EFFECTS OF FORM-FOCUSED INSTRUCTION ON L2 PRONUNCIATION DEVELOPMENT OF /ɹ/ BY JAPANESE LEARNERS OF ENGLISH

Kazuya Saito

Department of Integrated Studies in Education

McGill University, Montreal

May 2011

A Thesis Submitted to McGill University

in Partial Fulfillment of the Requirement of the

Degree of Doctor of Philosophy

ACKNOWLEDGEMENTS

It would have been impossible to complete this dissertation without several people who have always and patiently supported, encouraged, and inspired me throughout the process. First, I am truly grateful to Dr. Roy Lyster, my thesis supervisor, for guiding me into the field of second language acquisition, psycholingustics, and classroom-based research. Not only has he shared his extensive knowledge and thoughts on this topic based on his world-renown research experience, which has in turn helped improve the quality of my work tremendously, but also he has generously given me a number of opportunities for collaboration in academic presentations and papers, and provided financial support to actually make the project happen. Second, I would also like to thank my committee members, Dr. Pavel Trofimovich for sharing his countless ideas and huge passion towards this topic with me; and Dr. Carolyn Turner for her consistent support and advice on methodologies in my PhD dissertation. I am indebted to François-Xavier Brajot, a PhD candidate in the Department of Communication Sciences and Disorders at McGill University, for his top-notch advice on acoustic analyses and other related issues in phonology. Finally, I would like to express my gratitude towards my parents for their unconditional support for my ambition to pursue PhD studies at McGill University.

The studies reported herein were funded by a *Language Learning* Dissertation Grant and a McGill internal fellowship for dissertation research, both awarded to Kazuya Saito; and a grant from the Social Sciences and Humanities Research Council of Canada (410-2006-2212) awarded to Dr. Roy Lyster.

CONTRIBUTION OF AUTHORS

As described in the McGill thesis preparation guideline, "the research may be presented as a collection of papers of which the student is the author or co-author (i.e., the text of one or more manuscripts, submitted or to be submitted for publication, and/or published articles (not as reprints) but reformatted according to thesis requirements as described below)" (http://www.mcgill.ca/gps/students/thesis/programs/guidelines/preparation/#manu). In this respect, I chose a manuscript thesis option, and my PhD dissertation consists of three interrelated studies (Studies 1, 2, and 3) which investigate one main research question—whether and to what degree can a range of form-focused instruction can facilitate the L2 pronunciation development of /ɹ/ by Japanese learners of English? I am a co-author of Study 1 with Dr. Roy Lyster, my thesis supervisor. As the primary author, I designed the research framework, conducted data collection, wrote the initial manuscript, and took the main responsibility for subsequent revision. As second author, Dr. Roy Lyster introduced me to relevant literature (e.g., form-focused instruction and corrective feedback), funded the project via the Social Sciences and Humanities Research Council of Canada (410-2006-2212), and played a secondary role in writing and revising the manuscript for publication. This manuscript has been accepted as a co-authored publication to appear in *Language Learning* in 2012 (Saito & Lyster, in press). In Studies 2 and 3, I solely designed the research framework, conducted data collection, and wrote the manuscripts. These two manuscripts are currently under review by academic journals.

ABSTRACT

This manuscript-based thesis comprises three interrelated studies investigating the acquisitional value of various types of form-focused instruction (FFI) techniques—focused tasks (FT) with and without corrective feedback (CF) as well as with and without explicit instruction (EI). Study 1 investigates the relative effectiveness of FT with and without CF on L2 pronunciation development by conducting an experimental study with 65 Japanese learners of English in ESL settings. Acoustic analyses were conducted on frequency values of the third formant (F3) of English /ɹ/ tokens elicited by pre- and post-test measures targeting familiar items and a generalizability task targeting unfamiliar items. The results showed that (a) F3 values of the FT+CF group significantly declined after the intervention, regardless of the following vowel contexts; (b) change in F3 values of the FT-only group and the control group was not statistically significant; and (c) the generalizability of FT to novel tokens remained unclear.

Study 2 revisits the original database of Study 1 and examines in depth what factors contribute to FFI effectiveness by conducting a set of new analyses (a rating session with 20 native-speaking listeners, acoustic analyses on various speech properties, and individual interviews with the participating students). The results of ANOVAs showed that, whereas the F2 values of both the FT+CF and FT-only groups equally decreased, only the FT+CF group significantly lowered their F3 values, which, in turn, indicates improvement resulting from recasts. In addition, the results of multiple regression analyses showed that FFI effectiveness was related to (a) the learners' initial pronunciation levels, (b) the amount of explicit knowledge about relevant oral gestures, and (c) motivation.

Due to several limitations of FFI effectiveness emerging from Studies 1 and 2 (e.g., the moderate improvement within only familiar lexical contexts), Study 3 finally examines whether and to what degree providing EI at the beginning of FFI lessons can enhance the *generalizability* and *magnitude* of FFI effectiveness by assisting learners to notice the perceptual difference between a new sound (English /ɹ/) and its L1 counterpart (Japanese tap /ɾ/) and restructure/develop the new phonetic category in their long-term memory representation. Participants were 49 Japanese learners of English in EFL settings. The results of the ANOVAs showed that (a) the participants who received FFI without EI demonstrated small-to-medium improvement (a transition from hybrid to poor English /ɹ/ exemplars especially in familiar lexical contexts); and (b) those who received FFI and EI exhibited large improvement (a transition from hybrid to good English /ɹ/ exemplars) and their gain was generalizable to unfamiliar lexical contexts.

Taken together, the three studies show that L2 learners need EI at the beginning of FFI lessons to make the best of subsequent contextualized input- and output-based practice (i.e., FT and CF treatments) in order to establish the new phonetic representation as well as to proceduralize the newly-acquired phonetic knowledge in a wide variety of lexical, task, and phonetic contexts.

RÉSUMÉ

Cette thèse manuscrite comporte trois études interreliées portant sur la valeur acquisitionnelle de divers types de techniques d'enseignement centrées sur la forme (ECF) : tâches centrées (TC) avec ou sans rétroaction corrective (RC), et accompagnées ou non d'instruction explicite (IE). L'étude 1 se penche sur l'efficacité relative des TC, avec ou sans RC, sur le développement de la prononciation en L2 sur la base d'une étude menée auprès de 65 apprenants japonais de l'anglais langue seconde. Des analyses acoustiques ont été effectuées sur la fréquence du troisième formant (F3) du phonème /ɹ/ de l'anglais obtenue par des mesures prétest et postest ciblant des éléments familiers et une tâche de généralisabilité ciblant des éléments non familiers. Les résultats ont montré que (a) les valeurs de F3 du groupe TC+RC avaient baissé de manière importante à la suite de l'intervention, sans égard aux contextes vocaliques subséquents; (b) le changement des valeurs de F3 du groupe TC seulement et du groupe de contrôle n'était pas significatif sur le plan statistique et (c) la généralisabilité des TC à de nouvelles occurrences ne pouvait être clairement établie.

L'étude 2, utilisant la même base de données que l'étude 1, examine en profondeur quels facteurs contribuent à l'efficacité de l'ECF. Pour ce faire, de nouvelles analyses ont été effectuées (session d'évaluation avec 20 auditeurs locuteurs natifs, analyses acoustiques de diverses propriétés de la parole et entrevues individuelles avec les participants). Les résultats d'analyses de la variance ont montré que, tandis que les valeurs de F2 tant des groupes TC+RC et TC seulement avaient baissé dans une même mesure, seuls les apprenants du groupe TC+RC ont manifesté une baisse importante de leurs valeurs F3, ce qui indique une amélioration attribuable aux reformulations. De plus, les résultats de plusieurs analyses de régression ont

montré que l'efficacité de l'ECF était liée (a) aux niveaux de prononciation initiaux des apprenants, (b) au degré de connaissance explicite des mouvements oraux pertinents et (c) à la motivation.

En raison des nombreuses limites à l'efficacité de l'ECF constatées dans le cadre de ces deux études (c.-à-d. l'amélioration modérée dans des contextes lexicaux familiers seulement), l'étude 3 examine si, et à quel degré, la communication d'IE au début des leçons ECF peut relever la *généralisabilité* et l'*ampleur* de l'efficacité de l'ECF en aidant les apprenants à remarquer la différence perceptive entre un son nouveau (/ɹ/ de l'anglais) et sa contrepartie en L1 (/ɾ/ battu du japonais) et à restructurer/élaborer la nouvelle catégorie phonétique dans leur représentation en mémoire à long terme. Les participants à cette étude étaient 49 apprenants japonais de l'anglais dans le cadre d'un enseignement de l'anglais langue étrangère. Les résultats d'analyses de la variance ont montré que (a) les participants ayant reçu un ECF sans IE ont démontré une amélioration variant de faible à modérée (transition d'une prononciation hybride à une piètre prononciation du /ɹ/ de l'anglais, surtout dans des contextes lexicaux familiers) et (b) ceux ayant reçu un ECF et des IE ont manifesté une grande amélioration (transition d'une prononciation hybride à une bonne prononciation du /ɹ/ de l'anglais) et que leur gain pouvait être généralisé à des contextes lexicaux non familiers.

Considérées dans leur ensemble, ces trois études montrent que les apprenants en L2 ont besoin d'IE au début de leurs leçons d'ECF pour tirer pleinement profit de leur pratique contextualisée subséquente fondée sur l'écoute-production orale (c.-à-d. TC et RC) en vue d'établir la nouvelle représentation phonétique et aussi d'assurer la procéduralisation de la connaissance phonétique nouvellement acquise dans une grande diversité de contextes lexicaux, phonétiques et pragmatiques.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

ABSTRACT iv

RÉSUMÉ vi

TABLE OF CONTENTS viii

LIST OF TABLES xvii

LIST OF FIGURES xviii

INTRODUCTION 1

STUDY 1: EFFECTS OF FOCUSED-TASKS AND CORRECTIVE FEEDBACK ON L2 PRONUNCIATION DEVELOPMENT OF /ɹ/ BY JAPANESE LEARNERS OF ENGLISH 5

FFI in SLA 5

Focused-tasks 10

Corrective Feedback 11

Pronunciation Teaching 14

Paradigms in Pronunciation Teaching 14

Previous Intervention Studies 17

Methodological Problems in Pronunciation Teaching Studies 23

Types of Instruction 24

Measuring Pronunciation Development 26

English /ɹ/ 27

Pedagogical Importance of English /ɹ/ 28

Acoustic Properties of English /ɹ/ 29
Current Study 30
Research Questions 31
Method 31
Participants 32
Students 32
Instructors 33
Listeners 34
Procedure 35
FT Treatment 37
Integration of Pronunciation Targets into the English-Argumentative Skills Lessons 37
Critiquing arguments. 39
Debating topics. 39
Developing arguments and public speaking. 40
Warm-up Activities 40
English Karuta. 40
English card game. 41
Guessing game 41
FT+CF 42
Control Group 43
Teacher Training 43
Measures 44
Familiar Items 44

Word Reading 45
Sentence Reading 45
Timed Picture Description 45
Unfamiliar Items 47
Rating and Judgment Sessions 47
Acoustic Analysis 49
Results 51
Analysis of Acoustic Properties and Listener Ratings 51
Word-Reading, Sentence-Reading, and Picture-Description Tasks 53
Task Types 54
Following Vowel Backness 55
Following Vowel Height 56
Generalizability Task 57
Following Vowel Backness 57
Following Vowel Height 59
Personal Interview 59
Discussion 61
Acoustic Properties of English /ɹ/ 61
Task Types 62
Following Vowel Backness and Height 63
Other Phonetic Cues 63
FT with and without CF 64
Conclusion 66

Limitations and Pedagogical Implications 68

TRANSITION FROM STUDY 1 TO STUDY 2 70

STUDY 2: VARIABLES AFFECTING RECAST EFFECTIVENESS IN L2 PRONUNCIATION DEVELOPMENT 72

Recasts in SLA 72

Different Linguistic Features 73

Repair 74

Learners' Initial Pronunciation Level 75

Explicit Knowledge 75

The Amount of Motivation and L2 Experience 76

English /ɹ/ 76

Motivation for Study 2 79

Research Questions 80

Method 80

Design 80

Participants 82

Students 82

Teachers 82

Listeners 83

Interview 83

Instructional Treatment 83

Pronunciation-focused Recasts 84

Control Group 85

Measures 85
Familiar Tokens 86
Word reading 86
Sentence reading 86
Timed picture description 86
Target Words 87
Novel Tokens 87
NS Baseline 88
Listener Ratings and Acoustic Analyses 88
Token Preparation 89
Acoustic Properties and Contextual Factors 89
Rating Sessions 92
Group Comparisons and Post-Hoc Analyses 92
Dependent Variables 93
Independent Variables 93
Amount of repair 93
Learners' initial pronunciation levels 95
Explicit knowledge 95
Motivation 95
Length of residence 95
Results 96
Listener Ratings and Acoustic Analyses 96
F3 Values 97

F2 Values 98
Data Analyses 99
ANOVAs 99
Familiar Tokens 99
Novel Tokens 100
NS Baseline 101
Multiple Regression Analyses 101
Amount of Repair 101
Initial Pronunciation Levels 102
Explicit Knowledge 102
Motivation 102
Length of Residence 102
Results of Multiple Regression Analyses 102
F3 of singletons following front vowels. 103
F3 of singletons following back vowels 103
F2 of singletons following front vowels 103
F2 of singletons following back vowels 103
Summary of the Results 103
Discussion 104
NS perception of English /ɹ/ 104
Effects of Recasts on F3 and F2 Values 105
Variables Affecting Recast Effectiveness 106
Limitations 107

Conclusion 108

TRANSITION FROM STUDY 2 TO STUDY 3 110

STUDY 3: RE-EXAMINING THE EFFECTS OF FORM-FOCUSED INSTRUCTION ON L2 PRONUNCIATION DEVELOPMENT: THE ROLE OF EXPLICIT INSTRUCTION 112

L2 Phonological Development 113

Theoretical Issues in L2 Phonological Development 114

Decoding L2 Phonological Information 114

Phonetic Categories, Perception, and Production 117

Representation and Processing in L2 Phonological Development 121

Representation in L2 Phonological Development 122

Processing in L2 Phonological Development 126

Summary 130

FFI and L2 Phonological Development 132

Findings and Limitations of FFI 134

Adding EI to FFI 136

EI in L2 Morphosyntactic Development 136

EI in L2 Phonological Development 138

Current Study 141

Research Questions 141

Method 141

Design 141

Participants 144

Students 144

Instructors 145
Interview 146
FFI Treatment 146
EI Treatment 146
Control Group 148
Teacher Training 149
Measures 149
CP test. 150
SP test 151
Data Analyses 152
NS Baseline 153
Acoustic Analyses 153
Post-hoc Analyses 156
Amount of Repair 156
Final Interview 157
Results 157
Pre and Post Test Data 157
Pre-test Data 158
Familiar Items 159
Task types 159
Vowel backness 160
Vowel height 160
Unfamiliar Tokens 161

Task type 161
Vowel backness 162
Vowel height 162
NS Baseline 163
Post-hoc Analyses 165
Amount of Recasts and Repair 165
Endpoint Interview 165
Discussion 165
Re-examining FFI Effectiveness 166
Adding EI to FFI Lessons 168
Noticing 168
Proceduralization and Generalization 169
Amount of Repair 169
Conclusion and Future Directions 170
REFERENCES 176

LIST OF TABLES

Table 1. Summary of 10 classroom FFI studies 7

Table 2. Summary of 5 pronunciation teaching studies..... 18

Table 3. Participant information by Group 34

Table 4. Target minimally-paired words 38

Table 5. Average F3 values of the 100 speech tokens rated by 5 NS listeners 53

Table 6. Summary of results of pre- and post-tests targeting familiar items..... 58

Table 7. Summary of results of Generalizability Task targeting unfamiliar items..... 60

Table 8. The results of NS perception of English /ɹ/ 98

Table 9. Participant information by Group 146

Table 10. 15 tokens in CP tests in relation to following vowel conditions 150

Table 11. 10 tokens in SP tests in relation to following vowel conditions..... 152

Table 12. The summary of the significant changes in F3 values..... 164

LIST OF FIGURES

Figure 1. Types of CF from Lyster and Saito (2010a, p. 278) 12

Figure 2. Summary of the procedure 36

Figure 3. Summary of the acoustic analyses 50

Figure 4. Summary of the procedure in Study 2 81

Figure 5. Procedure for the acoustic analysis for English /ɹ/ 91

Figure 6. Decoding information during L2 phonological development 116

Figure 7. Summary of the SLM 121

Figure 8. Change in new phonetic representation during L2 phonological development 125

Figure 9. Change in processing during L2 phonological development 129

Figure 10. Revised model for L2 phonological development 131

Figure 11. The pedagogical sequence of FFI techniques adapted from Lyster (2004, 2007; Ranta & Lyster, 2007) 133

Figure 12. Summary of the procedure 143

Figure 13. Procedure for the acoustic analysis for English /ɹ/ 155

.

INTRODUCTION

Second language acquisition (SLA) studies have examined the effects of form-focused instruction (FFI) as "a set of psycholinguistically motivated pedagogic options" (Ellis, 2001, p. 12), which include focused-tasks (FT), corrective feedback (CF), and explicit instruction (EI) in the context of second language (L2) grammar teaching. They found that FFI can impact learners' developing system of L2 morphosyntax not only at a controlled level but also at a spontaneous level (e.g., Doughty, 2001, 2003; Ellis, 2001, 2002, 2006; Long, 2007; Long & Robinson, 1998; Spada, 1997; Spada & Lightbown, 2008; for meta-analyses, see Lyster & Saito, 2010a; Norris & Ortega, 2000; Spada & Tomita, 2010). However, this line of L2 grammar research has also revealed two crucial research topics that strongly call for future research. First, they have yet to tease apart and test various combinations of the three instructional components (i.e., FT with and without CF as well as with and without EI) in order to further examine how they interact to impact L2 learners' interlanguage development. Second, although it is assumed that the empirical findings of L2 grammar studies could be applicable to all types of language features such as L2 lexis (e.g., Schmitt, 2008) and L2 pragmatics (e.g., Rose & Kasper, 2002), it is surprising that little attention has been given to FFI research in the domain of L2 speech pronunciation.

Given that "the linguistic target of the instruction can be phonological, lexical, grammatical or pragmatic" (Ellis, 2006, p. 157), the time is ripe for SLA researchers to move ahead the FFI research agenda developed in L2 grammar studies to an underrepresented area of SLA research: How can a range of FFI techniques such as FT, CF, and EI, impact L2 development according to *different* aspects of language (i.e., L2 pronunciation development)? In

this regard, the proposed study expands the scope of SLA scholarship by fulfilling two main research aims: (a) testing whether effective/efficient L2 grammar teaching methodologies (i.e., FFI) can function in the area of pronunciation teaching contexts where FFI is tailored to highlight a particular group of L2 learners, namely, Japanese learners of English, to acquire L2 pronunciation development of English /ɹ/[1]; and (b) investigating in depth which combination of FFI techniques (FT ± CF ± EI) can most benefit their L2 phonological development. Thus, for my PhD dissertation, I report a series of experimental studies with a pre- and post-test design where Japanese learners of English received four hours of FFI lessons and their improvement was measured via a variety of production tests.

This manuscript-based thesis consists of three studies, each presenting an empirical study and each preceded by a relevant literature review. *Study 1* explores what types of FFI options (i.e., FT ± CF) appear most conducive to enhancing L2 pronunciation development of /ɹ/ by Japanese learners of English and what types of outcome measures most effectively assess learners' development resulting from instruction. In Study 1, I first propose a model research design in order to investigate the acquisitional value of FFI on L2 pronunciation development drawing on knowledge developed in L2 grammar studies as well as L2 phonology studies. Subsequently, I test the robustness of the research design by examining the results of an experimental study I conducted where 65 Japanese learners of English received four hours of meaning-oriented lessons in relation to two types of FFI (i.e., FT-only vs. FT+CF). This manuscript has been accepted as a co-authored publication to appear in *Language Learning* in 2012 (Saito & Lyster, in press). As the primary author, I designed and implemented the entire

[1] The current study concurs with Flege's (1995) theoretical orientation of L2 phonetics and phonology that "sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level" (p. 239). Although L2 sounds are considered and analyzed on the basis of phonetic units instead of phonemic units in the current study, a decision was made to use not [ɹ] but /ɹ/ in order to follow the norm in L2 phonetics and phonology studies of this kind.

study then played the lead role in writing the manuscript. As second author, Dr. Roy Lyster funded the project and played a secondary role in writing and revising the manuscript for publication.

Whereas Study 1 identified the relative effects of combining FT and CF in L2 pronunciation development, *Study 2* further investigates how a range of factors interact to determine FFI effectiveness from various perspectives. That is, drawing on the database of Study 1, Study 2 carried out three new secondary analyses: (a) re-examining the perception of English /ɹ/ via a new rating session with more listeners and tokens (20 native-speaking [NS] listeners rated 150 speech samples produced by 30 Japanese learners of English), (b) implementing acoustic analyses not only on the frequency values of the third formant (F3) but also on the frequency values of the second formant (F2) values for the entire dataset, and (c) analyzing the details of the individual interviews with 54 participating students. As sole author, I am responsible for the data collection and analysis as well as for writing the entire manuscript, which is currently under review.

Finally, due to the limitations of FFI effectiveness revealed by Studies 1 and 2 (e.g., the moderate improvement only within familiar lexical items), I decided to conduct a conceptual replication of the original research framework developed in Studies 1 and 2 with 49 Japanese learners of English. To this end, *Study 3* was designed to examine whether and to what degree providing *explicit phonetic instruction* (EI) at the beginning of FFI lessons can enhance the *generalizability* and *magnitude* of FFI effectiveness by assisting learners to notice the perceptual difference between a new sound (English /ɹ/) and its L1 counterpart (Japanese tap /ɾ/) and restructure/develop the new phonetic category in their long-term memory representation. As sole

author, I am responsible for the data collection and analysis as well as for writing the entire manuscript, which is currently under review.

Although it is claimed that pronunciation teaching likely falls into over-dependency on intensive decontextualized methods such as repetition and mechanical drills, and that the details of L2 pronunciation development have remained unclear (e.g., DeKeyser, 1998), the findings of Studies 1, 2 and 3 provide both pedagogical and theoretical accounts of how a range of FFI techniques (FT ± CF ± EI) can be facilitative of L2 pronunciation development by means of establishing new representations of phonetic categories in long-term memory as well as proceduralizing the newly-acquired phonetic knowledge in a variety of lexical, task, and phonetic contexts.

STUDY 1: EFFECTS OF FOCUSED-TASKS AND CORRECTIVE FEEDBACK ON L2 PRONUNCIATION DEVELOPMENT OF /ɹ/ BY JAPANESE LEARNERS OF ENGLISH

Research into form-focused instruction (FFI) has focused almost exclusively on morphosyntactic targets, in spite of calls for research into the roles FFI in phonological development and suggestions that the latter might be especially amenable to phonological recasts. *Study 1* will first present a literature review of relevant topics, focusing on (a) the pedagogical effectiveness of focused-tasks (FT) and corrective feedback (CF) in grammar studies, and (b) the history of previous pronunciation teaching studies and their methodological problems. Second, I will present the results of a quasi-experimental study which integrates FFI into pronunciation teaching contexts where FFI is tailored to target a particular group of L2 learners, namely, Japanese learners of English, to support them in their development of intelligible pronunciation of English specific sound /ɹ/.

FFI in SLA

Although naturalistic approach proponents assume that implicit acquisition dominates SLA and that simple exposure to positive evidence alone is sufficient and ideal for SLA processes (e.g., Krashen, 1981; Schwartz, 1993), several empirical studies revealed that L2 learners participating in focus-on-meaning programs (e.g., French immersion in Canada) failed to achieve the expected levels of L2 accuracy (for a review, Lyster, 2007). Subsequently, some SLA scholars emphasized the importance of negative evidence in SLA (Gass, 1997; Long, 1996; L. White, 1987), and one of the topics of much discussion is the role of FFI as a

psycholinguistically appropriate way to enhance the rate and ultimate attainment of SLA (e.g., Doughty, 2003; Ellis, 2001, 2002, 2006; Lyster, 2007, Spada & Lightbown, 2008).

FFI is defined as "any pedagogical effort which is used to draw the learners' attention to language form either implicitly or explicitly" (Spada, 1997, p. 73). Different from traditional grammar translation methods that introduce forms in a decontextualized manner, FFI is hypothesized to be most effective when implemented in content-based and communicative language classrooms in which conveying a meaningful message is a priority, arguably because integrating form in this way helps L2 learners to (a) develop their "form-meaning mappings" (Doughty & Williams, 1998; Long, 1996, 2007; Long & Robinson, 1998; VanPatten, 2002, 2004) and (b) promote a gradual transition from effortful to automatic use of rules (DeKeyser, 1998, 2003, 2007; Lyster, 2007; Lyster & Saito, 2010b). Among effective FFI activities identified by researchers, those of direct relevance to Study 1 include FT (i.e., communicative activities which are designed to create obligatory contexts and elicit learners' use of a specific linguistic feature in comprehension and production) and CF (i.e., providing corrective information on students' linguistic errors). In the next subsection, I will introduce a number of quasi-experimental studies conducted in a range of classroom settings, and discuss how previous studies confirmed the effectiveness of FFI on learners' interlanguage development in L2 grammar studies. Table 1 summarizes 10 widely-cited quasi experimental FFI studies that I will use to discuss how previous studies confirmed the effectiveness of FFI on learners' interlanguage development in L2 grammar studies.

Table 1. *Summary of 10 classroom FFI studies*

	participants	Target of instruction	Length	FT techniques	CF techniques	Outcome measures	Findings
L. White, Spada, Lightbown, & Ranta (1991)	239 grade 5 and 6 ESL students in Montreal	English WH interrogatives	6 hours	- Metalinguistic explanation - Focused tasks	- Explicit correction	- Grammaticality judgment test - Picture description (with prompts) - Picture description (through oral communication)	- Students who received explicit type of EEI demonstrated their improvement in all of the measures
VanPatten & Cadierno (1993)	80 university-level English learners of Spanish	Spanish direct object pronouns	2 hours	- Metalinguistic explanation - Structured input activities (processing instruction group only)	*n.a.*	- Sentence-picture matching (interpretation task) - Picture-prompted sentence-creation (production task)	- Processing instruction (explicit instruction + structured activity) benefited not only students' comprehension abilities but also their production abilities - Traditional instruction (explicit instruction + mechanical practice drills) had an effect only on students' production abilities
Doughty & Varela (1998)	34 grade 8 ESL students	English conditional past tense	4 weeks	- Focused tasks	- Repetition + recasts	- Accuracy rate of the target form in students' oral and science reports	- Students who received repetition followed by recasts significantly improved both in oral and written measures - Their improvement was retained until 4 months after the instruction especially in written measures

J. White (1998)	86 grade 5 and 6 ESL students	English possessive determiner	20 weeks	- Textually enhanced input - Focused tasks - Extensive reading	No feedback	- Passage correction -Multiple-choice - Picture description - Listening comprehension	- All three groups (focused task, ± typographic enhancement, ± extensive reading) improved between pre- and post-tests - No group difference was found
Muranoi (2000)	91 university-level Japanese learners of English	English indefinite article	1 hour and 30 minutes	- Metalinguistic explanation (form-focused group only) - Focused tasks	- Request for repetition + recasts	- Grammaticality judgment - Oral and written picture description - Oral story description	- FFI benefited learners' acquisition of English indefinite article - Students who received both metalinguistic explanation and CF significantly improved their interlanguage development more than those who received only CF
Lyster (2004)	179 grade 5 immersion students	French gender attribution	9 hours	- Textually enhanced input - Focused tasks	- Prompts - Recasts	- Binary-choice (written) - Text-completion (written) - Object identification (oral) - Picture description (oral)	- All of the three groups (proactive-FFI,± recasts and ± prompts)outperformed the control group - Students exposed to prompts showed more improvement than those exposed to recasts especially in written measures
Ammar & Spada (2006)	64 grade 5 or 6 ESL learners	English Possessive determiners	6 hours	- Focused tasks	- Prompts - Recasts	- Passage correction (written task) - Picture description task (oral task)	- Both experimental groups who received either recasts and prompts outperformed the control group - For low proficiency learners (developmentally-unready learners), prompts were more effective than recasts in both measures - For high proficiency learners (developmentally-ready learners), effects of prompts and recasts were comparable

Ellis, Loewen, & Erlam (2006)	34 adult ESL students in New Zealand	English regular past tense	1 hour	- Focused tasks	- Providing metalinguistic information - Recasts	- Oral imitation - Grammaticality judgment	- The explicit feedback group (explicit feedback + focused tasks) outperformed the recast group (recasts + focused tasks) and the control group (focused tasks only)
Sheen (2007)	80 adult ESL students	English article	90 minutes	- Focused tasks	- Explicit correction - Recasts	- Total scores of (a) speeded dictation, (b) writing task, and (c) error correction	- Only the explicit correction group (explicit correction + focused tasks) outperformed the control group (focused tasks only) in all measures
Yang & Lyster (2010)	72 university-level EFL students	English regular and irregular past tense	2 hours	- Focused tasks	- Prompts - Recasts	- Picture-retelling with word cues (oral) - Story composition (written)	- All of the three groups (focused tasks, recasts, prompts) demonstrated significant improvement between pre- and post-tests - For irregular past tense, both recasts and prompts were equally effective - For regular past tense, prompts were more effective than recasts

Focused-tasks

The definition of FT in Study 1 follows the well-known concept of *task-essentialness* by Loschky and Bley-Vroman (1993) as follows: "The most extreme demand a task can place on a structure . . . the task cannot be successfully performed unless the structure is used' (p. 132). Lyster (2007) referred to these FT treatment as a part of *proactive FFI* which involves "pre-planned instruction designed to enable students to notice and to use target language features that might otherwise not be used or even noticed in classroom discourse" (p. 44).

For example, a number of FFI researchers developed innovative FT techniques, such as (a) structured input (i.e., learners are required to process linguistic form in input for meaning without being pressured to produce output; VanPatten, 2002, 2004), (b) typographically enhanced input (i.e., the 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; Han, Park, & Combs, 2008; Sharwood Smith, 1993), (c) output-focused tasks (i.e., learners are required to produce linguistically accurate output to successfully complete meaning-oriented tasks; Ellis, 2001, 2006).

A number of experimental studies have been carried out in a variety of classroom contexts with different methodologies, in order to test the impact of one or a combination of these proactive FFI techniques on learners' morphosyntactic interlanguage development. For example, VanPatten and his colleagues (e.g., VanPatten & Cardierno, 1993; VanPatten & Oikennon, 1996) conducted a series of quasi-experimental studies to test whether or not processing instruction entailing structured input activities can alter learners' default strategies for morphosyntactic input processing (such as the First Noun Principle), finding that processing instruction significantly improved not only students' comprehension abilities but also their

production abilities (but see DeKeyser & Sokalski, 1996). Doughty and Varela (1998) integrated FTs in content-based classrooms by asking middle school ESL learners to complete oral and written science reports that were designed to elicit students' use of English conditional past tense. In Lyster's (2004a) quasi-experimental study in Grade 5 French immersion classrooms, students were asked to read passages about the history of Quebec with noun endings predictive of grammatical gender highlighted in bold (i.e., textually enhanced input) and to solve riddles in which they had to produce various target nouns with accurate grammatical gender attribution (i.e., FT).

In her quasi-experimental study on the acquisition of English possessive determiners in child ESL classrooms, J. White (1998), however, found that typographically enhanced input alone did not sufficiently draw learners' attention to form. She argued that this implicit kind of proactive FFI might need to be combined with more explicit pedagogical techniques such as metalinguistic explanation (see Spada, Lightbown, & J. White, 2005; also Muranoi, 2000; Vanpatten & Cadierno, 1993). Note that, in order to enhance the overall FFI efficacy, most of the quasi-experimental studies mentioned above (e.g., Doughty & Varela, 1998; Lyster, 2004a) also incorporated provision of CF to learners' linguistic errors (i.e., reactive FFI).

Corrective Feedback

CF is defined by Ellis (2006) as "responses to learner utterances containing an error" (p. 28). Since Lyster and Ranta's (1997) descriptive study, which consisted of 18 hours of teacher-student interaction in French immersion classrooms, provided a classification of different types of teachers' CF types, SLA research has descriptively and experimentally corroborated the role of CF in FFI in classrooms as well as lab settings. CF types can be categorized not only in terms of implicitness and explicitness (e.g., Ellis & Sheen, 2006) but also in terms of Ranta and

Lyster's (2007) distinction between reformulation, which "includes recasts and explicit correction because both these moves supply learners with target reformulations of their non-target output" (p. 152), and prompts, which "include a variety of signals, other than alternative reformulations, that push learners to self-repair" (p. 152). Lyster and Ranta (1997) defined recasts as "the teacher's reformulation of all or part of a student's utterance, minus the error" (p. 46). Explicit correction also provides the correct form but, unlike recasts, "clearly indicates that what the student had said was incorrect" (p. 46). Importantly, although prompts include a wide variety of CF types such as elicitation, metalinguistic clues, clarification requests, and repetition, unlike reformation moves, they always withhold correct forms and instead provide clues to prompt students to retrieve these correct forms from their existing knowledge (see, Lyster 2002, 2007). Figure 1 attempts to group CF types along a continuum that ranges from implicit to explicit and according to the dichotomous distinction between reformulations and prompts.

