure three spectral cues (F3, F2, F1) through linear predictive coding spectra[8] and one temporal cue (transition duration) as to word-initial /ɹ/. Two steps were taken as follows:

1. The beginning of the words (where /ɹ/ starts) was determined by using both the spectrographic representations and wave forms of the speech tokens. To identify the beginning of the tokens embedded in continuous speech (i.e., Sentence Reading, Timed Picture Description), the researcher focused on the endpoint of the falling F3 (i.e., local peak) as a reliable clue.[9] The frequency values of F1, F2 and F3 were then measured by putting the cursor on this point.
2. The duration of the F3 transition was measured by dragging the cursor from the beginning of its transition (where F3 started rising towards the following vowels) to the endpoint of the F3 transition (i.e., the increase in F3 values stopped when reaching the beginning of the following vowels). The procedures of the acoustic analyses are summarized in Figure 1.

In order to reduce spectra variations due to the nonlinear relationship between the formant frequencies and the corresponding perceived quality (more similar to human perception), the raw acoustic values in Hz were converted into Bark using the formula described in the *Praat* manual (Boersma & Weenik, 2009).[10]

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

c Rating sessions. The rating sessions took place individually with each listener in a quiet room for approximately one hour. First, a 9-point scale descriptor was adapted and modified from Flege et al.'s (1995) 6-point scale, and the rating criteria were explained as follows: 1 (*very good /ɹ/*) → 2 (*good /ɹ/*) → 3 (*probably /ɹ/*) → 4 (*possibly /ɹ/*) → 5 (*neutral exemplars, neither /ɹ/ nor /l/*) → 6 (*possibly /l/*) → 7 (*probably /l/*) → 8 (*good /l/*) → 9 (*very good /l/*). Because they were allowed to listen to the speech samples more than once (with an average of two or three times), it can be said, in accordance with the researcher's observations, that the listeners were confident about their ratings. It was also carefully explained to them that the purpose of the rating session, as well as its procedure, to judge only the quality of /ɹ/ and not the entire word unit. Also, to reduce the chance of having biased opinions towards choosing only /ɹ/, the listeners were encouraged to guess whether the speech tokens sounded like /ɹ/, /l/ or neither, using the entire range of the 9-point scale as much as possible. Finally, they rated five speech tokens (not included in the subsequent listening session) on a 9-point scale as practice to become familiarized with the procedure.

10 Post-hoc analyses

a Dependent variables. The differences between the pre- and post-test scores within and among the participants of the experimental group (i.e., F3 and F2 values at the post-test sessions minus F3 and F2 values at pre-test sessions) were used as the dependent variable to represent recast effectiveness. For example, if a learner produced F3 values of 13.27 Bark in the pre-test and 12.69 Bark in the post-test, he or she received a score of 0.58 Bark (13.27 Bark – 12.69 Bark = 0.58 Bark).

b Independent variables. Four independent variables were analyzed as follows:

1. *Amount of recasts and Repairs*: The learner-external factor was the number of recasts directed at a single learner and the number of times there was immediate learner repair. The author watched all video-taped lessons of the experimental group (4 hr of lessons × 5 classes = 20 hr) and noted (a) how many times each student received recasts from their teacher (i.e., the number of recasts), and (b) how many times he/she repeated the teacher's recast (i.e., the number of repairs).

   To code pronunciation error treatment, we used the following procedure in line with Lyster and Ranta's (1997) original notions of uptake, topic continuation, repair, and needs-repair. First, we paid attention to whether the learners showed uptake (i.e., any reaction immediately following the teacher's recasts) or topic continuation (i.e., the teacher's recasts went completely unnoticed, or the teacher did not provide opportunities for uptake). If there was uptake, we then checked whether the learners repeated the teacher's recasts (i.e., repair) or gave a simple acknowledgement (e.g., "yes" or "no") in response to the teacher's feedback. Due to the nature of pronunciation errors[11], we did not attempt to further analyze precisely how much the learners self-corrected their own errors (for a similar inclusive coding approach for phonological recasts, see also

Carpenter et al., 2006; Sheen, 2006 and many others). Some examples of topic continuation (Example 3), repair (Example 4) and needs-repair (Example 5) were as follows:

Example 3 – Topic Continuation (in Argument Critique)

S: Japanese cuisine always comes with rice /ɾaɪs /*

T: Rice /ɹaɪs/

S: It also comes with fish and a lot of vegetable. ← NO UPTAKE

Example 4 – Repair (in English Debate)

S: We believe that running /ɾʌnɪŋ/* inside is better than running /ɾʌnɪŋ/* outside.

T: Running /ɹʌnɪŋ /. → RECAST

S: Running /ɹʌnɪŋ /. Because running /ɹʌnɪŋ /inside is not influenced by the weather. ← REPAIR

Example 5 – Needs-repair (in Guessing Game)

S: Reading /ɾidɪŋ/*comic books promotes creativity

T: Reading /ɹidɪŋ/ ← RECAST

S: Yes. ← NO REPAIR

T: Okay, I see.

2. *Learners' initial pronunciation levels:* The learners' pre-test scores were used to indicate their initial pronunciation levels.
3. *Explicit knowledge*: At the endpoint interview, the learners were asked if they had any explicit knowledge of the articulatory gestures of /ɹ/. Their self-reported explicit knowledge was classified into two categories: (a) height and backness of tongue positions,

and (b) lip rounding. For each category, if they showed some knowledge[12], they were given a "1"; if not, they were given a "0."

III Results

1 Listener ratings and acoustic analyses

The interclass correlation between the 20 NS listeners was computed at .803 for the entire data set (n = 150), .789 for word reading (n = 60), .815 for sentence reading (n = 60), and .815 for picture description (n = 30). Given that the 20 NS listeners showed significantly high interrater agreement ($p < .001$), their rating scores were averaged for each speech token.

A multiple regression analysis was conducted using the rating scores as the dependent variable, and four speech properties (F1, F2, F3, F3 transition) and three contextual factors (task types, following vowel height and backness) as the independent variables. The simultaneous method was run to assess the relative contribution of the four predictors to the entire model, accounting for variance in the NS listeners' rating scores.

An ANOVA performed on the regression model showed that the predictive power of the model was statistically significant, F (7, 142) = 35.464, $p < .001$, and that it accounted for 63.6% of the variance in the rating scores. The model identified three variables as significant predictors: F3 (t = 8.409, $p < .001$), F2 (t = 1.992, p = .048) and following vowel backness (t = 2.112, p = .036). The standard coefficients of the three variables are as follows: F3 (β = .651), F2 (β = .141), and following vowel backness (β = .129). In order to further examine the relative weights of F3 and F2 values in NS listeners' perception of /ɹ/ according to following vowel backness, further analyses were conducted.

a F3 values. The partial correlation analyses showed that the correlation between F3 values and the NS rating scores remained highly significant when the effects of variation in F2 values and following vowel backness were removed, $r = .670$, $p < .001$; F3 values accounted for 44.9% of the variance in the NS listeners' rating scores. Interestingly, the simple correlation analyses did not find any significant correlation between F3 values and following vowel backness among 150 speech tokens, $r = .096$, $p = .2403$. Thus, any decline in F3 values could be interpreted as improvement regardless of following vowel backness, and so F3 values were treated as *a primary phonetic cue* in the present study.

