LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

The Acquisitional Value of Recasts in Instructed Second Language Speech Learning: Teaching the Perception and Production of English /ɹ/ to Adult Japanese Learners

Kazuya Saito
Waseda University

The current study investigated the impact of recasts together with form-focused instruction (FFI) on the development of second language speech perception and production of English /ɹ/ by Japanese learners. Forty-five learners were randomly assigned to three groups—FFI recasts, FFI only, and Control—and exposed to four hours of communicatively oriented lessons. Whereas many FFI activities including explicit instruction were embedded into the treatment in order for the experimental groups to notice and practice /ɹ/ in a meaningful discourse, an instructor provided recasts only to the FFI-recast group in response to their mispronunciation of /ɹ/. Perception was measured using a two-alternative forced choice identification task, while pronunciation performance was elicited using controlled and spontaneous production tests and assessed by 10 naïve native-speaking listeners. According to the statistical comparisons, whereas the FFI-only group attained perception and production improvement particularly under trained lexical conditions, the FFI-recast group demonstrated similar but generalizable gains both in trained and untrained lexical contexts. The results indicate that (a) FFI itself impacts various domains of L2 speech learning processes (perception, controlled, and spontaneous production) and (b) recasts promote learners' attentional shift away from lexical units as a whole to phonetic aspects of second language speech (i.e., vocabulary to sound learning).

Keywords recasts; second language speech acquisition; pronunciation teaching; listening teaching; English /ɹ/

The project was funded by a *Language Learning* Research Grant and a Government of Canada Post-Doctoral Research Fellowship. I would like to thank Murray Munro, Leila Ranta, Ron Thomson, and the anonymous *Language Learning* reviewers for their useful and constructive comments on the earlier versions of the manuscript. Finally, I gratefully acknowledge the administration, teacher, and students at ECC Foreign Language Institute, especially Noriaki Okamura and Yukio Nagato.

Correspondence concerning this article should be addressed to Kazuya Saito, Waseda University, School of Commerce, 1–6–1 Nishi Waseda, Shinjuku, Tokyo 169–8050, Japan. E-mail: kazuya.saito@waseda.jp

DOI: 10.1111/lang.12015

Introduction

Over the past 25 years, second language acquisition (SLA) research has extensively examined how recasts can help adult second language (L2) learners develop their functional use of an L2 in dyadic or classroom interactions (e.g., Ellis & Sheen, 2006; Goo & Mackey, 2013; Long, 2007; Lyster & Saito, 2010). Whereas previous relevant findings have been exclusively concerned with the morphosyntactic, lexical, and pragmatic aspects of language, some recent studies have begun to investigate the role of recasts in L2 phonological development (Saito, 2013; Saito & Lyster, 2012). Although L2 speech research has provided a detailed description of how L2 learners acquire new sounds over time in relation to the quantity and quality of received input in a naturalistic context (e.g., Best & Tyler, 2007; Flege, 1995, 2003), little research attention has been given to the role of instruction in enhancing the rate and ultimate attainment of L2 speech in an efficient and effective manner (Derwing & Munro, 2005). To this end, this article reports a quasi-experimental study which investigated the efficacy of recasts together with form-focused instruction (FFI) on the development of the speech perception and production of /ɹ/ by Japanese learners of English in a simulated classroom setting.

Recasts in SLA

According to Lyster and Ranta's (1997) classification of feedback types, recasts are defined as "the teacher's reformulation of all or part of a student's utterance minus the error" (p. 46). From a theoretical perspective, some SLA scholars strongly advocate recasts as the ideal type of corrective feedback, precisely because of their implicitness (Doughty, 2001; Long, 2007); that is, it is assumed that recasts can simultaneously provide both positive and negative evidence without interrupting the communicative flow. In this way, L2 learners can make cognitive comparisons between their nontarget forms and targetlike reformulations during meaningful discourse. Other researchers, however, have argued that recasts might not be the most effective type of feedback, at least for L2 morphosyntax development. It has been contentious to what degree learners can succeed in perceiving negative evidence available in recasts and consider it as correction (Ellis & Sheen, 2006; Lyster & Saito, 2010).

With respect to L2 phonological learning (the focus of this article), however, a series of descriptive studies have found that recasts can be quite salient to learners. That is, although teachers and interlocutors tend to provide more recasts on morphosyntactic errors than on phonological errors, learners generate more successful repair and accurate perceptions in response to recasts on phonological errors (e.g., Ellis, Basturkmen, & Loewen, 2001; Kim & Han, 2007; Lyster, 1998; Mackey, Gass, & McDonough, 2000; Sheen, 2006). This finding suggests that recasts might be relatively salient to L2 learners and thus facilitative of their phonological development. This tentative generalization seems to hold true across different contexts. For example, in his descriptive study of French immersion classrooms, Lyster (1998) noted that whereas phonological recasts occurred less frequently (14%) than morphosyntactic recasts (46%), young students showed a higher rate of successful repair in response to the former (62%) compared to the latter (22%). Similar asymmetric patterns have also been observed in adult English-as-a-second-language (ESL) classrooms in New Zealand (Ellis et al., 2001) and adult English-as-a-foreign-language (EFL) classrooms in Korea (Sheen, 2006). Furthermore, in lab settings, Mackey et al. (2000) found that when asked to watch video clips of their task-based interaction with native-speaking (NS) interlocutors (i.e., stimulated recall sessions), two groups of learners (learners of ESL and Italian as a foreign language) recognized phonological recasts more accurately than morphosyntactic recasts (see also Kim & Han, 2007, for similar results in a classroom setting). Mackey et al. (2000) speculated that the learners' sensitivity to their phonological errors might be due to the fact that inaccurate pronunciation has "more potential to seriously interfere with understanding" than morphosyntactic errors (p. 493) (see also Isaacs & Trofimovich, 2012). Despite accumulating suggestions for the amenability of recasts to L2 speech learning, few studies have tested the relationship between the amount of recasts, the rate of learners' repair, and its ultimate impact on the development of perception and production abilities of specific sound features.

Naturalistic L2 Speech Learning

Recent speech research has suggested that L2 phonological development is subject to similar patterns of sound and word learning found in first language acquisition (e.g., Best & Tyler, 2007, for Perceptual Assimilation Model in the L2; Bundgaard-Nielsen, Best, & Tyler, 2011, for Vocab Model; Walley, 2007, for Lexical Restructuring Model). That is, L2 learners first encode word-sized units of L2 input during their initial stage of vocabulary learning, focusing especially on the phonotactic and prosodic features of speech streams. Therefore, their representation of sounds is based on lexical items that they have fully or partially acquired without much phonological awareness of individual sounds.

As learners increase the amount of experience and proficiency in L2 via more exposure to input as well as interaction with native speakers, they rapidly expand their vocabulary size in what sometimes is known as a vocabulary spurt. At this stage, they will start paying equal attention to not only the lexical units but also the phonetic details of L2 input, to correctly perceive and produce a number of phonetically similar words such as minimal pairs. In fact, previous findings have shown that experienced L2 learners tend to have more fine-grained segmental representations, such that they generally perceive and produce L2 sounds with little variance, regardless of various lexical contexts (e.g., familiarity, frequency, and density) (Bradlow & Pisoni, 1999; Flege, Frieda, Walley, & Randazza, 1998; Imai, Walley, & Flege, 2005; Trofimovich, Collins, Cardoso, White, & Horst, 2012). Finally, major L2 speech theories have claimed that the developmental process of new sound categories is perception-based (i.e., change in learners' perception abilities activates relevant articulatory configurations: Flege, 1995, 2003), entailing various levels of processing abilities (i.e., learners access representation first at a controlled and subsequently at a spontaneous level: Major, 2008). Taken together, naturalistic L2 speech learning constitutes a transition from vocabulary to sound learning; from perception to production; and from controlled to spontaneous processing abilities.