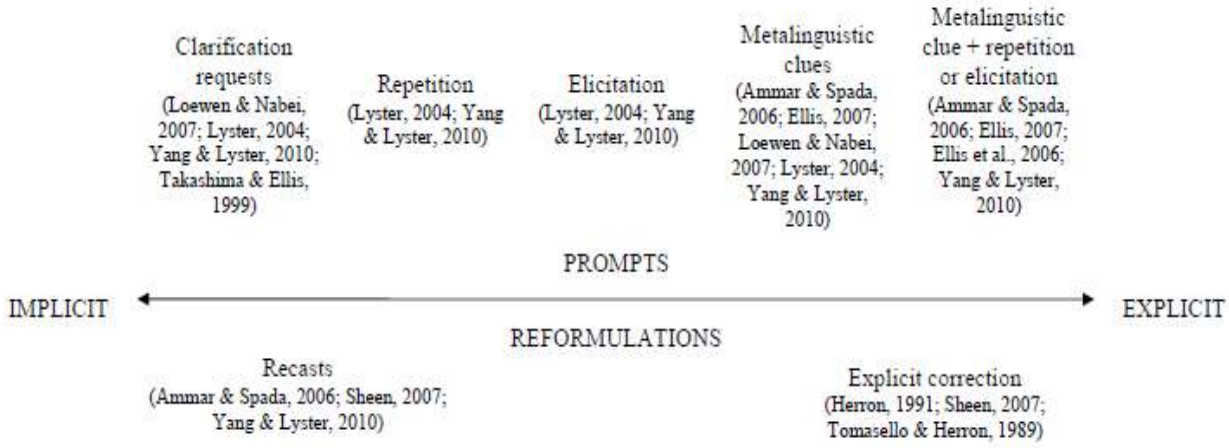


Figure 1. Types of CF from Lyster and Saito (2010a, p. 278)

Although recasts have been identified as the most frequent type of CF used by teachers in a wide range of instructional settings, some studies have pointed out that recasts of grammatical errors are potentially ambiguous for classroom learners accustomed to focusing more on communication, because such recasts might appear to be identical or alternative ways of saying the same thing in order to confirm message comprehensibility or veracity (Ellis & Sheen, 2006; Lyster. 1998a, 2007; Nicholas, Lightbown, & Spada, 2001). As a remedy, several recent empirical studies confirmed the importance of adopting more pedagogically-oriented CF techniques (i.e., prompts). On the one hand, in a series of quasi-experiment studies in adult ESL classrooms in New Zealand (Ellis, 2007; Ellis, Loewen, & Erlam, 2006), it was shown that prompts as a form of metalinguistic clue led learners to gain more control over their already-acquired knowledge of the English past tense and comparative than implicit feedback such as recasts. Similarly, Sheen (2007) demonstrated that explicit correction (i.e., explicitly reformulating learners' nontargetlike production with metalinguistic explanation) benefited adult ESL learners' acquisition of English articles more than recasts. On the other hand, the relative efficacy of prompts (i.e., pushing learners to make self-corrections by withholding correct forms) over recasts has been confirmed in the case of young immersion students' acquisition of French gender attribution (Lyster, 2004a), young ESL students' acquisition of English possessive determiners (Ammar & Spada, 2006), and university-level EFL students' acquisition of regular English past tense (Yang & Lyster, 2010). In sum, with respect to CF effectiveness on L2 morphosyntactic development, as Ellis and Sheen (2006) concluded, "there is no clear evidence that recasts work better for acquisition than other aspects of interaction such as models, prompts, or explicit corrective strategies" (p. 597).

Very intriguingly, many descriptive studies have suggested, however, that recasts might be relatively effective for L2 phonological development compared to other domains such as L2 morphosyntax, arguably because learners tend to perceive the corrective force of teachers' recasts on pronunciation errors (Carpenter, Jeon, MacGregor, & Mackey, 2006; Ellis, Basturkmen, & Loewen, 2001; Kim & Han, 2007; Han, 2008; Lyster, 1998b; Lyster, & Saito, 2010b; Mackey, Gass, & McDonough., 2000; Sheen, 2006). To the best of my knowledge, there exist no experimental studies with a pre- and post-test design that investigate the acquisitional value of FT with and without CF for L2 pronunciation development; the time is ripe to explore this new direction. In the next section, I will turn to a review of previous pronunciation teaching studies that have grown *independently* of all of the L2 grammar studies reviewed here.

Pronunciation Teaching

Whereas much attention has been directed to grammar instruction as discussed above, as Derwing and Munro (2005) pointed out, "much less research has been carried out on L2 pronunciation than on other skills such as grammar and vocabulary, and instructional materials and practices are still heavily influenced by commonsense intuitive notions" (p. 379). In this section, I will first describe the history of pronunciation teaching studies and summarize research findings of empirical studies in order to demonstrate the state of the art in this research field. Subsequently, I will carefully spell out their methodological problems that need to be revised in order to develop/refine a trustworthy framework for future FFI studies in the context of pronunciation teaching.

Paradigms in Pronunciation Teaching

In the 70's, pronunciation teaching was considered a priority in L2 classrooms by proponents of the audio-lingual approach to language teaching who emphasized the mastery of

native-like pronunciation (especially phonemic contrasts) through the use of minimal pair drills and the imitation of appropriate models (for a comprehensive review of various approaches to pronunciation teaching, see Celce-Murcia, Goodwin, & Briton, 1996, 2010). Their nativeness assumptions in the audio-lingual approach were, however, not well supported by recent L2 speech research evidence. That is, it has been convincingly shown that (a) L2 speech is in general foreign-accented, mainly due to the interaction between learners' age and L1 (e.g., Flege, 2003; Flege, Munro, McKay, 1995; Best et al., 2001), and (b) very few adult learners can achieve native-like pronunciation in their L2 (Moyer, 1999; Ioup, Boustagi, El Tigi, & Moselle, 1994). Derwing and Munro (2005) summarized that:

> We know of no study documenting a link between pronunciation instruction and the elimination of a foreign accent. Rather, most learners who strive for nativeness are likely to become disheartened…it may do more harm than good for teachers to lead learners to believe that they will eventually achieve native pronunciation or to encourage them to expend time and energy working toward a goal that they are unlikely to achieve. (p. 384)

As a result, unavoidable foreign accent led many researchers as well as practitioners to consider pronunciation as an *unteachable* subject, and, in extreme cases, to completely ignore the importance of pronunciation teaching from their L2 instructional syllabi (Celce-Murcia et al., 1996; Levis, 2005; Setter & Jenkins, 2005). With respect to researchers, recent meta-analyses on various topics in instructed SLA experimental studies published between 1980 and 2008 indeed did not include any single pronunciation teaching study, revealing surprisingly little research interest in the field (Lyster & Saito, 2010a; Mackey & Goo, 2007; Norris & Ortega, 2000;

Russell & Spada, 2006; Spada & Tomita, 2010). With respect to practitioners, despite some efforts still made by teachers to incorporate pronunciation instruction, their choices are either limited to simply teaching articulatory phonetics or are profoundly influenced by their ideology and intuition rather than research (Derwing, 2008; Derwing & Munro, 2005; Levis, 2005; Setter & Jenkins, 2005; Munro, 2008). Breitkreutz, Derwing and Rossiter (2001) administered a questionnaire to 67 ESL teachers in Canada, finding that 60% of them had received no professional training on pronunciation teaching, and that most of them did not have sufficient knowledge of the matter, although they reported a serious need to teach pronunciation in their programs.

After a long period of controversies in terms of the importance of pronunciation teaching in L2 classrooms, there exists, however, a growing revived interest in another perspective towards pronunciation teaching—the *communicative approach* (Celce-Murcia et al., 1996, 2010). Different from the audio-lingual approach, proponents of the communicative approach maintain that the ultimate goal for L2 speech learning is to achieve not only accurate but also fluent usage of "*intelligible*" pronunciation for the purpose of successful L2 communication. Instead of emphasizing the elimination of all kinds of pronunciation errors necessary to acquire accent-free speech, the communicative approach stresses that the focus of instruction should concern only certain pronunciation rules affecting intelligibility and comprehensibility which are at the heart of successful L2 communication (Derwing, 2008; Derwing & Munro, 2005, 2009, 2010; Field, 2005; Hahn, 2004; Levis, 2005; Munro, 2008; Munro & Derwing, 1995, 1999; Munro, Derwing, & Morton, 2006; Setter & Jenkins, 2005). Although some L2 learners who strongly strive for nativeness should not be discouraged from pursuing their ambitions, researchers as well as teachers should set realistic goals for L2 learners based on empirical research evidence, such as

intelligibility and comprehensibility rather than accentedness (Derwing & Munro, 1997, 2005; Levis, 2005; Venkatagiri & Levis, 2007).

Previous Intervention Studies

Some studies have examined the effects of explicit instruction on segmental aspects of L2 pronunciation via phonetic transcriptions and repetition practice (e.g., Rivers & Temperley, 1978) as well as suprasegmental aspects of L2 pronunciation through computer-mediated visual feedback (e.g., Anderson-Hsieh, 1992; de Bot, 1983; Levis & Pickering, 2004; for a review on audio-visual training studies, see Hardison, 2010). Others have investigated the effects of intensive perceptual training on L2 speech perception (e.g., Iverson, Hazan, & Bannister, 2005; Lively, Logan, & Pisoni, 1993, 1994; Logan, Lively, & Pisoni, 1991) and its impact on L2 speech production (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997, Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Hardison, 2003; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005). Yet, the relevance of these studies to real classroom settings can only be indirect at best, because they focused on the isolated teaching of "difficult" sound rules under strict laboratory conditions in which variables such as intensity and consistency of instruction were well controlled and the length of instruction on only one phonological target, in some cases, lasted for many hours (e.g., 15-22.5 hr in Bradlow et al., 1997 1999; 11.5 hr in Hardison, 2003).

Though few in number, other studies have further investigated the effects of pronunciation instruction by conducting quasi-experimental studies in actual classrooms (e.g., Couper, 2006; Elliott, 1997; Macdonald, Yule, & Powers, 1994; Neri, Mich, Gerosa, & Giuliani, 2008; for summaries of classroom studies of pronunciation teaching, see Derwing, 2008; Derwing & Munro, 2005; Setter & Jenkins, 2005). To contextualize the following discussion, I will use the five pronunciation teaching studies summarized in Table 2.

Table 2. *Summary of 5 pronunciation teaching studies*

	participants	Target of instruction	Instructional treatment	Outcome measures	Findings
Macdonald et al. (1994)	23 ESL learners 120 NS listeners	10 key vocabulary	Length: 10-30 min (a) teacher-led vocabulary choral repetition (10 minutes, *n* = 6) (b) self-study (listening to tape-recordings) (10 minutes, *n* = 6) (c) interaction with clarification requests from NS interlocutors (30 minutes, *n* = 6) (d) control group (*n* = 5)	<u>Task</u> giving mini lecture <u>Judgment</u> 120 undergraduate students were asked to listen to a pair of the same words and rate which one sounded more native-like	- No significant difference was found except for better performance by the self-study group than the control group - There was no apparent decline in learners' pronunciation performance between immediate post-tests and delayed post-tests (2 days after)
Elliott (1997)	66 English learners of Spanish 4 near-native NNS listeners	19 allophones in Spanish	Length: one semester (a) repeating model words and phrases, inductive rule discovery, metalinguistic explanation, word and sentence repetition exercise (15 minutes × 21 classes, *n* = 43) (b) control group (*n* = 23)	<u>Task</u> - repeating 19 words after a NS model - repeating 19 sentences after a NS model - reading 19 written words - describing a set of pictures <u>Judgment</u> 3 near-native NNS listeners were asked to rate nativelikeness of 8580 words with a 3-point scale and 132 picture descriptions with a 5-point scale	- Learners in the experimental group exhibited significant improvement in a word repetition task. a sentence repetition task, and a word reading task - No significant improvement was found in a picture description task - Learners improved, in particular, liquid and stop phonemes

Derwing et al. (1998)	48 ESL learners 48 naïve NS listeners 6 NS experienced listeners	Segmental and suprasegmental aspects of English	Length: 11 weeks (a) segmental focus (explicit instruction on individual sounds and syllable structures, *n* = 18) (b) supra-segmental focus (explicit instruction on speech rate, intonation, rhythm, and word and sentence stress, *n* = 18) (c) control group (*n* = 18)	<u>Task</u> - one sentence reading task (consisting of frequent lexicons) - one picture description task <u>Judgment</u> - 48 undergraduate students were asked to rate comprehensibility and accentedness of 96 statement sentences (2 sentences × 48 participants) - 6 experienced ESL teachers were asked to rate fluency, comprehensibility and accentedness of 96 picture descriptions (2 sentences × 48 participants)	- In the sentence reading task, both of the suprasegmental and segmental groups showed significant gains in comprehensibility ratings but only the segmental group showed significant gains in accentedness ratings. - In the picture description task, only the suprasegmental group significantly improved in both comprehensibility and fluency ratings.
Neri et al. (2006)	28 child Italian learners of English 3 NS experienced listeners	28 target words	Length: 120 minutes for CALL group and 240 minutes for teacher-oriented group (a) CALL treatment (students listened to recordings and were asked to practice target words, *n* = 13) (b) Teacher-led treatment (a teacher read books and encouraged students to practice target words, *n* = 15)	<u>Task</u> - 28 target words from reading materials <u>Judgment</u> - 3 experienced NS listeners were asked to rate1568 words with a 10 point scale	- Both groups significantly improved between pre-test and post-test sessions

Saito (2011a)	20 Japanese learners of English 4 NS experienced listeners	5 segmentals in English	Length: Four 60-minute sessions (a) Explicit phonetic instruction and repetitive practice for the experimental group (n = 10) (b) No pronunciation-focused instruction was delivered to the control group (n = 10)	<u>Task</u> - Four sentence reading tasks - One picture-description task <u>Judgment</u> - 4 experienced NS listeners were asked to rate comprehensibility and accentedness of 160 statement sentences and 40 picture descriptions (4 sentences × 20 participants × pre-/post-tests)	- The experimental group improved their pronunciation performance in the sentence reading tasks in the domain of comprehensibility. - No gain was found for the picture description tasks.

Macdonald, Yule, and Powers (1994) tested four different pedagogical activities which lasted for 10 and 30 minutes in the speech of 23 adult Chinese learners of English: traditional drilling activities, self-study with tape recording, interactive activities, and a no-intervention control condition. However, results showed that 120 NS listeners noticed significant improvement only in the self-study group. Elliot (1997) conducted a quasi-experimental study in university-level Spanish as a foreign language classroom over a semester period, and examined how explicit instructional treatment such as phonetic transcriptions followed by repetition practice promoted their acquisition of 19 Spanish allophones. Their improvement was measured through elicited imitation tasks (i.e., repeating a NS's model), isolated word and sentence reading tasks (i.e., reading isolated words/sentences that included the target allophonic sounds) and picture description tasks (i.e., freely describing a set of pictures), with all of the speech tokens being judged by three near-native nonnative speaking listeners. The results demonstrated that the experimental group exhibited significant improvement at a controlled speech level (i.e., elicited imitation tasks; word and sentence reading tasks) but not at a spontaneous speech level (i.e., picture description tasks).

In their classroom study, Neri, Mich, Gerosa, and Giuliani (2008) compared the effects of a Computer Assisted Language Learning tool (CALL, i.e., an automatic voice recognition system) and traditional pronunciation teaching methods (i.e., teacher-fronted instruction) on word-level pronunciation skills. While students in the traditional method group practiced correct pronunciation of several key words in reading in response to a teacher's model pronunciation, the others in the CALL group did the same activities but individually via the computer software program. Both groups significantly improved but by the same amount without any significant group difference. One of the interpretations could be, however, that, despite our high

expectations towards the possibility of a CALL tool in pronunciation skill development (e.g., Lambacher, 1999), the pedagogical possibility of a CALL tool needs further research. For instance, the discussion between Coniam (1999) and Derwing, Munro and Carbonaro (2000) regarding the accuracy of one prominent automatic speech recognition application software called Dragon System's Naturally Speaking is notable. Although Coniam (1999) promoted it as a CALL tool based on her research, Derwing, Munro and Carbonaro (2000) presented counter-evidence that the accuracy level was around 70% for NNS's speech because the software cannot fully achieve human-like recognition patterns.

Saito (2011a) examined the role of explicit phonetic instruction as an independent variable on two aspects of L2 pronunciation development (i.e., comprehensibility and perceived accentedness) by conducting a quasi-experimental study with 20 Japanese learners of English in ESL settings. After four hours of lab-based instruction on the target pronunciation features of English-specific segmentals /æ,f,v,θ,ð,w,l, ɹ/ which are hypothesized to be most difficult for Japanese learners of English (Saito, 2009), the comprehensibility and perceived foreign accent of the participants' oral production in English via a sentence reading task (designed to measure their improvement at a controlled-speech level) and a picture description task (designed to measure their improvement at a spontaneous-speech level) was evaluated by four experienced NS listeners. The results showed that effects of explicit instruction followed by repetition appeared to be significant only at a controlled-speech level in the domain of comprehensibility, but not at a spontaneous-speech level.

One of the most oft-cited studies is Derwing, Munro, and Wiebe (1998), who investigated how a 10-week instructional treatment targeting either segmentals or suprasegmentals differentially impacted not only accentedness but also comprehensibility of learners'

pronunciation. The gains made by students depended on the method of evaluation. With respect to acentedness rating scores, while students in both the segmental and suprasegmental groups showed significant improvement in a sentence reading task, they did not show any significant gains in a picture description task. With respect to comprehensibility rating scores, only students in the suprasegmental group demonstrated significant improvement in a picture description task. In sum, the researchers concluded that pronunciation teaching studies need to flexibly take into account not only the focus of instruction (e.g., segmentals, suprasegmentals) and units of measurement (e.g., sentence reading task, picture description) but also aspects of improvement (e.g., accentedness, comprehensibility, intelligibility).

Methodological Problems in Pronunciation Teaching Studies

As summarized in the preceding review, research evidence explaining the role of pronunciation instruction can be found within a research framework of experimental phonetics and, to a lesser degree, SLA. The results of these primary studies, however, have not been well acknowledged in either field. In their review of experimental phonetics studies, Piske, MacKay, and Flege (2001) concluded, "Many studies examining the influence of formal instruction on the degree of L2 foreign accent have not produced encouraging results for language teachers" (p. 200). In their research synthesis of 49 instructed SLA studies published between 1980 and 1998, Norris and Ortega (2000) excluded pronunciation teaching studies, because research designs of pronunciation teaching studies need to be critically reconsidered in order to provide any educational implications for ESL/EFL classroom (see also Spada & Tomita, 2010). In this section, I specifically address (a) the problem of the over-dependency on decontextualized instruction followed by repetition as well as (b) the challenge of designing reliable outcome measures to assess the impact of instruction on learners' intelligible pronunciation development

at a fine-grained level. I will point out some problems in these respects in previous studies and propose a solution that I adopted to conduct Study 1.

Types of Instruction

In terms of type of instruction, most pronunciation teaching studies have depended on the exclusive use of explicit instruction on phonetic transcriptions followed by decontextualized practice such as mechanical drills and repetition, and their focus was apparently on forms rather than meaning. As DeKeyser (1998) pointed out, "It is rather uncontroversial that pronunciation is relatively immune to all but the most intensive formS-focused treatments [i.e., decontextualized language-focused methods]" (p. 43). One of the reasons for the dominance of focus-on-formS practice in pronunciation teaching could be that pronunciation requires not only metalinguistic knowledge (i.e., pronunciation rules) but also physical action (i.e., motor activities). That is, L2 learners need to develop abilities to manipulate articulatory organs properly to produce correct L2 sounds (see Flege, 2003, for a discussion of peculiarities of L2 speech *production* compared to other L2 skills).

Some L2 researchers, however, are doubtful of the effects of decontextualized instruction on learners' communicative competence, and call for further research incorporating more psycholinguistically motivated instructional options in pronunciation teaching that would include tasks that are not only extensive and repetitive but also *authentically communicative* (Celce-Murcia et al., 1996; Pennington, 1996; Trofimovich & Gatbonton, 2006; Segalowitz, 2003). In her review of instructed SLA studies, Doughty (2003) noted:

> given the completely decontextualized nature of explicit focus on forms, this type of instruction promotes a mode of learning that is arguably unrelated to SLA, instructed or

> otherwise, in that the outcome is merely the accumulation of metalinguistic knowledge about language. (p. 271)

In fact, the results of previous pronunciation teaching studies have shown that the effects of explicit instruction followed by decontextualized practice on learners' improvement at a spontaneous speech level (measured by picture-description tasks) was rather *discouraging* (e.g., Elliott, 1997; Macdonald et al., 1994; Saito, 2011a); other studies did not even test learners' extemporaneous speech production at all (e.g., Neri et al., 2008). Notably, Derwing et al.'s (1998) study did yield some positive results (i.e., participants receiving suprasegmental-based instruction showed improvement in picture-description tasks). Their goal was to compare a focus on suprasegmentals and segmentals via a mixture of various teaching methods (e.g., pronunciation lessons and some meaningful activities such as group presentations) rather than isolate and test the effectiveness of a specific teaching method on one phonological target. In order to conduct a fine-grained analysis of how instruction facilitates L2 pronunciation development, intervention studies are called for that carefully spell out (a) what suprasegmental and segmental aspects of pronunciation are specifically targeted and (b) in what way and for how long the intervention is implemented to teach each of the target features.

L2 grammar studies have convincingly shown that psycholinguistically motivated instructional treatments integrating form and meaning (FFI, focus-on-form) are more effective than (a) decontextualized teaching methods (grammar-translation methods, focus-on-formS) and (b) mere exposure to the target language (naturalistic approach, focus-on-meaning) (Doughty, 2003; Ellis, 2001, 2002, 2006; Lyster, 2007; Lyster & Saito, 2010b; Spada, 1997; Spada & Lightbown, 2008; Sapada & Tomita, 2010). Given the overall effectiveness of FFI on

grammatical development (see meta-analyses by Norris & Ortega, 2000; Spada & Tomita, 2010), I consider it timely to explore the feasibility of FFI tasks that target pronunciation in meaning-oriented contexts and to assess their impact on L2 pronunciation development.

Measuring Pronunciation Development

Most of the previous pronunciation teaching studies cited above adopted human rating methods (i.e., asking NS listeners to rate NNS speech samples), and the validity of the method has been substantially confirmed as the golden standard to measure the quality of sentence-level speech samples in a holistic manner (i.e., listeners generally agree with their rating judgment on perceived accents with high inter-rater reliability; see Derwing & Munro, 2009). Whereas listening tasks can be the best fit for cross-sectional data, they might not, however, be the most appropriate outcome measure for time-series data (as in intervention studies) where a number of students need to be recruited and tested via various kinds of tasks (eliciting both controlled- and spontaneous-speech production) at several times (i.e., pre- and post-test sessions). Human rating methods arguably take time that risks causing listener fatigue and limits the number of participants and speech tokens within one study (e.g., Derwing et al., 1998 for 6 hours of listening; Elliot, 1997 for 6 hours of listening in total; Saito, 2011a for 3 hours).

As a reliable way to examine only acoustic properties of ample speech samples such as frequency values of formants, intensity, and duration at an individual word level, L2 phonology researchers tend to draw on an acoustic analysis (see Ladefoged, 2003); this kind of robust analysis enables researchers to measure change in the acoustic properties of L2 sounds between pre- and post-test sessions (Saito, 2007). One could argue, however, that it is unclear how such changes in acoustic properties can actually impact NS listeners' comprehension of L2 speech production (which is arguably the ultimate goal of pronunciation teaching). Thus, as optimal

outcome measures for pronunciation teaching studies of this kind, I propose one possible framework—a combination of an acoustic analysis and human rating method. That is, NS listeners are first recruited to rate a small subset of speech data randomly selected from the original data pool in order to find significant acoustic properties which positively influence NS listeners' rating scores. Second, acoustic analyses are conducted on the entire data set with a focus on these significant acoustic variables. The assumption here is that, given that some acoustic properties are significantly correlated to NS listeners' comprehension, changes in such crucial acoustic properties will either immediately or eventually enhance overall intelligibility of L2 speech production.

English /ɹ/

FFI in Study 1 targets one of the most well-researched cases of L2 speech sound learning—the acquisition of the English-specific sound /ɹ/ by Japanese learners of English (for a review, see Bradlow, 2008; Yamada, 1995). Due to the lack of any corresponding approximant sounds in the Japanese phonetic system (Japanese has only two approximants, /w/ and /j/), Japanese learners of English are predisposed to substituting the Japanese tap /ɾ/ for the English /ɹ/ and /l/ and thereby neutralize the contrast in their *production*, even after many years of residence in English-speaking countries (Larson-Hall, 2006). They also judge both English /ɹ/ and /l/ *perceptually* as poor exemplars of the Japanese tap /ɾ/ (Best & Strange, 1992; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003).

Importantly, Study 1 specifically focused on the acquisition of English /ɹ/ rather than /l/. A number of studies have demonstrated that English /ɹ/ is acquired more easily by Japanese learners, both in terms of perception and production, in contrast to English /l/ (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004; Bradlow et al., 1997; see also Flge, 1995, 2003,

2007 for theoretical arguments). The authors of these studies argue that the difference between /ɾ/ and /ɹ/ is more perceptually salient than /ɾ/ and /l/ to Japanese learners of English, who in turn have greater facility in learning the articulatory configuration for /ɹ/ (as distinct from the Japanese tap /ɾ/). NS listeners' perceptions of Japanese learners' speech sound production revealed a clear and significant improvement in learners' production of the sound /ɹ/ after both naturalistic and structured exposure to the language—in contrast, once again, to the sound /l/. Furthermore, recent perception research has revealed that, in Japanese learners of English, the phonemic representation of the Japanese flap overlaps with that for English /l/, whereas /ɹ/ is more easily distinguished from /ɾ/ (Hattori & Iverson, 2009; for a review, see Bradlow, 2008).

On the basis of these findings, I argue that Japanese learners of English may actually benefit from a focus on English /ɹ/ as an initial step in developing clear perceptual and articulatory representations of the /ɹ/-/l/ contrast, despite a noticeable accent which may otherwise affect their intelligibility. Doing so could provide a more efficient means of establishing an important sound contrast in English and help learners improve overall communicative success in their L2.

Pedagogical Importance of English /ɹ/

Some researchers emphasize the relative importance of suprasegmentals in successful L2 communication (e.g., Derwing et al., 1998; cf. Jenkins, 2000, 2002). From a pedagogical perspective, however, English /ɹ/ can be considered as a top priority especially for Japanese learners of English to improve overall intelligibility of their L2 speech production. In preparation for the proposed research, I (Saito, under revision) administered an "expert judgement" questionnaire (see Ellis, 2006; Robinson, 1996) to a sample of 120 teachers of English in Japan comprising both native and non-native speakers of English, asking them to rank 25 pronunciation

problems, which included a number of segmental problems (e.g., /ð/, /θ/, /f/, /v/, /æ/) as well as suprasegmental problems (e.g., lexical and sentential stress, speech rate, fluency). Results corroborated previous findings that English /ɹ/ is considered the most crucial teaching/learning target owing to its potential to affect the intelligible pronunciation of Japanese learners of English.

Furthermore, several pronunciation specialists also argue that the English /ɹ/ and /l/ contrast, which has a relatively large number of frequently occurring minimal pairs, needs to be considered as one of the top teaching/learning targets not only for Japanese learners of English but also other ESL/EFL students worldwide, because of its high functional load on listeners' comprehension[2] (see Munro & Derwing, 2006). Thus, examining this well-researched as well as pedagogically important topic is expected to benefit not only Japanese learners of English, insofar as the findings may be generalizable to other types of adult language learners (i.e., EFL learners in East Asia; ESL learners in North America) as well as different pronunciation features (i.e., segmentals vs. suprasegmentals).

Acoustic Properties of English /ɹ/

Previous L2 phonology studies have examined which acoustic properties (e.g., frequency values of the first formant [F1], second formant [F2], third formant [F3], and transitional duration of F3) determine NS listeners' categorical perceptions of English /ɹ/ and /l/ contrasts based on natural speech tokens (e.g., Espy-Wilson, 1992; Flege et al., 1995) as well as

[2] This hypothesizes that functional loads of English phonemic contrasts for listeners can be determined by some key factors such as (a) the frequency of minimally paired words, (b) the degree of hybridization between regional English dialects and (c) segmental positions within a word. For example, whereas the contrast of /l-ɹ/ in word initial positions distinguishes relatively many words (e.g., "lead" vs. "read" "lock" vs. "rock"), the contrast in word final positions (e.g., "wall" vs. "war" "tall" vs. "tore") remains unclear according to the regional varieties of English (e.g., the rhotic /ɹ/ is typical of General American but not Received Pronunciation).

synthesized samples (e.g., Hattori & Iverson, 2009; Iverson & Kuhl, 1996; Iverson et al., 2003). Their findings generally suggest that the acoustic difference between /ɹ/ and /l/ depends primarily on the frequency values of F3.[3] That is, NS listeners tend to perceive the sound as /ɹ/ when its F3 dips below 2000Hz and as /l/ when its F3 exceeds 2400 Hz or more (see also Ladefoged, 2003). In order to see if these findings (i.e., F3 as a primary phonetic cue) could be applicable to the Study 1 where speech tokens were naturally produced across different tasks with various ensuing vowel contexts, I included a rating session where five NS listeners evaluated a small subset of the data, allowing us to examine the extent to which the relevant acoustic properties varied according to task type and ensuing vowel contexts.

Current Study

I conducted a quasi-experimental study with a pre-test and post-test design to investigate the effects on L2 pronunciation of FT with and without CF (i.e., FT-only and FFI+CF). Based on previous L2 grammar studies, FT adopted in Study 1 refers to a set of comprehension and production tasks designed to develop participants' argumentative skills in English while drawing their attention to the target forms through: (a) structured input; (b) typographically enhanced input; and (c) output-focused tasks. CF involves pronunciation-focused recasts (i.e., partial, declarative type). In addition, I developed outcome measures through a combination of acoustic analysis and human ratings, and tested their validity by measuring the impact of instruction on the learners' L2 performance in various phonetic contexts not only at a controlled- but also a spontaneous-speech level.

[3] This line of research also showed that F2 and transition duration can be secondary phonetic cues (for details, see Hattori & Iverson, 2009)

Research Questions

The research questions to be addressed in Study 1 are as follows:

1. Which acoustic properties of /ɹ/ affect NS listeners' judgments and how do these properties vary relative to task type and to the backness and height of ensuing vowels?
2. Does FFI lead to improvement in learners' pronunciation of /ɹ/ and do the effects of FT vary according to whether or not learners also receive CF?

Method

Study 1 involved two experienced ESL teachers, five experienced NS listeners, and 65 adult intermediate-level Japanese learners of English in Montreal, Canada. The study comprised two phases. In the instructional phase, 65 learners were randomly divided into three groups (i.e., FT-only group, FT+CF group, and control group) and received four hours of meaning-based instruction about argumentative skills taught by two ESL teachers. In addition, the FT-only and FT+CF groups received instruction on the English /ɹ/ sound. In the assessment phase, a rating session first took place in which five NS listeners were asked to rate a small subset of speech samples randomly selected from the original data pool of speech tokens elicited at the pre-test sessions in order to ascertain which acoustic properties in English /ɹ/ (i.e., F1, F2, F3, transition duration) affected NS listeners' perceptional patterns according to task types and following vowel contexts. Subsequently, based on the results of a multiple regression analysis, an acoustic analysis was conducted on crucial speech properties of all of the speech data produced in the pre- and post-test sessions in order to see whether the learners showed any significant improvement in three tasks designed to tap different types of L2 speech production (i.e., controlled and automatic knowledge) as well as two following vowel contexts (English /ɹ/ followed by high/low vowels and front/back vowels).

Participants

Students

For the purpose of student recruitment, I created ads which advertised the four-hour free English argument project, specifying the proficiency levels required for participation (e.g., 450-700 for TOEIC scores, 50-80 for TOEFL iBT scores)[4]; our purpose was to recruit intermediate Japanese learners of English based on the assumption that they would still have problems producing /ɹ/. The ad was posted on several community websites specifically for Japanese people studying abroad, with hardcopy versions also distributed at many language institutes in Montreal. Interested participants contacted me through email or by phone, and set up a date for their first interview and pre-test sessions. The recruitment continued until the number of participants reached the maximum number, which had been set in advance at 72. However, because four participants did not complete the instructional treatment nor did they attended the post-test sessions, and three others were considered too advanced based on the pretest scores, there were a total of 65 participants included in the final analysis (age: $M = 29.7$, $SD = 6.9$).

During the first interview, a majority of participants reported that they attended either university-level English-speaking schools or private language institutes and had many opportunities to use English academically and socially on a daily basis. All of them had learned English for more than 10 years since their entrance to 7th grade in junior high school in Japan. Although most of the students had just arrived in Montreal, their length of residence (LOR) in Canada varied widely from one month to 13 years (LOR: $M = 15.5$ months, $SD = 31.8$ months).

[4] This proficiency test consists of listening, reading, and grammaticality judgment tasks, but without any speaking tasks. So, the test scores can indicate the learners' proficiency levels in listening skills as well as the amount of lexicogrammar knowledge, but they don't reflect their L2 pronunciation skills. Note that those who had taken a TOEIC test represented the upper range of proficiency. In fact, the remaining participants who did not report TOEIC scores confessed that they did not feel ready to take exams due to their lack of explicit grammatical knowledge and sufficient listening skills.