A closer examination of the relationship between the F3 and the listeners' 9-point rating scores revealed a benchmark for NS perception of /ɹ/. First, 150 tokens were categorized into five groups based on the NS rating scores: (a) *good English /ɹ/* ($1 \leq x \leq 2.5$), (b) *poor English /ɹ/* ($2.5 < x \leq 4$)), (c) *neutral exemplars* (i.e., neither English /ɹ/ nor English /l/) ($4 < x < 6$), (d) *poor English /l/* ($6 \leq x < 7.5$), and (e) *good English /l/* ($7.5 \leq x \leq 9$). Second, NS tokens were also calculated and labelled as *native-like English /ɹ/* for comparison reasons (6 talkers × 22 tokens = 132 tokens). The results of the descriptive statistics are summarized in Table 2. Finally, 95% confidence intervals (CI) were calculated to analyze whether the six groups of F3 values were significantly different. The results showed that the five groups (native-like English /ɹ/ vs. good English /ɹ/ vs. poor English /ɹ/ vs. neutral exemplars vs. good English /l/) proved to be independent of each other at a $p < .05$ level. However, F3 values of poor English /l/ did not significantly differ from those of neutral exemplars nor those of good English /l/. Taken together, the learners' performance was assessed on five levels: (a) native-like English /ɹ/ (CI = 11.45-11.79 Bark) (b) good English /ɹ/ (CI = 12.97-13.42 Bark), (c) poor English /ɹ/ (CI = 13.72-14.42 Bark), (d) confusing exemplars (CI = 14.50-15.14 Bark), and (e) English /l/ (CI > 15.29 Bark).

TABLE 2 AROUND HERE

----

b F2 values. The partial correlation analyses also found a significant correlation between F2 values and the NS rating scores when effects of variation in F3 values and ensuing vowel backness were removed, $r = .190$, $p = .02$. However, F2 values accounted for only 3.6% of the variance in the NS listeners' rating scores. Not surprisingly, F2 values were highly correlated with ensuing vowel backness, $r = .389$, $p < .001$: $M = 11.91$ Bark for front vowels ($SD = 3.98$) → $M = 10.79$ Bark for back vowels ($SD = 3.51$). Although F2 values did play a role in NS perception of /ɹ/ to some degree (the lower the F2 values were, the more positive the NS perceptions tended to be), their decline could be confounded with characteristics of the following vowel (the further back ensuing vowels were, the lower F2 tended to be). In this regard, F2 values were treated as *a less relevant phonetic cue* for English /ɹ/ in the subsequent analyses.

c Data analyses. Given that the results identified neither task type nor following vowel height as significant variables for NS listeners' perception of /ɹ/, the influence of these variables is not discussed further in the current study. That is, the rest of the entire data set ($n = 2376$) was analyzed by focusing only on F3 (i.e., the primary cue) and F2 (the less relevant cue) according to three phonetic contexts: English /ɹ/ following front, central, and back vowels.

2 Descriptive analysis

First, we descriptively examined how individual learners changed their individual interlanguage performance of /ɹ/ in response to FFI together with recasts. With respect to F3 (the extent of acquisition), a visual inspection of the profile plots for the experimental (Figure 2) and control (Figure 3) groups indicates that (a) most of the learners who received recasts and FFI showed equally declined F3 values (more targetlike exemplars of /ɹ/) regardless of following vowel conditions, and (b) those receiving FFI alone did not necessarily show F3 decline.

FIGURE 2 AROUND HERE

FIGURE 3 AROUND HERE

With respect to F2 (the use of interlanguage strategy), a visual inspection of profile plots for the experimental (Figure 4) and control (Figure 5) groups suggests that almost all learners in the current study equally reduced F2 values after instruction, irrespective of recast treatment.

FIGURE 4 AROUND HERE

FIGURE 5 AROUND HERE

----

3 General linear model ANOVAs

Next, we computed a set of analyses of variance (ANOVAs) to find whether the descriptive observations could be statistically significant as generalizeable patterns. F3 and F2 values of each participant's performance were averaged according to following vowel backness (10 tokens for front vowels, 2 tokens for central vowels, 10 tokens for back vowels) at two different times (i.e., pre- vs. post-test sessions), respectively. Three-way General Linear Model ANOVAs (Group × Time × Backness) were separately conducted for the F3 and F2 values. The alpha level was set at a $p < .05$ level for all statistical analyses.[13] Cohen's *d* was also calculated in order to measure the magnitude of instructional effectiveness between two contrast groups of means.[14]

With respect to F3, the ANOVA results revealed significant effects for the overall Group × Time interaction, F (1, 52) = 15.918, $p < .001$ regardless of following vowel conditions. A simple main effect of time was significant only for the experimental group (M = 14.46 Bark → M = 13.84 Bark), F (1, 52) = 40.416, $p < .001$, d = 0.61. In general, learners receiving recasts made a substantial transition from confusing exemplars (around 14.50-15.14 Bark) to intelligible exemplars of /ɹ/ (around 13.72-14.42 Bark) with medium effects. With respect to F2 values, however, the ANOVA results revealed only overall effects for Time (M = 10.79 Bark → M = 10.45 Bark), F (1, 52) = 8.355, $p < .001$, d = 0.31. F2 declined equally in both the experimental and control groups with small effects.

The raw acoustic data of Japanese learners as well as the NS baseline are summarized in Table 3.

TABLE 3 AROUND HERE

----

4 Mixed model ANOVAs

As post-hoc analyses, Mixed Model ANOVAs[15] were conducted to analyze the effects of recasts in the experimental group (i.e., within-group difference between pre- and post-test scores) as a dependent variable relative to six independent variables: (a) one repeated variable (/ɹ/ preceding front, central, and back vowels), (b) three covariates (the amount of recasts and repairs, initial pronunciation levels), and (c) two factors (explicit articulatory knowledge of tongue positions, explicit knowledge of lip rounding).

a Amount of recasts and repairs. In order to ensure the validity of the author's coding (watching 20 hours of video-taped lessons to count the number of recasts and repairs moves for each student), after some brief instruction, a NS of Japanese with near native-like proficiency in English watched 20% of the data for four hours and counted the number of recasts and repairs. The inter coder reliability was significantly high for the number of recasts ($r = .92$) and the number of repairs ($r = .83$). The results of the descriptive statistics showed that each learner received 32.59 recasts on average ranging from 16 to 57 ($SD = 11.10$), followed by a high number of repairs ($M = 26.68$, $SD = 9.60$), which indicates a relatively high rate of repair (i.e., 81.90%).