Instructed L2 Speech Learning

Over the past 20 years, pronunciation teaching has been typically characterized as "all but the most intensive formS-focused [decontextualized] treatments" (DeKeyser, 1998, p. 43), a reference to typical pronunciation instructional approaches such as the exclusive use of phonetic transcriptions, minimal-pair drills, and imitation of appropriate models (for a review, see Derwing & Munro, 2005). Several classroom pronunciation teaching studies have investigated the effects of these formS-focused methods on learners' improvement, particularly at a controlled-speech level (see Saito, 2012, for a research synthesis on 15 quasi-experimental studies published since 1990). Yet, it remains unclear to what extent an intervention of this kind actually allows students to transfer what they learn in the classroom to outside the classroom, especially at the spontaneous-speech level (Trofimovich & Gatbonton, 2006). Due to the lack of sufficient discussion on type of instruction (e.g., focus-on-form vs. focus-on-formS) and outcome measures (e.g., controlled vs. free constructed responses), the study of pronunciation teaching research has been excluded from all meta-analyses of instructed SLA to date (e.g., Norris & Ortega, 2000; Spada &

Tomita, 2010). Indeed, it appears to have been “marginalized within the field of applied linguistics” (Derwing & Munro, 2005, p. 379). As a result, teachers and material developers generally do not receive research-based guidance in terms of what and how to teach L2 pronunciation (Levis, 2005).

Whereas developing segmental-level perception abilities strongly relate to spoken word recognition and vocabulary growth (Bundgaard-Nielsen et al., 2011; Cutler & Broersma, 2005) and thus ultimately impact global listening skills (Field, 2008), relevant research findings have been limited to the laboratory-training paradigm. The most notable of these paradigms is High Variability Phonetic Training (HVPT). In this method, learners intensively listen to hundreds of natural exemplars of new sounds produced by various talkers (for a review, see Thomson, 2012). Previous studies have shown that HVPT can lead even adult L2 learners to attain improvement that is generalizable beyond trained lexical and talker contexts (e.g., Logan, Lively, & Pisoni, 1992), as well as to durable gains even after a few months (e.g., Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994). They have also demonstrated that such benefits can be transferred from perception to production levels (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). This vein of theoretically motivated research, however, provides few pedagogical implications, arguably because the audiolingual drill exercises consist of many hours of intensive listening (e.g., the training in Bradlow et al., 1997, which lasted for 15–22.5 hours) exclusively focused on a few difficult sounds, without any contextualized use of language. HVPT thus seems to be unrealistic in classroom settings, where teachers aim primarily to improve students’ communicative competence in the L2 with limited amount of time allocated to listening and pronunciation lessons.

English /ɹ/

The target of instruction in the current study is one of the most well-researched cases of L2 speech learning—the acquisition of word-initial /ɹ/ by Japanese learners of English. Because the Japanese phonetic system contains neither /ɹ/ nor /l/, Japanese learners of English initially substitute the closest Japanese counterpart (i.e., the Japanese tap /ɾ/) for both English /ɹ/ and /l/, neutralizing the nonnative contrast (Guion, Flege, Ahahane-Yamada, & Pruitt, 2000). According to cross-linguistic research, while the Japanese tap is similar to English /l/ to some degree, the tap is substantially different from English /ɹ/ (see Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004, p. 234). From an acoustic standpoint, English /ɹ/ can be described in relation to lower third formant (F3) frequencies (<2200 Hz) compared to the Japanese tap (>2400 Hz) (Hattori & Iverson, 2009). From an articulatory perspective, whereas the Japanese tap is characterized as a brief contact of the tongue against the alveolar ridge (Vance, 1987), producing /ɹ/ requires the simultaneous constrictions in the labial, palatal, and pharyngeal areas of the vocal tract with a sufficient amount of space in the sublingual cavity (Espy-Wilson, Boyce, Jackson, Narayanan, & Alwan, 2000). Given that the Japanese tap substitution greatly hinders successful communication (Riney, Takada, & Ota, 2000), L2 education researchers and practitioners alike have emphasized the importance of teaching English /ɹ/ as a priority for Japanese learners to attain intelligible pronunciation (Fraser, 2011).

To acquire English /ɹ/, however, Japanese learners reportedly overrely on several interlanguage strategies, rather than increasing their phonological awareness of the relevant acoustic (sensitivity to F3 variance) and articulatory (the labial/palatal/pharyngeal constrictions) features of /ɹ/. For example, they likely ignore F3 variation but instead depend on second formant (F2) frequencies[1] in order to discriminate /ɹ/ from /l/ (Iverson et al., 2003), and they produce /ɹ/ with an over-reliance on tongue retractions (i.e., /w/-like production, see Saito & Brajot, 2013). In fact, it has been shown that adult Japanese learners tend to have extreme difficulty learning English /ɹ/ in terms of both perception (Aoyama et al., 2004) and production (Larson-Hall, 2006), and that this is true even after years of residence in English-speaking countries.

Motivation for the Current Study

Given the significant absence of research attention in instructed L2 speech learning and the tremendous difficulty in acquiring the English /ɹ/, Saito (2013) and Saito and Lyster (2012) set out to investigate how a range of FFI techniques, including recasts, can promote Japanese learners' abilities to produce /ɹ/ in the context of meaning-oriented classrooms. These studies found that FFI can lead to change in L2 pronunciation performance both at a controlled and spontaneous level when it is combined with the provision of (a) explicit information before exposure to FFI (Saito, 2013) and (b) recasts in response to learners' mispronunciation during FFI treatment (Saito & Lyster, 2012). The results concur with similar findings in regards to the role of recasts in L2 morphosyntactic development (e.g., Doughty & Varela, 1998) and they lend empirical support to the amenability of recasts to pronunciation errors, as suggested by relevant descriptive studies (e.g., Mackey et al., 2000).

Nevertheless, Saito (2013) and Saito and Lyster (2012) raised further questions which future research needs to answer to obtain a full-fledged understanding of the complex mechanism of recasts in L2 speech learning. First, the previous studies tested and confirmed the overall potential of FFI by employing composite instructional options, including explicit information, interactional tasks, and recasts; however, it still remains unclear the extent to which the pure effects of recasts can account for such general FFI effectiveness. Second, the previous findings in the precursor studies have been limited to the production domain. Notably, adult L2 learners can consciously produce correct pronunciation of forms by drawing on their explicit articulatory knowledge, regardless of the present state of their perceptual representations (Sheldon & Strange, 1982). Thus, researchers need to adopt both production and perception measures in order to examine whether recasts can simply reinforce students' monitoring skills or simultaneously promote the concurrent development of perception and production abilities, in a fashion that is more similar to naturalistic L2 speech learning.