TOEIC scores were reported by 27 of the participants (M = 577, SD = 168.12). As a group, therefore, the participants were considered intermediate ESL learners. After the interviews, the 65 students were first randomly divided into 12 classes (six students per class), and then these classes were designated as one of three groups: (a) FFI-only group (five classes, n = 29), (b) FFI+CF group (five classes, n = 25) and (c) control group (n = 11). Table 3 provides the details of the 65 participants' information according to the three groups.

Six students per class might sound relatively small. However, especially in Asian EFL settings such as in Japan, China and Korea, the number of students in conversation English classes offered by private language institutes is typically between 6 and 10. For example, Sheen (2004, 2006) conducted her descriptive study about adult EFL learners' focus on form in private language institutes in Korea, and the number of students per class she observed was eight. Importantly, the population of these adult EFL learners (even including university-level students) is dramatically growing, especially because English is now being used as International Language and high levels English proficiency are necessary especially in business contexts. Thus, it is possible that the current study that simulated small L2 classrooms of these kinds has pedagogical implications particularly for adult EFL education common in private language institutes. Our assumption is that these findings can be ultimately applied to relatively bigger L2 classrooms in public school settings such as high school and University.

Instructors

The two ESL teachers were both female and had several years of L2 teaching experience, including ESL/EFL instruction, and had worked at private language institutes in Montreal prior to the time of the project. Both were certified teachers with undergraduate degrees in L2 teaching and both were completing MA degrees in L2 education. They were selected on the basis of their

professional and academic backgrounds and on their willingness and availability to participate. Both teachers followed training sessions, which will be described below. One instructor taught the first 6 classes (3 FT-only classes and 3 FT+CF classes), and the other taught the other 6 (2 FT-only classes, 2 FT+CF classes, 2 control classes).

Table 3. *Participant information by Group*

	FT+CF Group (n = 29)	FT-only Group (n = 25)	Control Group (n = 11)
Gender	8 males / 21 females	3 males / 22 females	2 males / 9 females
Age	M = 29.2 (SD = 6.0)	M = 29.7 (SD = 7.1)	M = 30.9 (SD = 9.1)
LOR (months)	M = 18.5 (SD = 31.5)	M = 11.5 (SD = 28.7)	M = 18.0 (SD = 40.8)
Age of Arrival	M = 28.4 (SD = 4.9)	M = 28.8 (SD = 7.2)	M = 29.3 (SD = 6.4)
TOEIC	M = 550.4 (SD = 135.6)	M = 630.4 (SD = 192.0)	M = 425.0 (SD = 176.7)

Listeners

Five native speakers of English (three males, two females) were recruited as NS listeners to rate the quality of 100 speech tokens randomly selected from the data pool of 1,430 speech tokens produced at the pre-test sessions. The five NS listeners participating in Study 1 were selected on the basis of two crucial variables: their L1 variety of English and their familiarity with Japanese-accented English speech. Thus, although all five were undergraduate students studying at an English-speaking university in Montreal at the time of the study, they were all originally from the US and spoke north-eastern American English as their L1. All of them took Japanese classes and reported having frequent contact with Japanese learners of English in

Montreal and being familiar with Japanese-accented English speech, including mispronunciation and unclear pronunciation of English /ɹ/. I thus considered them as "experienced" listeners (for the influence of accent familiarity on NS listeners' intelligibility judgment, see Bradlow & Bent, 2008; Kennedy & Trofimovich, 2008).

Procedure

Students in the experimental groups received four hours of FFI, which was designed to encourage them to notice and practice the target feature in the context of meaning-oriented instruction. Although those in the control group received comparable instruction (English argumentative skills), the target pronunciation feature of their FFI was different (English vowel sounds). The instructors gave CF only to students in the FT+CF group by recasting their mispronunciation or unclear pronunciation of /ɹ/ while no CF was directed at those in the FT-only group.

Each class consisted of four one-hour lessons and took place twice a week, finishing within two weeks (1 hour × 2 lessons per week × 2 weeks = 4 hours). The entire project (12 classes) took place over seven months between March and September. All of the classes were conducted in a classroom located on the campus of an English-speaking university in Montreal. All instructional treatments were videotaped (4 hours of instruction × 12 classes = 48 hours), while the researcher sat at the back of the classroom to ensure that the consistency of the instruction was maintained within groups by the two instructors. Two weeks after the 4 hours of instruction, the students individually completed post-tests as well as final interviews.[5] Figure 2 summarizes the design of the study and the procedures followed.

[5] The post-test sessions in the current study could be considered as "short-delayed post-tests" rather than "immediate post-tests" according to the FFI research standards in L2 grammar studies (e.g., Mackey &

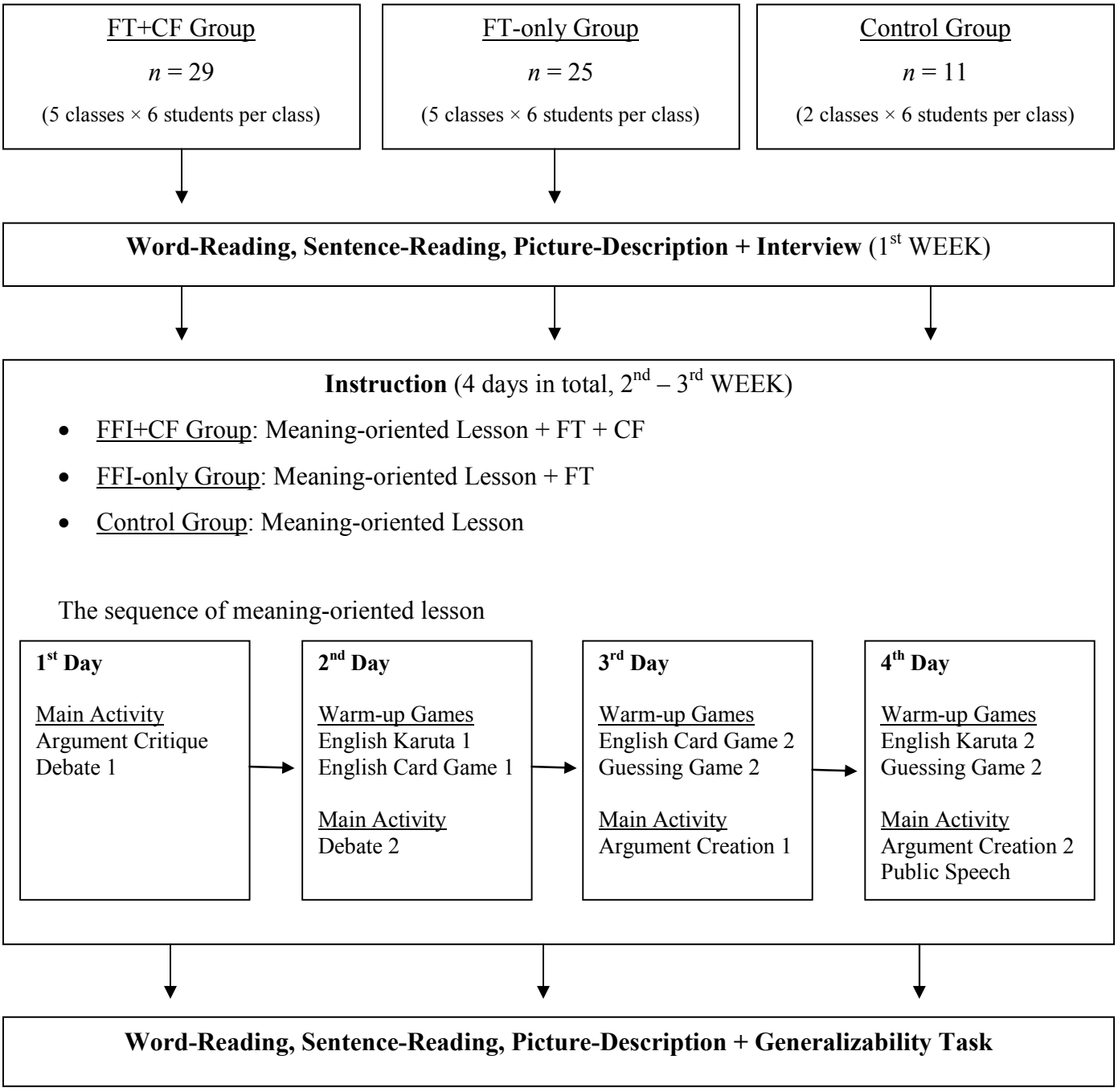


Figure 2. Summary of the procedure

Goo, 2007; Spada & Tomita, 2010). To our knowledge, none of the previous pronunciation teaching studies adopted delayed post-testing measures (see Derwing & Munro, 2005).

It is important to note here that, although the students were made aware of the main purpose of the project (i.e., English argumentative skills) at the first interview, the other focus of the project—the pronunciation-related FFI part—was purposely not explained to them at all until the introspective interview was conducted at the endpoint of the experiment This is because one of the purposes of the study was to investigate whether or not and in what ways FFI with or without CF induces students to *notice* the target feature and *practice* it during meaningful discourse without any explicit explanations.

FT Treatment

Thirty-eight minimally-paired words (including near minimally-paired words) were used as target words in the FFI treatment (see Table 4). Among these 38 words, English /ɹ/ appeared in various positions: 25 occurrences in word-initial positions, 3 in word-medial position, 10 in consonant clusters. All words appeared frequently in various tasks, and they were italicized and highlighted in red so that the learners could notice the target feature during meaning-oriented tasks. The 4-hr FFI treatment comprised four main activities known as the "English Argument Project," which was supplemented by a set of warm-up games to be played twice or at least once per lesson (for a summary, see Figure 2).

Integration of Pronunciation Targets into the English-Argumentative Skills Lessons

I created all of the instructional materials, which involved "developing English argumentative skills." This topic was chosen for two reasons. First, acquiring English argumentative skills is not only motivating but also necessary, especially for adult ESL/EFL learners who likely have high expectations for developing not only their communicative abilities in daily English conversations but also professional English proficiency to achieve academic and career-related goals. That is, in the near future, these highly motivated adult learners will tend to

use English in the real world in ways that involve cognitively demanding tasks such as making oral presentations, quickly responding to questions during formal interviews and meetings, and writing logical/concise reports and papers in business and academic contexts. Second, according to several cross-cultural educational reports, it has been

Table 4. *Target minimally-paired words*

/ɹ/ in word-initial positions		/ɹ/ in word-medial	/ɹ/ in consonant cluster
*race	*road	arrive	bread
*rain	*rock	correct	crab
*ram	rocket	pirate	crime
rate	Rome		crowds
*read	*room		fries
*red	round		fruit
reef	*row		grass
rent	*rule		green
*right	*run		free
rice	*Ryan		pray
ring	*wrong		
rink	wrap		
river			

Note. * denotes words included in outcome measures

shown that Japanese individuals use logic and argument in different ways than Westerners do (e.g., Oi, 2005) and teaching Western standards could be beneficial especially for Japanese learners of English (e.g., Toulmin, Rieke, & Janik, 1984). Given that the FT treatments on the

pronunciation targets were integrated in these activities, I describe them in some length in this section. Meaningful activities in the FFI materials included (a) critiquing arguments, (b) developing arguments, (c) debating topics, and (d) public speaking.

Critiquing arguments. Learners were first given several arguments that had some sort of logical problem related to examples, analogies, or causality. The instructors presented one argument after another, and asked the students to find the problems so that they could develop logical ways of critiquing each others' arguments. Each of the arguments used in this task included at least one target word, which was italicized and highlighted in red. One of the arguments included in the materials was as follows:

> "Whenever I eat Japanese ***rice*** I have digestion problems. So, eating Japanese ***rice*** causes digestion problems"

Here the causality relationship between eating rice and digestion problems is unclear unless the participant describes what happened when he or she did not eat Japanese rice. Note that "rice" was the target word here.

Debating topics. The students were divided into two groups and asked to debate certain topics. Each topic included at least one target word. The purpose of this task was to encourage students to use the rules of logic of which they had become aware in the previous activity (critiquing arguments) and to be more critical of others' arguments as well as their own opinions. When their responses to others conveyed logical reasoning, the students received one point and received an additional point if they added their own convincing argument. Total points were summed and the instructor decided which team won the debate. Among the topics used in this activity were "***Run***ning inside is better than ***run***ning outside?" and "Is it good to have a ***rain***y day?"

Developing arguments and public speaking. In these activities, the instructors first drew students' attention to a logical progression for making an argument (i.e., Introduction → Evidence → Objection → Defending → Conclusion) and illustrated how to use clear reasoning to develop a convincing argument. Next, the students were paired and asked to elaborate their own logical opinions about various topics. Finally, after sufficient practice, they were given one final topic and then asked to develop their own argument to present in front of the class. These topics included at least one target word. 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)?"

Warm-up Activities

For use as warm-up activities prior to the main activities, teachers implemented three communicatively-oriented games designed to create opportunities for participants to hear and to use target words: (a) English Karuta, (b) an English Card Game, and (c) a Guessing Game. These games were designed to promote the proceduralization of declarative knowledge by drawing on Lyster's (2007) proposal for form-focused instructional tasks that interweave opportunities for noticing, awareness, and practice.

English Karuta. In this noticing activity based on a traditional Japanese card game called *Karuta*, 36 cards were prepared, each representing a different word (minimally-paired contrasts for English /ɹ/ and /l/ such *lace* vs. *race* and *cloud* vs. *crowd*) via picture and the first letter of the word. First, the instructor randomly placed 36 cards on a table around which students were gathered. After the instructor read one of the words from a word list, the students tried to find the right card as soon as possible. When students chose the wrong card, they had to return it to its original place, but kept it if it was the right one. In order to obtain many cards, the students had

to listen carefully, without being pressured to produce any sounds and instead focusing on perceiving the instructors' pronunciation of the English /ɹ/ and /l/ contrast.

English card game. In this awareness activity, students were paired and given a set of 36 cards. Each card included two sentences which were identical except one word, as follows:

"Children are ***playing***" vs. "Children are ***praying***"

"My teacher will ***collect*** our papers" vs. "My teacher will ***correct*** our papers"

One of the two students was asked to read aloud one of the sentences; the other student was to guess which sentence his or her partner was reading. All of the words were minimally-paired for the English /ɹ/ and /l/ contrast. If they succeeded in their guess, they kept the card. If they failed, they had to put the card at the bottom of the pile. Because the game ended when all cards were gone, the instructors always encouraged participants to complete the game as quickly as possible and to compete with each other as a fun game. In order to get the correct card, the student needed to pay attention to correctly producing the English /ɹ/ in particular[6] so that his or her partner could distinguish English /ɹ/ from /l/ contrast and choose the right card. In this activity, the students were required to produce the intelligible pronunciation of English /ɹ/ but within controlled contexts. Importantly, given that both students in a pair were nonnative speakers of English, the main purpose of this activity was not to have students listen to correct pronunciation models but to become more aware of the acoustic difference in the English /ɹ/ and /l/ contrast.

Guessing game. This practice activity was based on a communicative riddle game used in Lyster's (2004a) L2 grammar study. In pairs, one learner was given a card on which a word was written and asked to explain what the word was without saying the name of it, while the other tried to guess the vocabulary item his or her partner was trying to describe. These words were

[6] Because English /l/ could be substituted by Japanese tap /ɾ/, intelligibly producing English /ɹ/ is crucial for Japanese learners of English to discriminate English /ɹ/ and /l/ contrast.

from 37 target words (e.g., *rock music, rent, red, road, room, ring*). The students were required to actually produce the target feature in somewhat complex task contexts (the students had to think about how to describe in addition to paying attention to the correct pronunciation). Teachers were encouraged to oversee each pair of students and asked to give CF.

FT+CF

The FT+CF group did the same activities involving the pronunciation targets as just described for the FFI group, plus CF. CF was operationalized as *pronunciation-focused recasts*, referred to as "partial recasts" by Sheen (2006). That is, the instructors were asked to recast only one word in which an error occurred, with falling intonation and without adding any additional meaning. Arguably, recasts of this kind are so explicit that their corrective function is likely to be quite obvious. Two examples of pronunciation-focused recasts used in Study 1 follow (S = students; T = teachers; asterisks denote mispronunciation or unclear pronunciation). Example 1 occurred during the main activity of developing arguments and Example 2 occurred during student-student interaction in the context of the guessing game.

Example 1:

S: Children spend too much time in read[lead]*ing...

T: Reading

S: Reading. Too much time in reading comic books rather than novels.

Example 2:

S1: car or people walk and drive on this…

S2: lane?

S1: no

S2: road [load]*?

T: road

S2: road

By contrast, the instructors in the FT-only group were encouraged to give feedback on (a) the content of the argument (e.g., "your arguments need more evidence, right?") and (b) vocabulary (e.g., "the word *sufficient* is the same as *enough*"), but not their pronunciation forms. I observed and counted the number of pronunciation-focused recasts that the instructors gave to their students during all classes and was thus able to ascertain their consistent use in one group and not the other.

Control Group

The students in the control group received four hours of comparable instruction (also on the topic of "developing English argumentative skills") but without form-focused instruction on the English /ɹ/. In the main activities targeting argumentative skills in English all of the target words were replaced by non-target words. For example, "running inside is better than running outside" became "jogging inside is better than jogging outside." In the warm-up activities, students in the control group played the same three games but with different target words featuring English vowel sounds, while the teacher, to the best of her ability, provided pronunciation-focused recasts.

Teacher Training

The instructional materials were sorted out and labeled as one package for each lesson (e.g., Day 1, Day 2). Each instructor received one hour of teacher training for each one-hour lesson. During training sessions, I explained to the instructor about the content and purpose of each FFI activity. In order to help the instructors understand how to deliver FFI and provide

pronunciation-focused recasts, I also demonstrated model lessons. Finally, the instructors were encouraged (a) to ask any questions until they clearly understood the intent of the research as well as (b) to practice the FFI materials with the researcher until they felt comfortable enough to teach.

Measures

In order to examine the impact of instruction on learners' oral production of English /ɹ/ from various perspectives, three tests targeting familiar items were administered as pre and post-test measures before and after the instructional treatment, and a fourth test targeting unfamiliar items was administered on only one occasion after the treatment. The testing sessions were completed in a quiet room in one-on-one meetings with me, a NS of Japanese. All of the communication and instruction about the procedure was always done in Japanese so that the learners never heard any model pronunciation of the target words from the researcher.[7] Their speech tokens were carefully recorded by means of speech analysis software, *Praat* (Boersma & Weenik, 2009; downloadable at www.praat.org), at 44.1 kHz sampling rate and a 16-bit resolution. A unidirectional microphone was used (DM-20SL) and all of the recordings were stored on the hard drive of a TOSHIBA Satellite U400 laptop computer.

Familiar Items

Three tests targeting 14 of the 38 words that had appeared in the instructional treatment were used to evaluate the learners' improvement in their pronunciation of English /ɹ/ between pre- and post-tests: a word-reading task, a sentence-reading task, and a timed picture-description task. The word- and sentence-reading tasks were designed to measure controlled performance,

[7] In addition, to avoid any lexical familiarity effect (i.e., L2 speech production is relatively better when learners are familiar with target words; see Flege, Takagi, & Mann, 1996; Flege, Frieda, Walley, & Randazza, 1998), all participants were asked if any words in the test materials were unfamiliar to them. All of the target words were found to be quite familiar to the learners.

whereas the picture-description task was designed to elicit learners' spontaneous speech production. The same tasks were used at both pre- and post-tests (hence the importance of a control group to assess possible test-retest effects).

Word Reading

Learners read 25 individual words which included 10 target words and 15 distracters. All target words were CVC singleton tokens beginning with /ɹ/ (i.e., *read, room, root, rule, red, race, rough, row, ram*, *right*).

Sentence Reading

In addition to four distractor sentences, participants read the following five sentences, in which eight target words were embedded:

He will read my paper by the time I arrive there.

She left her red bicycle on the side of the road.

The race was cancelled because of the rain.

I can correct all wrong sentences tonight.

Ryan does not like to run in the snow.

All target items were CVC singleton tokens (i.e., *read, red, road, race, rain, wrong*, *run*) with one exception of CVVC (i.e., *Ryan*).[8]

Timed Picture Description

Learners were asked to describe four pictures designed to elicit four CVC singleton tokens (i.e., *read, rain, rock, road*) as well as four distracter pictures that did not include any target words at all. Adjacent to each picture were three word cues to prompt learners to use the

[8] Given that learners' performance of minimally-paired words including English /ɹ/ in word medial positions was not the focus of the current study, the decision was made to exclude two minimally-paired words (i.e., "arrive" and "correct") from the original analysis.

target word while describing the scene. For example, a picture of a table left on a driveway in the rain was accompanied by three word cues (i.e., *table, driveway, rain*) and was used to elicit the target word *rain* in the participant's description. Participants were given only five seconds to prepare before being prompted by the researcher to begin their description so that they were required to perform the task under time pressure while their main focus was on meaning (see Ellis, 2005).

Whereas the 38 minimally-paired words targeted by the instruction included English /ɹ/ in various positions (word initial, word medial, consonant clusters), /ɹ/ occurred only in word-initial positions in the 14 test items. All but one were consonant-vowel-consonant (CVC) singletons (see asterisked words in Table 4). This decision was made because Japanese learners of English tend to have an especially difficult time in their perception and production of English /ɹ/ in word-initial positions compared to sylalble-final positions (Bradlow et al., 1997; Goto, 1971, Lively et al., 1993; Logan et al., 1991; Mochizuki, 1981; Sheldon & Strange, 1982) and because the assumption in Study 1 was that measuring the learners' performance of English /ɹ/ on relatively difficult positions (word-initial) would reveal their current proficiency levels.[9]

Twenty-two word-initial singleton tokens (out of 14 words, 8 were tested twice but in different tasks while the others appeared only once) were elicited per participant at pre- and post-test sessions respectively. The results were analyzed separately based on (a) three different tasks (n = 10 for word reading, n = 8 for sentence reading, and n = 4 for picture description); (b) following vowel backness (n = 10 for singletons with front vowels, n = 2 for singletons with central vowels, n = 10 for singletons with back vowels); and (c) following vowel height (n = 5

[9] Lively et al. (1993) actually showed that their tailored perceptual training, which focused on English /ɹ/ and /l/ occurring only in its most difficult position in words (pre-vocalic), improved the perception abilities of Japanese learners of English and confirmed their transferred effects to other relatively easy contexts such as English /ɹ/ and /l/ on post-vocalic positions.

for singletons with high vowels, *n* = 14 for singletons with mid vowels, *n* = 3 for singletons with low vowels).

According to previous research, Japanese learners of English display more difficulty in perceiving English /ɹ/ preceding rounded vowels (i.e., /o/, /u/) than any other context (e.g., Hardison, 2003; Ingram & Park, 1998; Sheldon & Strange, 1982). Study 1 further pursued this topic by investigating differential effects of training based on various following vowel contexts: (a) backness (front, central, back vowels) and (b) height (high, mid, low).

Unfamiliar Items

At the post-test only, a generalizability task was administered to 60 participants to investigate whether or not they could generalize their newly acquired knowledge of English /ɹ/ to unfamiliar items. Participants were asked to read a list of four non-minimally-paired words that had not appeared during the instructional treatment (i.e., *real, roll, rumor, regular*) along with four distracter items. The results were analyzed to detect any between-group differences according to vowel contexts (i.e., backness and height of vowels immediately following /ɹ/).

Rating and Judgment Sessions

Given the relatively large number of participants in Study 1 (*n* = 65), asking NS listeners to rate a huge number of speech samples (*n* = 3,100 tokens) could cause listeners' fatigue, which in turn would force me to limit either the number of test tokens or the number of participants. Alternatively, in order to achieve a precise description of the nature of L2 speech development under *instructed* conditions, Study 1 adopted a combination of human rating methods and acoustic analysis.

One hundred speech tokens were randomly chosen from the original data pool of the learners' performance in the pre-test sessions and then presented to the five NS listeners to rate.

To select the 100 speech tokens, 20 learners were first chosen randomly (4 participants from the control group, 7 from the FFI-only group, and 9 from the FFI+CF group) and each of them contributed 5 words (2 from the word-reading task; 2 from the sentence-reading task; 1 from the picture-description task). These words were carefully edited from the original sound files by me by means of the sound recognition software (*Praat*) (Boersma & Weenik, 2009). In the case of all 100 tokens—especially those extracted from the sentence-reading and picture-description tasks—every effort was made to isolate the tokens by following the procedure described next.

First, a spectrographic representation of each word was displayed on the computer screen using *Praat*. Then I listened to each token several times and tried to locate the beginning and end of the word without including any trace of the preceding and following sounds. As a reliable clue, the starting point of English /ɹ/ was identified by the endpoint of the gradually falling transition of F3. The F3 of the preceding vowel and consonant sounds tend to gradually fall, because the F3 for English /ɹ/ is relatively low (e.g., Bradlow, 2008; Hattori & Iverson, 2009; Iverson et al., 2003; see for details of acoustic properties of English /ɹ/ Espy-Wilson, 1992; Flege et al., 1995; Ladefoged, 2003). At the onset of the target word, the researcher placed a cursor, which he then moved to the end of the word in order to cut and paste it into a separate sound file. In this manner, all 100 tokens were prepared and put on one data CD to be used in the rating session.

The rating sessions took place individually with each listener in a quiet room and lasted about one hour. First, after a briefing about the purpose of the research project, each NS listener completed a training session with five speech samples. After the training, the NS listener was presented 100 speech tokens in a randomized order and asked to rate them on a 9-point scale, with 1 as "very good English /ɹ/" and 9 as "very poor English /ɹ/," using the whole scale as much as possible. Following this was an intelligibility judgment task in which the listener was asked to

decide whether the speech token sounded like the English /ɹ/ or the English /l/, again following a brief training session. The NS listeners were allowed to listen to each speech token as many times as they wanted until they felt confident about their ratings and judgments.[10] In addition, I was always next to the listeners during their sessions so that he could answer any questions they had regarding the rating procedure and ensure consistency in the rating and judgment procedures.

Acoustic Analysis

To conduct the acoustic analysis of English /ɹ/ across various contexts, I adopted the procedures used by Flege et al. (1995) in their analysis of the /ɹ/ and /l/ contrast produced in various tasks by Japanese learners of English. Accordingly, I measured F1, F2 and F3 values in hertz (Hz) and transitional duration of F3 in milliseconds (ms) for all speech tokens. He also categorized speech tokens based on task types (word reading, sentence reading, picture description), and also according to the height (high, mid, low) and backness (front, central, back) of subsequent vowels.

As for F1, F2, and F3, as described earlier, the beginnings of the English /ɹ/ were first carefully identified by the endpoint of falling F3, and then a cursor was put on the location where energy was clear for all three formants and F3 was starting to rise (i.e., the end of the steady state). The procedure is visually summarized in Figure 3. As for the transitional duration of F3, the researcher first put a cursor on the starting point of rising F3 and moved it to the beginning of the following vowels (measured in ms).

[10] Although conducting listening sessions in this way does not correspond to a real life situation where listeners have only one opportunity to listen and understand interlocutors, note that it would have been otherwise very difficult for listeners to rate only the quality of English /ɹ/ rather than basing their judgment on the whole word. In fact, all of the raters reported that the talkers tended to mispronounce not only English /ɹ/ but also many other segmentals such as /æ/ in *ram.*

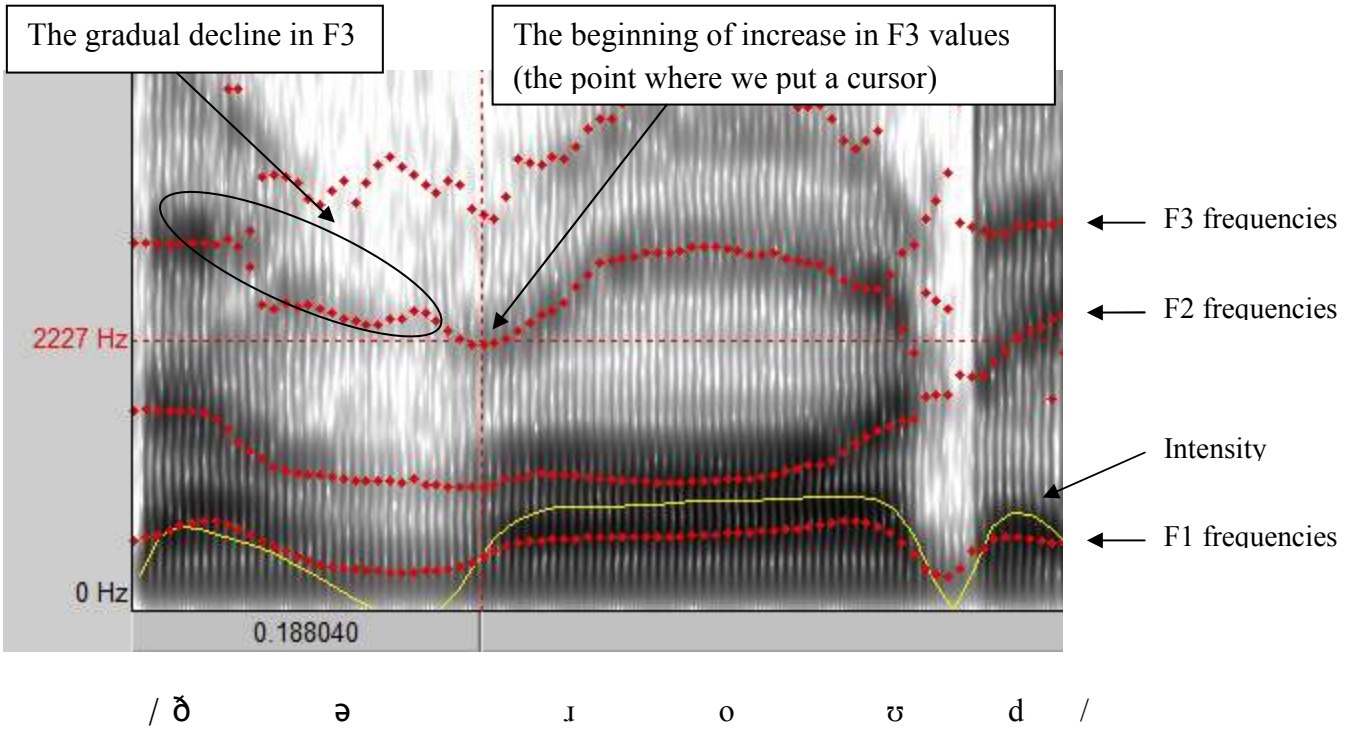


Figure 3. Summary of the acoustic analyses

Finally, to determine which speech and non-speech factors predicted listeners' rating and judgment patterns, two multiple regression analyses were conducted on the rating and judgment scores with two types of listening tasks as dependent variables (i.e., 9-point scale and intelligibility judgment) and seven predictors as independent variables (i.e., F1, F2, F3 transition duration, subsequent vowel height, subsequent vowel backness, and task types). After identifying crucial acoustic properties as significant predictors of listeners' 9-point scores and intelligibility scores alike, we moved on to the acoustic analyses of the remaining speech tokens (n = 3,100) in

order to examine the learners' improvement in conjunction with relevant acoustic properties between the pre- and post-test sessions.

For comparison purposes, I asked the five NS listeners to do the same four tasks (i.e., word reading, sentence reading, picture description, and generalizability task) so that I could obtain baseline data conveying NS production of the four acoustic properties (F1, F2, F3, and transition duration). They contributed a total of 130 speech tokens (5 NS talkers × 26 tokens = 130 tokens).

Results

I will first present (a) which acoustic properties were significantly correlated to listeners' rating scores, and then further examine (b) to what extent the learners improved in their pronunciation of familiar and unfamiliar tokens.

Analysis of Acoustic Properties and Listener Ratings

Because the interclass correlation among the five NS listeners on the 9-point rating scale proved strongly correlated, $r = .78$, $p < .0001$, their scores were averaged for each speech token. To identify which factors predicted their rating scores, a multiple regression analysis was conducted on the listeners' average scores as dependent variables and seven predictor factors as independent variables: (a) F1, (b) F2, (c) F3, (d) transition duration, (e) subsequent vowel height, (f) subsequent vowel backness, and (g) task types. Although the original model proved significant, $F(7, 92) = 12.593$, $p < .0001$, colliniarity statistics found that both F2 and F3 exhibited relatively high variation inflation factor (*VIF*) values of more than 2.0 (*VIF* = 2.26 and 2.01, respectively). In fact, these two factors were highly correlated with one another ($r = .64$, $p < .0001$). Following previous research findings whereby F3 is a *primary* phonetic cue for NS listeners, only the F3 factor was taken into account; the second multiple regression analysis was

conducted with six predictor factors excluding F2: (a) F1, (b) F3, (c) transition duration, (d) subsequent vowel height, (e) subsequent vowel backness, and (f) task types. This second model proved to be significant, F (6, 93), p <. 0001, revealing only F3 as a significant predictor factor, t = 6.269, p < .0001.