b Explicit articulatory knowledge. Whereas 23 out of 29 participants in the experimental group (79.3%) reported explicit knowledge about tongue movement (e.g., unlike the Japanese tap /ɾ/, the tip of the tongue does not touch the alveolar ridge in English /ɹ/), only nine out of 29 participants (31.0%) actually knew about the lip rounding.

c Results of ANOVAs. According to the first ANOVA, using F3 change as the dependent variable, significant fixed effects were found for (a) the amount of recasts, $F(1, 73.442) = 4.539$, $p = .036$, (b) the amount of repairs, $F(1, 73.292) = 4.804$, $p = .032$, and (c) learners' initial pronunciation levels in F3, $F(1, 74.730) = 13.935$, $p < .001$ respectively. According to the second ANOVA, using F2 change as the dependent variable, only learners' initial pronunciation levels exhibited significant fixed effects, $F(1, 61.763) = 55.104$, $p < .001$. That is, whereas learners with higher F3 and F2 values at the beginning of the project (less-advanced learners) tended to attain larger reduction in F3 and F2 values, the amount of recasts and repairs was significantly related to F3 values (the extent of acquisition) but not to F2 values (the use of the interlanguage strategy).

IV Discussion

1 NS perception of English /ɹ/

To answer the first research question, which asked how the acoustic properties of /ɹ/ (F1, F2, F3 and duration) and contextual factors (task type, ensuing vowel backness and height) relate to NS listeners' perception of /ɹ/, 20 NS listeners rated 150 speech tokens produced by 30 out of the 54 Japanese learners during the pre-test session. The results showed that their positive judgments were highly correlated with lower F3 values, regardless of the ensuing vowel contexts. Although lower F2 values were very weakly correlated with the positive judgments, a decline in F2 values was rather highly related to ensuing vowel backness.

Thus, any decline in F3 could be considered as *improvement* (i.e., beneficial effects of instructional treatment), whereas a change in F2 could be considered as either an effect of the following vowel (F2 values following back vowels tend to be lower) or the interlanguage strategy (i.e., Japanese learners are sensitive to F2 variation). The results corroborated similar findings using synthesized speech tokens (e.g., Iverson et al., 2003) as well as natural speech tokens (e.g., Flege et al., 1995). The learners' performance was roughly categorized as follows (this benchmark was used to assess improvement resulting from instruction):

- Native-like English /ɹ/ (F3 = 11.45-11.79 Bark [1672-1759 Hz])
- Good English /ɹ/ (F3 = 12.97-13.42 Bark [2028-2174 Hz])
- Poor English /ɹ/ (F3 = 13.72-14.42 Bark [2283-2529 Hz])
- Neutral exemplars (F3 = 14.50-15.14 Bark [2550-2797 Hz])
- English /l/ (F3 > 15.29 Bark [2850 Hz])

2 Effects of recasts

In response to the second research question, which asked whether recasts influence Japanese learners' L2 pronunciation development of /ɹ/, the General Linear Model ANOVAs revealed differential effects of recasts according to F3 and F2 values. As shown in the original study (Saito & Lyster, 2012), only participants in the experimental group significantly lowered their F3 values (M = 14.38-14.58 Bark → 13.11-14.14 Bark). This in turn indicates that their improvement from confusing to intelligible exemplars of /ɹ/ was a result of receiving recasts. By contrast, given that both the experimental and control groups equally exhibited a decline in their F2 values, learners receiving FFI regardless of recasts started to improve their production of /ɹ/, but by drawing on L1-related articulatory characteristics (i.e., tongue backness).

In short, when Japanese learners were guided to notice /ɹ/ in L2 input via FFI activities, they first turned to the default strategy for altering their L2 performance (i.e., retracting movement of the tongue: low F2). Subsequently, recasts played a key role in triggering and promoting their gradual adjustment to F3 values, which in turn indicates an increasing amount of learner attention drawn to new articulatory configurations (i.e., lip/palatal/pharyngeal constrictions).

3 Variables affecting recast effectiveness

With respect to the third research question, the results of the Mixed Model ANOVAs shed light on the detailed description of recast effectiveness according to four affecting variables: the amount of recasts and repairs, initial proficiency levels, and explicit articulatory knowledge.

a Amount of recasts and repairs. While the students noted considerably high rates of repair in response to recasts (i.e., 81%) (for similar results, e.g., Sheen, 2006), the amount of recasts and repairs were significantly related to change in F3 values (the extent of acquisition), but not to a change in F2 values (the use of the interlanguage strategy). That is, the more recasts learners received, the more repairs they generated, and the more they became aware of the primary acoustic correlate and articulatory configurations of /ɹ/ (low F3 and lip/palatal/pharyngeal constrictions).

In terms of the amount of recasts, the results concurred with those of most L2 grammar studies, in that recast effectiveness stems from the exposure to targetlike exemplars of recasts (see e.g., Loewen & Philp, 2006). From an L2 phonology perspective, recasts can make L2 sounds quite salient to Japanese learners (i.e., acoustic enhancements of F3 distinction), and as a result, stimulate the establishment and development of a new phonetic category in their common phonological space (see Kuhl, 2000, p. 11855, for a discussion of modified input in L2 speech acquisition).

Unlike the precursor studies in L2 morphosyntax (e.g., Loewen & Philp, 2006), however, the current study identified learners' repetition of recasts as a significant predictor for recast effectiveness. This could be due to the peculiarities of L2 speech *production* compared to other L2 skills. It is only L2 pronunciation learning that requires both learners' accurate perceptions of speech properties in L2 input (i.e., *cognitive* phase) and their actual usage of articulators to produce correct sounds (i.e., *physical* domain) (see Flege, 2003)[16]. This in turn indicates the relative weight of output-based practice (i.e., repetition) to stimulate both L2 speech perception and production systems. Pronunciation-focused recasts may play an appreciative role in leading L2 learners to self-modify their mispronunciation of /ɹ/ and practice the productive use of the newly-acquired sound in a meaningful manner.

b Initial proficiency levels. The degree of recast effectiveness on F3 (and F2) was significantly related to initial pronunciation levels. The results support the view that beginners have greater room for improvement (Derwing & Munro, 2005). In line with previous L2 phonology studies in naturalistic settings (e.g., Larson-Hall, 2006), the results of the current study also suggest that additional L2 input might be highly beneficial, especially in the early stages of L2 pronunciation development, but that it might not help advanced learners to become more nativelike. In other words, although pronunciation-focused recasts can be effective to help Japanese learners reach intelligible exemplars of /ɹ/ (F3 < 14.50 Bark), other explicit interventions might be necessary to attain a very clear production of /ɹ/ (F3 < 13.50 Bark).

c Explicit articulatory knowledge. According to the results, the existence of explicit knowledge about how to produce /ɹ/ did not necessarily help the students make the best of recast effectiveness. This could be used as indirect evidence to prove that phonetic categories are based on perception rather than articulatory gestures: Perceiving new sounds through recasts is hypothesized to activate relevant sensorimotor skills and thus result in improved production ability, regardless of explicit articulatory knowledge (i.e., a perception-first view: Flege, 2003; Kuhl, 2000). The findings also coincide with previous perception training studies which showed that L2 learners who were intensively exposed to an ample amount of modified input of new sounds not only improved their perception performance, but also transferred this gain to production levels, even without explicit articulatory instruction on various aspects of L2 phonological development (Bradlow et al., 1997).Yet, this conclusion must remain speculative due to the exploratory nature of the methodology (i.e., self-report). Given that there still does not exist any standardized measure for explicit phonological knowledge (Venkatagiri & Levis, 2007), we need to wait for future studies to elaborate and validate reliable assessments.