Last, the precursor studies have exclusively depended on acoustic analyses of /ɹ/ with a focus on F3 height as an index of learners' production improvement (i.e., F3 $<$ 2200 Hz for intelligible exemplars of English /ɹ/). Although F3 has been described as a primary acoustic correlate of /ɹ/ (e.g., Espy-Wilson et al., 2000), NS listeners also turn to multiple acoustic parameters (e.g., first and second formant frequencies, formant bandwidth, phoneme duration, voice quality) to attain a unitary phonetic percept (Hattori & Iverson, 2009). For instance, Underbakke, Polka, Gottfried, and Strange (1988) found that NS listeners could identify English /ɹ/ and /l/ from synthetic tokens which varied only in temporal dimensions (/ɹ/ for more than 50ms), regardless of F3 information—a well-known phenomenon referred to as phonetic trading relation. Furthermore, certain experienced NS listeners (e.g., ESL/EFL teachers and phonetically trained judges) might be able to perceive even a small change in such spectral and temporal domains of /ɹ/ between pre- and posttest sessions (for a review of the role of listener experience, see Kennedy & Trofimovich, 2008). Yet, given that the purpose of L2 pronunciation teaching is arguably to lead L2 learners to get ready for future communicative settings where they interact with a wide variety of NSs (and nonnative speakers) with different amounts of experience and tolerance towards foreign-accented speech, it is important to examine whether, how, and to what degree instructional gain can sufficiently impact the perception of phonetically untrained judges who do not have much familiarity with Japanese-accented speech (i.e., naïve listeners). The current study is designed to respond to these concerns.

Method

The present study investigated the acquisitional value of recasts and FFI in L2 speech learning in a quasi-experimental, pre- and posttest design. The instruction phase took place at a private language institute in Osaka, Japan. The participants were adult Japanese learners in an EFL setting. After completing a pretest session, two experimental groups received four hours of instruction that included an FFI treatment on pronunciation of English /ɹ/, while a control group received comparable instruction in terms of length and nonpronunciation (argumentative skills) content, without any focus on form regarding pronunciation. The same instructor taught all three groups and provided recasts only to the FFI-recast group in response to their unclear or mispronunciations of /ɹ/. A posttest session took place two weeks after instruction.[2] In order to ensure consistency between different instructional treatments, the researcher always sat at the back of the classroom as an observer, and all classes were video-recorded for post hoc analyses. The learners' pre- and postinstructional performance of /ɹ/ was assessed via three outcome measures (perception, controlled production, and spontaneous production tests) as well as analyzed separately in various lexical contexts (trained vs. untrained items) in order to examine the impact of instruction on the three developmental levels of L2 speech learning posited in the literature: vocabulary → sound learning; perception → production; controlled → spontaneous processing. Finally, 10 NS listeners were recruited at a Canadian university to evaluate the foreign accentedness of the production data of the pre- and posttest sessions.

Participants

Students

For recruitment purposes, the researcher distributed ads to students who belonged to the language institute and its branch schools as well as university-level schools near the research site. Interested participants contacted the researcher to set up a date for pretests via email or phone. Of the 48 students who initially participated in the current study, three failed to complete the project for various personal reasons. The performances of 45 students were included in the final analysis.

Participants' mean age was 30.9 years old, with a range from 19 to 54 years old. Twenty-two students reported some length of residence (LOR) in English-speaking countries (United States, England, Australia) from 1 month to 3 years ($M = 6.8$ months) and 23 students reported that they had never been abroad. Overall, participants' use of spoken English was limited to a few hours

Table 1 Participant information by group

	FFI-recasts ($n = 17$)	FFI-only ($n = 18$)	Control ($n = 10$)
Age	$M = 31.5$ ($SD = 10.6$)	$M = 36.1$ ($SD = 7.8$)	$M = 29.7$ ($SD = 6.7$)
Gender	3 males, 14 females	3 males, 15 females	1 male, 9 females
LOR in months	$M = 4.5$ ($SD = 7.8$)	$M = 8.7$ ($SD = 10.9$)	$M = 7.2$ ($SD = 11.3$)
$x = 0$	11 learners	7 learners	5 learners
$0 < x \leq 12$	1 learners	6 learners	4 learners
$12 < x \leq 36$	5 learner	5 learner	1 learner

of conversational lessons at schools (i.e., EFL); thus, their proficiency level could be considered as beginner-intermediate. Table 1 provides the details of the 45 learners by experimental condition.

After taking the pretests, all students were randomly assigned to eight classes of about six students each and then categorized into three groups: (a) FFI-recast (three classes, $n = 17$), (b) FFI-only (three classes, $n = 18$), and (c) Control (two classes, $n = 10$).[3] Some acoustic information of the production data by the FFI-recast group (i.e., F3) has already been partially reported in another study (Saito, 2013). Because in order to examine the relative effectiveness of recasts with FFI versus FFI alone on L2 speech perception and production development, the same production evidence needed to be assessed by human listeners and analyzed in the present study in comparison with the FFI-only and control groups; all the perception data of all groups is reported for the first time in the present study.

Instructor

One experienced instructor (male, a NS from California) who worked at the participating language institute taught all classes.

Listeners

In order to assess the production test data, 10 listeners (5 males, 5 females) were recruited from undergraduate linguistics and psychology courses at a Canadian university. All participants were NSs of northwestern English with a mean age of 22.5 years. They reported having little frequent contact with Japanese learners of English and being unfamiliar with Japanese-accented English speech. All passed a pure-tone screening test at octave frequencies between 250 and 4000 Hz.

FFI Treatment

Content of Instruction

As mentioned, the FFI-only and FFI-recast instructional treatments were embedded in meaning-oriented lessons that were also taught to the control group by the same teacher. The primary goal of the meaning-oriented lessons was to teach English argumentative skills entailing logical thinking, negotiation and debating skills, and public speaking abilities (for details of instruction, see Appendix S1 in the Supporting Information online).

Instructional Conditions

Among the FFI activities previously identified by researchers, the two FFI treatments in the current study included: (a) explicit instruction, when an instructor provides exaggerated pronunciation of target sounds and explains relevant articulatory configurations (as in Saito, 2013); (b) structured input, when learners are required to process linguistic form in input for meaning without being pressured to produce output (as in VanPatten, 2004); (c) typographically enhanced input, when target structures are highlighted by means of emphatic stress or visual changes, such as italics, to induce learners to notice the forms in oral and written L2 input (as in Han, Park, & Combs, 2008); and (d) focused tasks, when learners are required to produce linguistically accurate output to successfully complete meaning-oriented tasks (as in Ellis, 2006). A control group received no pronunciation instruction, an experimental group received FFI-only instruction, and another experimental group received the same FFI instruction plus recasts. Recasts were operationalized as partial recasts in the current study (for more details, see below).

Target Words

The target sound of instruction (i.e., English /ɹ/) appeared in various phonetic positions (pre-, inter-, postvocalic positions) in the form of 39 words minimally (or near-minimally) paired with /l/ (see Table 2). All of these words were encountered by the FFI-only and FFI-recast groups italicized, highlighted in red, and frequently used as key words in the teaching materials, so that the learners could notice the target words in meaning-oriented lessons and become aware of the target feature of /ɹ/ at a lexical level (i.e., learners experienced typographically enhanced input, see Han et al., 2008).

Recast Treatment and Fidelity of Implementation

For the FFI-recast group, the instructor consistently provided recasts when students pronounced /ɹ/ in the target words in an unclear or unintelligible

Table 2 39 target words included in the instructional treatment

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

Note. * indicates the 15 words included in the pre/post tests (15 for perception, 7 for production).

manner (e.g., the substitution of the Japanese tap).[4] In particular, he was asked to consistently recast only one word with a falling intonation (i.e., partial recasts), which has been identified as the most perceptually salient type of recasts (Sheen, 2006). To ensure the students' self-correction opportunities, the instructor always provided an opportunity for uptake (Mackey, Oliver, & Leeman, 2003).