With respect to listeners' intelligibility judgment, the interclass correlation was also strongly correlated, r = .72 p < .0001. Therefore, a multiple regression analysis was again conducted on their average scores as a dependent variable and seven predictor factors as independent variables (F1, F2, F3, transition duration, vowel height, vowel backness, task types), revealing colliniarity problems based on the *VIF* values of F2 and F3 (*VIF* = 2.26 and 2.01, respectively). Consequently, a second multiple regression analysis was conducted with six independent variables excluding F2: (a) F1, (b) F3, (c) transition duration, (d) vowel height, (e) vowel backness and (f) task types. The results showed that the model was significant, F (6, 93) = 9.775 p < .0001, revealing only F3 as a significant predictor factor, t = 6.09, p < .0001.

In their goodness of English ratings on a 9-point scale, the listeners perceived F3 values around 2230 Hz as very good English /ɹ/ (1 < x < 3), F3 values around 2363 Hz as hybrid English /ɹ/ (4 < x < 6), and F3 values around 2780 Hz as very poor English /ɹ/. Similarly, in their intelligibility judgments (i.e., the speech token sounds like /ɹ/ or like /l/), all 5 listeners judged sounds with F3 values around 2270 Hz as definitely /ɹ/ and sounds with F3 values at around 2800 as definitely /l/. Sounds with F3 values around 2400 and 2600 were judged as either English /ɹ/ or English /l/. Details of these results are displayed in Table 5.

Following the finding that F3 frequency values were the only significant predictor of NS listeners' perception of English /ɹ/, I proceeded to conduct acoustic analyses on the rest of the entire data focusing on F3 values, in order to investigate whether or not and to what degree the

learners' F3 values changed as a result of the FT treatment. In order to interpret these results, I first considered any reduction in F3 values as *improvement* (positive impacts on NS listeners' perception), and then examined the degree of improvement by measuring the amount of decline between pre- and post-test sessions (e.g., a transition of F3 values from 2600 Hz to 2200 Hz suggests that the learners' unclear pronunciation of English /ɹ/ becomes more intelligible).

Table 5. *Average F3 values of the 100 speech tokens rated by 5 NS listeners*

Goodness of English (9-point scale)	*n*	*M*	*SD*
Very good /ɹ/ (1 < x < 3)	32	2230 Hz	354 Hz
Hybrid /ɹ/ (4 < x < 6)	30	2363 Hz	370 Hz
Very poor /ɹ/ (7 < x < 9)	38	2780 Hz	383 Hz
Intelligibility judgment	*n*	*M*	*SD*
Judged as /ɹ/ by 5 listeners	50	2270 Hz	382 Hz
Judged as /ɹ/ by 4 listeners	10	2416 Hz	359 Hz
Judged as /ɹ/ by 3 listeners	5	2611 Hz	295 Hz
Judged as /ɹ/ by 2 listeners	4	2765 Hz	199 Hz
Judged as /ɹ/ by 1 listener	9	2546 Hz	342 Hz
Judged as /ɹ/ by no listeners	22	2804 Hz	248 Hz

Word-Reading, Sentence-Reading, and Picture-Description Tasks

The 65 learners each produced 44 tokens (22 at each teaching session) for a grand total of 2,860 tokens. First, as for the pre-test scores (n = 22) and post-test scores (n = 22), I averaged

across each participant's F3 values according to (a) task types (word reading, sentence reading, and timed picture description), (b) following vowel backness (singletons with front vowels, central vowels and back vowels) and (c) following vowel height (singletons with high vowels, mid vowels, and low vowels) respectively. Second, for each contextual factor (task type, following vowel backness and height), I conducted separate three-factor ANOVAs in order to identify statistically significant differences between pre- and post-test sessions (within-group comparison) and among the three groups at the post-test sessions (between-group comparison). An alpha level was set at a $p < .05$ level for all statistical analyses. Cohen's *d* was also calculated in order to measure the magnitude of instructional effectiveness between two contrast groups of means.[11]

Task Types

To assess effects of instruction on learners' speech production of English /ɹ/ both at controlled and spontaneous speech levels, three-factor ANOVA was conducted: Group (FFI+CF, FFI-only, Control) × Task (word reading, sentence reading, timed picture description) × Time (pre-/post-tests). The ANOVA results revealed significant effects for the overall Group × Task × Time interaction, F (4, 124) = 2.635, p = .0373. A simple main effect of Time was found significant only for the FFI+CF group at all contexts: (a) word reading (M = 2511 → 2339 Hz), F (1, 186) = 15.647, $p < .00001$, d = 0.59; (b) sentence reading (M = 2542 → 2342 Hz), F (1, 186) = 21.030, $p < .00001$, d = 0.76; and (c) timed picture description (M = 2622 → 2385 Hz), F (1, 186) = 29.676, $p < .00001$, d = 0.81. In addition, a simple main effect for Group was significant for the timed picture-description task at post-test sessions. A Tukey test was

[11] According to Cohen (1988) effect sizes are roughly classified as small (0. 20 ≤ d < 0.50), medium (0.50 ≤ d < 0.80), or large (0.80 ≤ d). In all cases, whereas control group means were used to calculate between-group contrasts, pre-/post-test scores were used to calculate within-group contrasts.

conducted as a post hoc analysis with its alpha level set at a $p < .05$ level, revealing that only the FT+CF group ($M = 2385$ Hz) exhibited significantly lower F3 values than the control group ($M = 2695$ Hz) on the timed picture-description task with large effects ($d = 1.14$). Their F3 values generally declined from 2600-2700 Hz (confusing English /ɹ/) to 2300 Hz (clear English /ɹ/), which indicates significant effects of FT with CF on their L2 speech production of English /ɹ/. Interestingly, the ANOVA results also found main effects for Task, F (2, 124) = 12.245, $p < .000001$. According to a Tukey test, F3 values significantly differed at a $p < .001$ level in the following order: word reading ($M = 2451$ Hz) < sentence reading ($M = 2495$ Hz) < timed picture description ($M = 2547.45$ Hz).

According to the NS baseline data, the mean of F3 values was 1648 Hz ($SD = 212$ Hz) for word reading task, 1677 Hz ($SD = 226$ Hz), and 1692 Hz ($SD = 221$ Hz) for timed picture description. Note that the NS F3 values (around 1700 Hz) differed substantially from those produced by Japanese learners of English (around 2500 Hz). Not surprisingly, different from those of Japanese learners of English, the F3 values of the NS baseline data did not significantly differ across the three tasks, F (2, 8) = 3.101, $p = .100$, (i.e., NS talkers produced English /ɹ/ with little variance regardless of different tasks).

Following Vowel Backness

To examine how instruction effectiveness varied according to types of tokens (English /ɹ/ preceding front, central, and back vowels), three-factor ANOVA was conducted: Group (FT+CF, FT-only, Control) × Backness (singletons with front vowels, central vowels, and back vowels) × Time (pre-/post-tests). The ANOVA results revealed significant effects for the overall Group × Time interaction, F (2, 124) = 7.337, $p = .00014$. A simple main effect for Time was found significant only for the FT+CF group ($M = 2538 \rightarrow 2321$ Hz), F (1, 62) = 31.090, $p < .00001$, d

= 0.63. The decline in their F3 values can be interpreted as their improvement from confusing English /ɹ/ to clear English /ɹ/. Furthermore, an overall main effect for Backness was also identified as significant, F (2, 124) = 37.495, p = .00014. According to a Tukey test, F3 values significantly differed between tokens with back and central vowels (M = 2426 Hz and 2427 Hz respectively) and front vowels (M = 2554 Hz).

The results of the NS baseline data showed that the mean of their F3 values was 1717 Hz (199 Hz) for singletons with front vowels, 1640 Hz (189 Hz) for singletons with central vowels, and 1661 Hz (174 Hz) for singletons with back vowels. Although their F3 values (around 1700 Hz) were substantially different from those of Japanese learners of English (around 2500 Hz), they did not significantly differ according to following vowel backness, F (2, 14) = 2.801, p = 0.134.

Following Vowel Height

To examine how instruction effectiveness varied according to types of tokens (English /ɹ/ preceding high, mid and low vowels), three-factor ANOVA was conducted: Group (FFI+CF, FFI-only, Control) × Height (singletons with high vowels, mid vowels, and low vowels) × Time (pre-/post-tests). The ANOVA results revealed significant effects for the overall Group × Time interaction, F (2, 124) = 8.472, p = .0006. A simple main effect of Time was significant only for the FFI+CF group (M = 2554 → 2355 Hz), F (1, 62) = 30.596, p < .00001, d = 0.81. A simple main effect of Group also proved significant at the time of post-tests, F (2, 1224) = 3.444, p < .05. A Tukey test further showed that the FFI+CF group (M = 2329 Hz) outperformed the control group (M = 2529 Hz) with large effects (d = 0.93). The change in their F3 values (2500-2600 Hz → 2300 Hz) could be considered as evidence that their unclear pronunciation of English /ɹ/ became less ambiguous. An overall main effect for Height was identified as

significant, F (2, 124) = 20.581, p < .00001. According to a Tukey test, F3 values significantly differed according to following vowel height in the following manner: mid vowels (M = 2455 Hz) < low vowels (M = 2496 Hz) < high vowels (M = 2558 Hz).

The mean of the NS F3 values was 1682 Hz (SD = 183 Hz) for singletons with high vowels, 1630 Hz (SD = 234 Hz) for singletons with mid vowels, and 1646 Hz (SD = 184 Hz) for singletons with low vowels. As was the case with the other contexts, the mean of their F3 values (around 1700 Hz) was greatly different from that of Japanese learners of English (around 2500 Hz) and did not significantly differ according to following vowel height, F = (2, 14), p = 0.438. The results of pre- and post-tests are summarized in Table 6.

Generalizability Task

In order to assess participants' pronunciation of English /ɹ/ in words that had not appeared during the instructional treatment, I conducted a separate two-way ANOVA on the group means of the F3 values yielded by the generalizability task according to two contextual factors: (a) following vowel backness (singletons with front and back vowels) and (b) following vowel height (singletons with high and mid vowels).

Following Vowel Backness

Although the two-way ANOVA (Group × Backness) found an overall main effect for Backness, F (1, 57) = 36.704, p < .00001, it did not reach statistical significance for the overall Group effect, F (2, 57) = .034, p = .362. Singletons with back vowels (M = 2247 Hz) exhibited

Table 6. Summary of results of pre- and post-tests targeting familiar items

1. Group × Task × Time

Contextual factors (F3 values)

- Word reading (*M* = 2451Hz) < sentence reading (*M* = 2495 Hz) < timed picture description (*M* = 2695 Hz)

Between-group difference (F3 values)

- For timed picture description, FFI+CF (*M* = 2385 Hz) < Control (*M* = 2695 Hz)

Within-group differences

- For word reading: FFI+CF at post-tests (*M* = 2339 Hz) < FFI+CF at pre-tests (*M* = 2511 Hz)
- For sentence reading: FFI+CF at post-tests (*M* = 2342 Hz) < FFI+CF at pre-tests (*M* = 2542 Hz)
- For timed picture description: FFI+CF at post-tests (*M* = 2385 Hz) < FFI+CF at pre-tests (*M* = 2622 Hz)

NS baseline

- No significant difference between tasks (*M* = 1648 Hz for word reading; *M* = 1677 Hz for sentence reading; *M* = 1692 Hz for timed picture description)

2. Group × Backness × Time

Contextual Factors (F3 values)

- Low and mid vowels (*M* = 2426 Hz and 2427 Hz respectively) < front vowels (*M* = 2554 Hz)

Within-group difference (F3 values)

- For all contexts: FFI+CF at post-tests (*M* = 2355 Hz) < FFI+CF at pre-tests (*M* = 2544 Hz)

NS baseline

- No significant difference between tasks (*M* = 1717 Hz for front vowels; *M* = 1640 Hz for mid vowels; *M* = 1661 Hz for back vowels)

3. Group × Height × Time

Contextual Factors (F3 values)

- Mid vowels (*M* = 2455 Hz) < low vowels (*M* = 2496 Hz) < high vowels (*M* = 2558 Hz)

Between-group difference (F3 values)

- For all contexts: FFI+CF (*M* =2329) < Control (*M* = 2529 Hz)

Within-group difference (F3 values)

- For all contexts: FFI+CF at post-test (*M* = 2355 Hz) < FFI+CF at pre-tests (*M* = 2554 Hz)

NS baseline

- No significant difference between tasks (*M* = 1682 Hz for high vowels; *M* = 1630 Hz for mid vowels; *M* = 1646 Hz for low vowels)

significantly lower F3 values than those with front vowels (M = 2401 Hz), but there existed no significant difference between F3 values of the FFI+CF group (M = 2289 Hz), the FT-only group (M = 2313 Hz), and the control group (M = 2434 Hz). In spite of the lack of statistical significance, effect size analyses showed that both the FT-only and FT+CF groups began to exhibit small-to-medium effects compared to the control group (d = 0.45 for FFI-only group, d = 0.72 for FT+CF group) in their pronunciation of English /ɹ/ preceding front vowels (where F3 values are relatively high). The results of the NS baseline data exhibited no significant difference in mean F3 values for front vowels (1754 Hz) and back vowels (1635 Hz), F (1, 4) = 3.989, p = 0.1185.

Following Vowel Height

The two-way ANOVA (Group × Height) indentified an overall main effect for Height, F (1, 57) = 19.330, p < .00001, but not for Group, F (2, 57) = 1.035, p < .3617. That is, the F3 values of all three groups were significantly different between high vowels (M = 2392 Hz) and mid vowels (M = 2250 Hz). Nevertheless, small-to-medium effects were found for both the FFI-only group (d = 0.33 for high vowels and 0.38 for mid vowels) and the FT+CF group (d = 0.51 for high vowels and 0.47 for mid vowels). The results of the NS baseline data exhibited no significant difference in the mean F3 values for high vowels (1658 Hz) and for mid vowels (1631 Hz), F (1, 4) = 0.241, p = 0.6495. The results of generalizability task are summarized in Table 7.

Personal Interview

At the onset and endpoint of the project, the learners were interviewed by the researcher in face-to-face meetings. Among several questions asked that were not the focus of this paper, one was highly relevant to Study 1 and will thus be reported here. Namely, after they finished the

Table 7. *Summary of results of Generalizability Task targeting unfamiliar items*

1. Group × Backness × Time

Contextual Factors (F3 values)

- Low vowels (M = 2247 Hz) < front vowels (M = 2401 Hz)

Effect size analysis (compared to control group)

- FFI-only (small-to-medium effects)
- FFI+CF (small-to-medium effects)

NS baseline

- No significant difference between vowel backness (M = 1635 Hz for back vowels; M = 1754 Hz for front vowels)

2. Group × Height × Time

Contextual Factors (F3 values)

- Mid vowels (M = 2250 Hz) < high vowels (M = 2392 Hz)

Effect size analysis (compared to control group)

- FFI-only (small-to-medium effects)
- FFI+CF (small-to-medium effects)

NS baseline

- No significant difference between vowel height (M = 1658 Hz for high vowels; M = 1631 Hz for mid vowels)

post-test sessions, the learners were asked what they had learned the most from the four-hour instruction. Out of 65 learners who completed the project, 63 learners reported "English argumentative skills" such as debating and public speaking skills as their *primary* concerns and "the importance of an English /ɹ/ and /l/ contrast" as their *secondary* concern. Only two learners (both from the FT+CF group) reported that their focus was always on form (i.e., English /ɹ/) because the content of the lesson (English argumentative skills) was beyond their English proficiency.

Discussion

As for the first research question, which asked which acoustic property of /ɹ/ affect NS listeners' judgments and whether they vary according to task type, following vowel backness, and following vowel height, Study 1 identified F3 values for English /ɹ/ as the most crucial speech properties but with some variance according to task type and differences in vowel backness and height). This finding resulted from outcome measures that combined human rating and acoustic analysis. As for the second research question, which asked whether FFI improves learners' pronunciation of /ɹ/ and whether its effects increase through provision of CF, results revealed that learners receiving FT without CF did not show any significant change in their F3 values whereas those receiving FT in conjunction with CF generally decreased their F3 values from 2600-2500 Hz (unclear English /ɹ/) to 2200-2300 Hz (clear English /ɹ/), which in turn suggests a significant improvement in their pronunciation of English /ɹ/ as a result of the FT treatment with CF. A detailed discussion of the results is now presented.

Acoustic Properties of English /ɹ/

Based on 100 speech tokens randomly selected from the 1,430 speech tokens produced at the pre-test sessions, a multiple regression analysis confirmed that, among various independent

variables, only F3 values were a significant predictor factor for NS listeners' rating scores (for similar results, see Espy-Wilson, 1992; Flege et al., 1995; Hattori & Iverson, 2009; Iverson et al., 2003; Ladefoged, 2003). Post hoc analyses revealed several patterns as to the relationship between NS listeners' perception of English /ɹ/ and F3 values: (a) Speech tokens with F3 values around 2200-2300 Hz tend to be considered as both "good-enough" exemplars of English /ɹ/ and definitely English /ɹ/ rather than English /l/; (b) those with F3 values around 2400-2600 Hz proved to be confusing to NS listeners, who judged them as either English /ɹ/ or English /l/; and (c) speech tokens with F3 values above 2600 Hz were judged either as poor exemplars of English /ɹ/ or definitely as English /l/. Interestingly, learners' F3 values varied significantly according to task type, following vowel backness, and following vowel height, whereas those of the NS talkers did not show any significant variance associated with contextual factors.

Important to emphasize is that the listeners' judgment in Study 2 (i.e., speech tokens with F3 values around 2200-2300 Hz are close enough to English /ɹ/) is relatively lenient rather than strict, because their judgment of good exemplars (F3 values around 2200-2300 Hz) is still significantly different from the NS baseline data with F3 values at around 1600-1700 Hz. In other words, not only did the current study identify F3 as a crucial speech property which significantly influences listeners' judgment of English /ɹ/, but also set realistic goals for L2 learners in terms of intelligible pronunciation of English /ɹ/ (F3 around 2200-2300 Hz) rather than nativelike pronunciation of English /ɹ/ (F3 around 1600-1700).

Task Types

The results showed that the Japanese learners of English in the current study tended to produce significantly higher F3 values (less nativelike production of English /ɹ/) when cognitive demands increased according to the three different task types: word reading (M = 2451 Hz) <

sentence reading (*M* = 2495 Hz) < timed picture description (*M* = 2695 Hz). Although the learners exhibited the highest F3 values (unclear English /ɹ/) in the timed picture-description task, one could argue that the latter might not be the most appropriate way to tap learners' spontaneous speech production of English /ɹ/ (i.e., the learners were asked to describe pictures, but they were provided orthographic representations of target words as cues they could then read). Thus, we call for future research that further develops robust elicitation techniques appropriate for assessing spontaneous speech production (see Piske et al., 2001).

Following Vowel Backness and Height

The learners had difficulties in producing English /ɹ/ preceding high front vowels /i, ɪ, ɛ, æ/, demonstrating relatively high F3 values for front vowels (*M* = 2554 Hz) and high vowels (*M* = 2558 Hz). Intriguingly, previous research found that it is most difficult for Japanese learners of English to *perceive* the English /ɹ/ and /l/ contrast preceding round vowels such as /u/ and /o/ (e.g., Hardison, 2003). Taken together, it might be the case that learners' processing of the target language is different in production (i.e., English /ɹ/ preceding high front vowels is the most difficult instance for production) and perception (English /ɹ/ preceding round back vowels is the most difficult instance for perception), which might in turn contribute to the asymmetry in production and perception abilities of Japanese learners of English for the /ɹ/ and /l/ contrast (i.e., some learners produce English /ɹ/ better than perceive it, and vice versa; Goto, 1972; Mochizuki, 1981; Sheldon & Strange, 1992). Again, more research is needed to further pursue this topic.

Other Phonetic Cues

The findings about F3 as a primary phonetic cue for the English /ɹ/ and /l/ contrast in Study 1 support previous research findings in experimental phonetics and L2 phonology studies (Espy-Wilson, 1992; Flege et al., 1995; Hattori & Iverson, 2009; Iverson et al., 2003; Ladefoged,

2003). Noteworthy is that several studies found not only F3 but also other speech properties as significant predictors for NS listeners' perceptional patterns for the English /ɹ/ and /l/ contrast (e.g., Polka & Strange, 1985 for F1 transition length). Although I acknowledge the possibility of considering other speech properties such as F1 transition duration as independent variables, the types of speech tokens and the listening procedure adopted in Study 1 were substantially different from previous studies, which might have influenced our finding that only F3 values were significant predictors. First, compared to synthetic speech samples typically used in previous studies, Study 1 used natural speech tokens, which were less systematically controlled, and so the listeners might have failed to capture the subtle influence of so-called secondary phonetic cues such as F1 transition. Second, I carefully recruited only experienced NS listeners who had familiarity with Japanese accented English /ɹ/ and /l/, which might have influenced their tendency to judge even relatively high F3 as English /ɹ/. However, I still need to wait for more L2 speech research to be done, further investigating how NS listeners use various phonetic cues (e.g., F1, F2, F3, transition duration) to discern the English /ɹ/ and /l/ contrast in the context of *natural* speech tokens as well as synthesized speech.

FT with and without CF

In Study 1, whereas the frequency and saliency of the target form was enhanced through FT that drew attention to the target pronunciation form of English /ɹ/ in the context of meaning-oriented tasks, participants in the FT+CF group were also given pronunciation-focused recasts by their teachers in response to their mispronunciation and unclear pronunciation. The effects of instruction will be discussed based on the results of tests targeting familiar items and unfamiliar items, respectively.

When FT was combined with CF, the learners' F3 values generally changed on familiar items from 2500-2600 Hz to 2200-2300 Hz between pre- and post-test sessions in all contexts (task types, following vowel backness, and following vowel height). The FT+CF group also outperformed the control group on familiar items in the picture-description task. Taken together, I interpret these results as a significant improvement in their pronunciation of English /ɹ/, which suggests that FT in conjunction with CF might be an effective and efficient way to promote not only L2 grammar but also L2 pronunciation development, at both a controlled- and spontaneous-speech level, irrespective of vowel contexts. Thus, given that previous pronunciation teaching studies, drawing for the most part on decontextualized instructional methods, influenced students' performance only at a controlled-speech level, the findings of Study 1 reveal the benefits of communicatively oriented pronunciation teaching. In order to impact their communicative competence, it might be the case that learners need to process form jointly with meaning and practice it repetitively in communicatively authentic contexts (Celce-Murcia et al., 1996; Pennington, 1996; Trofimovich & Gatbonton, 2006; Segalowitz, 2003).

In contrast, the results revealed that the FT-only group did not demonstrate any significant F3 decline in any context. Although the FT activities were designed to promote learners' noticing and awareness of the target pronunciation form (through structured input, typographically enhanced input, and focused tasks), they proved insufficient on their own, without CF, to significantly impact learners' developing L2 system. These results point up one crucial theoretical issue in SLA—the role of both positive and negative evidence in L2 pronunciation development. Namely, L2 pronunciation development might require not only *enhanced positive evidence* (i.e., FT) but also *immediate negative evidence* from the teachers (i.e., CF). Even after they begin to gain awareness of the target pronunciation feature of English /ɹ/ as

well as intensive exposure to the teachers' model pronunciation, the learners might still need negative evidence in the form of CF from their teachers in order to (a) double-check whether or not their own pronunciation of English /ɹ/ is intelligible enough and (b) revise their own output in response to the teachers' model pronunciation. In conjunction with several research findings that Japanese learners of English gradually improve their production and perception abilities of English /ɹ/ over substantial length of residence in English-speaking counties (Aoyama et al., 2004; Larson-Hall, 2006), it is possible that FFI that provides both enhanced positive evidence and immediate negative evidence in an effective and efficient manner can expedite the rate of L2 pronunciation development (Doughty, 2003).

Although previous intensive perception training studies confirmed that learners' improvement in their use of familiar tokens could be transferred to novel tokens (e.g., Bradlow et al., 1997; Hardison, 2003; Iverson et al., 2005; Lively et al., 1993), Study 1 showed no group differences in the generalizability task in which the learners were asked to read four new non-minimally-paired words containing the English /ɹ/ in word-initial positions. Despite the lack of significance, both the FT+CF group and the FT-only group, however, noted small-to-medium effect sizes in comparison with the control group, suggesting that participants who received FFI treatments (i.e., ± CF) might have begun to apply their improved abilities to new contexts (to words that did not appear during the FT), but with considerable individual variance. This topic needs future research that will include more speech samples and more free-constructed measures to elicit L2 speech production at a spontaneous-speech level.

Conclusion

By integrating a FFI research framework developed in L2 grammar studies into the context of pronunciation teaching in an interdisciplinary manner, Study 1 took a first step toward

investigating the acquisitional value of FFI for L2 pronunciation development, providing a number of noteworthy findings. First, communicative focus on phonological form can benefit L2 pronunciation development when FT is provided in tandem with CF. This is the case even with English /ɹ/, supposedly the most difficult sound for adult Japanese learners of English. Second, the impact of FFI on learners' interlanguage development is apparent not only at a controlled-speech level but also at a spontaneous-speech level, suggesting that FFI can promote not only development of a new metalinguistic representation of English /ɹ/ but also its internalization in the learners' L2 developing system. Third, although it is important to develop learners' selective attention towards the target pronunciation feature of English /ɹ/ through enhanced positive evidence in instructional input (i.e., FT), the learners still need immediate negative evidence (i.e., CF) in order to ascertain whether or not their output is perceived as sufficiently intelligible. The relative importance of CF could be due to the fact that, with respect to L2 pronunciation development, it is difficult for L2 learners alone to make online judgments about the extent to which their interlanguage form is good enough (i.e., intelligible pronunciation).

In addition, Study 1 adopted unique methodological features with respect to outcome measures. Instead of adopting only human rating methods, Study 1 adopted both human rating methods and acoustic analyses. This technique enhanced the validity of pronunciation teaching studies of this kind, by enabling us (a) to include a sufficient number of participants and test materials to measure students' improvement at various levels and (b) to track with relative precision any changes over time in the speech properties of the target features. As shown in Study 1, with respect to reliable outcome measures, future intervention studies should (a) first ask NS listeners to rate a small subset of speech tokens from the original data pool, (b) carefully find which speech properties relate to their rating scores, and (c) then conduct acoustic analyses

on the entire set of speech tokens to measure the effects of instruction on these relevant speech properties at a fine-grained level.

Limitations and Pedagogical Implications

First, because the instructors were encouraged to give priority to developing their students' clear understanding of meaning (i.e., English argumentative skills) rather than form (i.e., English /ɹ/), the instructional treatment in Study 1 could not be as consistent as that in lab-based studies. However, in light of the research goal—the *pedagogical* effectiveness of FFI on L2 pronunciation development—it seems ecologically valid to conduct research in classroom contexts even though many variables cannot be controlled as much as in lab settings. Second, use of the timed picture-description task as the only means of eliciting learners' spontaneous-speech production in Study 1 is also problematic. I call for more studies to refine and develop ecologically valid outcome measures specific to the context of pronunciation teaching studies. Third, the design of the generalizability task needs to be refined; a wide range of tasks should have been included to investigate how learners could generalize newly-acquired phonetic knowledge to novel contexts. Fourth, as perception training studies have generated positive results, it would be important for future research to investigate the sustainability of FFI effectiveness over a longer period of time (see Bradlow et al., 1998; Lively et al., 1993).

Study 1 first showed that teaching pronunciation forms through meaningful contexts can enhance not only accuracy but also fluency of the learners' L2 speech performance at various levels. By contrast, the effects of more explicit strategies such as explicit instruction on articulatory gestures remain unclear, especially in the light of previous research (e.g., Elliott, 1997; Macdonald et al., 1993). Second, the role of teachers' immediate feedback might be

relatively important for pronunciation teaching, because students need to (a) receive the teachers' feedback on the intelligibility of their output (negative evidence) and (b) practice the correct form in response to their teachers' model pronunciation (positive evidence). In this respect, I suggest that pronunciation recasts might be especially effective for L2 pronunciation development, but I recommend further research to compare other CF types. Given that the findings of Study 1 were based on the specific case of L2 speech acquisition of English /ɹ/ by Japanese learners, I call for future research to replicate and extend the current research framework; in particular, it would be intriguing if future experimental studies investigated the effects of FFI on L2 phonological development but with respect to other less salient sound contrasts such as the /æ-ɛ/ and /i-ɪ/ distinctions (Derwing & Munro, 2005) or suprasegmentals such as word stress (Field, 2005) and speech rate (Munro & Derwing, 2001).

TRANSITION FROM STUDY 1 TO STUDY 2

In Study 1, I took a first step toward testing how a range of form-focused instructional (FFI) techniques, including CF (i.e., recasts), can promote the acquisition of the English sound /ɹ/ by adult Japanese learners of English. According to the results, the learners who received form-focused tasks and recasts demonstrated significant decline in the frequency values of the third formant (F3) in their pronunciation of English /ɹ/, which indicates the relative importance of pronunciation-focused recasts in FFI effectiveness. However, the results also raised several relevant questions concerning which variables interacted to lead to the instructional gain. First, it still remains unclear how native-speaking (NS) listeners perceive English /ɹ/. Whereas Study 1 employed only five NS listeners and found that only F3 values were significantly correlated with their positive perception of English /ɹ/, it is still necessary to replicate the listening session with more NS listeners and speech tokens. Second, we need to conduct more robust acoustic analyses not only on F3 values but also on the frequency values of the second formant (F2) in order to examine the influence of recasts on interlanguage development of /ɹ/ by Japanese learners of English from various perspectives. Previous L2 phonology research has shown that Japanese learners of English are sensitive to F2 values rather than F3 values as their *default strategy,* suggesting that change might occur in the former acoustic property (F2 values) rather than in the latter acoustic property (F3 values) (see Bradlow, 2008). Third, whereas L2 morphosyntax studies have identified a range of individual variables that significantly influence recast effectiveness, such as the amount of repair in response to corrective feedback (Loewen & Philps, 2006) and developmental readiness (MacKay & Philps, 1998), L2 phonology research has demonstrated that the quality of adult L2 pronunciation development can be determined by

learners' length of residence (LOR) in the target language country (Flege & Liu, 2001) and motivation (Flege, Yeni-Komshian, & Lin, 1999). In this respect, more research is called for in order to examine how these affecting factors explain recast effectiveness. In order to answer these above-mentioned questions arising from Study 1, Study 2 revisits the original database and conducts a set of new analyses (a rating session with 20 NS listeners, acoustic analyses on various speech properties, and individual interviews with 54 participating students), focusing on the interrelationship between a set of individual variables (i.e., the amount of recasts and repair, initial pronunciation levels, explicit knowledge of articulatory gestures, motivation, and LOR) and two acoustic properties (i.e., F2 and F3 values).

STUDY 2: VARIABLES AFFECTING RECAST EFFECTIVENESS IN L2 PRONUNCIATION DEVELOPMENT

Drawing second language (L2) learners' attention to form via recasts is a complex phenomenon, arguably because L2 learners do not always succeed in perceiving implicit feedback such as recasts as language-focused, especially when feedback is directed to morphosyntactic errors. Several observational studies, however, have found that recasts can be quite salient to learners when their targets are L2 pronunciation errors, suggesting that recasts might be relatively facilitative of L2 pronunciation development. Study 2 revisits the original database of Study 1, and conducts a set of new analyses in order to examine in depth (a) how recasts impact L2 pronunciation development in relation to two acoustic properties of English /ɹ/ (i.e., the frequency values of the third and second formants), and (b) to what degree recast effectiveness varies according to a range of individual variables which include the amount of immediate repair, initial pronunciation levels, explicit knowledge, motivation, and length of residence (LOR)..

Recasts in SLA

According to Lyster and Ranta's (1997) classification of feedback types, recasts were defined as "the teacher's reformulation of all or part of a student's utterance minus the error" (p. 46); an example from the current study follows[12]:

Example 3

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

T: Reading /ɹidɪŋ/

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

[12] S is used as students, T as teachers and * as mispronunciation or unclear pronunciation. All examples are taken from the current study.

From a theoretical perspective, some second language acquisition (SLA) researchers assume that recasts can simultaneously provide both positive and negative evidence without interrupting the communicative flow so that L2 learners can make cognitive comparisons between their nontarget forms and targetlike reformulations during meaningful discourse (Doughty, 2001; Long, 2007). Other researchers, however, have argued that recasts might not be the most effective feedback type, at least for L2 morphosyntax development, because it has been contentious to what degree learners can succeed in perceiving the negative evidence available in recasts (Ellis & Sheen, 2006; Lyster, 2007; Lyster & Saito, 2010a; Nicholas, Lightbown, & Spada, 2001). Subsequently, a number of primary studies have descriptively and experimentally examined a range of variables associated with the noticeability and saliency of recasts. Among the variables identified by researchers, those of direct relevance to Study 2 include (a) different linguistic targets (i.e., L2 pronunciation), (b) the amount of repair, (c) learners' initial pronunciation levels, (d) explicit knowledge, (e) motivation, and (f) LOR.