V Conclusion and Future Directions

In conjunction with the complexities of the interlanguage development of /ɹ/ and multiple functions of recasts, the current study investigated the acquisitional value of pronunciation-focused recasts according to two developmental stages (i.e., change in F2 → change in F3) as well as four affecting variables (i.e., the amount of recasts and repairs, initial pronunciation levels, explicit articulatory knowledge). The comparison between the experimental group (FFI with recasts) and the control group (FFI without recasts) showed that although FFI led both groups to equally reduce F2 values (the use of the interlanguage strategy), recasts did play a pivotal role in inducing only the experimental group to lower their F3 values (the extent of acquisition). These findings indicate that pronunciation-focused recasts are an effective way to draw the learners' attention away from their default interlanguage strategy (i.e., F2 variation) and push them towards new phonetic cues (i.e., F3 variation). In addition, a set of post-hoc analyses found that (a) the learners benefited from recasts not only because of the exposure to acoustically-enhanced positive exemplars (the amount of recasts), but also due to the subsequent output opportunities (the amount of repairs); and (b) recasts were beneficial especially at the initial stage of L2 development (i.e., confounding → intelligible exemplars) but not so at the later stage (intelligible → nativelike exemplars).

To close, future directions must be pursued for our further understanding of recasts in instructed L2 phonological learning. First, the amount of recast effectiveness in the current study was considered as small-to-medium (d = 0.30-0.60) rather than large (d > 0.80). One might suggest simply increasing the amount of instruction (more than 4hr) as a remedy to boost the magnitude of recast effectiveness. However, instructed SLA studies have convincingly shown that the effects of L2 instruction are highly related to its quality and intensity rather than quantity (see Norris & Ortega, 2000). To trigger more tangible change resulting from recasts, future research is called for to test other remedial techniques, such as the providing of metalinguistic information before FFI (cf. Saito, 2013a). Next, all findings in the current study were exclusively limited to production domains. Given that a change in L2 phonetic categories initially occurs in perception prior to production (Flege, 2003), and adult L2 learners carefully monitor their accurate production forms regardless of the present state of their perceptual representation (Sheldon & Best, 1982), it would be intriguing to adopt both perception and production measures to examine recast effectiveness at different processing levels (cf. Saito, 2013b).

Acknowledgement

The research was made possible through Waseda University Grant for Special Research Projects and the Government of Canada Post-Doctoral Research Fellowship. I am grateful to Norie Moriyoshi and Kim van Poeteren for their help with data analysis. I also would like to thank Roy Lyster and anonymous *Language Teaching Research* reviewers for their insightful comments on earlier versions of the manuscript. Address correspondence to: Kazuya Saito, School of Commerce, Waseda University, 1-6-1, Nishi Waseda, Shinjuku, Tokyo 169-8050, Japan; email: kazuya.saito@waseda.jp

References

Bradlow, A. R. 2008. Training non–native language sound patterns. In Hansen, J and Zampini, M., editors, *Phonology and second language acquisition.* John Benjamins, 287–308.

Bradlow, A., Pisoni, D., Akahane–Yamada, R., and Tohkura, Y. 1997. Training Japanese listeners to identify English /ɹ/ and /l/. *Journal of the Acoustical Society of America* 101. 2299–2310.

Carpenter, H., Jeon, K. S., MacGregor, D., and Mackey, A. 2006. Learners' interpretations of recasts. *Studies in Second Language Acquisition* 28. 209–236.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). New York. Academic Press.

DeKeyser, R. 1998. Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In Doughty, C. and Williams, J., editors, *Focus on form in classroom second language acquisition.* Cambridge University Press, 42–63.

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

Ellis, R., Loewen, S., and Erlam, R. 2006. Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition* 28. 339–368.

Espy–Wilson, C. Boyce, S., Jackson, M., Narayanan, S., and Alwan, A. 2000. Acoustic modeling of American English /ɹ/. *Journal of the Acoustical Society of America* 108. 343–356.

Flege, J. 2003. Assessing constraints on second-language segmental production and perception. In Meyer, A. & N. Schiller., editors, *Phonetics and phonology in language* *comprehension and production, differences and similarities.* Mouton de Gruyter, 319–355.

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

Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. 2003. Rapid access to speech gestures in perception. Evidence from choice and simple response time tasks. *Journal of Memory and Language,* 49. 296–314.

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

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

Loewen, S., and Philp, J. 2006. Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal* 90. 536–556.

Long, M. H. 2007. *Problems in SLA.* Mahwah, NJ. Lawrence Erlbaum.

Lyster, R. 1998. Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. *Language Learning* 48. 183–218.

Lyster, R. 2007. *Learning and teaching languages through content. A counterbalanced approach*. John Benjamins.

Lyster, R., & Ranta, L. 1997. Corrective feedback and learner uptake. *Studies in Second Language Acquisition* 19. 37–66.

Mackey, A., Gass, S., and McDonough, K. 2000. How do learners perceive interactional feedback? *Studies in Second Language Acquisition* 22. 471–497.

Mackey, A., and Philp, J. 1998. Conversational interaction and second language development: Recasts, responses and red herrings? *Modern Language Journal* 82. 338–356.

Rozeboom, W. 1960. The fallacy of the null–hypothesis significance test. *Psychological Bulletin* 57. 416‑428.

Segalowitz, N. 2003. Automaticity and second languages. In Doughty, C. and Long, M., editors, *The handbook of second language acquisition.* Blackwell Publishers, 382–408.

Saito, K. 2012. Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi–experimental intervention studies. *TESOL Quarterly* 46. 842–854.

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

Saito, K. 2013b. Recasts in instructed L2 speech learning: Teaching the perception and production of English /r/ to adult Japanese learners. *Language Learning* 63.

Saito, K., and Brajot, F. 2013. Scrutinizing the role of length of residence and age of acquisition in the interlanguage pronunciation development of English /r/ by late Japanese bilinguals. *Bilingualism. Language and Cognition*16.

Saito, K., and 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.

Sheen, Y. 2006. Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research* 10. 361–392.

Sheldon, A., and Strange, W. 1982. The acquisition of /ɹ/ and /l/ by Japanese learners of English. Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3. 243–261.

Trofimovich, P., & Gatbonton, E. 2006. Repetition and focus on form in L2 Spanish word processing: Implications for pronunciation instruction. *The Modern Language Journal* 90. 519–535.