To examine how often students received recasts and repaired their utterances, the researcher watched 12 hours of video recordings of all the FFI-recast classes (4 hours × 3 classes), and analyzed the instructor's recasts and students' repairs. To code pronunciation-related error treatment, I used Lyster and Ranta's (1997) original analysis scheme: (a) Repair as shown in Example 1 (i.e., students repeated the teacher's recasts) and (b) Needs-repair, as shown in Example 2 (i.e., students simply reacted to the teacher's recasts via some acknowledgement).

Example 1 for Repair in English Debate:

S: If I run /ɾʌn/*, I like running /ɾʌnɪŋ/* inside.

T: Running /ɹʌnɪŋ /. ← RECAST

S: Running /ɹʌnɪŋ/. ← REPAIR

Example 2 for Needs-repair in Argument Critique:

S: I eat rice /ɾaɪs/* every day.

T: Rice /ɹaɪs/. ← RECAST

S: Yes, I love Japanese food. ← NO REPAIR (some acknowledgement)

Teacher Training

The instructor received four 1-hour training sessions over 2 days. After the researcher carefully explained the purpose and procedure of each FFI task and the recast treatment with an actual teaching demonstration, the instructor practiced until he felt comfortable with the procedure.

Control Group

The students in the control group received instruction on English argumentative skills which did not include any FFI component on /ɹ/ but was identical in all other aspects. That is, all target words in the instructional materials were replaced with comparable words ("rain" → "snow," "run" → "jog"). For the warmup games, they did other purely communicative activities with no emphasis on pronunciation or listening practice.

Tests

The effects of FFI on various domains of L2 phonological development (perception → production, controlled → spontaneous processing) were examined using: (a) a perception test, (b) a controlled production test, and (c) a spontaneous production test. In addition, these tests included both lexical items which appeared in FFI materials (i.e., trained items) and also some which did not (i.e., untrained items), both in the pre- and posttest materials. Assessing students' test scores in the two lexical contexts was assumed to reveal whether their improvement at a segmental level could be generalized to novel lexical contexts beyond instructional materials (i.e., vocabulary to sound learning). The testing sessions took place individually before and after instruction in a quiet room at the language institute. In order to avoid too much focus on form, especially in the spontaneous production test, the pretests and posttests were conducted in the following order: spontaneous production, followed by controlled production, and finally perception.

Perception Test

The impact of instruction on learners' perception abilities of /ɹ/ was measured via a two-alternative forced-choice identification task: They listened to a total of 70 minimally paired words (e.g., "read" vs. "lead") and identified exemplars with low F3 (<2400 Hz) as English /ɹ/ and those with high F3 (>2400 Hz) as English /l/.[5] The test consisted of 25 /ɹ/-/l/ minimal pairs (15 trained + 10 untrained items) together with 10 distracter ones (e.g., /s/-/θ/ as in "sink" vs. "think"). These words were Consonant-Vowel-Consonant (CVC) singletons contrasting word-initial /ɹ/ and /l/ in three phonetic contexts: 20 singletons with

Table 3 Fifty tokens in the perception tests in relation to following vowel conditions

Vowel height and backness		Front vowels	Central vowels	Back vowels
High vowels	(Trained items)	*"rink, link"*		*"roof, Loof"*
		"reef, leaf"		*"rule, lure"*
		"read, lead"		*"room, loom"*
	(Untrained items)	*"ring, ling"*		*"rude, lude"*
		"reach, leach"		*"root, loot"*
Mid vowels	(Trained items)	*"race, lace"*		*"road, load"*
		"rent, lent"		*"wrong, long"*
		"rain, lane"		*"roan, loan"*
	(Untrained items)	*"rate, late"*	*"rough, laugh"*	*"roll, loll"*
		"red, led"	*"rush, lush"*	*"rope, lope"*
Low vowels	(Trained items)		*"right, light"*	
			"ride, lied"	
	(Untrained items)		*"rice, lice"*	

front vowels, 10 singletons with central vowels, and 20 singletons with back vowels (see Table 3).

All speech samples were spoken by one male NS and one female NS of North American English and recorded on a Rolad-05 Wave recorder in isolation in a quiet room at an English-speaking university in Montreal. The tokens were digitized at a 40-kHz sampling rate and normalized for peak intensity by means of the *Praat* (Boersma & Weenik, 2012) speech analysis software.

The test items were presented through a headset, and the learners responded to each speech token by marking one of two orthographic choices on a prepared answer sheet. Given that some of these words were infrequently encountered ones (e.g., loll, lure), the learners were told that all items were real minimally paired words, and were asked to focus on the contrasting sounds (e.g., /ɹ/-/l/, /θ/-/s/, /v/-/b/), rather than basing their judgments on word-nonword status and possible lexical meaning.

Production Tests

Two production tests (word reading and timed picture description) were designed to measure the learners' pronunciation performance of /ɹ/ at two different levels of processing: controlled vs. spontaneous speech. The 15 words used in the two production tasks included eight trained and eight untrained items with the following vowel conditions (height, backness) carefully controlled (see Table 4). All of these words were CVC singletons which included /ɹ/ in

Table 4 Fifteen tokens in the controlled and spontaneous production tests in relation to following vowel conditions

Vowel height and backness		Front vowels	Back vowels
A. *Controlled Production Test*			
High vowels	(Trained items)	*Rink*	*rule*
	(Untrained items)	*Reach*	*rude*
Mid vowels	(Trained items)	*Race*	*road*[a]
	(Untrained items)	*Rate*	*roll*
B. *Spontaneous Production Test*			
High vowels	(Trained items)	*Read*	*Roof*
	(Untrained items)	*Ring*	*Route*
Mid vowels	(Trained items)	*Rain*	*road*[a]
	(Untrained items)	*Red*	*Rope*

[a]"Road" was tested twice both in the controlled and spontaneous production tests.

the word-initial position. According to the results of the vocabulary profiling (Cobb, 2012), 13 of them fall into the first 2,000 most frequent words. None of the participants reported their unfamiliarity with the remaining words (i.e., *rink, route*), probably because Japanese adopts many English lexical items as loan words (i.e., Katakana).[6] Thus, the effects of lexical frequency and familiarity factors on Japanese learners' /ɹ/ production were minimized in the current study. Words were distributed as either trained or untrained items.

The controlled production test elicited learners' production of /ɹ/ under controlled conditions, by asking participants to read a list of 40 words, which included eight target words in isolation, once at normal speed. The spontaneous production test, on the other hand, measured participants' ability to use certain linguistic structures under spontaneous conditions. This was in response to SLA researchers, who have emphasized the importance of communicative use of language (i.e., learners are required to pay equal attention to morphosyntactic, lexical, pragmatic, and phonological aspects of language) (Spada & Tomita, 2010) under time pressure (i.e., learners are not given much planning time for accessing their explicit knowledge stored in general memory) (Ellis, 2005). Often included in such cognitively demanding tasks in L2 phonology is a picture description task (with either written or oral prompts), whereby learners are guided to use the target phonological features while performing a cognitively demanding task (i.e., describing pictures) (see Rau, Chang, & Tarone, 2009, for the case of English /θ/ by Chinese EFL learners).

The picture description task was adapted in the current study to elicit learners' spontaneous production of /ɹ/ as follows:

(1) The learners were first given 10 seconds to memorize a written list of four key words.
(2) Immediately after the list was taken away, they described two pictures in a row, with no planning time, using two of the key words for each picture, one of which was a target word that included /ɹ/ in word-initial position.

In total, the learners described eight pictures with target words and eight distracter pictures such that each participant contributed eight target words.

Listener Judgment

This section presents the stimuli and procedures followed by the 10 naïve native-speaking listeners in order to assess participants' pronunciation performance in the controlled and spontaneous production tests.