Different Linguistic Features

Whereas the studies cited above exclusively focused on the effects of recasts on L2 morphosyntax development, with respect to L2 pronunciation learning (the focus of this paper), several observational studies have found that (a) learners tend to generate more successful repair following pronunciation-focused recasts than morphosyntax-focused recasts (Ellis, Basturkmen, & Loewen, 2001; Lyster, 1998b; Sheen, 2006), and to perceive the corrective intention of pronunciation-focused recasts (e.g., Carpenter, Jeon, MacGregor, & Mackey, 2006; Kim & Han, 2008; Mackey, Gass, & McDonough, 2000). In his descriptive study of French immersion classrooms, Lyster (1998b) noted that students showed a higher rate of successful repair in response to pronunciation-focused recasts than to grammar-focused recasts; similar patterns have

been also observed in adult ESL classrooms in New Zealand (Ellis et al., 2001) and adult EFL classrooms in Korea (Sheen, 2006). Furthermore, in lab settings, Mackey et al. (2000) found that, when asked to watch the video clips of their task-based interaction with NS interlocutors (i.e., stimulated recall sessions), two groups of learners (learners of ESL and Italian as a foreign language) recognized pronunciation-focused corrective feedback more accurately than morphosyntax-focused corrective feedback (see also Carpenter et al., 2006; Kim & Han, 2007). Importantly, Mackey et al. (2000) argued 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) (for empirical evidence, Varonis & Gass, 1982; Saito, 2011b). Taken together, these descriptive studies have pointed out the amenability of recasts to L2 pronunciation development owing especially to their perceived saliency.

Repair

Whereas the direct link between the amount of repair (i.e., repetition of recasts) and the subsequent impact of immediate repair on acquisition remains open to debate, especially with respect to L2 morphosyntax development (Leeman, 2003; Loewen & Philp, 2006; Lyster & Izquierdo, 2009; McDonough & Mackey, 2006), L2 pronunciation research has paid little attention to repetitive practice because it is viewed as reminiscent of decontextualized practice typical of traditional pronunciation teaching (Celce-Murcia, Brinton, & Goodwell, 1996; Trofimovich & Gatobonton, 2006); some have argued strongly that this focus-on-forms practice does not promote learners' communicative competence at all (Doughty, 2001, 2003). However, repetition of recasts that occurs during genuinely communicative L2 interaction could be considered as contextualized repetitive practice, which Segalowitz (2003) argued impacts not

only accuracy but also fluency. To my knowledge, no empirical studies have tested the relationship between the amount of repetition following recasts and L2 pronunciation development.

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 a review, see Pienemann, 2007). 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. In contrast, with respect to L2 pronunciation development, some argue that beginner learners tend to reveal rapid improvement compared to more advanced learners (Derwing & Munro, 2005), stressing learners' initial pronunciation levels as a significant variable for the effects of pronunciation instruction. Study 2 further pursues this topic by closely investigating recast effectiveness on L2 pronunciation in relation to learners' initial performance at the pre-test sessions.

Explicit Knowledge

Much L2 education research has examined the role of explicit information in the acquisition of target linguistic features, leading to a consensus that explicit knowledge plays a key role in SLA processes by expediting the rate of interlanguage development and enhancing the ultimate attainment of SLA in instructed settings (Ellis, 2006; Norris & Ortega, 2000; Spada & Lightbown, 2008). Although Ellis, Loewen, and Erlam (2006) found that recasts led adult ESL learners with a great deal of explicit knowledge about regular English past tense (measured via metalinguistic tests) to manifest significant gains in their implicit knowledge (measured via

timed oral imitation tests), it still remains unclear how *differential amounts* of explicit knowledge influence recast effectiveness. In this regard, Study 2 takes a first step towards examining the relationship between learners' explicit knowledge about articulatory gestures of the English specific sound /ɹ/ (i.e., lip rounding, tongue backness and height) and recast effectiveness on L2 speech pronunciation development.

The Amount of Motivation and L2 Experience

L2 speech studies have found some positive effects of motivation and LOR on degree of L2 foreign accent in *naturalistic* settings. On the one hand, although some studies identified the extent of learners' concern for L2 pronunciation accuracy as a significant predictor of accentedness (Elliott, 1995; Suter, 1976), other studies showed that motivation factors accounted for only a small portion of the variance (1-3%) in foreign accent ratings (Flege, Munro, & MacKay, 1995; Flege, Yeni-Komshian, & Liu, 1999). On the other hand, Flege and his colleagues have shown that additional LOR experience positively influences adult L2 pronunciation development when (a) L2 learning begins at an early stage (e.g., < 1 year) (Flege et al., 1995; Meador, Flege, & MacKay, 2000; cf. Flege, 1988) and (b) L2 learners have many opportunities to use the L2 (e.g., university-level international students) (e.g., Flege & Liu, 2001). Study 2 is a first attempt to examine in depth how these significant variables interact to impact L2 pronunciation development under *instructed* conditions (i.e., FFI treatment).

English /ɹ/

FFI in Study 2 targets one of the most well-researched cases of L2 speech learning—the acquisition of English /ɹ/ by Japanese learners of English. Given that there exists no counterpart for English /ɹ/ in the Japanese phonetic system, Japanese learners of English tend to continue to use Japanese tap /ɾ/which lies in "a position in a phonological space that is somewhere between

English /ɹ/, /l/, and /d/" (Flege, Takagi, & Mann, 1995, p. 25; see also Best & Strange, 1992; Goto, 1971; Mochizuki, 1981; Sheldon & Strange, 1982). From an acoustic standpoint, the English /ɹ/ can be discriminated by NS listeners in relation to F3: They perceive the sounds with F3 values below 2000 Hz as English /ɹ/ and those with F3 values above 2400 Hz as English /l/ (Flege et al., 1995; Ladefoged, 2003). In contrast, several cross-linguistic perception studies have also found that, as their *default* cross-linguistic perceptional strategies, Japanese learners of English ignore F3 values and instead draw on F2 values to perceive English /ɹ/, resulting in their non-nativelike perceptual representation of the sound (Hattori & Iverson, 2009; Iverson & Kuhl, 1996; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann, & Siebert, 2003; Yamada, 1995). For example, 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 boundary between English /l/ and /ɹ/ around 2067 Hz and 2523 Hz (i.e., categorical perception), and (b) Japanese listeners exhibited sensitivity to F2 variation that is irrelevant to the English /ɹ/-/l/ distinction rather than the crucial differences in F3. Similarly, in Hattori and Iverson's (2009) experiment, both Japanese and English listeners were asked to listen to a number of synthesized stimuli (i.e., /ɹa/-/la/ tokens embedded in a natural carrier sentences) whose acoustic properties—F1, F2, F3, closure/transition duration—were carefully manipulated. They were then asked to rate the extent to which the token they heard represented the best exemplar of English /l/, English /ɹ/ or Japanese /ɾ/ (the last one was only for Japanese listeners) by mouse clicking on a contiguous bar. Whereas English listeners attended only to F3 and transition duration to decode the English /ɹ/-/l/

contrast, Japanese listeners differentiated English /l/, English /ɹ/ and Japanese /ɾ/ in relation to not only F3 and transition duration but also F2.

Intriguingly, Bradlow (2008) attributed these cross-linguistic perceptual patterns of Japanese learners of English (i.e., relying on F2 variation) to the markedness of articulatory gestures in Japanese and English phonetic systems. That is, whereas both Japanese and English mark *tongue backness* (related to F2 variation), a combination of constriction in the pharyngeal and velar regions of the vocal tract and lip rounding (related to F3 variation) are marked in English but not in Japanese (for details of articulatory gestures of English /ɹ/, see Campbell, Gick, Wilson, & Vatikiotis-Bateson, 2010).[13] Taken together, their sensitivity to F2 values rather than F3 values suggests that Japanese learners of English might (a) initially produce English /ɹ/, simply by drawing on L1-related articulatory characteristics (i.e., tongue backness); and (b) gradually increase their awareness of new oral gestures (i.e., pharynx/velar constriction and lip rounding) in relation to their developing phonetic category for English /ɹ/. In order to further pursue this topic, the current study first examines how Japanese learners of English produce English /ɹ/ via acoustic analyses on F3 and F2 values of their speech production of English /ɹ/, and then investigates how recasts promote their attentional shift away from F2 values (irrelevant to English /ɹ/) and towards F3 values (highly relevant to English /ɹ/) by analyzing their productive use of English/ɹ/.

[13] This argument also concurs with the motor theory of speech perception whereby objects of perception are articulatory rather than acoustic events and humans have a specialized module for perception and production in terms of articulatory gestures (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; for a comprehensive review, Diehl, Lotto, & Holt, 2004)

Motivation for Study 2

Study 2 originates in the larger research project described in Study 1 where I conducted a quasi-experimental study in simulated classroom settings in order to investigate how a range of FFI techniques (FT ± CF) affect the acquisition of English /ɹ/ by Japanese learners of English. Given that the purpose of Study 2 is to extend and reanalyze the original database, several highly relevant findings need to be reported here. First, in Study 1, five NS listeners were asked to rate the quality and intelligibility of 100 speech stimuli produced by 20 Japanese learners of English (i.e., Consonant-Vowel-Consonant singleton tokens including English /ɹ/ in word-initial positions). Based on the multiple regression analyses which identified only F3 values as a significant predictor of their rating scores, the acoustic analyses were conducted on the entire data set; the results revealed that only learners receiving recasts during FFI demonstrated significant decline in their F3 values in a range of tasks, which in turn indicated improvement in their production of English /ɹ/ at both a controlled and spontaneous speech level.

Importantly, the results of Study 1 raised several questions worthy of further pursuit in order to obtain a more precise picture of recast effectiveness in the acquisition of English /ɹ/. That is, given such complex acoustic characteristics of English /ɹ/, do recasts impact not only F3 but also other phonetic cues such as F2, and, if so, how? Furthermore, to what degree does recast effectiveness vary according to five specific variables (i.e., the amount of repair, initial pronunciation levels, explicit knowledge, motivation, and LOR)? In this regard, drawing on the database of Study 1, Study 2 adopted three new analyses: (a) re-examining the perception of English /ɹ/ via a new rating session with more listeners and tokens (20 NS listeners rated 150 speech samples produced by 30 Japanese learners of English), (b) implementing acoustic

analyses not only on F3 but also on F2 values for the entire data set, and (c) analyzing the details of the individual interviews with 54 participating students.

Research Questions

The research questions to be addressed follow:

1. How do acoustic properties of English /ɹ/ (F1, F2, F3 and duration) and contextual factors (task types, ensuing vowel backness and height) relate to NS listeners' perceptions of English /ɹ/?
2. How do recasts impact L2 pronunciation development of English /ɹ/ in relation to F3 and F2 values?
3. To what degree does recast effectiveness vary according to (a) the amount of immediate repair, (b) initial pronunciation levels, (c) explicit knowledge, (d) motivation, and (e) LOR?

Method

Design

Study 2 employed a quasi-experimental design with a pre-test and post-test design in simulated classroom settings. First, after individually completing pre-tests and the initial interview, student participants received four hours of FT treatment with and without recasts provided by two NS teachers, and, two weeks after the last day of classes, the students came back to the researcher's office to complete post-tests and 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 at the pre-test sessions in order

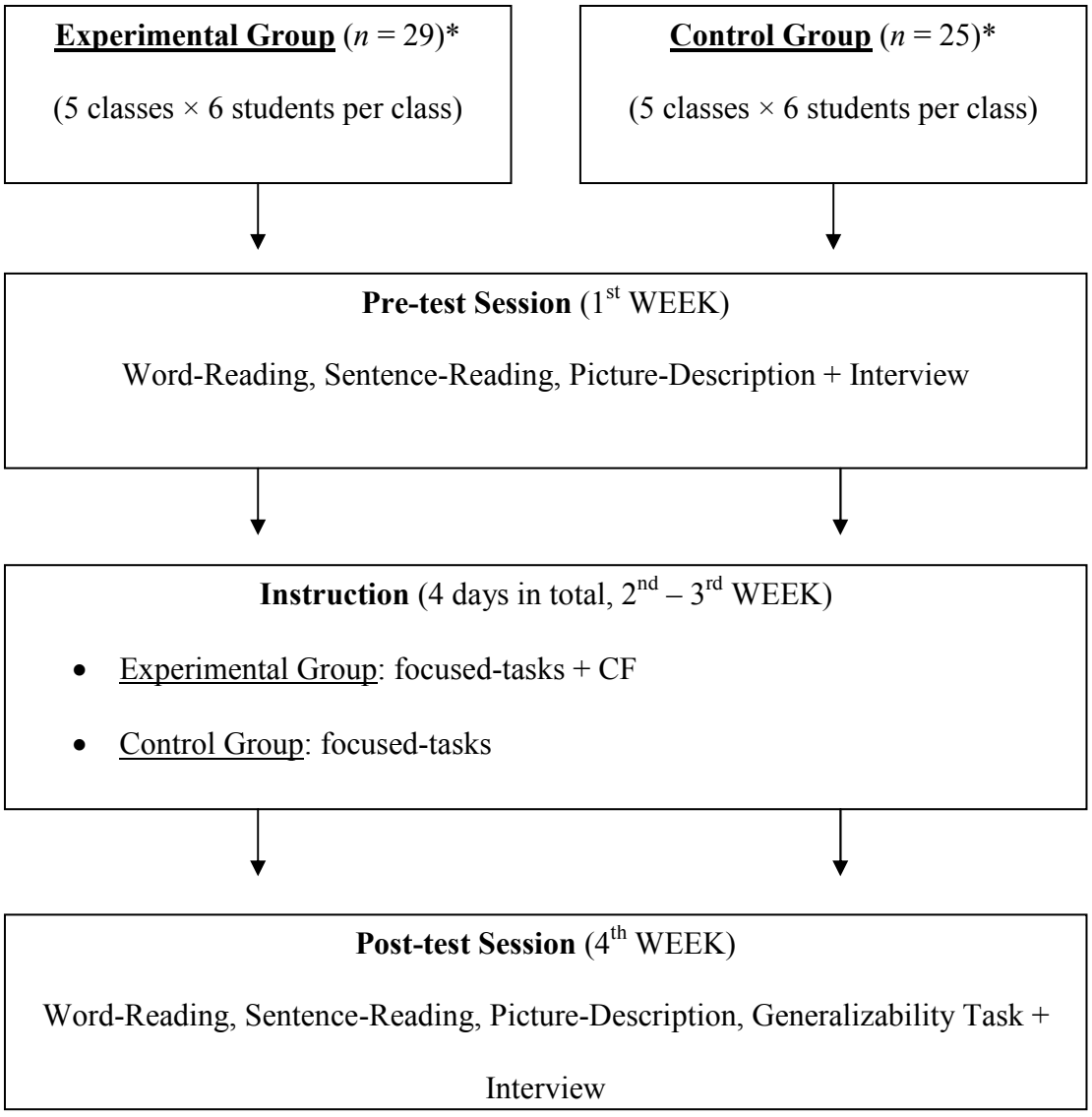


Figure 4. Summary of the procedure in Study 2

Note. * shows the number of the participants included for the final analyses.

to find significant acoustic properties and contextual factors affecting their rating scores. Finally, based on these relevant acoustic properties and contextual factors, the students' improvement in English /ɹ/ was carefully assessed. The research design is summarized as Figure 4.

Participants

The participants in Study 2 were two experienced ESL teachers, 20 NS listeners, and 54 adult intermediate-level Japanese learners of English in Montreal, Canada.

Students

Among 64 Japanese learners of English who participated in Study 1, the performance of 54 learners in the FT+CF group (n = 29) and the FT-only group (n = 25) was included for the secondary analyses in Study 2. In order to tease apart recast effectiveness from the overall effects of FFI, a decision was made to label the former as the experimental group (who received recasts during FT) and the latter as the control group (who did not receive recasts during FT). The assumption in Study 2 is that comparing the experimental and control groups in this way (FT ± recasts) will enable me to examine which factors contribute to the impact of recasts on L2 pronunciation development of English /ɹ/ by Japanese learners of English. Although most participants had arrived in Montreal just before the project, their LOR varied widely (M = 15.05 months, SD = 30.13). A majority of 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).

Teachers

Two experienced ESL teachers participated. Whereas the first teacher taught six classes (3 classes in the experimental group, 3 classes in the control group), the other teacher taught four classes (2 classes in the experimental group, 2 classes in the control group).

Listeners

Whereas Study 1 recruited five NS listeners to rate 100 speech tokens produced by 20 Japanese learners of English, Study 2 included 20 NS listeners to rate 150 speech tokens produced by 30 Japanese learners of English with a view to obtaining a more robust analysis of which acoustic properties of /ɹ/ influence NS listeners' perception. Twenty listeners were carefully selected based on (a) their familiarity with Japanese-accented speech (they reported that they had frequent contact with native Japanese speakers as their friends) and (b) L1 background (their L1 was a north-eastern dialect of American English).

Interview

I individually interviewed all participants in my office right after the pre-test sessions (the initial interview) as well as the post-test sessions (the final interview). The purposes of the interviews were (a) to elicit personal information including motivation and LOR at the onset of the project and (b) to ascertain whether learners had had any pre-existing explicit knowledge of English /ɹ/ in a retrospective manner after the project was completed. All communication was in Japanese and audio-recorded.

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, and 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 (for the details of the project, see Study 1). Each activity included at least one target word which has English /ɹ/ in various phonetic contexts (see Table 4), and these target words were highlighted in red in the power point presentations used to deliver the lessons (i.e.,

typographically enhanced input: Han, Park, & Combs, 2008). For example, in English debating, the students debated several topics such as "***Run***ning inside is better than ***run***ning outside?" and "Is it good to have a ***rain***y day?"

In addition to the main activities, three activities were used as warm-up games requiring students to distinguish /ɹ/ in perception and production at least from Japanese flap /ɺ/ in order to win. For example, in *English Karuta,* 36 cards are placed on a table. Each card represents one lexical item, and has a picture depicting the item along with the first letter of the word. As the teacher reads out from a list of the words, students try to find and pick up the right card as soon as possible. In order to get many cards, the students have to pay attention to the perceptual difference between /ɹ/ and /l/.

The instructional treatment highlighted 38 target words including English /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 English /l/, which required the students to make a clear distinction between /ɹ/ and /l/. These words were presented in Table 4 in Study 1.

Pronunciation-focused Recasts

In the experimental group, the teachers were asked to reformulate students' mispronunciation and unclear pronunciation of English /ɹ/ without altering the original meaning of utterances—pronunciation-focused recasts. In order to control the linguistic characteristics of recasts (length number of changes), which greatly influence the saliency and efficacy of recasts (Sheen, 2006), teachers were instructed to consistently recast only one word and to use falling intonation, without adding any additional meaning, as in the following example.

Example 4 in Argument Critique

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

T: Rain /ɹeɪn/

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

Such recasts can be defined as "partial recasts" (Sheen, 2006), which have been found to increase learners' awareness of the corrective nature of feedback. In order to ensure the consistency of recasts for the entire project, the researcher observed all classes after which he gave comments to teachers on the way they delivered recasts.

Control Group

Twenty-five learners in the control group also received comparable lessons in English argumentative skills with form-focused practice activities. However, they received no feedback on their pronunciation errors, only on ungrammaticality and inappropriate lexical choice (e.g., (e.g., "you should say, 'she is manipulative' instead of 'she is tricky'") as well as the contents of the lessons (e.g., "you need more evidence to support your idea").

Measures

In order to investigate the effects of recasts on L2 pronunciation development from various perspectives, two types of learners' speech tokens were elicited: *familiar* tokens and *novel* tokens. Because three participants neither completed the instructional treatment nor attended the post-test sessions, and three others were considered too advanced based on the pretest scores (e.g., their F3 values were always less than 2000 Hz), there were a total of 54 participants included in the final analysis. All of the test sessions took place individually in a quiet room, and I explained and led the testing procedure in Japanese. Recordings were made employing *Praat*, speech analysis application downloadable at www.praat.org (Boersma &

Weenik, 2009), at a 44.1 kHz sampling rate and a 16-bit resolution, as well as a unidirectional microphone (DM-20SL).

Familiar Tokens

As for familiar tokens that the learners were exposed to during the FT treatment, the learners completed three tasks both at the pre- and post-test sessions (n = 2,376 tokens).

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).

Sentence reading. In this task, the learners read five sentences where eight target words were included together with three distracter sentences:

He will read my paper by the time I arrive there.

She left her red bicycle on the side of the road.

The race was cancelled because of the rain.

I can correct all wrong sentences tonight.

Ryan does not like to run in the snow.

All of these words were CVC singletons except "*Ryan*" (CVVC). In total, 864 token were created (8 words × 54 students × 2 test sessions = 864 tokens).

Timed picture description. In this task, the learners described eight pictures each of which was accompanied by three words that learners were thereby prompted to use in their descriptions. Whereas four pictures served as distracters without any 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 (Ellis, 2005). In total, 432 tokens were created (4 words × 54 students × 2 test sessions = 432 tokens).

Target Words

Out of 38 words featured in the form-focused activities, 14 words /ɹ/ in word-initial positions were used in the pre-/post-test sessions. All of these words were consonant-vowel-consonant (CVC) singletons except one word (*Ryan* as CVVC; see the words with asterisks in Table 1). Given that Japanese learners of English experience difficulties in producing and perceiving /ɹ/ particularly in word-initial positions (Bradlow et al., 1997; Goto, 1972; Sheldon & Strange, 1982), the assumption underlying the use of word-initial tokens in the test materials was that, by examining the most difficult phonetic contexts, significant improvement would imply the likelihood of learners' ability to generalize to other relatively easier contexts (e.g., word-medial and syllable-final tokens). Furthermore, Study 2 also takes into account three contextual factors affecting the learners' production of /ɹ/: (a) task types (controlled vs. spontaneous speech levels) (e.g., Flege et al., 1995), (b) following vowel backeness (singletons with front/back vowels), and (c) following vowel height (singletons with high/low vowels) (e.g., Hardison, 2003). Each participant generated 22 word-initial singleton tokens at the pre- and post-test sessions respectively (out of 14 words, eight were tested twice but in different tasks).

Novel Tokens

To assess learners' production of novel tokens that never appeared during the four hours of FFI lessons, they completed a generalizability task only at the post-test session. The purpose was to examine whether or not learners could generalize their newly acquired knowledge of English /ɹ/ to unfamiliar items. These words were single and multisyllabic words beginning with

English /ɹ/ (i.e., *real, roll, rumor, regular*).[14] Because five students could not complete the generalizability task for several personal reasons, 196 tokens were created in total (4 words × 49 students × 1 test session = 196 tokens).

NS Baseline

For comparison purposes, five native speakers of English (three males, two females) who were undergraduate students at an English-speaking university in Montreal completed the same tasks (Word Reading, Sentence Reading, Timed Picture Description, and Generalizability Task). All of them reported that they had grown up in the north-eastern regions of the USA and moved to Montreal in Canada at the time of their university entrance. In total, 110 tokens were generated as baseline data (26 words per talker × 5 NS talkers = 130 tokens), in order to demonstrate to what degree NS tokens differ from those produced by Japanese learners of English.

Listener Ratings and Acoustic Analyses

Twenty NS listeners rated 150 tokens produced at 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 reseracher in order to see which variables significantly influence NS listeners' rating scores of English /ɹ/.

Although previous pronunciation research has exclusively adopted the human rating method (i.e., asking NS listeners to rate L2 speech samples) in order to measure the quality of sentence-level speech samples in a holistic manner, its methodological difficulty especially in intervention studies of this kind (with large data sets collected both at pre- and post-test sessions)

[14] These words in Generalizability Task included not only English /ɹ/ but also /l/ in order to increase the learners' awareness towards English /ɹ/-/l/ contrast and measure their best production abilities.

is a time-consuming procedure: Asking NS listeners to rate all of the speech tokens by 54 learners (n = 2,376) would have taken a great deal of time, which entails risks causing listener fatigue. As a remedy, a decision was made, first, to ascertain which acoustic properties (e.g., F3, F2) play an important role in NS perception of /ɹ/ based on a small subset of speech data randomly selected from the original data pool and, second, to conduct acoustic analyses on the entire data set with a focus on these significant acoustic variables (i.e., F3).

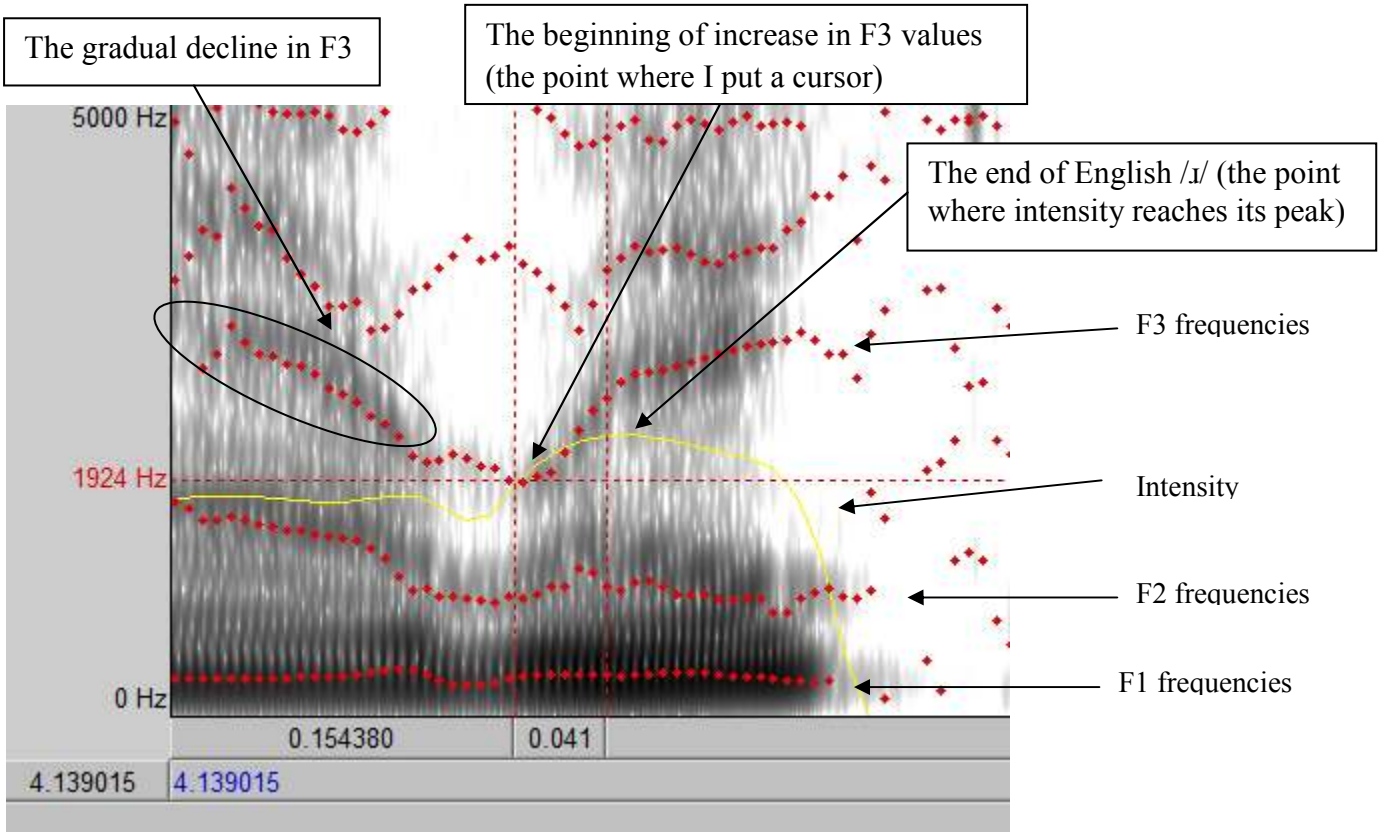
Token Preparation

Drawing on the set of 1,188 tokens produced by learners at 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 each learner contributed five words (two from Word Reading, two from Sentence Reading, and one word from the Timed Picture Description). Second, while listening to the original speech samples several times, the reseracher selected/cut only the target words without including any preceding and following sounds, and pasted them into separate sound files by means of the speech analysis software (i.e., *Praat*).

Acoustic Properties and Contextual Factors

The researcher measured F1, F2, and F3 values in hertz (Hz) and transitional duration of F3 in milliseconds (ms) of the 150 speech tokens by means of *Praat*, and categorized them based on task types (word reading, sentence reading, picture description), and according to the backness (front, central, back) and height (high, mid, low) and of subsequent vowels, respectively. Given that Flege et al.'s (1995) study is one of very few studies that acoustically analyzed natural L2 speech tokens, Study 2 followed their procedure in order to measure three spectral cues (F1, F2 and F3) and one temporal cue (duration of F3 transition) as to word-initial English /ɹ/. Three steps were taken as follows:

1. The beginning of the words (where English /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 falling F3 and the onset of the steady state of F1, F2, and F3 as a reliable clue (i.e., F3 of the preceding sounds tends to continue to decline towards the beginning of the word because F3 of English /ɹ/ is relatively low).
2. The frequency values of F1, F2 and F3 were then measured by putting the cursor on the endpoint of the steady state where F3 started rising again (the beginning of F3 transition towards the following vowels).
3. Finally, the duration of F3 transition was measured by dragging the cursor from the onset to the endpoint of F3 transition (i.e., the increase in F3 values stopped when reaching the beginning of the following vowels). Not only the spectrographic and waveform representations of the tokens were used as reliable clues for the endpoint of the F3 transition, but also an intensity analysis (the beginning of the following vowels are likely characterized by the peak of intensity). The procedure of the acoustic analyses is summarized in Figure 5.



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

Figure 5. Procedure for the acoustic analysis for English /ɹ/

Rating Sessions

The rating sessions took place individually with each listener in a quiet room for approximately one hour. First, after a brief explanation about the purpose of the current project, they rated five speech tokens (not included in the subsequent listening session) on a 9-point scale as practice. The 9-point scale 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 (*hybrid exemplars, neither /ɹ/ nor /l/*) → 6 (*possibly /l/*) → 7 (*probably /l/*) → 8 (*good /l/*) → 9 (*very good /l/*). At the end of the practice, the researcher explained how other listeners had rated the same samples and asked them to compare their ratings with those of others. Second, the listeners rated 150 speech tokens of English /ɹ/. They were allowed to listen to speech samples as many times as they wanted until they felt confident about their ratings (averaging two or three times according to the researcher's observation).[15] The researcher always sat next to the listeners so that they could ask whatever questions they had during their rating sessions.

Group Comparisons and Post-Hoc Analyses

First, a set of ANOVA analyses were conducted on pre- and post-test scores to examine the extent to which the two groups of learners (FT with and without recasts) differed in their pronunciation of familiar as well as unfamiliar tokens. Next, multiple regression analyses were conducted as post-hoc analyses to explore the extent to which the *dependent variables* of recast

[15] Although one could argue that the results of the rating procedure adopted in the current study might not reflect a real life situation (listeners have to understand their interlocutors without asking too many repetitions), note that it was tremendously difficult for NS listeners who had never had any phonetic training to judge only the quality of English /ɹ/. That is, they reported that they would have otherwise based their judgement on the entire words, because Japanese learners of English tend to mispronounce not only English /ɹ/ but also other pronunciation features such as segmentals (/ʌ/ in "rum") as well as suprasegmentals (sentence and lexical stress, speech rate).

effectiveness in the experimental group can be related to five *independent variables* (i.e., the amount of repair, initial pronunciation levels, explicit knowledge, motivation, and LOR).

Dependent Variables

The within-group difference between pre-test and post-test scores in 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 dependent variables to represent *recast effectiveness*. For example, if a learner produced F3 values of 2550 Hz at the pre-test and those of 2210 Hz at the post-test, he or she receives a score of 340 Hz (2550Hz – 2210 Hz = 340 Hz). These scores were grouped relative to four different contexts: F3 and F2 values of familiar tokens with front vowels and back vowels (n = 290 for each context). However, due to the lack of samples ($n < 100$) and post-tests (for novel tokens)[16], the following were eliminated: the data regarding novel tokens following front vowels (n = 58) and back vowels (n = 58) as well as familiar tokens following central vowels (n = 58) from the multiple regression analyses.

Independent Variables

Five independent variables were analyzed as follows:

Amount of repair: 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 researcher watched all video-taped lessons of the experimental group (4 hours of lessons × 5 classes = 20 hours), and noted (a) how many times each student received recasts from the teacher (i.e., the number of recasts) and (b) how many times he/she repeated the teacher's recast (i.e., the amount of repair). Due to the high correlation between the amount of recasts and repair (which inevitably causes multicollinearity problems), the researcher decided to take into account "the amount of

[16] For the purpose of robust analysis, the number of observations needs to be more than 100 for multiple regression analyses.

repair" instead of "the amount of recasts" as an independent variable, because the former was better suited for the purpose of Study 2 (i.e., the impact of the whole interactional sequence—recasts followed by repair—on acquisition).

Unlike morphosyntax-focused recasts where we can clearly distinguish "repair" and "no-repair" based on the orthographic transcription of learners' utterances (i.e., whether or not their utterances are still ungrammatical), it was difficult to determine to what degree learners repaired their unclear pronunciation and mispronunciation of English /ɹ/. For example, some learners repeated recasts without any perceptible modifications, despite their obvious noticing and conscious efforts to improve their mispronunciation. In addition, whether the students' reproduction of /ɹ/ preceding recasts sounded *intelligible enough* or still *heavily accented* depended on listeners' *subjective* judgement. Thus, all learners' repetitions of recasts were taken into account as "repair" in Study 2, regardless of the degree of their accentedness, whereas instances of no attempt at repetition were coded as "no repair," as in the following examples.

Example 5 in English Debate

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

T: Running /ɹʌnɪŋ /. ← RECAST

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

Example 6 in Guessing Game

S: Reading /lidɪŋ/comic books promotes creativity

T: Reading /ɹidɪŋ/* ← RECAST

S: Yes. ← NO REPAIR

T: Okay, I see.

*Learners' initial pronunciation levels.*The learners' pre-test scores were used as indicators for their initial pronunciation levels.