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

Table 1. *38 Target Words*

Phonetic contexts	Target words
Word–initial	*race, *rain, *ram, rate, *read, *red, reef, rent, *right, rice, ring, rink, river, *road, *rock, rocket, Rome, *room, round, *row, *rule, *run, *Ryan, *wrong, wrap
Word–medial	arrive, correct, pirate
Consonant cluster	bread, crab, crime, crowds, fries, fruit, grass, green, free, pray

* indicates the words included in the pre/post tests

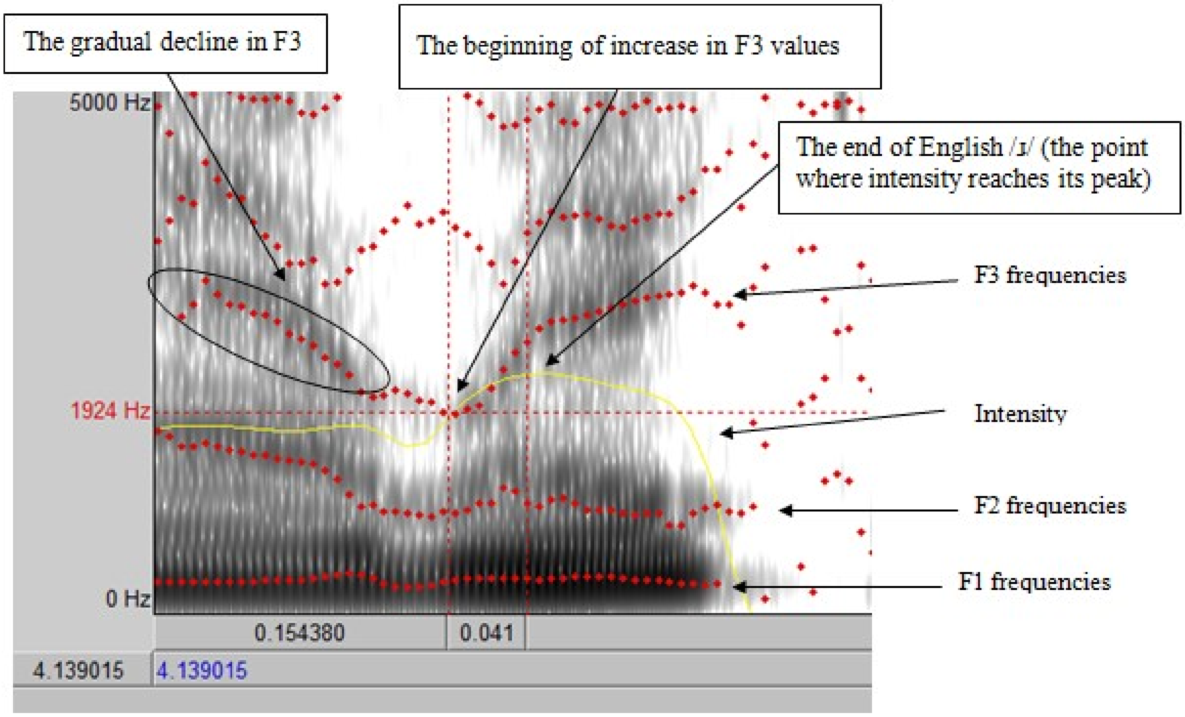
Table 2. *The Results of NS Perception of /ɹ/*

Category	9–point scale	No. of tokens	Bark				Hertz			
			M	*SD*	95% CI		*M*	*SD*	95% CI	
					Lower	Upper			Lower	Upper
Good English /ɹ/	$1 \leq x \leq 2.5$	38	13.19	0.71	12.97	13.42	2101	229	2028	2174
Poor English /ɹ/	$2.5 < x \leq 4$	32	14.07	1.01	13.72	14.42	2406	356	2283	2529
Neutral exemplars	$4 < x < 6$	20	14.82	0.73	14.50	15.14	2673	281	2550	2797
Poor English /l/[a]	$6 \leq x < 7.5$	20	15.22	0.51	15.00	15.45	2830	209	2739	2921
Good English /l/	$7.5 \leq x \leq 9$	40	15.45	0.52	15.29	15.61	2928	213	2850	3004
NS baseline of English /ɹ/		110	11.82	0.96	11.69	11.99	1716	234	1672	1759

Notes. [a] Poor English /l/ was not significantly different from neutral exemplar nor good English /l/ at a $p < .05$ level

Table 3. *The Results of the Raw Acoustic Data in F3 and F2*

		F3 values (Bark)					F2 values (Bark)			
	Following vowels	Pre		Post		Following vowels	Pre		Post	
		M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*
Experimental Group (n = 29)	Front	14.58	0.78	14.14	0.78	Front	11.46	1.16	10.87	1.04
	Central	14.42	0.85	13.61	0.97	Central	10.87	1.06	10.41	0.83
	Back	14.38	0.71	13.77	0.86	Back	10.31	0.95	9.98	0.79
Control Group (n = 25)	Front	14.40	0.85	14.36	0.75	Front	11.37	1.30	11.10	1.13
	Central	14.06	1.04	13.98	0.87	Central	10.50	1.03	10.39	0.92
	Back	14.09	0.78	14.01	0.76	Back	10.17	0.91	9.99	0.89
NS Baseline (n = 6)	Front	11.95	0.90			Front	9.67	0.96		
	Central	11.91	1.01			Central	9.26	1.00		
	Back	11.62	0.79			Back	9.12	0.72		



/ð ə ɹ ɒ k /

"the rock" from the picture description task (spontaneous speech)

Figure 1. Acoustic properties of English /ɹ/ and the analysis procedure.