Stimulus Preparation

The 45 Japanese learners produced 1,440 words from the pre- and posttest sessions (45 students × 16 words [8 controlled tokens + 8 spontaneous tokens] × 2 sessions [pre- and posttests]). All speech tokens were recorded using a Roland-05 Wave recorder at a 44.1 kHz sampling rate and 16-bit resolution and a unidirectional microphone (DM-20SL) and normalizing for peak amplitude. With respect to the words particularly embedded in continuous speech streams (i.e., spontaneous production), the researcher took care to listen to the speech samples multiple times in order to put a cursor on the onset of the word (where any component of /ɹ/ could be heard), and move toward its offset by 5-millisecond steps. To avoid significant distortion of extracted words, inflected endings (rain<u>ed</u>, read<u>ing</u>) were included. All speech tokens were randomized and divided into six blocks (240 tokens per block).

Procedure

The listening sessions were conducted individually in a quiet room at the Canadian university, and all speech tokens were presented to the listeners on a laptop computer screen. To reduce listener fatigue (evaluating the entire dataset consisting of six blocks required three hours in total), the 90-minute meetings took place over 2 days. A 9-point scale descriptor was adapted and modified from Flege, Takagi, and Mann's (1995) 6-point scale. Upon hearing each token, the listeners clicked one of the nine rating criteria: 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/*).

A "repeat" button was available to allow the listeners to hear an item up to three times before making a judgment. These listeners were explicitly asked to base their judgments as much as they could on only the quality of /ɹ/—instead of the entire words— because their ratings would have otherwise been influenced by other pronunciation errors typical of Japanese learners, such as mispronunciation of vowels (e.g., /ɹi:ŋ / for "ring") and epenthesis insertion (/ɹutə/ for "route"). They rated five speech tokens (not included in the subsequent listening session) on a 9-point scale as practice.

Results[7]

Perception Test

To avoid any ceiling effects for advanced learners without much room for improvement, the decision was made to eliminate seven learners who scored more than 85% (43 out of 50 points) during the pretest session ($n = 2$ from FFI recast, $n = 5$ learners from FFI only); the total number of the participants for the statistical analyses was 38 (for a similar decision, see Iverson, Hazan, & Bannister, 2005, and other auditory training studies). The learners' correct identification scores in trained items ($n = 30$) as well as in untrained items ($n = 20$) are visually displayed in Figure 1. According to the descriptive results, whereas all of the learners scored just above chance level at the beginning of the project ($M = 62.7\%$), those in the experimental groups (FFI recast, FFI only) reached around 70% after instruction.

To find any preexisting difference according to group factors, participants' total pretest scores (30 trained + 20 untrained = 50 items per participant) were first submitted to a one-way analysis of variance (ANOVA) with one between-group factor (Group [FFI recasts, FFI only, Control]). No main effects for Group ($p > .05$) were found, suggesting that the learners were comparable at the time of the pretest sessions.

The results were analyzed with a two-way ANOVA with Group as a between-group factor and Time (pre-/posttests) as a within-group factor. It yielded a significant main effect of Time, $F(1, 35) = 10.104$, $p = .003$, but neither a significant Group effect nor a significant Group Time interaction effect ($p > .05$). Bonferroni multiple comparisons found significant improvement over time for the FFI-recast group ($M = 60.5 \rightarrow 68.4\%$, $p = .003$, $d = 0.87$) and FFI-only group ($M = 68.0 \rightarrow 73.8\%$, $p = .035$, $d = 0.47$).

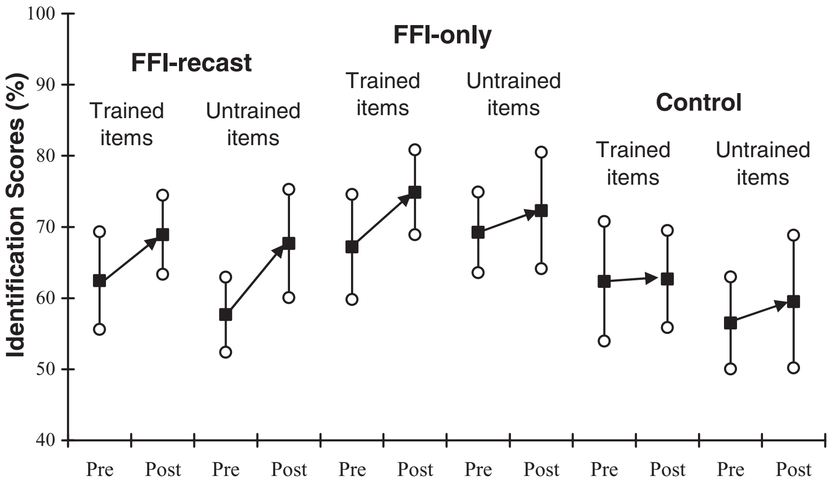

Figure 1 95% confidence intervals and mean values of the learners' pre–post test scores.

To further examine if gain patterns differed according to the lexical conditions (trained vs. untrained items), scores were submitted to a three-way Group × Time × Lexis ANOVA. Only a significant main effect was found for Time, $F(1, 35) = 9.820$, $p = .003$. Neither the effects of the two-way Group × Time interaction nor the three-way Group × Time × Lexis were significant ($p > .05$). According to Bonferroni multiple comparison,[8] the FFI-recast group demonstrated significant improvement over time both in trained items ($M = 62.4 \rightarrow 68.9\%$, $p = .033$, $d = 0.66$) and untrained items ($M = 57.7 \rightarrow 67.7\%$, $p = .009$, $d = 0.86$). In contrast, the FFI-only group significantly increased their identification scores only in trained items ($M = 67.2 \rightarrow 74.9\%$, $p = .018$, $d = 0.57$).

Taken together, these results indicate the following patterns for the perception test results: (a) both of the experimental groups significantly improved their perception abilities; (b) the instructional gain of the FFI-only group was limited only to trained lexical contexts; and (c) the FFI-recast group demonstrated improvement in their perception abilities of /ɹ/ in both trained and untrained lexical contexts.

Production Test

Inter-Rater Reliability

The interclass correlation between the 10 NS listeners was computed at .723 for the entire data set ($n = 1{,}440$), .736 for the controlled production tokens

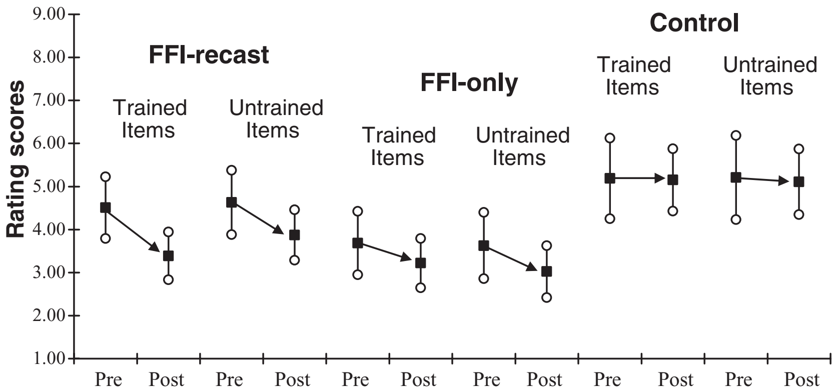

Figure 2 95% confidence intervals and mean values of the learners' pre–post test scores (1: English /ɹ/ – 9: English /l/).