Explicit knowledge. At the endpoint interview, the learners were asked if they had any explicit knowledge of the articulatory gestures of English /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[17], they were given "1"; if not, they were given "0."

Motivation. In order to measure motivational variables, previous L2 speech studies asked to rate a variety of factors affecting motivation from "very important" to "not important at all" (see Flege et al., 1995, 1999). Given that most learners in Study 2 came to Canada to improve conversational English skills, a decision was made to measure their ESL-related motivation. At the initial interview, the learners were asked to rate their motivation to improve four linguistic skills (grammar, lexis, listening, and pronunciation) on a 5-point scale (1. highly motivated to improve – 5. not very motivated to improve). They were first asked to use the whole scale as much as they could. Yet, if they were equally-motivated for all linguistic domains, they could, of course, assign "1" or "2" to all linguistic areas. Their rating scores only for pronunciation skills were used for Study 2.

Length of residence. At the initial interview, the learners were asked their total length of residence in Canada.

[17] As illustrated in many pronunciation textbooks in Japan, many participants reported that, unlike the Japanese flap, the tip of tongue does not touch alveolar ridge, which I categorized as explicit knowledge about tongue positions.

Results

First, I will present the results of the rating sessions conducted to determine which acoustic properties (F1, F2, F3, F3 transition) and contextual factors (task types, following vowel backness and height) are related to NS listeners' perception of English /ɹ/. Second, I will present the results of ANOVA analyses undertaken to examine the extent to which the two groups of learners (FT with and without recasts) differed in their pronunciation of familiar as well as unfamiliar tokens. Finally, I will present the results of the multiple regression analyses that explored the extent to which recast effectiveness is influenced by the amount of repair, initial pronunciation levels, explicit knowledge, motivation, and LOR.

Listener Ratings and Acoustic Analyses

Given that the 20 NS listeners showed significantly high interclass correlation ($r = .803$, $p < .001$), their rating scores were averaged for each speech token. A multiple regression analysis was conducted on the rating scores as dependent variables, and four speech properties (F1, F2, F3, F3 transition) and three contextual factors (task types, following vowel height and backness) as independent variables. The results showed that the model was significant, F (7, 142) = 35.857, $p < .00001$, and it accounted for 63.8% of the variance in the NS listeners' rating scores. The model identified three variables as significant predictors: F3 values ($t = 8.006$, $p < .0001$), F2 values ($t = 2.376$, $p < .01$) and following vowel backness ($t = 2.425$, $p < .05$). The standard coefficients of the three variables are as follows: F3 ($\beta = .583$), F2 ($\beta = .191$), and following vowel backness ($\beta = .149$). In order to further examine the relative weights of F3 and F2 values in NS listeners' perception of English /ɹ/ according to following vowel backness, several post-hoc analyses were conducted.

F3 Values

The partial correlation analyses showed that the simple correlation between F3 values and the NS rating scores remained highly significant when effects of variation in F2 values and following vowel backness were removed, $r = .645$, $p < .001$; F3 values accounted for 41.6% 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 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 F3 values were treated as *a primary phonetic cue* in Study 2.

A closer examination of the relationship between the F3 and the listeners' 9-point rating scores revealed a benchmark for NS perception of English /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) *hybrid 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 (5 talkers × 22 familiar tokens = 110 tokens). The results of the descriptive statistics are summarized in Table 8. Finally, 95% confidence intervals (CIs) 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. hybrid 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 hybrid exemplars nor those of good English /l/. Taken together, the learners' performance was roughly assessed at the five levels: (a) native-like English /ɹ/

(1600-1700 Hz) (b) good English /ɹ/ (2000-2280 Hz), (c) poor English /ɹ/ (2280-2550 Hz), (d) confusing exemplars (2550-2800 Hz), and (e) English /l/ (> 2800 Hz).

Table 8. *The results of NS perception of English /ɹ/*

Category	9-point scale	No. of tokens	*M*	*SD*	95% CI	
					Lower	Upper
Good English /ɹ/	$1 \leq x \leq 2.5$	38	2101 Hz	229 Hz	2028 Hz	2174 Hz
Poor English /ɹ/	$2.5 < x \leq 4$	32	2406 Hz	356 Hz	2283 Hz	2529 Hz
Hybrid exemplars	$4 < x < 6$	20	2673 Hz	281 Hz	2550 Hz	2797 Hz
Poor English /l/[a]	$6 \leq x < 7.5$	20	2830 Hz	209 Hz	2739 Hz	2921 Hz
Good English /l/	$7.5 \leq x \leq 9$	40	2928 Hz	213 Hz	2850 Hz	3004 Hz
NS baseline of English /ɹ/		110	1664 Hz	234 Hz	1621 Hz	1709 Hz

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

F2 Values

Although the partial correlation analyses 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, $r = .263$, $p < .01$, F2 values accounted for only 6.9% of the variance in the NS listeners' rating scores. Not surprisingly, F2 values were significantly correlated with ensuing vowel backness, $r = .389$, $p < .0001$: $M = 1722$ Hz for front vowels ($SD = 391$) → $M = 1448$ Hz for back vowels ($SD = 340$). Although F2 values did play a role in NS perception of English /ɹ/ to some degree (the lower 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.

Data Analyses

Given that the results identified neither task types nor the following vowel height as significant variables for NS listeners' perception of English /ɹ/, the influence of these variables is not further discussed in Study 2. That is, the rest of the entire data set (n = 2572) was analyzed focusing only on F3 (i.e., the primary cue) and F2 (the less relevant cue) according to three phonetic contexts: English /ɹ/ preceding front, central, and back vowels.

ANOVAs

Familiar Tokens

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. Separate three-way ANOVAs were conducted for F3 and F2 values in order to identify statistically significant differences within groups (following vowel contexts, pre-/post-test) and between groups (the experimental group vs. the control groups). An alpha level was set at a $p < .05$ level for all statistical analyses. Cohen's d was also calculated in order to measure the magnitude of instructional effectiveness between two contrast groups of means.[18]

With respect to F3 values, the ANOVA results revealed significant effects for the overall Time × Group interaction, F (1, 52) = 9.737, p = .0029 for /ɹ/ preceding front vowels, F (1, 52) = 15.449, $p < .001$ for /ɹ/ preceding central vowels, and F (1, 52) = 15.082, $p < .001$ for /ɹ/

[18] According to Cohen (1988) effect sizes are roughly classified as small (0. 20 ≤ d < 0.50), medium (0.50 ≤ d < 0.80), or large (0.80 ≤ d).

preceding back vowels. A simple main effect of time was significant only for the experimental group (M = 2538 Hz, SD = 287 Hz → M = 2321 Hz, SD = 302 Hz), F (1, 52) = 40.416, p < .001, d = 0.74. In general, learners receiving recasts made a substantial transition from confusing exemplars (around 2500-2600 Hz) to good-poor English /ɹ/ (around 2300-2400 Hz) with medium effects. With respect to F2 values, however, the ANOVA results revealed only overall effects for Time (M = 1469 Hz, SD = 276 Hz → M = 1389 Hz, SD = 229 Hz), F (1, 52) = 18.256, p < .001, d = 0.31. F2 values declined equally in both the experimental and control groups with small effects; however, F2 values are not related to NS listeners' perception of English /ɹ/.

Novel Tokens

F3 and F2 values of each participant's performance were averaged according to following vowel backness (2 tokens for front vowels and 2 for back vowels) at one time condition (i.e., post-test sessions). In order to determine whether the two groups differently generalized their ability to produce /ɹ/ in new lexical contexts, two-way ANOVAs (Group × Backness) were conducted to compare only their post-test scores. An alpha level was again set at a p < .05 level for all statistical analyses. With respect to F3 values, the ANOVA did not reach statistical significance for the overall effect for Group, F (1, 47) = 0.066, p = .798, nor the overall Group × Backness interaction effects F (1, 47) = 0.476, p = .493. Both the experimental and control groups exhibited relatively poor English /ɹ/ in front vowel contexts (M = 2347 Hz for the experimental group, 2393 Hz for the control group) and good English /ɹ/ in back vowel contexts (M = 2231, 2233 Hz respectively). With respect to F2 values, the ANOVA did not detect any significant group difference either: F (1, 47) = 1.111, p = .2973 for the overall Group effects and F (1, 47) = 0.125, p = .7248 for the overall Group × Backness interaction effects.

NS Baseline

For familiar tokens, the mean of their F3 values was 1682 Hz for singletons with front vowels, 1630 Hz for those with central vowels, and 1645 Hz for those with back vowels. For novel tokens, the mean F3 value was 1753 Hz for singletons with front vowels and 1535 Hz for those with back vowels. In contrast, their F2 values (around 1000-1200 Hz) did not appear to be different from those of Japanese learners of English (around 1000-1200 Hz). For familiar tokens, the mean of their F2 values was 1223 Hz for those with front vowels, 1117 Hz for those with central vowels, and 1124 Hz for those with back vowels. For novel tokens, the mean of their F2 values was 1207 Hz for singletons with front vowels and 1074 Hz for those with back vowels.

Multiple Regression Analyses

As post-hoc analyses, Study 2 also analyzes the effects of recasts in the experimental group (i.e., within-group difference between pre- and post-test scores) relative to five independent variables: the amount of repair, initial pronunciation levels, explicit knowledge, motivation, and LOR.

Amount of Repair

In order to ensure the validity of the researcher's coding (watching 20 hours of video-taped lessons to count the number of recasts and repair moves for each student), after a brief training, one near native-like NNS speaker of English (a NS of Japanese) watched 20% of the data (4 hours) and counted the number of recasts and repair moves. The interrater reliability was significantly high for the number of recasts (r = .92) and the amount of repair (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 amount of repair (M = 26.68, SD = 9.60), which indicates a relatively high rate of repair (i.e., 81.90%).

Initial Pronunciation Levels

Learners' pre-test scores were used as an indicator for their initial pronunciation levels. Not surprisingly, the results of a one-way ANOVA showed that their pre-test performance significantly differed between front and back vowels both for F3 values, F (1, 289) = 13.547, p = .0003, (M = 2582 Hz for front vowels; M = 2506 Hz for back vowels), and F2 values F (1, 289) = 149.257, p < .0001, (M = 1627 Hz for front vowels; M = 1354 Hz for back vowels).

Explicit Knowledge

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

Motivation

The results of the initial interview showed that the learners' generally had high motivation for learning/improving pronunciation skills (M = 1.89, SD = 0.93), ranging from 1 (highly motivated to improve) to 5 (not very motivated to improve).

Length of Residence

According to the results of the initial interview, their LOR greatly varied between 1 month and 86 months (M = 19.44 months; SD = 31.77 months).

Results of Multiple Regression Analyses

Four sets of forward stepwise multiple regression analyses were carried out to explore the relationship between four dependent variables (i.e., F3 and F2 values of familiar tokens following front and back vowels) and the five independent variables.

F3 of singletons following front vowels. A two-factor model with (a) initial pronunciation levels (the beta weight *ß* = .556), and (b) explicit knowledge regarding tongue movement (*ß* = -.187) accounted for a significant 37.3% of the variance in F3 scores, F (6, 283) = 28.118, p <. 001.

F3 of singletons following back vowels. A three-factor model with (a) initial pronunciation levels (*ß* = .453), (b) their knowledge regarding tongue movement (*ß* = -.166), and (c) motivation (*ß* = .122) significantly explained 26.9% of the variance in F3 scores, F (6, 283) = 17.356, p < .001.

F2 of singletons following front vowels. A four-factor model with (a) initial pronunciation levels (*ß* = .381), (b) LOR (*ß* = -.164), (c) explicit knowledge of tongue movement (*ß* = -.121), and (d) explicit knowledge of lip rounding (*ß* = -.120) accounted for 27.2% of the variance in F2 scores, F (6, 283) = 17.800, p < .001.

F2 of singletons following back vowels. A two-factor model with (a) initial pronunciation levels (*ß* = .509) and (b) repair (*ß* = .199) accounted for 35.5% of the variance in F2 scores, F (6, 283) = 25.934, p < .001.

Summary of the Results

First, the results of the rating sessions with 20 NS listeners found that their perception of English /ɹ/ can be significantly determined by F3 values (the primary cue); the role of F2 variation in their perception of English /ɹ/ remains unclear. Second, whereas all students changed their F2 values after receiving FT treatment, those receiving recasts led to a significant decline in F3 (i.e., improvement resulting from instruction). Finally, the results of the multiple regression analyses on students in the experimental group found that such recast effectiveness (FT+CF) was

significantly related to learners' initial pronunciation levels, and the amount of their explicit knowledge and motivation.

Discussion

NS perception of English /ɹ/

To answer the first research question, which asked how acoustic properties of English /ɹ/ (F1, F2, F3 and duration) and contextual factors (task types, ensuing vowel backness and height) relate to NS listeners' perception of English /ɹ/, 20 NS listeners rated 150 speech tokens produced at the time of the pre-test session by 30 out of 54 Japanese learners of English. 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 correlated with the positive judgment, decline in F2 values was also highly related to ensuing vowel backness. Thus, any decline in F3 could be considered as *improvement* (i.e., beneficial effects of instructional treatment), but a change in F2 values could be considered as either an effect of the following vowel (F2 values following back vowels tend to be lower) or L1-related effects (i.e., Japanese learners of English are sensitive to F2 variation). Similar findings were found with synthesized speech tokens (e.g., Hattori & Iverson, 2009; 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 category was used to assess improvement resulting from instruction):

- Native-like English /ɹ/ (F3 = 1600-1700 Hz)
- Good English /ɹ/ (F3 = 2000-2280 Hz)
- Poor English /ɹ/ (F3 = 2280-2550 Hz)
- Hybrid exemplars (F3 = 2550-2800 Hz)
- English /l/ (F3 > 2800 Hz)

Important to emphasize is that, especially for F3 values, good exemplars (i.e., F3 values = 2000-2200 Hz) still substantially differ from those of NS baseline data (F3 values around 1600 Hz). This result provides a possible benchmark of *intelligible* pronunciation that Japanese learners of English can realistically strive for (i.e., F3 values around 2000-2200 Hz) as opposed to *native-like* pronunciation of English /ɹ/ (i.e., F3 values around 1600-1700 Hz).

Effects of Recasts on F3 and F2 Values

In response to the second research question which asked whether recasts influence L2 pronunciation development of /ɹ/ by Japanese learners of English, the ANOVAs revealed differential effects of recasts according to F3 and F2 values. On the one hand, given that both the experimental and control groups equally exhibited a decline in their F2 values, learners receiving FT regardless of recasts started to improve their production of English /ɹ/, but drawing on L1-related articulatory characteristics (i.e., tongue backness). On the other hand, only participants in the experimental group significantly lowered their F3 values (M = 2550-2800 Hz → 2280-2550 Hz), which in turn indicates that their improvement from hybrid to poor-to-good exemplars of English /ɹ/ resulted from receiving recasts. This gain could be attributed to the fact that recasts in the context of FT enable the learners to (a) carefully listen to individualized model pronunciation by teachers (*positive evidence*), and (b) double-check the intelligibility of their own production based on the teachers' signal (*negative evidence*). Ultimately, such feedback techniques can perceptually stimulate their gradual adjustment to F3 values, which in turn indicates an increasing amount of learner attention drawn to new oral gestures (i.e., pharynx/velar constriction and lip rounding). Thus, the results suggest that, when L2 learners are encouraged to process the target phonological form of English /ɹ/ for meaning via recasts, they can start paying more attention to even new phonetic cues in the L2 input which they would otherwise ignore (for

the role of language awareness in L2 pronunciation development, see Kennedy & Trofimovich, 2011).

Variables Affecting Recast Effectiveness

With respect to the third research question, the results of the multiple regression analyses shed light on the relationship between recast effectiveness and the amount of repair, initial pronunciation levels, explicit knowledge, motivation, and LOR. First, the degree of recast effectiveness on F3 values can be explained by initial pronunciation levels, arguably because the beginners (whose F3 values were relatively high) had greater room for improvement (Derwing & Munro, 2005). In line with previous L2 speech studies in natural settings (Flege, 1988; Meador et al., 2000), the results of Study 2 suggest that recasts (i.e., additional exposure to enhanced L2 input) might be highly beneficial especially in the early stages of L2 learning (e.g., F3 values > 2700-2800 Hz), but recasts might not lead to a significant improvement for advanced learners (e.g., F3 values < 2400-2500 Hz). Given that LOR did not appear to be a significant variable, LOR might not be a good index for learners' proficiency level (i.e., some learners with short LOR can produce English /ɹ/ with low F3 values).

Second, the results also showed that the existence of explicit knowledge about tongue movement contributed to a decline in F3 values to some degree (β = -.187 and -.166 for front and back vowel contexts respectively). From a theoretical perspective, the results support the view that explicit knowledge can be facilitative of SLA processes as way to help learners notice target features in L2 input (Spada & Lightbown, 2008). From a pedagogical perspective, it seems reasonable to speculate that teachers should provide explicit information about relevant oral gestures at the beginning of lessons in order to make the best of the pedagogical potential of pronunciation-focused recasts. This recommendation echoes other L2 pronunciation researchers

who claim the importance of "explicit phonetic instruction" (e.g., Derwing & Munro, 2005) and "phonological awareness" (Venkatagiri & Levis, 2007). To this end, a pedagogically (and theoretically) intriguing topic to pursue would be to examine the effects of different types of corrective feedback such as metalinguistic correction, which not only provides a reformulation but also metalinguistic information (e.g., Ellis et al., 2006; Sheen, 2007). Last, Study 2 confirmed some influence of motivation on recast effectiveness, which concurs with similar findings in naturalistic settings (e.g., Flege et al., 1995, 1999). Given that very few studies have examined the relationship between individual differences and recast effectiveness (cf. Sheen, 2008), more research is called for to further examine this topic.

Notably, although the rate of repair in response to recasts was considerably high (i.e., 81%), which concurs with other descriptive studies (e.g., Ellis et al., 2001; Lyster, 1998b), the amount of output the learners actually produced (the amount of repair) did not significantly relate to recast effectiveness on F3 values, supporting the view in L2 grammar studies that recast effectiveness might not stem from learner repetition of recasts (e.g., Leeman, 2000; Loewen & Philp, 2006; Lyster & Izquierdo, 2009). Interestingly, the amount of repair proved to be a significant predictor for a change in F2 values. That is, the more repair the learners generated, the more attention they paid to tongue backness (marked in the Japanese phonology system). Therefore, it is possible that, when pushed to repair their mispronunciation of English /ɹ/, the learners might have reinforced their default strategy (i.e., moving the tongue backwards to produce English /ɹ/), at least at the initial stage of L2 pronunciation development.

Limitations

Given that the original study (Studies 1 and 2) is the very first attempt to examine the acquisitional value of pronunciation-focused recasts, several shortcomings must be

acknowledged with an eye toward future replication. First, all of the relevant findings need to be interpreted with caution because it remains to be determined to what degree the learners can generalize a change in F3 values for novel tokens. That is, these promising results could be due to the possibility that the learners benefitted from many opportunities to imitate, without much attention to sound-sized units of L2 information. After all, they did not improve on novel tokens (i.e., *item-based* learning instead of *rule-based* learning). At the same time, however, the lack of statistical significance in the generalizability task could be attributed to the construct validity of its design. For example, one might wonder how we can really measure generalizability simply by asking learners to read a list of four words only at the post-test sessions. That is, the results may have been different if the learners had read a number of unfamiliar tokens in a wide variety of tasks (e.g., Word Reading, Sentence Reading and Timed Picture Description) both at pre- and post-test sessions. Second, Study 2 did not provide clear evidence that recasts actually impacted not only learners' controlled but also spontaneous production of /ɹ/. For example, although the timed picture description task was originally designed to measure learners' spontaneous speech production, the learners might have simply but carefully read the orthographic representations of the target words while describing the pictures with a great deal of focus on form. Future research needs to further elaborate outcome measures of this kind.

Conclusion

The purpose of Study 2 was to investigate how recasts impact L2 pronunciation development of English /ɹ/ by Japanese learners of English and whether and how recast effectiveness varies according to five independent variables—amount of repair, initial pronunciation levels, explicit knowledge, motivation, and LOR. First, the comparison between the experimental group (FT with recasts) and the control group (FT without recasts) showed that,

although both the experimental and control groups equally reduced F2 values (the less relevant cue for English /ɹ/) regardless of whether or not they received recasts, recasts did play a pivotal role in inducing only the experimental group to lower their F3 values (the primary cue for English /ɹ/). Overall, the learners who received recasts significantly enhanced their performance of English /ɹ/ from confusing exemplars to poor-to-good exemplars. These findings indicate that pronunciation-focused recasts are effective ways to draw the learners' attention to new oral gestures (i.e., pharyngeal and velar constriction and lip rounding), which appear to effect a significant decline in F3 values. Finally, a set of multiple regression analyses identified three predictors of recast effectiveness on F3 values: initial pronunciation levels, explicit knowledge, and motivation. Thus, recasts might be effective especially for the beginner learners with relatively high F3 values who continue to neutralize the English /ɹ/-/l/ contrast despite their explicit knowledge of relevant oral gestures and high motivation for L2 pronunciation learning; the influence of the amount of repair and LOR on recast effectiveness could be very minor at best.

TRANSITION FROM STUDY 2 TO STUDY 3

With a general absence of research specifically investigating the pedagogical capabilities of FFI in L2 pronunciation development, Study 1 took a first step towards testing which combination of FFI techniques (FT ± CF) can be relatively effective in the area of pronunciation teaching contexts. The results showed that (a) exposing learners to FT alone is not sufficient to promote acquisition and (b) providing CF (i.e., recasts) to learners' mispronunciation during FT treatment plays an important role in leading to change in their L2 pronunciation performance. Then, in order to portray a precise picture of which individual variables interact to determine FFI effectiveness, Study 2 carried out a set of secondary analyses on the original dataset. The results revealed that (a) FFI especially with CF (i.e., recasts) significantly impacted both primary and secondary acoustic properties of English /ɹ/ produced by Japanese learners of English; (b) FFI (FT+CF) effectiveness on F3 values (the primary acoustic property of English /ɹ/) was related to the learners' initial pronunciation levels, the amount of explicit knowledge about relevant articulatory gestures, and motivation. Building on the results of Studies 1 and 2 whereby two types of FFI (i.e., FT+CF > FT-only) had differential effects , Study 3 will operationalize FFI as a combination of FT and CF as a potentially ideal way to lead learners to notice and practice L2 pronunciation features in meaning-oriented lessons.

However, although Studies 1 and 2 both highlight the critical role of FFI (FT+CF) in L2 pronunciation development, they also revealed several methodological limitations as well as empirical questions which future research needs to answer to obtain a better understanding of the relationship between FFI techniques and L2 pronunciation development. These issues include (a) the *generalizability* of FFI effectiveness (i.e., whether learners can extract segmental aspects of

L2 phonological information from FT and CF treatments and apply them to new lexical contexts), (b) the amount of improvement (whether learners can proceduralize more targetlike representations), (c) the impacts of FFI on L2 pronunciation development at a spontaneous-speech level (i.e., whether learners can spontaneously process their newly-acquired phonetic knowledge in production), and (d) the possibility of other remedial techniques (i.e., whether adding *explicit phonetic instruction* [EI] to the beginning of FFI lessons can enhance the overall FFI effectiveness). Study 3 was designed to respond to all these relevant concerns.

STUDY 3: RE-EXAMINING THE EFFECTS OF FORM-FOCUSED INSTRUCTION ON L2 PRONUNCIATION DEVELOPMENT: THE ROLE OF EXPLICIT INSTRUCTION

Second language acquisition (SLA) studies have examined the effects of form-focused instruction (FFI) as "a set of psycholinguistically motivated pedagogic options" (Ellis, 2001, p. 12), finding that FFI can impact learners' developing system of second language (L2) morphosyntax not only at a controlled level but also at a spontaneous level (e.g., Doughty, 2003; Ellis, 2006; Norris & Ortega, 2000; Spada & Tomita, 2010). With a general absence of research specifically investigating the pedagogical capabilities of FFI in L2 pronunciation development, the original study (Studies 1 and 2) took a first step towards testing how a range of form-focused instructional (FFI) techniques can promote the acquisition of the English sound /ɹ/ by adult Japanese learners of English. Although the results showed that L2 pronunciation development can be amenable to FFI, they also revealed several limitations which future research needs to answer (e.g., the moderate improvement only within familiar lexical items). To this end, *Study 3* was designed to examine whether and to what degree providing *explicit phonetic instruction* (EI) at the beginning of FFI lessons can enhance the *generalizability* and *magnitude* of FFI effectiveness by assisting learners to notice the perceptual aspects of a new sound (English /ɹ/) from L2 input, and restructure/develop the new phonetic category in their long-term memory representation.

In Study 3, I will first provide theoretical accounts and empirical evidence for two crucial topics in L2 pronunciation development: (a) how L2 learners become aware of sound-sized units of phonological information as *phonetic rules* and generalize them to unfamiliar lexical contexts (i.e., *representation* issues), and (b) how they make a gradual transition from effortful to

automatic use of the newly-acquired phonetic knowledge in production (i.e., *processing* issues). Subsequently, I will offer several predictions as to which combination of FFI techniques (i.e., FFI ± EI) can be facilitative of L2 pronunciation development at various levels of representation and processing. Finally, I will present the results of a quasi-experimental study with 49 Japanese learners of English designed to assess the generalizability and effects of two types of FFI (i.e., FFI vs. FFI+EI) on learners' L2 pronunciation performance.

L2 Phonological Development

In L2 speech research, many studies have investigated how L2 learners create new phonetic categories as representations in long-term memory through an interaction between the L1 phonetic system and L2 input; not much attention, however, has been paid to how they develop these categories over time and process the newly-acquired phonetic knowledge in a productive mode (e.g., Flege, 1995, 2003, 2007, 2009; Logan, Lively, & Pisoni, 1993; Piske, Flege, & MacKay, 2001). In contrast, psycholinguistic studies of SLA have examined the developmental mechanism of L2 morphosyntax, focusing on (a) how L2 learners establish and adapt abstract linguistic knowledge in memory based on received L2 input (i.e., *representation*) and (b) how they retrieve information in comprehension and production in an effortful or automatic manner (i.e., *processing*) (e.g., de Bot, 1996; de Bot, Paribakht, & Wesche, 1997; Levelt, 1989; VanPatten, 2002, 2004); missing from this SLA-oriented research, however, is a discussion of the phonological domain of language (cf. N. Ellis, 2002). In this section, I first synthesize these two domains in an interdisciplinary manner and provide theoretical accounts and empirical evidence with respect to how L2 phonological development takes place both in naturalistic and instructed settings. This literature review will be used to support my predictions

as to the relationship between a range of FFI techniques and L2 pronunciation development at segmental levels.[19]

Theoretical Issues in L2 Phonological Development

Decoding L2 Phonological Information

The first question to address concerns the level at which L2 phonological development takes place. That is, do L2 learners decode L2 phonological information not only at a lexical level but also at phonetic level? From a theoretical perspective, answering this question directly relates to the fundamental assumptions underlying L2 speech acquisition theories whereby "sound-sized units of speech are important units not only during the process of L2 speech acquisition but also, later, in the online regulation of L2 speech" (Flege, Frieda, Walley, & Randazza, 1998, p. 177; see also Flege, 1995, 2003, 2007, 2009). From a pedagogical perspective, the question of L2 learners' ability to detect new L2 sounds in familiar lexical contexts and generalize their newly-acquired phonetic knowledge to unfamiliar lexical contexts can inform the role of pronunciation instruction in vocabulary teaching contexts, and vice versa (Nation, 2001; Schmitt, 2008).

According to the L1 acquisition literature, infants initially use computational strategies to detect prosodic patterns and then start recognizing words (Jusczyk, 1997; Kuhl, 2000, 2004). Importantly, after words are learned as whole phonological units, "the resulting increased vocabulary could result in sufficient pressure to fill in finer phonetic detail in the lexical representations in order to avoid confusion between similar sounding, known words" (Werker, &

[19] Given that the focus of the current study is to investigate L2 pronunciation development of English /ɹ/ by Japanese learners of English, I do not intend to discuss here L2 suprasegmental learning. However, other researchers have pursued this point (e.g., Trofimovich & Gatbonton, 2006 for instructed settings; Trofimovich & Baker, 2006 for naturalistic settings).

Tees, 1999, p. 523; see also Pierrehumbert, 2003), and L1 phonetic development continues to take place up to adolescence (Walley & Flege, 2000).

Interestingly, similar phenomena can be observed in L2 phonological development. For example, whereas some studies showed that lexical factors such as lexical familiarity (i.e., the degree to which learners have familiarity with L2 words) and lexical density (i.e., whether L2 words are phonologically similar or dissimilar to other words[20]) affected L2 learners' perception abilities (e.g., Bradlow & Pisoni, 1999; Flege, Takagi, & Mann, 1996; Imai, Walley, & Flege, 2005; Meador, Flege, & MacKay, 2000), other studies evidenced that they could also decode segmental aspects of L2 phonological information as "new phonetic realization rules" and apply them to new lexical contexts (Flege et al., 1998, p. 117; see also Flege, Takagi, & Mann, 1995). Baker and Trofimovich (2008) stated:

> L2 learners are sensitive to phonological regularities at two levels of abstraction. They are sensitive to *sound-level regularities* in variation of individual L2 segments, making phonological generalizations across specific phonetic, syllabic, phonotactic, or prosodic contexts… L2 learners are also sensitive to higher-order *word-level regularities* within L2 lexicon, making phonological generalizations both within and across L2 lexical items. (*Emphasis added*, p. 48)

Taken together, it seems reasonable to assume that, like L1 acquisition, L2 learners' decoding of L2 phonological information at a lexical level for the purpose of successful L2 communication is

[20] Neighborhood density (i.e., phonological similarity) between words can be determined (a) substitution (e.g., "late" vs. "rate"), (b) deletion (e.g., "late" vs. "ate"), and (c) addition (e.g., "late" vs. "plate") (for detail, see Bradlow & Pisoni, 1999; Imai et al., 2005)

at first part of their vocabulary learning. L2 phonological development at this stage could be considered as *item-based learning*. As their vocabulary size increases, however, they will be forced to attend to fine-grained phoneme discrimination and identification in order to accurately comprehend and produce a large lexicon containing many confusing minimal pairs. Ultimately, while they are sensitive primarily to word-sized units of L2 phonological information, they concurrently become more capable of detecting new sounds in L2 input at a phonetic level. Many L2 speech researchers claim that this phonetic-level restructuring, in particular, leads learners to create new phonetic categories and to generalize the newly-acquired phonetic knowledge from familiar to new lexical contexts. L2 phonological development at this stage could be considered as *rule-based learning* (Best, 1995; Best et al., 2001; Flege, 1995, 2003, 2007, 2009; Kuhl, 2000). In this respect, segmental pronunciation instruction should be designed to (a) raise learners' noticing and awareness of L2 phonological information not only at a lexical level but also at a phonetic level, and (b) promote the formation of new phonetic categories in relation to their increasing vocabulary sizes. This bidirectional decoding procedure is visually summarized in Figure 6.

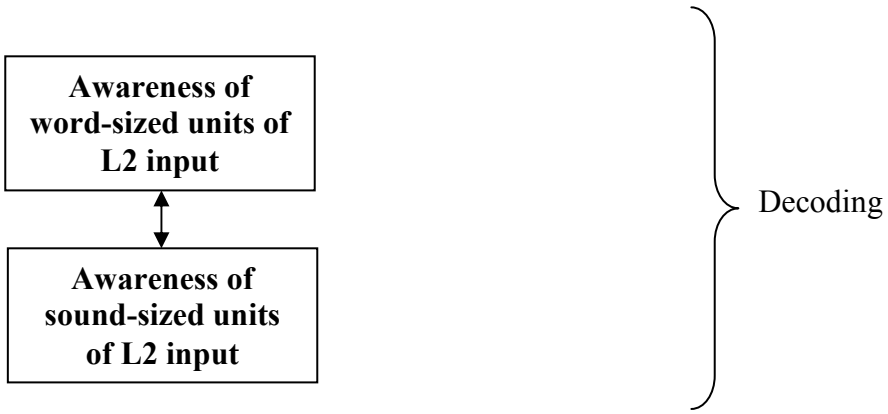


Figure 6. Decoding information during L2 phonological development

Phonetic Categories, Perception, and Production

I will turn my review next to how L2 learners establish new phonetic categories in their long-term memory representations and develop them over time in naturalistic and instructed settings. Among several well-known L1 and L2 speech learning models, such as the Perceptual Assimilation Model (PAM; Best, 1995; Best et al., 2001) and the Native Language Magnet (NLM; Kuhl, 1991, 2000; Kuhl et al., 1992), only Flege's Speech Learning Model (SLM) specifically relates to L2 speech *perception* and *production* at a phonetic level (Flege, 1995, 2003, 2007, 2009; Piske, MacKay, & Flege, 2001).[21] The main claim about the SLM is that speech learning capacity used in successful L1 speech acquisition remains active throughout life and can be applied to L2 speech acquisition. This controversial view sharply contrasts with other L2 researchers who claim that age-related effects in SLA are due to a loss of plasticity resulting from neural maturation after adolescence (e.g., DeKeyser, 2000; Johnson & Newport, 1989; Scovel, 1988). Whereas Flege agrees that early bilinguals (who immigrate to L2 speaking countries at an early age) can perceive and produce L2 sounds better than late bilinguals (who immigrate to L2 speaking countries at a later age) (e.g., Flege, Munro, & MacKay, 1995a, 1995b; Flege, Yeni-Komashian, & Liu, 1999), he attributes the advantage of early bilinguals to the interference between L1 and L2 phonetic systems in a common phonological space instead of to neurobiological changes in the brain system (i.e., the critical period).