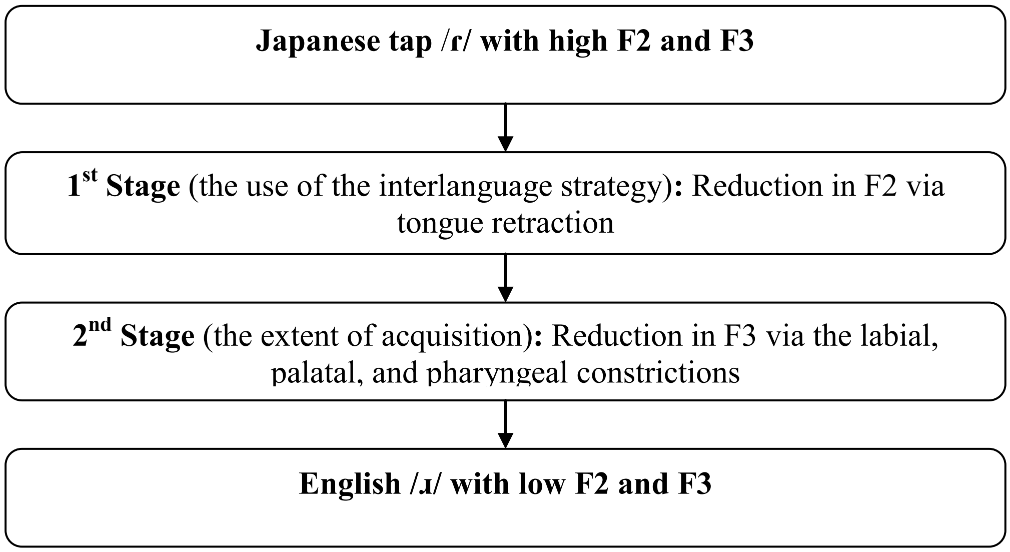


Figure 2. The acquisitional sequence of /ɹ/ by Japanese learners

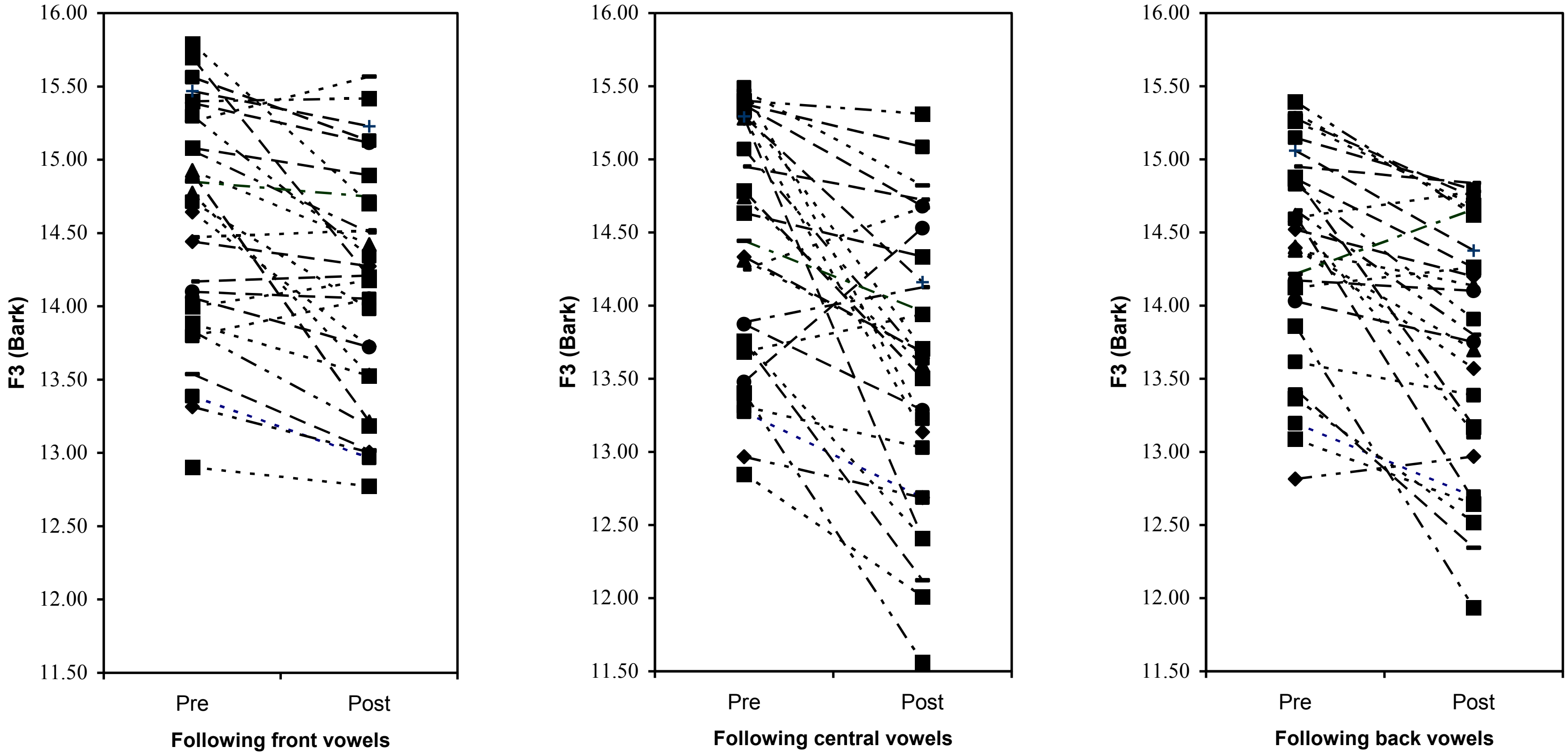


Figure 3. Profile plots of the experimental group in F3 dimension according to following vowel conditions