(n = 720), and .710 for the spontaneous production tokens (n = 720). The results indicate a significantly reliable level of inter-rater agreement ($p < .001$). By pooling over listeners, we assigned mean rating scores to each /ɹ/ token produced by the Japanese learners.

Controlled Production

To avoid ceiling effects, two learners in the FFI-only group who received average rating scores below 2.5 during the pre-test session were eliminated from subsequent analysis (for a similar decision, see Derwing, Munro, & Wiebe, 1998). The descriptive results of the controlled production test are summarized in Figure 2. A visual inspection of the figure indicates a range of overall rating scores from unintelligible exemplars of /ɹ/ (M = 5.0–6.0) to poor exemplars of /ɹ/ (M = 3.5–4.0) at the beginning of the project. After instruction, however, the experimental groups appeared to reach poor to good exemplars of /ɹ/ (M = 2.5–3.5).

Pretest scores were first submitted to a one-way ANOVA (Group [FFI-recast, FFI-only, Control]). The results yielded a significant main effect for Group, $F(2, 40) = 3.762$, $p = .032$, suggesting some group difference at the beginning of the study. According to Bonferroni multiple comparisons, the FFI-only group demonstrated significantly higher production pretest scores ($M = 3.66$) than the control group ($M = 5.20$) ($p = .035$). As a way to control this preexisting difference, an analysis of covariance (ANCOVA) was computed with pretest scores as a covariate.

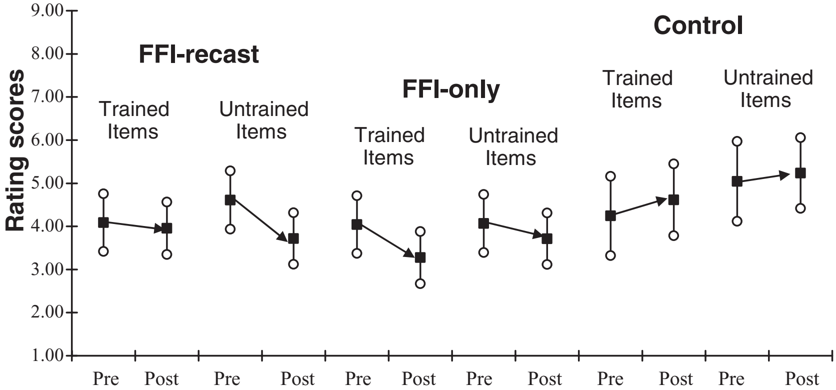

Figure 3 95% confidence intervals and mean values of the learners' pre–post test scores (1: English /ɹ/ – 9: English /l/).

The results of the ANCOVA yielded a significant main effect for Group, $F(2, 39) = 8.969$, $p = .001$. Bonferroni post hoc comparisons, performed on adjusted means, revealed that compared to the control group ($M = 4.68$), the experimental groups obtained significantly lower scores (more targetlike exemplars of /ɹ/) at the time of the posttest session ($M = 3.51$ for FFI recast, $d = 0.87$, $p < .001$ $d = 0.91$; $M = 3.53$ for FFI only, $p = .003$, $d = 1.02$). To further investigate the role of the lexical variable in their instruction gain (i.e., generalization of /ɹ/), pre–post results for trained and untrained items were also submitted to a two-way ANCOVA (Group × Lexis) with pretest scores as a covariate. Once again, there was only a main effect for Group, $F(2, 38) = 8.969$, $p < .001$, but no main Lexis effect nor a significant Group × Lexis interaction effect ($p > .05$).

To summarize, the results of the controlled production tests indicate that the experimental groups (FFI recast and FFI only) equally improved in the intelligibility of their /ɹ/ production at a controlled speech level, and this was true regardless of lexical context.

Spontaneous Production

To avoid ceiling effects, two learners whose average rating scores were below 2.5 during the pretest session ($n = 1$ from FFI only, $n = 1$ from Control) were eliminated from subsequent analysis. The descriptive results of the spontaneous production test are plotted in Figure 3. A visual inspection of the figure reveals that whereas all learners' production of /ɹ/ was generally considered as neutral at the beginning of the project ($M = 4.0$–5.0), the experimental groups began to produce more targetlike exemplars of /ɹ/ ($M = 4.0$–3.0) after the intervention.

Pretest scores were submitted to a one-way ANOVA (Group), which found no significant main effect for Group, $F(2, 40) = .646$, $p = .530$. This in turn suggests that there was no preexisting group difference at the beginning of the study. Next, the pre–post results were analyzed using a two-way ANOVA (Group $\times$Time). There was a significant Group $\times$ Time interaction effect, $F(2, 40) = 4.120$, $p = .024$. According to Bonferroni multiple comparisons, both of the experimental groups significantly improved their production of /ɹ/ over time: FFI recast ($M = 4.35 \rightarrow 3.84$, $p = .008$, $d = 0.40$) and FFI only ($M = 4.05 \rightarrow 3.49$, $p = .004$, $d = 0.44$).

To further examine generalizability of instructional gains, their pre–post results for trained and untrained items were submitted to a three-way ANOVA (Group $\times$ Lexis $\times$ Time). There was a significant Group $\times$ Lexis $\times$ Time interaction effect, $F(2, 40) = 5.828$, $p = .006$. Bonferroni multiple comparisons revealed different improvement patterns according to lexical condition: Whereas the FFI-recast group significantly improved their production of /ɹ/ in untrained lexical contexts ($M = 4.61 \rightarrow 3.72$, $p < .001$, $d = 0.65$), the FFI-only group noted a significant decline in their rating scores in trained lexical contexts ($M = 4.04 \rightarrow 3.28$, $p = .002$, $d = 0.60$). Thus, these results suggest that (a) instruction led both of the experimental groups to produce more intelligible exemplars of /ɹ/ in the spontaneous production test, but (b) their improvement was particularly observed in trained items for the FFI-only group, and in untrained items for the FFI-recast group.

Fidelity of Implementation Analysis

During the four hours of instruction, 303 recasts were directed to 17 students in the FFI-recast group ($M = 17.8$ recasts, ranging from 8 to 25 recasts per participant), and 277 of them were repaired (91.4% repair rate).[9] These data are in line with other descriptive studies (e.g., Sheen, 2006), which suggests that students in the FFI-recast group successfully (a) heard a great deal of their teacher's model pronunciation, (b) perceived recasts as corrections, and (c) practiced correct forms through self-modified output.

Discussion

The current study aimed to provide a comprehensive understanding of how recasts facilitate L2 speech learning in the context of FFI on the acquisition of /ɹ/ by Japanese learners of English. In this section, I first discuss how FFI itself can promote the development of L2 speech perception and production (Experimental vs. Control) and then address the extent to which adding recasts to FFI can enhance instructional gain (FFI-recast vs. FFI-only).

Effects of FFI

Whereas instructed L2 speech learning has been traditionally studied in the paradigm of focus-on-formS interventions, such as pronunciation drill activities and a massive amount of listening to target sounds via computer without much contextualized use of language, the current study tested the pedagogical potential of integrating a range of FFI activities into meaning-oriented classrooms. In this way, the learners were guided to notice and practice the target sound feature (English /ɹ/) with their primary focus on the main content of the lessons (English argumentative skills). Given the learners' beginner-intermediate proficiency level, explicit information about English /ɹ/ before the FFI instruction was also provided. In this way, instruction aimed at first enhancing participants' phonological awareness of the target feature of /ɹ/ at a segmental level free of communicative pressure, so as to make the best use of the subsequent FFI treatment during the lessons (Saito, 2013). Although most of the learners tended to use the Japanese counterpart at the beginning of the project (their identification accuracy was slightly above chance level; their production was generally rated by the 10 listeners as poor exemplars of /ɹ/), FFI exerted a substantial amount of impact on the examined L2 speech learning dimensions. For perception, four hours of FFI led to an average increase of 7–10% in the accuracy of the forced-choice identification task (cf. Logan et al., 1992, for an 8% gain after 10 hours of laboratory training). For production, the learners improved the intelligibility of their /ɹ/ pronunciation both at a controlled and a spontaneous speech level, and such improvement was perceived by the naïve NS English listeners, who did not have much familiarity with Japanese-accented English.