> The phonetic elements making up the L1 sound system and the phonic elements comprising the L2 system (either newly established categories or adaptations of L1 categories) exist in a "common phonological space," and so will mutually influence one

[21] In this respect, Flege (1995) hypothesized that "sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more abstract phonemic level" (p. 239).

> another… [A]s L1 phonetic categories develop, they become more likely to perceptually "assimilate" the vowels and consonants encountered in an L2 (Flege, 2007, p. 365)

According to the SLM, the degree of L1 and L2 interference can be determined by the perceived phonetic distance of L2 sounds from L1 counterparts. That is, faced with non-native sounds, L2 learners resort to either *dissimilation* or *assimilation.* When L2 learners identify a new L2 sound as perceptually distant from the closest L1 counterpart, *dissimilation* occurs: They make a new phonetic category and strive to maintain its phonetic contrast relative to L1 counterpart. When a new L2 sound is perceived as similar to the closest L1 sound, *assimilation* occurs: L2 learners make a merged intermediate category that includes both L1 and L2 phonetic features (see also Best et al., 2001; Kuhl, 2000).

The best example is the acquisition of the non-native contrast between English /ɹ/ and /l/ by Japanese learners of English. Japanese has only one alveolar tap /ɾ/ and its acoustic features are rather close to English lateral /l/ (Hattori & Iverson, 2009). A number of cross-language mapping studies have also shown that Japanese learners of English tend to judge English /l/ tokens as more similar to the Japanese tap /ɾ/ than to English /ɹ/ tokens (e.g., Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Sekiyama and Tohkura, 1993; Takagi, 1993). As a result, it has been found that Japanese learners of English tend to make greater learning strides in their perception and production of /ɹ/ than /l/ because the former is perceptually distinguishable for Japanese learners of English (*dissimilation*) and the latter is likely confused with Japanese tap /ɾ/ (*assimilation*). For example, Sheldon and Strange (1982) found that Japanese learners of English in the USA identified word-initial /ɹ/ better than word-initial /l/ (96% vs. 82%). Similarly, Flege, Takagi, and Mann (1996) found that adult Japanese learners of English regardless of length of

residence (21 years and 2 years) identified /ɹ/ more correctly than /l/ (96% vs. 81% and 87% vs. 53%, respectively). Aoyama et al. (2004) examined whether child and adult Japanese learners of English exhibited a different patterns in their acquisition for /ɹ/ and /l/ over the course of one year of residence in the USA. The results showed that child learners demonstrated more learning in their abilities to perceive and produce /ɹ/ than /l/.

Last, the SLM hypothesizes that L2 speech learning first occurs in the perception domain, which, in turn, activates relevant sensorimotor skills for production abilities. Namely, as Flege (1995) pointed out, "many L2 production errors have a perceptual basis" (p. 238) and this *perception-first view* is well-established in both L1 and L2 speech studies. In the L1 acquisition literature, whereas babies perceptually map critical aspects of their L1 phonetic inventories from the ambient language input in the first year before they speak, "vocal learning critically depends on hearing the vocalizations of others and hearing oneself produce sound" (Kuhl, 2000, p, 11854; see also Kuhl. 2004, Kuhl et al., 1992). In the L2 acquisition literature, Bradlow and her colleagues tested the perception-production link, finding that Japanese learners of English who received only intensive perception training on the /ɹ/-/l/ contrast not only enhanced their perception abilities but also transferred the gain to the production phase both in the short and long term (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999) (for similar behaviour results, see also Hardison, 2003; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Trofimovich, Lightbown, Halter, & Song, 2009). More recently, Zhang, Kuhl, Imada, Iverson, Pruitte, Stevens, Kawakatsu, Tohkura, and Nemoto (2009) used magnetoencephalography to examine how intensive perceptual training (English /ɹ/-/l/ contrast for Japanese learners of English) enhances neural activities relevant for L2 speech production. The results provided some evidence (i.e., increases in the left inferior

frontal regions) that "phonetic learning may strengthen the perceptual–motor link by recruiting the Broca's area" (p. 237).

Intriguingly, the input-driven view of SLA has been also extensively discussed in the domain of L2 morphosyntax studies (for review, Gass, 1997). For example, VanPatten (2002, 2004) maintains that L2 learners extract linguistic information from L2 input in order to enhance the quality of the L2 developing system, which is responsible for their performance in output at a later stage of SLA processes. VanPatten theorizes the relationship between input, intake (a subset of input learners comprehend), L2 developing system, and output as follows:

> Input provides the data, IP (input processing) makes (certain) data available for acquisition, other internal mechanisms accommodate data into the system (often triggering some kind of restructuring or a change of internally generated hypothesis), and output helps learners become communicators and, again, may help them become better processors of input. (VanPatten, 2002, p. 762)

Following his theoretical model of SLA, VanPatten and his colleagues conducted a series of empirical studies, providing some evidence that L2 learners who received only input-based practice as a form of processing instruction without any pressure to actually produce L2 output improved not only their comprehension abilities but also production abilities (e.g., VanPatten, 2004; VanPatten & Cardierno, 1993; VanPatten & Oikennon, 1996; cf. DeKeyser, Salaberry, Robinson, & Harrington, 2002; DeKeyser & Sokalski, 1996).

The SLM (and the relevant empirical evidence) posits that (a) when L2 learners hear *perceptual dissimilarity* between a new L2 sound and its L1 counterpart, they start creating the

novel phonetic category in the common phonological space, and (b) when the category is specified as a representation in long-term memory, change occurs first in the perception domains and, subsequently, in the production phase. The perception-first view has been well recognized in the domain of L2 morphosyntax studies (VanPatten, 2002, 2004). The SLM is visually summarized in Figure 7.

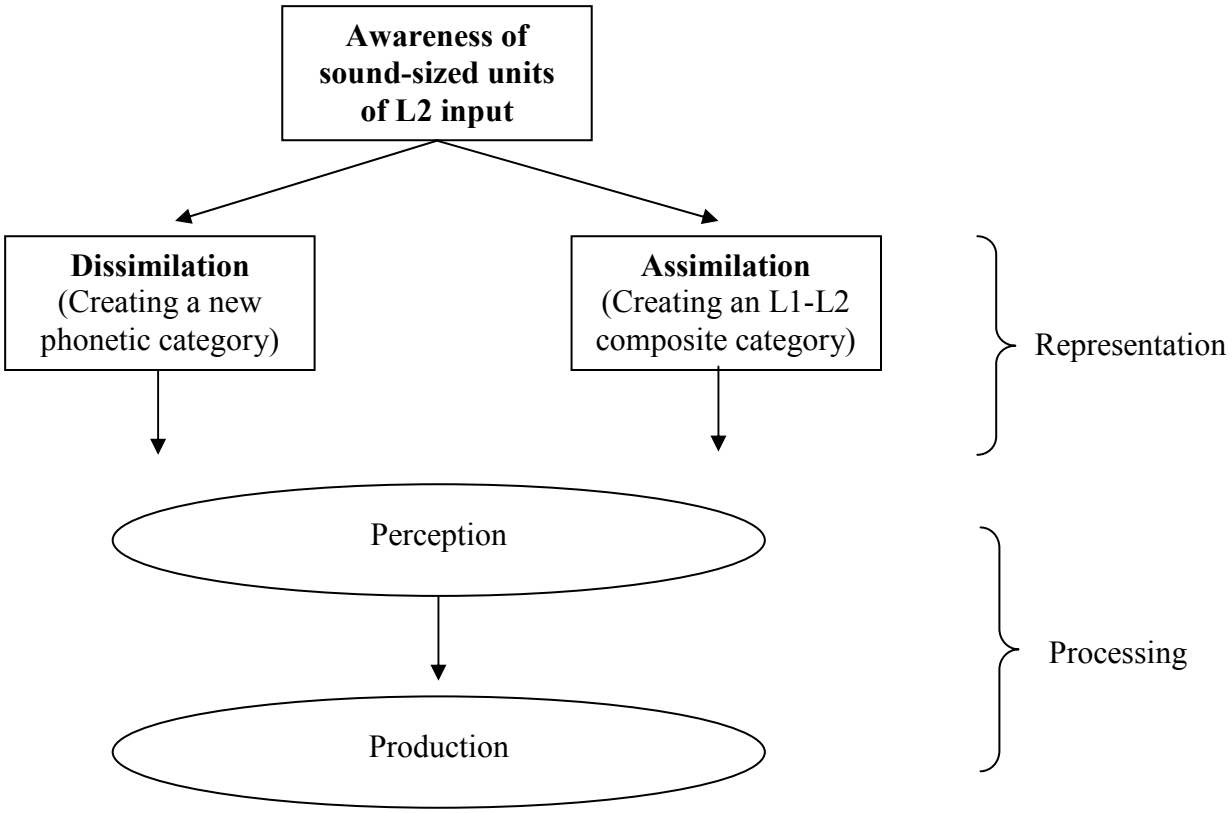


Figure 7. Summary of the SLM

Representation and Processing in L2 Phonological Development

The SLM depicts L2 phonological development in naturalistic settings as follows: noticing perceptual dissimilarity between a new L2 sound and its L1 counterpart → establishing

new phonetic categories in long-term memory representation → change in perception → change in production. However, it is noteworthy that this model lacks information with respect to two crucial aspects of L2 phonological development: (a) *the representational nature of L2 phonetic categories*, and (b) *different types of processing in the production phase*. Whereas the former point relates to what types of *representation* learners develop and internalize over time in relation to the quality and quantity of L2 input (i.e., context-specific vs. context-invariant representation), the latter point concerns how they process their newly-acquired phonetic knowledge in production (controlled vs. spontaneous speech production). To elaborate, I will further examine these representation and processing issues in L2 phonological development by interfacing relevant empirical evidence both in L2 speech and psycholinguistics studies.

Representation in L2 Phonological Development

Although the SLM predicts that L2 learners become aware of a new L2 sound in relation to its L1 counterpart, and start restructuring and creating a new phonetic category in the phonological space, it does not specify how the nature of the category changes over time according to the quality and quantity of received L2 input. This could be due to the fact that Flege's research has exclusively focused on immigrants and measured their L2 experience in a retrospective way (e.g., self reports of frequency of L1 and L2 use). Recently, Flege (2009) referred to this point:

> In many cases, researchers have not attempted to measure L2 input because they assumed that doing so is impossible. Indeed, practical and ethical limitations would prevent researchers from videotaping, and then subjecting to quantitative analysis, the input received over years of a person's daily life (p. 188)

In this regard, some L2 speech researchers have conducted intensive perception training studies where they can control the quality and quantity of L2 input and make precise descriptions of how different types of L2 input enhance the nature of perceptual representations (see Bradlow, 2008 for review). On the one hand, Strange and Dittman (1984) found that Japanese learners of English who were trained to perceive the /ɹ/-/l/ contrast via a number of synthesized stimuli along a "rock"-"lock" continuum (with F3 value variation from 1477 to 2594 Hz) could generalize their gain to a novel lexical context (i.e., a different synthetic continuum of "rake-lake") but not to a novel task (i.e., identifying natural minimally-paired /ɹ/-/l/ tokens). Similarly, Lively, Logan, and Pisoni (1993) showed that Japanese learners of English who were exposed to natural /ɹ/-/l/ tokens by a single NS talker could generalize their gain to new phonetic and lexical contexts to some degree but not to a new talker at all. In sum, these types of L2 input (synthesized stimuli, natural stimuli by a single talker) tend to lead L2 learners to develop *context-specific representations* according to lexical contexts (i.e., trained tokens vs. novel tokens), phonetic contexts (e.g., English /ɹ/-/l/ in word-initial, -medial, and -final positions) and talker contexts (familiar talkers vs. new talkers).

On the other hand, Lively and his colleagues conducted a series of studies to test effects of High Variability Phonetic Training (HVPT; i.e., intensive exposure to natural L2 tokens produced by many talkers) on the acquisition of the /ɹ/-/l/ contrast by Japanese learners of English. The results showed that (a) learners who received HVPT (30 minutes × 15 sessions = 7.5 hours) could identify English /ɹ/-/l/ contrast at post-test sessions more successfully than at pre-test session; (b) the gain resulting from HVST was generalized to novel lexical contexts and new talkers; and (c) improvement was sustainable for six months without any additional training (Lively et al., 1993, 1994; Logan et al., 1991). Finally, the extended HSVT (30 minutes × 45

sessions = 22.5 hours) allowed learners to achieve perfect generalization (Yamada, 1993) and transfer the learning to the production domains (Bradlow et al., 1997, 1999).

Given that exaggerated speech stimuli are hypothesized to be facilitative of discriminating non-native contrasts as shown in the L1 acquisition literature (Kuhl, 2000), other researchers investigated the acoustic enhancement technique (i.e., learners first receive synthesized tokens whose acoustic difference between English /ɹ/ and English /l/ is enhanced, and gradually move on to listening to natural speech tokens) (e.g., McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). Furthermore, Iverson, Hazan, and Bannister (2005) found that this technique was as effective as HSVT in terms of the amount of improvement and generalizability. In sum, intensive exposure to certain types of structured input (i.e., HSVT and the acoustic enhancement) leads L2 learners to achieve "robust and highly generalized improvements in speech perception and production" (Bradlow, 2008, p. 300), and this learning could be evidence for how L2 phonological development entails a transition from context-specific representation to *context-invariant representation*. In this regard, the SLM could be slightly revised in order to take into account such change in new phonetic representation during L2 phonological development and the model is visually summarized in Figure 8.

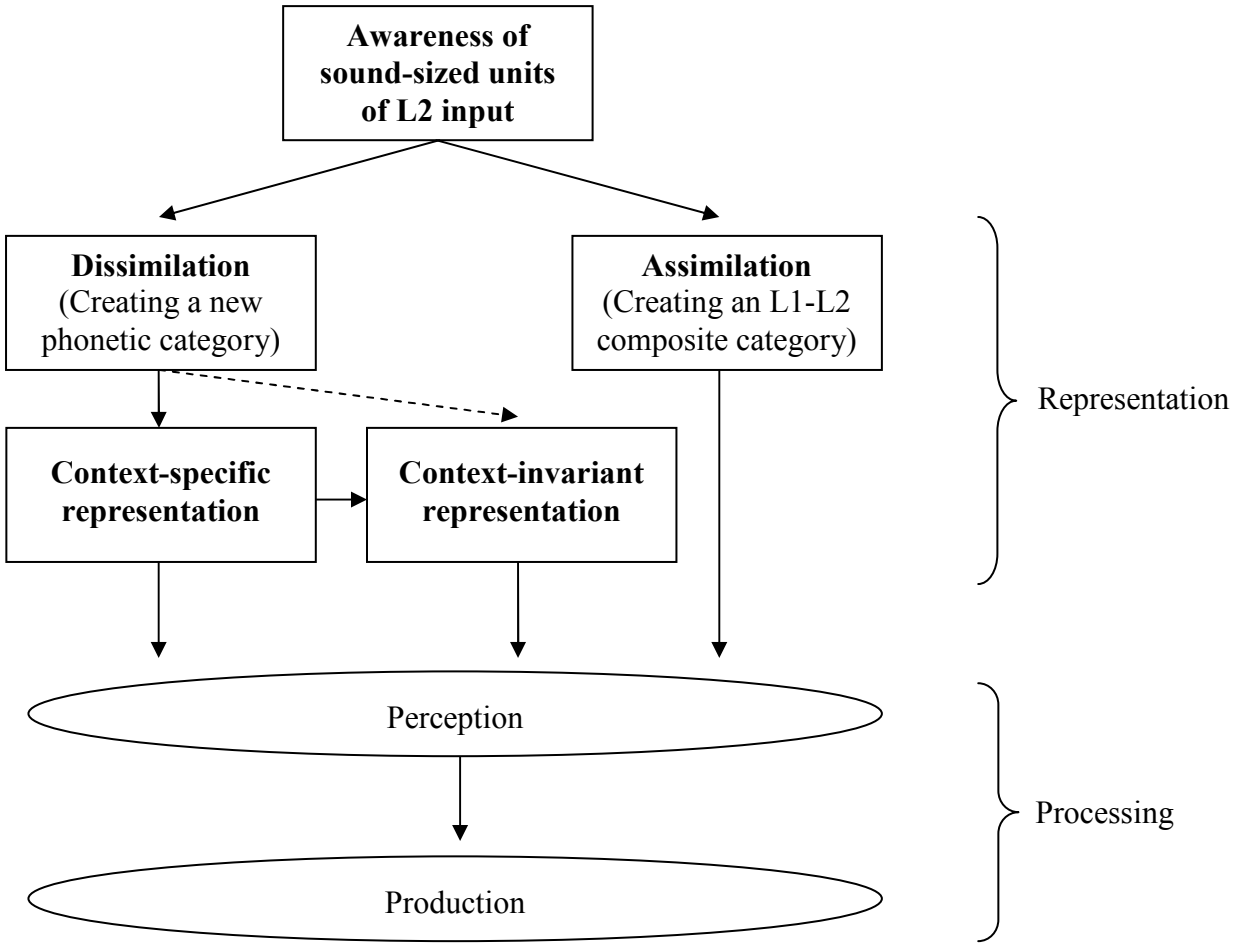


Figure 8. Change in new phonetic representation during L2 phonological development

Processing in L2 Phonological Development

Although L2 speech researchers agree that L2 learners' processing of new L2 sounds in the perception phase is relatively automatic and that perception test scores are thought to mirror their mental representation, it still remains open to debate to what degree their productive use of new L2 sounds is related to their developing L2 phonology system. For example, whereas L2 learners are assumed to produce new L2 sounds if change occurs in perception abilities (Flege, 1995, 2003, 2007, 2009), cognitive psychology research posits that L2 learners produce output through a gradual transition from effortful to automatic use of newly-acquired L2 knowledge (e.g., DeKeyser, 1998, 2001, 2003, 2007). Very importantly, this line of SLA research has also demonstrated that L2 learners can carefully produce language (and even speed up retrieval) without fully integrating linguistic knowledge into their mental representation (for a detailed discussion, see Segalowitz, 2003). In the context of L2 segmental production, L2 learners can consciously activate relevant articulatory gestures in order to produce these sounds especially at a controlled-speech level where they are given a sufficient amount of time to access their explicit knowledge stored in general memory. This could be especially true in the case of adult instructed SLA processes (e.g., these learners are explicitly taught about language through rule presentation but without many opportunities to proceduralize this declarative knowledge in authentically communicative contexts). However, little research attention has been given to the mechanism of processing with respect to L2 speech *production*. Thus, I first review how psycholinguistic studies have examined these processing issues in L2 morphosyntax and then apply these empirical findings to discuss processing within the paradigm of L2 speech acquisition.

According to the Lexical Hypothesis proposed by Levelt (1989), L2 speech production is *lexically-driven*. That is, Levelt stated that "the preverbal message triggers lexical items into

activity. The syntactic, morphological, and phonological properties of an activated lexical item trigger, in turn, the grammatical, morphological, and phonological encoding procedures underlying the generation of an utterance" (p. 189). Notably, Levelt argued that the internal structure of the mental lexicon consists of (a) morphological and phonological information as "lexeme" (i.e., form) and (b) syntactic and semantic information as "lemma" (i.e., meaning). Jian (2000, 2004, 2007) further conceptualized how L2 learners succeed or fail in integrating morphological information into the lexeme domain of their mental lexicon. When morphological information is specified in the mental lexicon, L2 learners can achieve "a one-step, automatic process of morphological production" that can be retrieved and applied spontaneously (Jian, 2000, p. 57), which, he argued, could be the ultimate goal of L2 learning and teaching.

In contrast, when L2 learners learn and remember morphological information as explicit knowledge, they follow *a two-step process*: selecting both a root form (e.g., "like") and an appropriate morphological form based on one's morphological knowledge (e.g., adding "-ed" in order to mark past tense). Jian (2000) described the two-step process as a typical characteristic of performance deficiency in interlanguage development:

> Such a two-step morphological process is not only less automatic. Morphological errors occur as well. The application of morphological knowledge is most likely to be a conscious process requiring attentional resources. When one's attention is focused on the message to be communicated, morphological errors result. (p. 58)

In order to examine the fundamental difference between the one-step process (i.e., integrated knowledge) and two-step process (i.e., non-integrated knowledge) in the morphosyntactic

domain of language, Jian (2004) conducted a series of experiments to investigate how morphological errors affect the performance of sentence-level comprehension tasks by advanced Chinese learners of English with considerable explicit knowledge. The results showed that, whereas native speakers (NSs) of English slowed down their reading speed for number disagreement sentences (including morphological errors) compared to number agreement sentences (including no morphological errors), Chinese learners of English demonstrated no speed difference between the two types of sentence reading tasks especially in terms of inflectional morphemes (e.g., third person singular -*s*, plurals, and past tense in English). Jian interpreted their insensitivity to inflectional morphemes even under no pressure for productive use of L2 (i.e., sentence-level comprehension tasks) as evidence that L2 learners with perfect performance in written grammar tests have not necessarily internalized the knowledge into the mental representation (see also Bialystok, 1978). However, he also claimed that, once morphological information is integrated into the mental representation, it can be retrieved automatically in both comprehension and production phases (see also Jian, 2007 for other behavioural evidence and further discussion; Guo, Guo, Yan, Jiang, & Peng, 2009 for neurological evidence).

As opposed to L2 morphosyntax research mentioned above, spontaneous production has not been discussed sufficiently in the context of L2 phonological development. Surprisingly, L2 speech research has traditionally measured learners' production abilities by asking them to read a list of words or sentences without any communicative use of language. In fact, it has been found that some Japanese learners of English could produce the /ɹ/-/l/ contrast very well at a controlled-speech level but could not perceive the non-native contrast (see Goto, 1971 for intermediate learners in Japan; Sheldon & Strange 1982 for advanced learners in the US). Although these

researchers claimed, based on the results, that production precedes perception, I would argue that these learners could have carefully produced these words drawing on their explicit knowledge of relevant articulatory gestures for the /ɹ/-/l/ contrast (the two-step process) rather than having accessed new phonetic categories in long-term memory (the one-step process).

From a more sociolinguistic perspective, some L2 phonology researchers have examined how L2 learners pronounce certain sounds differently according to *style* (formal and less formal tasks). Given that Chinese learners use different interlanguage strategies to overcome consonant clusters in English, such as deletion (a less advanced strategy with more obtrusiveness to interlocutors' comprehension) and schwa insertion (a more advanced strategy with less obtrusiveness to interlocutors' comprehension), Lin (2001, 2003) found that they tended to insert schwa vowels (the more advanced strategy) in formal tasks such as minimal-pair reading and sentence reading tasks, but were likely to delete one of two consonants in a cluster (the less advanced strategy) in less formal tasks such as free interview tasks (see also Rau, Chang, & Tarone, 2009 for similar results in regards to interdental fricatives).The change in processing during L2 phonological development is visually summarized in Figure 9.

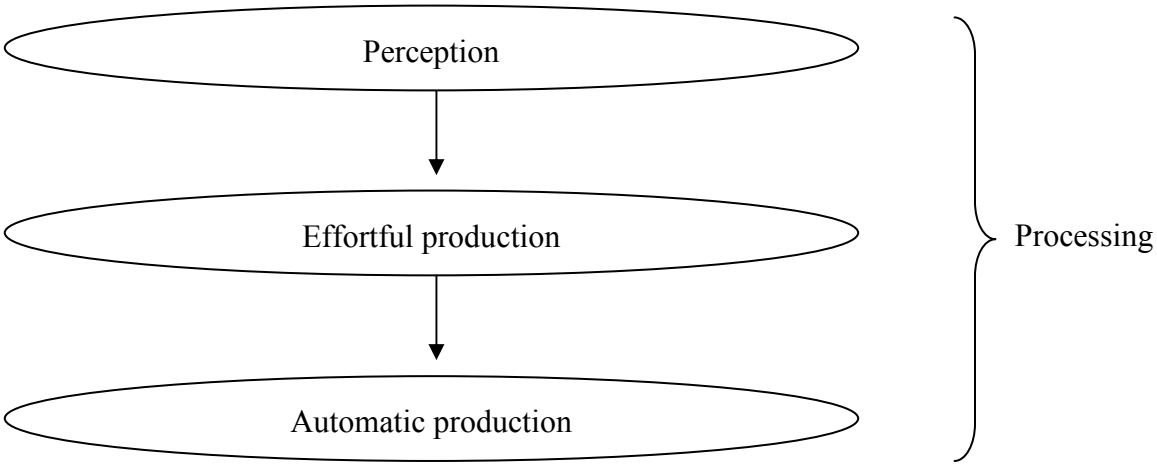


Figure 9. Change in processing during L2 phonological development

Summary

Future L2 speech research needs to examine whether and to what degree L2 learners develop new phonetic categories via a range of *perception measures* because mental representations and perception performance are directly related to each other. Yet, it is also necessary to track the development of mental representations; L2 learners slowly shift from context-specific representation to context-invariant representation according to the quantity and quality of received L2 input (e.g., Logan et al., 1991). Given that L2 learners transfer the learning in the perception phase to their productive use of L2 (e.g., Bradlow et al., 1997), a next step is to adopt various types of *production measures* to assess what type of processing they have access to: (a) effortful production abilities (the two-step process) and (b) spontaneous production abilities (the one-step process). Importantly, L2 morphosyntax studies have demonstrated that the latter process is arguably more related to learners' mental representations than the former process (e.g., Jian, 2000, 2004, 2007). Although some empirical research has shown that L2 learners' controlled production performance can precede their perception abilities (Goto, 1971; Sheldon & Strange, 1982), it remains unclear whether it is possible to retrieve knowledge requiring a great deal of attentional resources especially under communicative pressure (e.g., Segalowitz, 2003). Therefore, it is important to examine both controlled and spontaneous production abilities by elaborating innovative production tasks where learners are induced to spontaneously use these target features attending primarily to communication. The revised model for L2 phonological development is visually summarized in Figure 10.

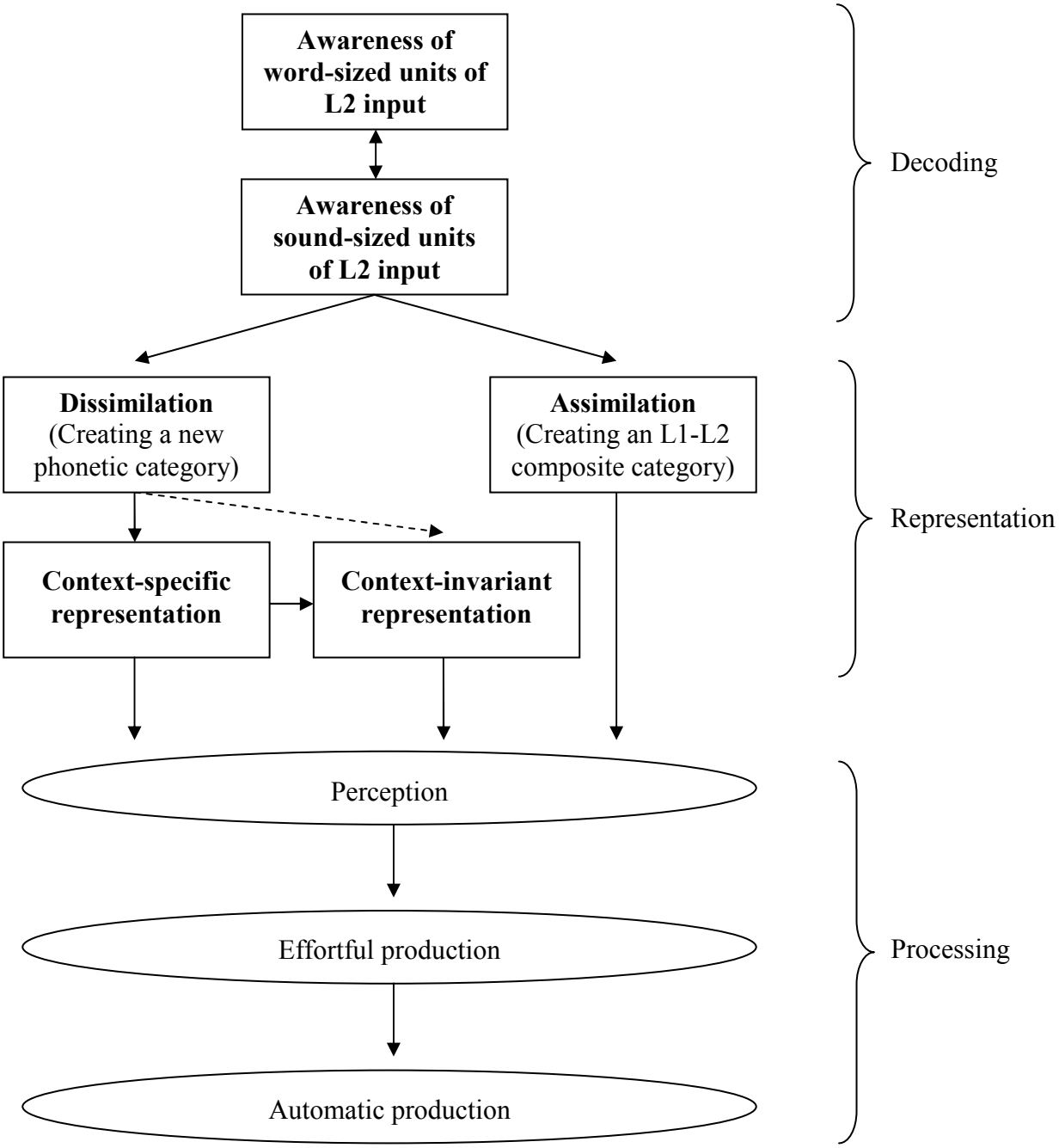


Figure 10. Revised model for L2 phonological development

FFI and L2 Phonological Development

Building upon the previous detailed discussion of L2 phonological development, I will now outline a pedagogical model as to *which*, *when* and *how* FFI techniques should be used to expedite the rate of SLA as well as enhance the ultimate attainment in SLA in the context of pronunciation teaching. FFI is defined as "any pedagogical effort which is used to draw the learners' attention to language form either implicitly or explicitly" (Spada, 1997, p. 73), and it is hypothesized to be most effective when integrated into communicative-oriented and content-based classrooms, because L2 learners can notice and practice target linguistic features during meaningful discourse, which in turn enhances (a) their "form-meaning mappings" (Doughty & Williams, 1998; Long, 1996, 2007; Long & Robinson, 1998; VanPatten, 2002, 2004) and (b) their gradual transition from effortful to automatic use of rules (DeKeyser, 1998, 2003, 2007; Lyster, 2007). In particular, Lyster (2004b, 2007; Ranta & Lyster, 2007) developed a pedagogical sequence of FFI in relation to three stages of interlanguage development: (a) *noticing phase*, (b) *awareness phase*, and (c) *practice phase*.

According to Lyster, FFI activities should be designed first to promote learners' noticing of a target language feature in L2 input especially at the initial stage of interlanguage development (*noticing phase*) and then to push learners to analyze the target feature with some degree of *elaboration* (*awareness phase*). Finally, after learners successfully restructure and develop interlanguage representations, they are ready to engage in FFI activities to repetitively practice the target feature in production under communicatively authentic contexts (*practice phase*). The main goal at this stage is to proceduralize their declarative knowledge, and providing output-prompting CF could be a very effective technique. Several suggested FFI activities in relation to the three stages are summarized in Figure 11.