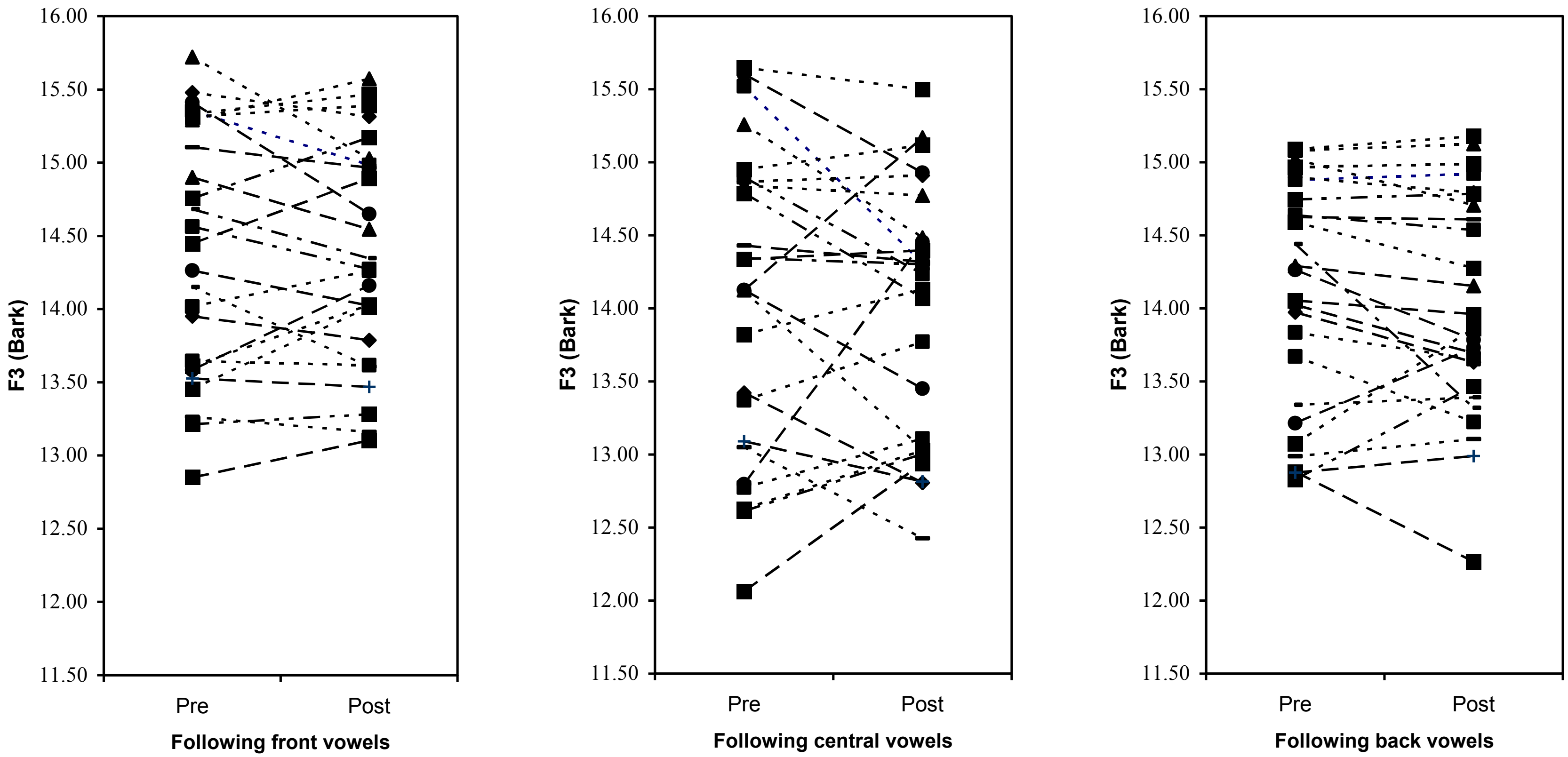


Figure 4. Profile plots of the control group in F3 dimension according to following vowel conditions

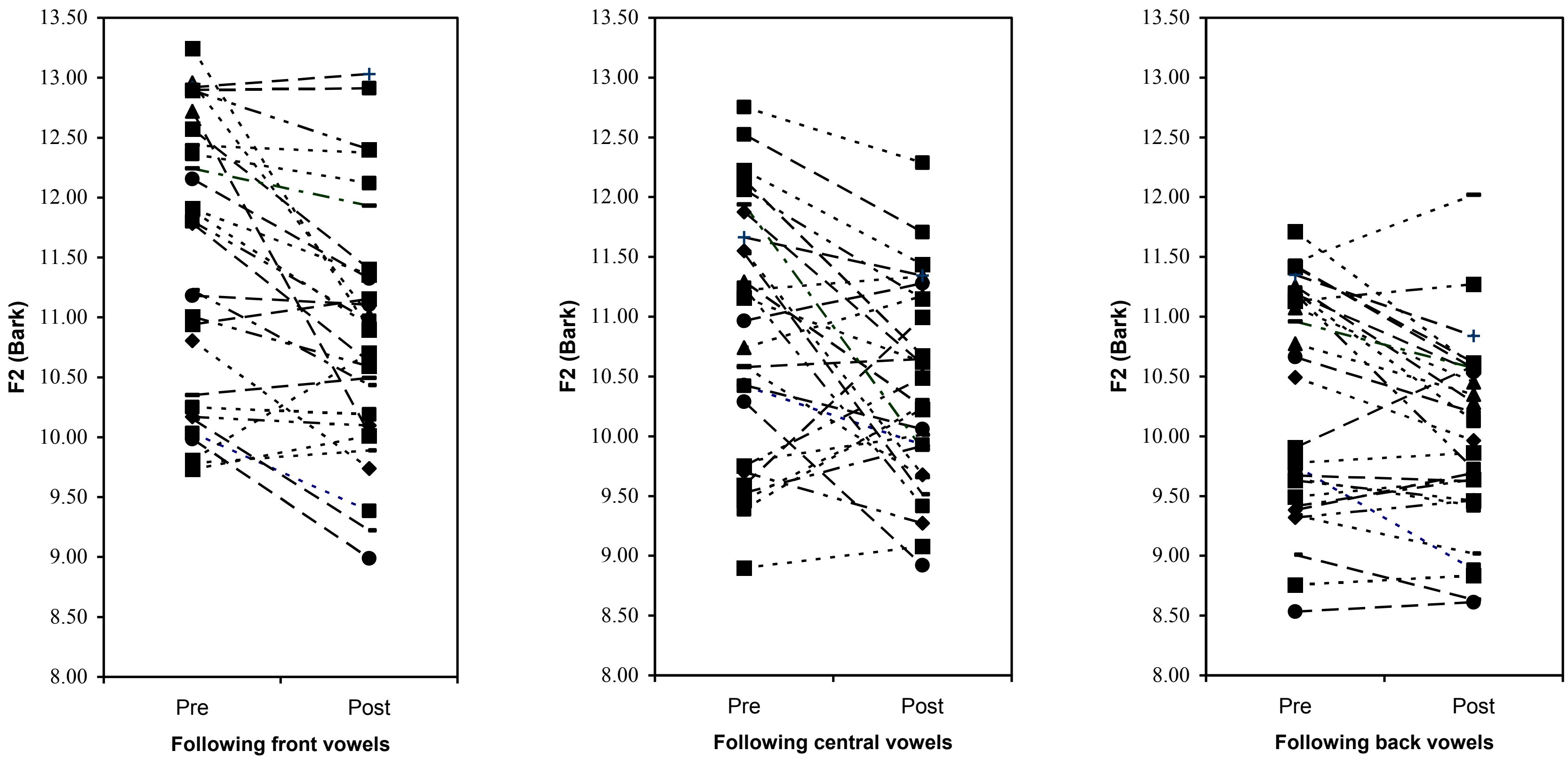


Figure 5. Profile plots of the experimental group in F2 dimension according to following vowel conditions