In line with previous FFI research, the overall effectiveness of instruction could be attributed to L2 use in a real sense of meaningful communication (Spada & Tomita, 2010). That is, FFI in the current study was able to successfully direct the learners' attention to using language accurately and fluently at the same time in order to get the message across to their interlocutors during teacher–student interactions. From the cognitive perspective of SLA, a communicative focus on form of this kind is hypothesized to promote students' form-meaning mappings (VanPatten, 2004) as well as the proceduralization of declarative knowledge (Ranta & Lyster, 2007), which ultimately leads learners to attain more robust representations of L2 linguistic features with various levels of processing abilities (i.e., automaticity) (Segalowitz, 2003).

Effects of Recasts

I now turn the discussion to the primary goal of the study: to investigate the role of recasts in FFI effectiveness on the acquisition of /ɹ/. Recasts were operationalized as the partial reformulation of target words where any mispronunciation or unclear pronunciation of /ɹ/ occurred. According to the results of the fidelity of implementation analyses, the learners received a fairly large amount of recasts (17.8 recasts per student) and repaired most of them (91.4% repair rate). The two experimental groups (FFI recast, FFI only) demonstrated a similar amount of improvement both in perception and production, but they showed slightly different gain patterns according to the lexical conditions: (a) the effects of FFI without recasts was evident particularly in trained lexical items and (b) a combination of FFI and recasts enabled the learners in the FFI-recast condition to generalize the medium-to-large scale of the FFI effectiveness from trained to untrained lexical items. The results in turn suggest two possible interpretations. First, FFI itself can increase learners' phonological awareness of the words they encountered during instruction (i.e., vocabulary learning). Second, recasts can further help learners to focus on not only word-sized but also sound-sized units of L2 input with the view of establishing new phonetic categories at a segmental level (i.e., sound learning). In conjunction with the earlier discussion about the multiple functions of recasts as well as the results of the fidelity of implementation analyses, the efficacy of phonological recasts can be ascribed to its three facilitating roles in the context of listening and pronunciation teaching. That is, phonological recasts can provide students with an explicit signal of errors (i.e., *enhanced negative evidence*) and with teacher pronunciation models (i.e., *enhanced positive evidence*), while, at the same time, eliciting self-modified output (i.e., *enhanced output practice*).

The importance of enhanced negative and positive evidence in pronunciation instruction via phonological recasts cannot be overstated. According to the Speech Learning Model (Flege, 1995), which specifically concerns adult L2 speech perception and production development at a segmental level, "a new phonetic category can be established for an L2 sound that differs from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds" (p. 239). In this regard, a combination of enhanced negative and positive evidence may greatly influence three key constructs of L2 segmental learning: (a) the phonetic-level noticing of new sounds, (b) the restructuring of the existing categories in the common phonological space, and (c) the establishment and internalization of new categories into long-term memory representation.

On the one hand, enhanced positive evidence in recasts (i.e., teachers' exaggerated pronunciation) may provide optimal linguistic input, such as higher pitch, simplified prosody, and vowel and consonant hyperarticulation, allowing adult learners to focus on the perceived dissimilarity of the L2 sound relative to its L1 counterpart (see Kuhl, 2000, p. 11855), particularly when preceded by explicit instruction, as was the case in the present study. Iverson et al. (2005) found that intensive exposure to synthesized speech stimuli with enhanced primary and secondary acoustic features (i.e., F3 difference maximized and duration lengthened) significantly improved /ɹ/-/l/ identification by Japanese listeners. On the other hand, the kind of enhanced negative evidence that is made available in recasts (i.e., the explicit signal of errors) may unambiguously help students detect their own nontargetlike production and draw their selective attention to the instructor's reformulation (Mackey et al., 2003). McCandliss, Fiez, Protopapas, Conway, and McClelland (2002) teased apart and tested the role of negative evidence in the efficacy of intensive auditory training on the learning of the English /ɹ/-/l/ contrast for Japanese learners. The results showed that those receiving negative evidence throughout their training sessions (i.e., an error signal of whether their answer was correct or not) not only improved their discrimination and identification abilities of the nonnative contrast in a trained synthetic continuum, but also transferred this gain to an untrained continuum; those who did not receive negative evidence failed to do so.

The opportunities for enhanced output that phonological recasts made available in the present study are another important dimension in explaining the observed learning benefits. The results of the fidelity of implementation analyses showed that the learners in the current study produced a great deal of self-modified output immediately following the instructor's recasts. According to psycholinguistic accounts of SLA, this recast-repair sequence could be considered as a repetitive output practice in the context of authentically communicative interaction. Many researchers have argued that this interactional move fosters a transfer from information-processing requirements at the time of language learning to use (Trofimovich & Gatbonton, 2006), as well as a gradual transition from effortful use to more automatic use of the target language (DeKeyser, 2007; Ranta & Lyster, 2007).

Another advantage of enhanced output practice may lie in its provision of many self-perception opportunities: The more the learners repaired the instructor's recasts, the more they heard their own production of /ɹ/. In the L2 speech literature, the ability to accurately perceive one's own speech is found to be a necessary condition for the development of accurate L2 production. Namely, after L2 learners sufficiently develop their perceptual abilities of new sounds, they start tuning their production to their perceptions, based on the comparison of one's own and others' speech (i.e., a perception-first view; see Baker & Trofimovich, 2006; see also Flege, 1995, 2003; Kuhl, 2000). In the recast-repair sequence, therefore, the many opportunities for self-perception of /ɹ/ may have pushed learners to compare their own articulatory configurations of /ɹ/ in relation to the model auditory targets of /ɹ/ (i.e., the instructor's production of /ɹ/). As a result, this parameter adjustment was able to ultimately help the learners attain a more closely matched auditory and articulatory relationship and by extension may have served as a strong template for improving superior production abilities (Sheldon & Strange, 1982).

Conclusion and Future Directions

Situated within instructed L2 speech learning, the current study investigated how recasts together with FFI can facilitate the acquisition of /ɹ/ by Japanese learners. Different from traditional focus-on-formS instruction, which typically introduces target sound features in a decontextualized manner, resulting in somewhat limited improvement at a controlled-speech level, FFI in the present investigation enabled learners to notice and practice the target feature of /ɹ/ via a set of engaging communicative practice activities. According to the results of perception and production tests, whereas the FFI-only group attained significant improvement, particularly under trained lexical conditions, the FFI-recast group demonstrated similar but generalizable gains, both in trained and untrained lexical contexts. The results in turn indicate that (a) FFI itself impacts various domains of L2 speech learning processes (perception → controlled production → spontaneous production) and (b) recasts promote learners' attentional shift from lexical units as a whole to phonetic aspects of L2 speech (i.e., vocabulary → sound learning). This could be because recasts provide students with a great deal of enhanced positive evidence, enhanced negative evidence, and enhanced output practice. In this way, recasts enable them to selectively attend to the perceptual difference between their own nontargetlike exemplars of /ɹ/ and the teacher's model pronunciation of /ɹ/ at a segmental level, while proceduralizing their productive use of /ɹ/.

To close, I would like to point out several topics that future studies of this kind should examine. First and foremost, although some efforts were made to examine participants' controlled and spontaneous production abilities, the two-alternative forced choice identification task was adopted out of necessity as the only way to measure their perception performance of /ɹ/. With the goal of obtaining a more detailed description of the relationship between FFI and L2 speech perception, future studies will need to adopt and elaborate on various types of perception tasks, such as identification and discrimination tests with the synthetic continua with the F2 × F3 domain (Iverson et al., 2003) and duration of first formant transition (Underbakke et al., 1988) under various lexical contexts (i.e., /ɹ/ in frequent words vs. infrequent words; see Flege, Takagi, & Mann, 1996) with reaction time instruments (Lively et al., 1994). Next, if one accepts that recast effectiveness is attributed to its three remedial functions—enhanced positive evidence, enhanced negative evidence, and enhanced output practice—it would be intriguing to tease apart and test precisely which components of recasts might be relatively facilitative of the acquisition of /ɹ/. One way to do so is to compare different types of corrective feedback (e.g., recasts vs. clarification requests vs. metalingusitic clues) (Lyster & Saito, 2010). Last, future studies need to test the applicability of the FFI framework to other areas of instructed L2 speech learning, such as vowels (Thomson, 2012), syllable structures (Cardoso, 2011), and suprasegmentals (Wang, Jongman, & Sereno, 2003), and then investigate the pedagogical potential of various FFI options (e.g., interactional tasks with and without recasts as well as with and without explicit instruction) with a range of L2 learners with different L1s, ages, and proficiency profiles.

Final revised version accepted 1 April 2013

Notes

1 The second frequency range of energy concentration (i.e., F2) is one type of acoustic signals that movements of articulator organs (e.g., tongue, lips, jaw) generate. It has traditionally been used as an index of to what degree talkers retract the tongue body to produce sounds.

2 This testing interval was determined based on psycholinguistic research evidence that the integration of novel phoneme sequences into the mental lexicon takes around a week or so rather than immediately after exposure (e.g., Gaskell & Dumay, 2003). It could be considered as "short-delayed" rather than an "immediate" interval according to the FFI research standards in L2 grammar studies (e.g., Spada & Tomita, 2010).

3 Although I originally planned to have nine classes (i.e., three classes for each experimental and control group), one Control class was cancelled for practical reasons such as teacher and classroom availability.
4 As an experienced EFL teacher, he reported that he was highly familiar with their Japanese tap substitution error.
5 According to the acoustic analysis of F3 values on the 50 speech tokens used in the perception test, the male NS speaker averaged 1555 Hz (SD = 126 Hz) for /ɹ/ and 2704 Hz (SD = 168 Hz) for /l/, while the female NS speaker averaged 1982 Hz (SD = 121 Hz) for /ɹ/ and 3149 Hz (SD = 188 Hz) for /l/.
6 According to British News Corpora, "rink" and "route" were categorized as belonging to the first 2000 (Cobb, 2012).
7 Given that the parametric statistics analyses reported in this section drew on the dataset with unequal numbers of participants per group, assumptions were checked and met in terms of equal variance (via Levine tests) and sphericity (via Mauchly's test) ($p > .05$).
8 In the current study, all Bonferroni-adjusted significance tests were conducted via SPSS by multiplying an unadjusted p value by the number of comparisons.
9 The high rate of repair could be due not only to the salience of pronunciation-focused recasts, but also to the fact that the instructor always provided students with an opportunity for repair.

References

Aoyama, K., Flege, J., Guion, S., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /ɹ/ and English /l/ and /ɹ/. *Journal of Phonetics*, *32*, 233–250.

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

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

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

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

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

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*, 433–461.

Cardoso, W. (2011). Teaching foreign sC onset clusters. In K. Dziubalska-Kolaczyk, M. Wrembel, & M. Kul (Eds.), *Achievements and perspectives in second language acquisition of speech* (pp. 29–40). New York, Oxford, UK: Peter Lang.

Cobb, T. (2012). *The compleat lexical tutor*. Retrieved September 1, 2012, from http://www.lextutor.ca/vp/

Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle & J. Mackenzie Beck (Eds.), *A figure of speech: A Festschrift for John Laver* (pp. 63–91). London: Erlbaum.

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

DeKeyser, R. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge, UK: Cambridge University Press.

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

Derwing, T., & Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393–410.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206–257). New York: Cambridge University Press.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). Cambridge, UK: Cambridge University Press.

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

Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. *AILA Review*, *19*, 18–41.

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, *51*, 281–318.

Ellis, R., & Sheen, Y. (2006). Re-examining the role of recasts in L2 acquisition. *Studies in Second Language Acquisition*, *28*, 575–600.

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

Field, J. (2008). Emergent and divergent: A view of second language listening research. *System*, *36*, 2–9.

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

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

Flege, J., Frieda, E., Walley, A., & Randazza, L. (1998). Lexical factors and segmental accuracy in second-language speech production. *Studies in Second Language Acquisition*, *20*, 155–188.

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

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

Fraser, H. (2011). Teaching teachers to teach /r/ and /l/ to Japanese learners of English: An integrated approach. *Proceedings of the Phonetics Teaching and Learning Conference*, pp. 11–15. Retrieved January 1, 2013, from https://www.ucl.ac.uk/psychlangsci/ptlc/proceedings_2011/ptlc2011_fraser-002

Gaskell, G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*, 105–132.

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, *35*, 127–165.

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

Han, Z., Park, E., & Combs, C. (2008). Textual enhancement of input: Issues and possibility. *Applied Linguistics*, *29*, 597–618.

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.

Imai, S., Walley, A., & Flege, J. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *Acoustical Society of America*, *117*, 896–907.

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*, 475–505.

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

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

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, *64*, 459–489.

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

Kim, J., & Han, Z. (2007). Recasts in communicative EFL classes: Do teacher intent and learner interpretation overlap? In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 269–297). Oxford, UK: Oxford University Press.

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

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

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 367–377.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /ɹ/and /l/: III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, *96*, 2076–2087.

Logan, J., Lively, S., & Pisoni, D. (1992). Training Japanese listeners to identify English /ɹ/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: 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., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, *19*, 37–66.

Lyster, R., & Saito, K. (2010). Corrective feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, *32*, 265–302.

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

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning*, *53*, 35–66.

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

McCandliss, B., Fiez, J., Protopapas, A., Conway, M., & McClelland, J. (2002). Teaching the [r]-[l] discrimination to Japanese adults: Behavioral and neural aspects. *Physiology & Behavior*, *77*, 657–662.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.

Ranta, L., & Lyster, R. (2007). A cognitive approach to improving immersion students' oral language abilities: The Awareness-Practice-Feedback sequence. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141–160). Cambridge, UK: Cambridge University Press.

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

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

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. (2013). 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., & 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*. doi: 10.1017/S1366728912000703.

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.

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

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

Sheldon, A., & 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.

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

Thomson, R. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*, 1231–1258.

Trofimovich, P., Collins, L., Cardoso, W., White, J., & Horst, M. (2012). A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *TESOL Quarterly*, *46*, 176–186.

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

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

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

VanPatten, B. (2004). *Input processing and grammar instruction in second language acquisition*. Norwood, NJ: Ablex.

Walley, A. (2007). Speech learning, lexical reorganization, and the development of word recognition by native and non-native English speakers. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 315–330). Amsterdam: John Benjamins.

Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, *113*, 1033–1043.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Content of Instruction