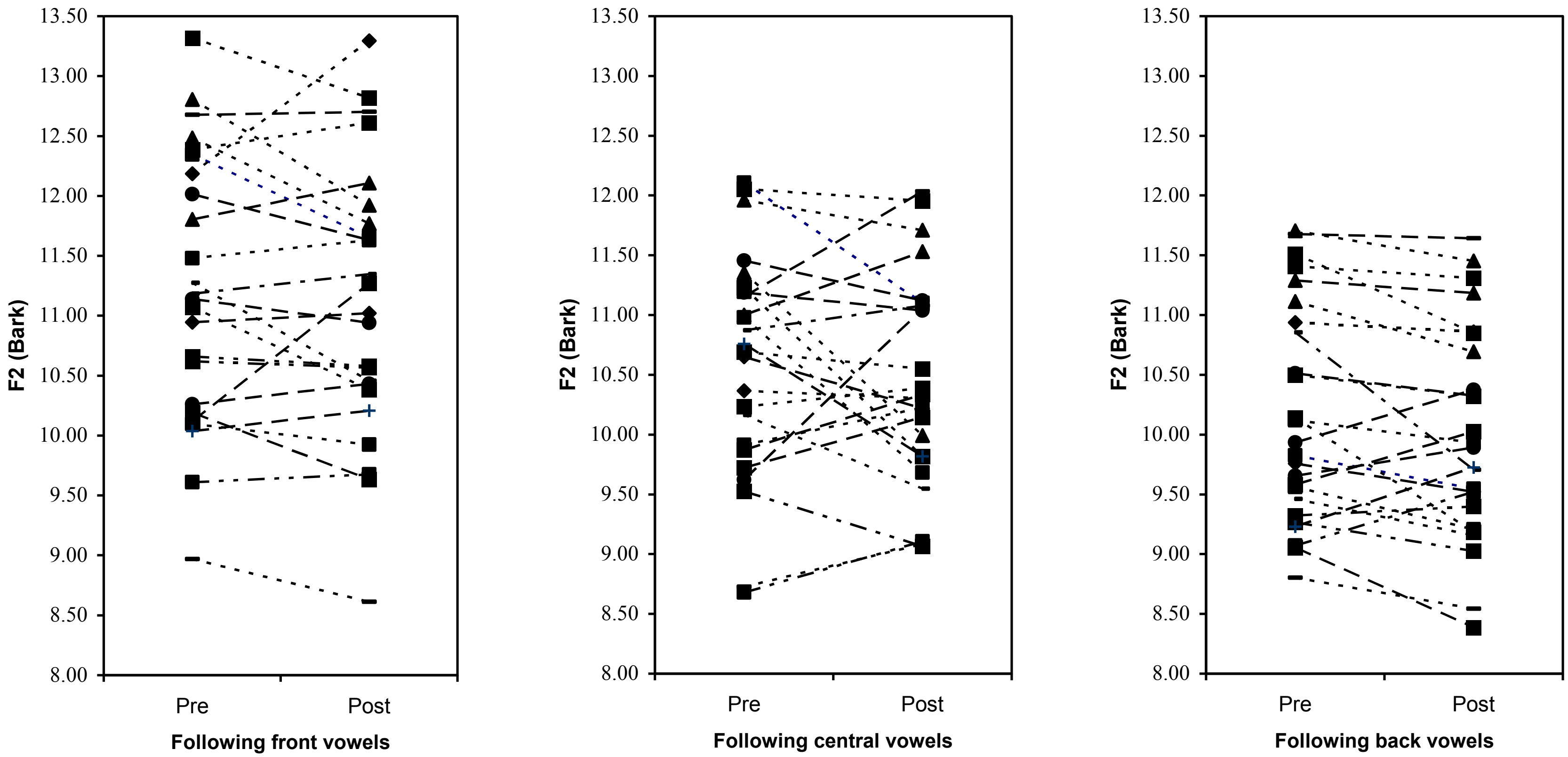


Figure 6. Profile plots of the control group in F2 dimension according to following vowel conditions

Notes

[1] S is used for students, T for teachers and * for mispronunciation or unclear pronunciation.

[2] Speech constitutes acoustic signals that movements of articulator organs (e.g., tongue, lips, jaw) generate, and such acoustic signals have traditionally been described as the first three frequency ranges of energy concentration in Hz (F1, F2, F3). These signals can be used as an index of how talkers actually use their articulators to produce sounds (F1 for degree of tongue height, F2 for tongue retraction, F3 for labial, palatal, and pharyngeal constrictions).

[3] Iverson et al. (2003) investigated how both English and Japanese listeners perceive various types of synthesized /ɹa/ and /la/ stimuli which varied both in F2 (744-1301 Hz) and F3 (1325-3649 Hz). Their results showed that (a) English listeners demonstrated a clear peak in sensitivity to F3 variation with marked perceptional space stretching at the English /ɹ/-/l/ boundary around 2067 Hz and 2523 Hz (i.e., categorical perception), and (b) Japanese listeners exhibited sensitivity to F2 variation that was irrelevant of the English /ɹ/-/l/ distinction, rather than the crucial differences in F3.

[4] Their F3 range generally overlaps with that of their L1 counterpart, Japanese tap /ɾ/ (Saito & Brajot, 2013).

[5] According to our casual classroom observation, the instructors appeared to hyper-articulate /ɹ/ when recasting their students' errors, which coincides with the nature of pedagogically-oriented classroom discourse and the saliency of partial recasts.

[6] Following the norm in experimental phonetics research and the *Praat* manual (Boersma & Weenik), *Praat* was set to a 44.1 kHz sampling rate and a 16-bit resolution for high-quality recording ("44.1 kHz" for a sufficient range of frequency values, and "16-bit resolution" for amplitude levels).

[7] Five words (i.e., 2 from word reading, 2 from sentence reading, 1 from picture description) were randomly chosen to represent how the learners originally produced 22 /ɹ/ words (i.e., 10 from WR, 8 from SR, 4 from TPD).

[8] The linear predictive coding function in *Praat* automatically detects the center frequency values of the predetermined number of formants (n = 4 in the current study) within the maximum formant search range (e.g., < 5500Hz for adult females, < 5000 for adult males).

[9] The F3 of the preceding sounds tends to continue to decline towards the beginning of the word because the F3 of /ɹ/ is relatively low.

[10] Bark is a conversion formula used to adjust raw acoustic values in Hz to human perception range, because frequency range is not linear (e.g., a change from 400 to 500 Hz is much more perceptible than a change from 2000-2100 Hz). In this sense, we converted all raw acoustic values into Bark so that any change in Bark values (e.g., 5 to 6 Bark, 10 to11 Bark) can have the same impact on human perception.

[11] Unlike grammatical and lexical errors which can easily and clearly be transcribed for subsequent analyses, it is relatively difficult to judge the degree of how successful learners actually repaired their pronunciation errors. As discussed earlier, Japanese learners tend to use several strategies to improve their /ɹ/ production skills (e.g., tongue retraction, lip rounding, palatal [and/or pharyngeal] constriction). In addition, given that all of the recast-repair moves occurred in the middle of the meaning-oriented lessons, the video recording was not consistent nor appropriate for refined listener judgement for obvious reasons (e.g., some repaired their errors while other students were talking).

[12] As illustrated in many pronunciation textbooks in Japan, many participants reported that unlike the Japanese tap, in pronouncing English /ɹ/, the tip of tongue does not touch alveolar ridge; this I categorized as explicit knowledge about tongue positions.

[13] Despite debates on the convention of the 95% (or 99%) significance level (e.g., Rozeboom, 1960), the standard has long been adopted in traditional statistics and applied to a range of academic fields (e.g., medicine, psychology, applied linguistics). Thus, a decision was made to use a cut-off point of alpha = $p < .05$ in the study.

[14] According to Cohen (1988) effect sizes are roughly classified as small ($0.20 \leq d < 0.50$), medium ($0.50 \leq d < 0.80$), or large ($0.80 \leq d$).

[15] Different from General Linear Model ANOVA, Mix Model ANOVA takes into account random effects of subjects. In the context of the current study, whereas the former approach focused on learners as a *group*, the latter approach allowed us to examine how a range of categorical and continuous variables interacted to influence *each* learner's pronunciation performance.

[16] See also DeKeyser (1998) for relevant discussion on the peculiarities of L2 morphosyntax development relative to other domains of language—i.e., pronunciation and vocabulary—learning.