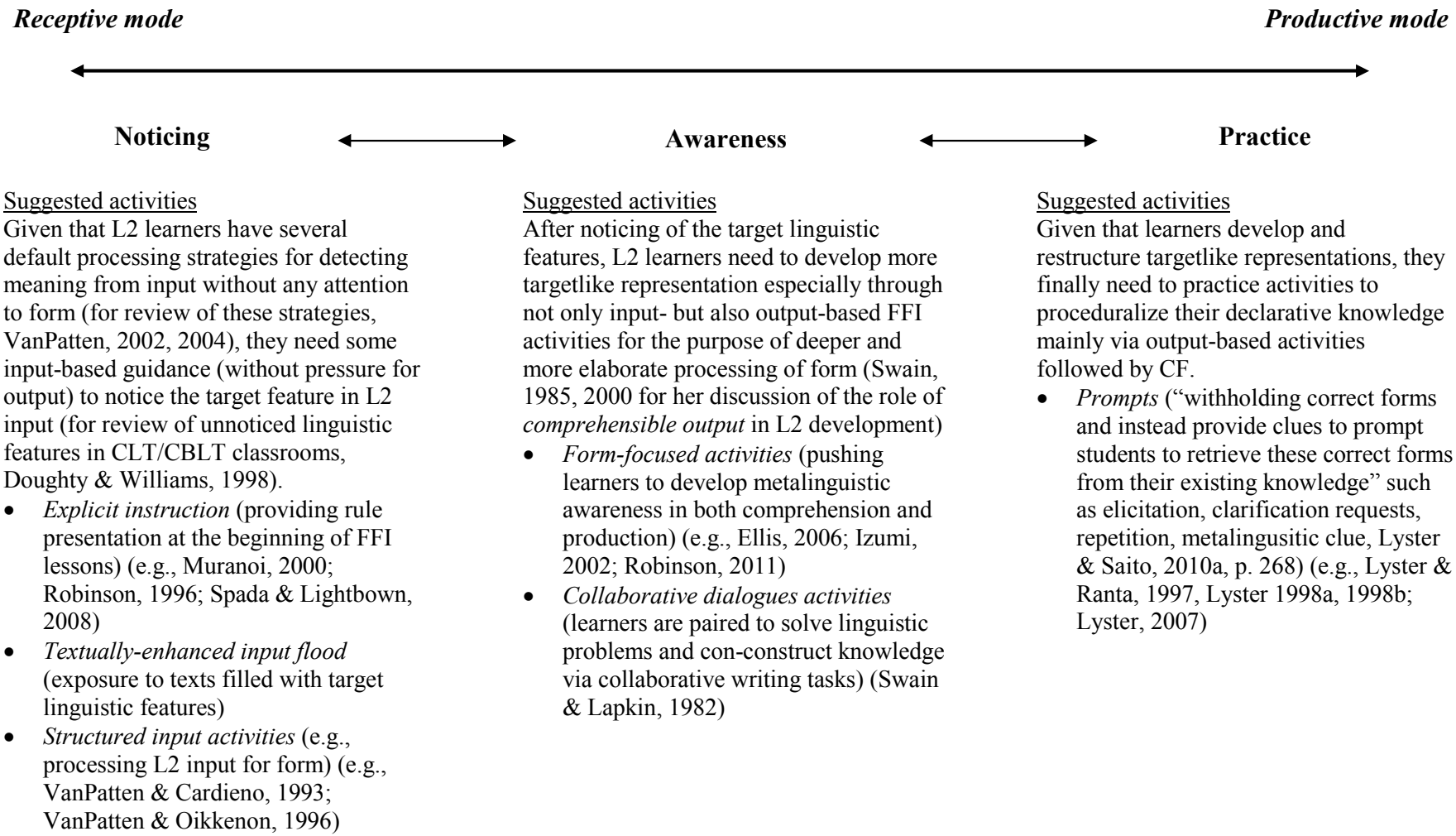


Figure 11. The pedagogical sequence of FFI techniques adapted from Lyster (2004, 2007; Ranta & Lyster, 2007)

Given that Studies 1 and 2 were the first attempt at testing Lyster's FFI model in the area of pronunciation teaching contexts, in this section, I will first summarize the findings as well as limitations of FFI effectiveness on L2 pronunciation development techniques in conjunction with the results of the original study, and subsequently discuss the possibility of adding EI to the beginning of FFI lessons as a remedial technique (the focus of Study 3).

Findings and Limitations of FFI

Whereas a number of observation studies have shown that L2 learners are sensitive to pronunciation-focused CF in that it generates a great deal of immediate repair (Carpenter et al., 2007; Ellis et al., 2001; Kim & Han, 2007; Han, 2008; Lyster, 1998b; Lyster & Saito, 2010b; Mackey et al., 2000; Sheen, 2006), Studies 1 and 2 examined how two types of FFI (FT ± CF) can be facilitative of L2 pronunciation development of English /ɹ/ by Japanese learners of English. One example of the CF treatment during the Argument Critique activity (i.e., FT treatment) from Study 3 is as follows:

Example 7

S: I love to eat rice /laɪs/*.

T: Rice /ɹaɪs/.

S: Rice /ɹaɪs/. I have tried many kinds of rice /ɹaɪs/.

The results showed that only the FT+CF group outperformed the control group, but the FT-only group did not. The critical role of CF in FFI effectiveness was attributed to the dual pedagogical function of pronunciation-focused CF: L2 learners are pushed to practice correct pronunciation forms (i.e., pronunciation practice) while carefully listening to a teacher's model pronunciation of English /ɹ/ (i.e., listening practice). However, the original study also generated several questions and was constrained by certain limitations worthy of further research attention.

First, all of the relevant findings need to be interpreted with much caution because it still remains unclear whether and to what degree the learners could generalize changes in F3 values to unfamiliar lexical tokens; learners' gains were found only when their performance was tested via familiar items (which they practiced during four hours of FFI treatment). This point could be used as evidence that the learners receiving FT and CF succeeded in restructuring mental representations at a lexical level (i.e., *lexically-driven* L2 phonological development) but failed to do so at a phonetic level (i.e., *phonetically-driven* L2 phonological development), the latter of which is hypothesized to be necessary for the development and generalization of new phonetic categories to unfamiliar lexical contexts (Best, 1995; Best et al., 2001; Flege, 1995, 2003, 2007, 2009; Kuhl, 2000). At the same time, however, the lack of statistical significance in the generalizability task could be attributed to the construct validity of its design. Arguably, generalizability cannot be ascertained simply by asking learners to read a list of four words only at the post-test sessions. In this respect, a decision was made in Study 3 to ask learners to read a number of familiar and unfamiliar tokens via controlled and spontaneous production tests both at pre- and post-test sessions (this point will be described in detail later).

In addition, the amount of improvement resulting from FFI in Studies 1 and 2 could be considered *moderate* rather than *large*. A close examination of the results showed that the FT+CF group changed their mean F3 values from 2500-2600 Hz to 2300-2400 Hz only within the range of the hybrid-to-poor exemplars of English /ɹ/ (F3 = 2250-2500 Hz) (see Tables 6 and 8). To this end, Study 3 examines the generalizability and magnitude of the effects of FFI on both familiar and unfamiliar lexical items across various task and phonetic conditions.

Last, the construct validity in the spontaneous production test also prevented us from drawing any clear conclusion about whether FFI led learners to create a new phonetic category

for English /ɹ/. For example, although the timed picture description task was originally designed to measure learners' spontaneous speech production, the learners might have simply read the orthographic representations of the target words while describing the pictures with a great deal of focus on form. Thus, Study 3 adapts and modifies the original design of the timed picture description test by eliminating word cues from the pictures and conducting the tests in a more interactive way with test takers (this point will be described in detail later).

Adding EI to FFI

In conjunction with the above-mentioned questions and limitations related to Study 1, Study 3 tests the pedagogical possibility of providing EI at the beginning of FFI lessons as a way to boost FFI effectiveness on L2 pronunciation development of English /ɹ/. I will first explain how EI has been examined in L2 morphosyntax studies and then provide the rationale for investigating this technique in L2 pronunciation teaching contexts.

EI in L2 Morphosyntactic Development

Over the past 25 years, instructed SLA studies have arrived at a consensus that EI allows L2 learners to notice and understand linguistic information from input, and makes a significant contribution to L2 development (Norris & Ortega, 2000; Schmidt, 2001; Spada & Tomita, 2010). However, it still remains unclear to what extent EI as *an independent variable* impacts SLA processes, arguably because the role of EI has been "conflated with the issue of explicit versus implicit learning" (Henry, Culman, & VanPatten, 2009, p. 561). Robinson (1996) compared whether explicit or implicit instructional approaches could be facilitative of the acquisition of pseudoclefts of location (complex rule) and subject-verb inversion following adverbial fronting (simple rule). However, EI was tested in tandem with decontextualized drill activities in his study. More empirical research, therefore, is called for which teases apart the role of EI in more

meaning-oriented FFI lessons where learners are to process language for meaning so that their explicit understanding of problematic structures activates the subsequent acquisition of implicit knowledge (Ellis, 2001, 2002, 2006). Recently, Spada and Lightbown (2008) referred to this point as follows:

> Both research and teaching experience have led to a growing consensus that instruction is most effective when it includes attention to both form and meaning. As a result, the most engaging questions and debates in L2 pedagogy are no longer about whether CLT should include FFI but rather how and when it is most effective…Researchers have not directly compared the effects of integrating or isolating form-focused and meaning-focused practice in CLT and CBI programs. (p. 184)

Very few primary studies have actually further examined the role of EI in meaning-oriented FFI lessons. Muranoi (2000) found that Japanese learners of English who received EI in addition to FT (interaction enhancement) and CF (repetition and recasts) significantly improved their use of the English indefinite article compared to those who received FT and CF without EI. Furthermore, the role of EI has been examined as an independent variable in the paradigm of processing instruction (i.e., input-based FFI), but generating slightly confounding results. VanPatten and Oikkenon (1996) examined how FFI can facilitate the acquisition of direct pronouns by English learners of Spanish, finding no significant difference between learners who received both EI and input-based practice and those who received only input-based practice. They concluded that EI might not be necessary (for similar results, see Bentai, 2004; Sanz & Morgan-Short, 2004). In contrast, other processing instruction studies have found that EI did

play an important role in processing instruction with respect to some linguistic structures (Fernandez, 2008 for Spanish subjunctive; Henry et al., 2009 for German word order and case markings).

In short, the role of EI in FFI effectiveness depends on types of language features. For example, Spada and Lightbown (2008) speculated that EI should be given before (or after) meaning-oriented lessons when the target language features are almost *unnoticeable* especially during ordinary communicative interaction, "either because they are acoustically imperceptible (e.g., most grammatical morphology in English) or redundant and unlikely to affect comprehension (e.g., word order in English questions)" (p. 186). In addition, Henry et al. (2009) pointed out that the effects of EI are related to the *portability* of EI: "Whether the information [included in EI] can be kept in working memory while the learner is simultaneously processing novel incoming data" (p. 572).

EI in L2 Phonological Development

Derwing and Munro (2005) emphasized the importance of *explicit phonetic instruction* (i.e., explicitly teaching segmental and suprasegmental aspects of sounds), claiming that "students learning L2 pronunciation benefit from being explicitly taught phonological form to help them notice the difference between their own productions and those of proficient speakers in the L2 community" (p. 388). For teaching new segmental sounds, EI consists of multiple exposures to a teacher's model pronunciation of the target sounds followed by explanation on relevant articulatory gestures in order to "raise learners' consciousness" (Celce-Murcia et al., 1996, p. 36). Given that pedagogical options for pronunciation teaching have been exclusively limited to focus-on-formS techniques such as minimal-pair drill activities (Celce-Murcia et al., 1996; DeKeyser, 1998; Derwing & Munro, 2005, 2009, 2010), the role of EI has been tested

only in relation to these decontextualized practice opportunities. The results of the previous studies have generally revealed that the combination of EI and decontextualized instructional approaches impacts learners' pronunciation development only at a controlled-speech level but not at a spontaneous-speech level (e.g., Elliott, 1997; Macdonald et al., 1994; Saito, 2011a; cf. Derwing et al., 1998). However, little research attention has been given to the relationship between EI and FFI (i.e., FT ± CF) in the context of pronunciation teaching; Study 3 takes a first step to examine this relationship.

EI could be an important variable for L2 phonological development especially in conjunction with more meaning-oriented FFI for several reasons. First, given that L2 learners are generally sensitive to *word-sized units* of L2 phonological information, providing EI at the beginning of FFI lessons will immediately draw learners' attention to *sound-sized units* of L2 phonological information. In this way, EI is hypothesized to promote their noticing of the perceptual difference between a new L2 sound and its L1 counterpart sound, which, many researchers argue, could be a first step towards L2 phonological development (Best, 1995; Best et al., 2001; Flege, 1995, 2003, 2007, 2009; Kuhl, 2000) (for the role of noticing in L2 morphological development, see DeKeyser, 1998, 2003, 2007; Ellis, 2001, 2002, 2006; Schmidt, 2001). Similarly, teaching articulatory gestures with listening discrimination activities could also help L2 learners notice the perceptual aspects of the new L2 sound (Celce-Murcia et al., 1996; Pennington, 1996). Ultimately, L2 learners can establish and internalize new phonetic categories, and generalize the newly-acquired phonetic knowledge from familiar to new lexical contexts.

Importantly, given the limited effectiveness of EI with decontectualized practice, EI should be embedded in meaning-oriented practice where L2 learners are encouraged to develop phonological awareness with their primary focus on meaning (Ellis, 2002; Henry et al., 2009;

Spada & Lightbown, 2008). Kennedy and Trofimovich (2010) investigated how different types of language awareness impacted ESL learners' pronunciation development during a 13-week pronunciation training course. The results of NS listening judgments showed that ESL learners who considered language as a tool for conveying meaning rather than as a set of discrete linguistic items significantly improved the quality of overall pronunciation performance between the beginning and end of the intensive ESL classes.

Another factor affecting EI is the differential learnability of L2 sounds. That is, L2 learners can quickly create new phonetic categories even without much modified input according to the *acoustic characteristics* of new L2 sounds (i.e., temporal vs. spectral differences), and EI might be even unnecessary for these relatively "easy" features (e.g., see Flege, 1989 for the impacts of intensive perceptual training on the acquisition of the final stop deletion by Chinese learners of English; Jamieson & Morosan 1986, 1989 for the acquisition of English interdental fricatives /θ/-/ð/ by French learners of English). Some L2 sounds, however, are extremely difficult and time-consuming to acquire, such as the non-native English /ɹ/-/l/ contrast for Japanese learners of English (Best et al., 2001; Kuhl, 2000). Intriguingly, Study 2 showed that FFI effectiveness was correlated with the learners' self-reported explicit articulatory knowledge of English /ɹ/ to some degree (i.e., a combination of constriction in the pharyngeal and velar regions of the vocal tract and lip rounding). In short, I would argue that EI is highly valuable for L2 pronunciation development of English /ɹ/ by Japanese learners of English, and Study 3 carefully explores this point.

Current Study

Motivated by the literature review, the primary goal of the current study is to carefully examine to what degree adding EI (triggering phonetically-driven L2 phonological learning) to FFI (triggering lexically-driven L2 phonological learning) can enhance the generalizability and magnitude of the overall instructional impact on interlanguage development of /ɹ/ by Japanese learners of English. To this end, I conducted a quasi-experimental study with a pre- and post-test design where 49 Japanese learners of English in EFL settings received two types of FFI (i.e., FFI-only vs. FFI+EI). Their performance resulting from instruction was assessed via measures of controlled and spontaneous production of both familiar and unfamiliar lexical items.

Research Questions

The research questions to be addressed follow:

1. To what degree is FFI facilitative of L2 pronunciation development of /ɹ/ by Japanese learners of English in both familiar and unfamiliar lexical contexts?
2. To what degree is FFI+EI facilitative of L2 pronunciation development of /ɹ/ by Japanese learners of English in both familiar and unfamiliar lexical contexts?

Method

Design

Student participants in the experimental groups received a range of FT activities embedded in four hours of meaning-oriented lessons in order to encourage learners to notice and practice the target sound feature (English /ɹ/) during meaningful discourse. During these activities, the instructors also provided CF following students' mispronunciation or unclear pronunciation of English /ɹ/. Only students in the FFI+EI group received EI at the beginning of FFI lessons. For the control group, students received meaning-oriented lessons that were

comparable in terms of duration and content but without any focus on form (i.e., English /ɹ/). Instructional treatments consisted of four one-hour lessons distributed over two weeks (1 hour lesson × 2 times per week × 2 weeks = 4 hours). All classes were videotaped and observed by the researcher who always sat at the back of the classroom to ensure the consistency of the FFI treatment for the entire project. Two weeks after the end of the lessons, all students took post-tests and were interviewed. Figure 12 summarizes the design of the study and the procedures followed. The project was conducted at a private language institute in Osaka, Japan, for a four-month period between April and August 2010. The researcher was offered a quiet room where he individually conducted the interview and testing sessions with all participating students. All classes took place in a classroom located in the institute.

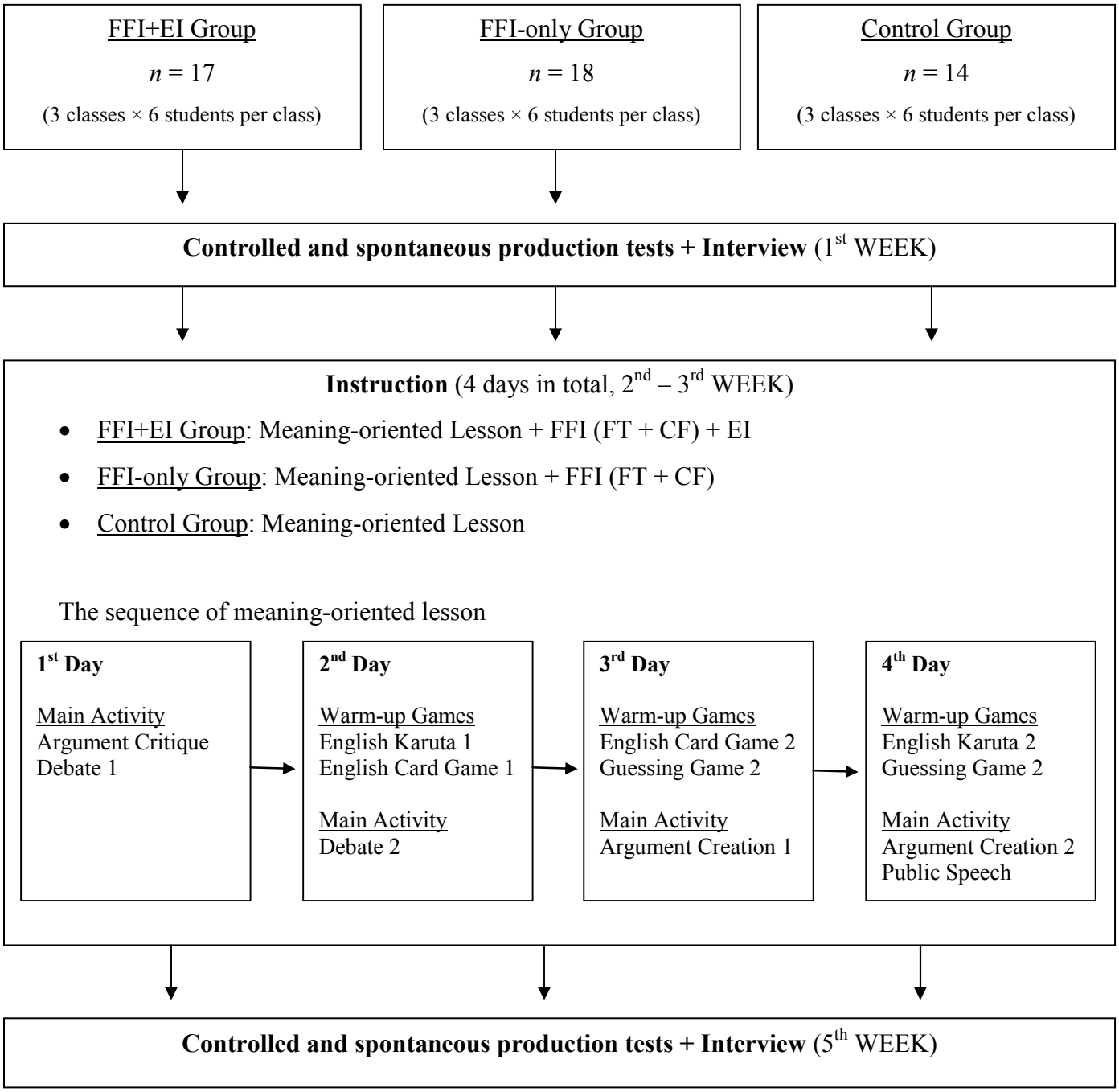


Figure 12. Summary of the procedure

Participants

Students

For the purpose of student recruitment, the researcher created ads which advertised four free hour-long English lessons with a focus on developing English argumentative skills. The ads specified the proficiency levels required for participation (e.g., 400-600 for TOEIC scores, 50-70 for TOEFL iBT scores); the purpose was to recruit beginner-to-intermediate Japanese learners of English based on the assumption that they would still have problems perceiving and producing /ɹ/. The ads were posted online under the McGill University domain (www.mcgill.ca/english-argumentative-project/), on a number of English education websites, and a few social communication network websites. The private language institute also posted the ads on its own website and distributed them to all their students at branch schools near the research site. Finally, the researcher contacted many university-level schools in the Osaka area, asking them to distribute the ads to their students or let them know about the relevant websites.

Interested participants contacted the researcher through email or by phone to set up a date for their first interview and pre-test sessions. The recruitment continued until the number of participants reached the maximum number, which had been set in advance at 54. Because five participants did not complete the instructional treatment nor attend the post-test sessions, a total of 49 participants were included in the final analysis (age: $M = 29.04$, $SD = 8.64$). The participating students were paid $10 for their extra trips to the research site to take pre- and post-tests beyond the four hours of lessons.

According to the first interview, although some participants reported previous experiences of staying in English-speaking countries for a few years (the USA, Canada, Australia, and England), most had never been abroad (Length of residence: $M = 4.44$ months, SD

= 8.66). While all of the learners had received 6-10 years of English education at public schools in Japan and attained relatively high TOEIC scores (M = 586.9, SD = 125.19), they stated that they had no opportunities to actually use English in communication at the time of the project except a few hours of English lessons either at the language institute or their college-level schools. In this respect, unlike Studies 1 and 2 which involved intermediate Japanese learners of English with a mean LOR exceeding one year in ESL settings (LOR: M = 15.5 months, SD = 31.8 months), Study 3 with beginner/intermediate Japanese learners of English in EFL settings could isolate and measure the pure impact of FFI on their phonological development at the initial stage (for similar discussion in the perception training studies, see Logan et al., 1991 with Japanese learners of English in ESL settings vs. Lively et al., 1993 with Japanese learners of English in EFL settings). The 49 students were randomly assigned to nine classes of six students each. Two treatment groups and one control group each of which comprises three classes were formed as follows: (a) FFI+EI group (3 classes, n = 17), (b) FFI-only group (3 classes, n = 18), and (c) control group (3 classes, n = 18). Table 9 provides the details of the 49 participants' information according to the three groups.

Instructors

Two experienced NS teachers (one male from California and one female from Ontario, Canada) who were employed by the language institute participated in Study 3. Both teachers were selected by the language institute based on their extensive EFL teaching experience in Japan. One instructor taught the first 5 classes (2 FFI-only classes and 3 control classes), and the other taught the other 4 (1 FFI-only classes, 3 FFI+EI classes). For 36 hours of teaching (4 hours × 9 classes), the language institute was paid $50 per hour (36 hours × $50 = $1,800 in total).

Table 9. *Participant information by Group*

	FFI+EI Group (n = 17)	FFI-only Group (n = 18)	Control Group (n = 14)
Gender	3 males / 14 females	3 males / 15 females	1 males / 13 females
Age	M = 26.7 (SD = 6.2)	M = 31.5 (SD = 10.6)	M = 28.7 (SD = 7.9)
LOR (months)	M = 4.6 (SD = 7.7)	M = 3.3 (SD = 8.8)	M = 5.5 (SD = 9.9)
TOEIC	M = 558.3 (SD = 119.4)	M = 569.2 (SD = 140.1)	M = 663.5 (SD = 88.3)

Interview

The researcher individually interviewed all participants in his office right after the pre-test sessions (the initial interview) as well as the post-test sessions (the final interview). The purposes of these interviews were to (a) elicit their personal information including age and English learning experiences at the onset of the project as well as (b) ascertain to what degree learners had focused on form and meaning during the FFI lessons in a retrospective manner after the project was completed. All communication was in Japanese and audio-recorded.

FFI Treatment

The FFI treatment in Study 3 consisted of the FT and CF treatments developed and implemented in Studies 1 and 2.

EI Treatment

For the FFI+EI group, during the first 10 minutes on Day 1 and 5 minutes on Days 2, 3, and 4, the instructor started with explicit phonetic instruction (EI) on how to perceive and produce English /ɹ/. Given that the combined FT and CF treatment is hypothesized to promote lexically-driven L2 phonological development, the primary purpose of EI in Study 3 was to

induce learners' awareness and noticing of the perceptual difference between /ɹ/ and /l/ at a phonetic level, the latter of which is acoustically similar to Japanese tap /ɾ/ (e.g., Iverson et al., 2003). In this regard, the instructor provided his exaggerated pronunciation model of the English /ɹ/-/l/ contrast, highlighting the perceptual difference between /ɹ/ and /l/ (i.e., acoustic enhancement). At the same time, the instructor taught relevant articulatory gestures about /ɹ/ to enable learners to actually produce the new sound, and encouraged them to listen to perceptual aspects of their own production to help them notice the acoustic (and articulatory) difference between English /ɹ/ and Japanese tap /ɾ/ (see the motor theory of speech perception, Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; for a comprehensive review, Diehl, Lotto, & Holt, 2004)[22]. The specific procedures of the EI treatment on Day 1 were as follows:

1. First, the instructor asked learners to carefully listen to his exaggerated pronunciation of both /ɹ/ and /l/ at a segmental level so that they become aware first of the perceptual characteristics of English /ɹ/, and then of its perceptual difference relative to English /l/ (which is acoustically similar to Japanese tap /ɾ/).
2. Subsequently, the relevant articulatory gestures (lip rounding, tongue backness, obstruction in the pharyngeal and glottal areas of the vocal tract) were explained for English /ɹ/ (but not English /l/) with the aid of an articulatory diagram adapted from Saito (2010). In particular, the instructor emphasized the relative importance of lip rounding following Bradlow's (2008) recommendation[23] as well as recent L2 speech research

[22] Catford and Pisoni (1970) showed that teaching relevant articulatory gestures helped learners perceive new sounds, concluding that "'exotic sounds' can generally be more readily and unerringly identified after one has learned to produce them" (p. 481).

[23] For pedagogic standpoints, Bradlow (2008) commented that "the lip rounding feature of English /ɹ/ production can be a useful characteristic to stress when teaching English pronunciation" (p. 292). Interestingly, the results of the interview with 18 learners in the FFI-only group who did not receive EI reported that only two of them already had explicit knowledge of lip rounding (16 out of them reported

findings (see also the ultrasound and optical tracking of supralaryngeal gestures in producing English /ɹ/, Campbell, Gick, Wilson, & Vatikiotis-Bateson 2010).

3. Finally, learners also performed three types of production practice: The learners first repeated only the instructor's model pronunciation of English /ɹ/ (*phonemic* practice), and then produced and perceived minimally-paired words (*lexical* practice) and tongue twisters (*sentence* practice).

For the rest of the lessons (Day 2 – Day 4), this practice phase was eliminated due to time constraints. Instead, the instructor always gave very short metalinguistic explanations at the beginning of the class, by reminding students of the relevant articulatory gestures for /ɹ/, modeling it with exaggeration and asking them to carefully listen and repeat.

Instead of receiving metalinguistic explanation, the FFI-only and control groups spent more time on self-introduction and small talk at the beginning of the first lesson on Day 1 in order to ensure that all groups received the same amount of instruction (i.e., four hours).

Control Group

Fourteen learners in the control group also received comparable meaning-oriented lessons on English argumentative skills but with neither FFI nor EI; the students received feedback not on any pronunciation errors but rather on ungrammaticality and inappropriate lexical choices (e.g., "you should say, 'I dropped a pen' instead of 'I fell a pen'") as well as the content of the lessons (e.g., "your opinion could be more convincing if you touched upon the opponent's critique"). As for warm-up games, the learners in the control group were given different communicative games without any focus on pronunciation/listening practice which the instructor usually used in his/her regular English conversation classes.

---

their knowledge of tongue movement). In sum, as many researchers pointed out (e.g., Derwing & Munro, 2005), pronunciation instruction has been left out from EFL/ESL classrooms for many years.

Teacher Training

Two instructors participated in a total of four hours (4 hrs) of teacher training led by the researcher over a two-day period. First, the instructors were given (a) a 20-page set of guidelines specifying the objectives and procedures for all instructional materials as well as (b) a list of 39 target words textually highlighted in the FFI materials. Next, the researcher carefully explained (a) how to deliver each activity in the FT treatment (English argumentative activities + communicative games) as well as when and how to provide CF to learners' mispronunciation of English /ɹ/ (i.e., pronunciation-focused recasts). In order to help their understanding of the contexts, the researcher also demonstrated model lessons when necessary. Last, only the instructor that taught all of the three FFI+EI classes also received special materials for EI and practiced demo lessons with the researcher.

Measures

To measure the effects of two types of FFI (i.e., FFI+EI vs. FFI-only) in comparison with the default performance by the control group, all learners were asked to complete two types of production tests both at pre- and post-test sessions: (a) the controlled production (CP) test (i.e., reading a list of words) and (b) the spontaneous production (SP) test (i.e., describing a set of pictures). To measure generalizability of FFI effectiveness, these CP and SP tests included (a) *familiar* items that the learners were exposed to during the FFI treatment and (b) *unfamiliar* items that never appeared in the FFI materials. Both at pre- and post-test sessions, all students first completed the SP test in order to measure their spontaneous production abilities without too much awareness of English /ɹ/; they subsequently moved onto the CP test. The researcher (a NS of Japanese) individually administered all of the pre- and post-test sessions with 49 learners in a sound-proof room. All communication about the testing procedure was in Japanese in order to

avoid any confusion. Their speech tokens were recorded by means of Rolad-05 Wave MP3 recorder, at 44.1 kHz sampling rate and a 16-bit resolution. A unidirectional microphone was used (DM-20SL) and all of the recordings were stored on the hard drive of a TOSHIBA Satellite U400 laptop computer.

Test Types

CP test. In order to measure their L2 pronunciation performance of /ɹ/ at controlled-speech levels, learners were asked to read a list of 40 words in total out of which 15 were target tokens (10 familiar items + 5 unfamiliar items) whose following vowel contexts were carefully controlled. These words are listed in Table 10. In total, 147 tokens were produced both at the pre- and post-test sessions (15 words × 49 students × 2 test sessions = 1,470 tokens).

Table 10. *15 tokens in CP tests in relation to following vowel conditions*

		Front vowels	Mid vowels	Back vowels
High vowels	(Familiar items)	rink, reef		rule, room
	(Unfamiliar items)	reach		rude
Mid vowels	(Familiar items)	race, rent		road[a], wrong
	(Unfamiliar items)	rate		roll
Low vowels	(Familiar items)		Ryan, right	
	(Unfamiliar items)		rough	

Note. [a] "Road" was tested twice both in the CP and SP tests.

SP test. Following Jian's (2000, 2004, 2007) distinction between the one-step process (i.e., automatic processing) and the two-step process (effortful processing), the spontaneous production measures in Study 3 assess to what degree L2 learners can accurately produce new L2 sounds (i.e., English /ɹ/) when one's attention is focused on communicating the message. As a trustworthy outcome measure of spontaneous production abilities, Ellis (2002, p. 225) recommended communicative free production defined as "an activity that calls for unplanned language use directed at fulfilling some communicative purpose."[24] Timed picture description tasks have been empirically validated as cognitively demanding (Derwing, Rossiter, Munro, & Thomson, 2004) and commonly used in L2 morphosyntax studies (e.g., Erlam, 2003; Lyster, 2004). In this regard, Study 3 adopted timed picture description tasks as the SP test; learners were asked to describe 10 pictures with six distracter pictures, and each of the 10 pictures led the learners to pronounce one target word including word-initial /ɹ/ (5 familiar items + 5 unfamiliar items). These words are listed in Table 11. In total, 147 tokens were produced both at the pre- and post-test sessions (10 words × 49 students × 2 test sessions = 980 tokens).

Although I used timed picture description tasks in Studies 1 and 2, several methodological constructs were modified; the revised timed picture description task in Study 3 was operationalized as follows:

1. Learners were first given 10 seconds to memorize four key words on a sheet of paper which related to two pictures they were to describe; one of the two key words for each picture was a target word including English /ɹ/ at a word-initial position while the other was a distracter.

[24] This time-pressure approach is not without problems. For example, Ellis (2002) himself acknowledged that "free-production tasks make it difficult but not impossible for learners to perform on the basis of explicit knowledge" (p. 234). Similarly, DeKeyser pointed out that the use of time pressure "merely made the use of explicit knowledge more difficult, and not impossible" (p. 326).

2. Right after the card was taken away, they were given one picture after another to describe by using all of the key words they just memorized. Importantly, in order to minimize their use of explicit knowledge, the learners were prompted to complete the task without any planning time (they described the pictures as soon as they received them).
3. After describing the pictures, they moved on to the next four key words for another set of two pictures.

Table 11. *10 tokens in SP tests in relation to following vowel conditions*

		Front vowels	Mid vowels	Back vowels
High vowels	(Familiar items)	read		roof
	(Unfamiliar items)	ring		route
Mid vowels	(Familiar items)	rain		road[a]
	(Unfamiliar items)	red	rush	rope
Low vowels	(Familiar items)		rice	
	(Unfamiliar items)			

Note. [a] "Road" was tested twice both in the CP and SP tests.

Data Analyses

All target words were consonant-vowel-consonant (CVC) singletons except one word Ryan as CVVC. Furthermore, with respect to familiar and unfamiliar items respectively, any instructional impact on the learners' performance of English /ɹ/ was carefully examined in relation to three affecting variables (a) task types (controlled vs. spontaneous speech levels) (e.g.,

Flege et al., 1995), (b) following vowel backeness (singletons with front/back vowels), and (c) following vowel height (singletons with high/low vowels) (e.g., Hardison, 2003).

NS Baseline

For comparison purposes, 10 NSs (six males, four females) were asked to complete the CP and SP tests following the same procedure as the Japanese learners of English. All of them were undergraduate students studying at an English-speaking university in Montreal at the time of the project, and spoke north-eastern American English as their L1 (age: $M = 21.54$, $SD = 1.64$). Their performance was used to demonstrate to what degree NS tokens differ from those produced by Japanese learners of English. In total, they generated 150 tokens for the CP tests (15 words × 10 NSs = 150 tokens) and 100 words for the SP tests (10 words × 10 NSs = 100 tokens).

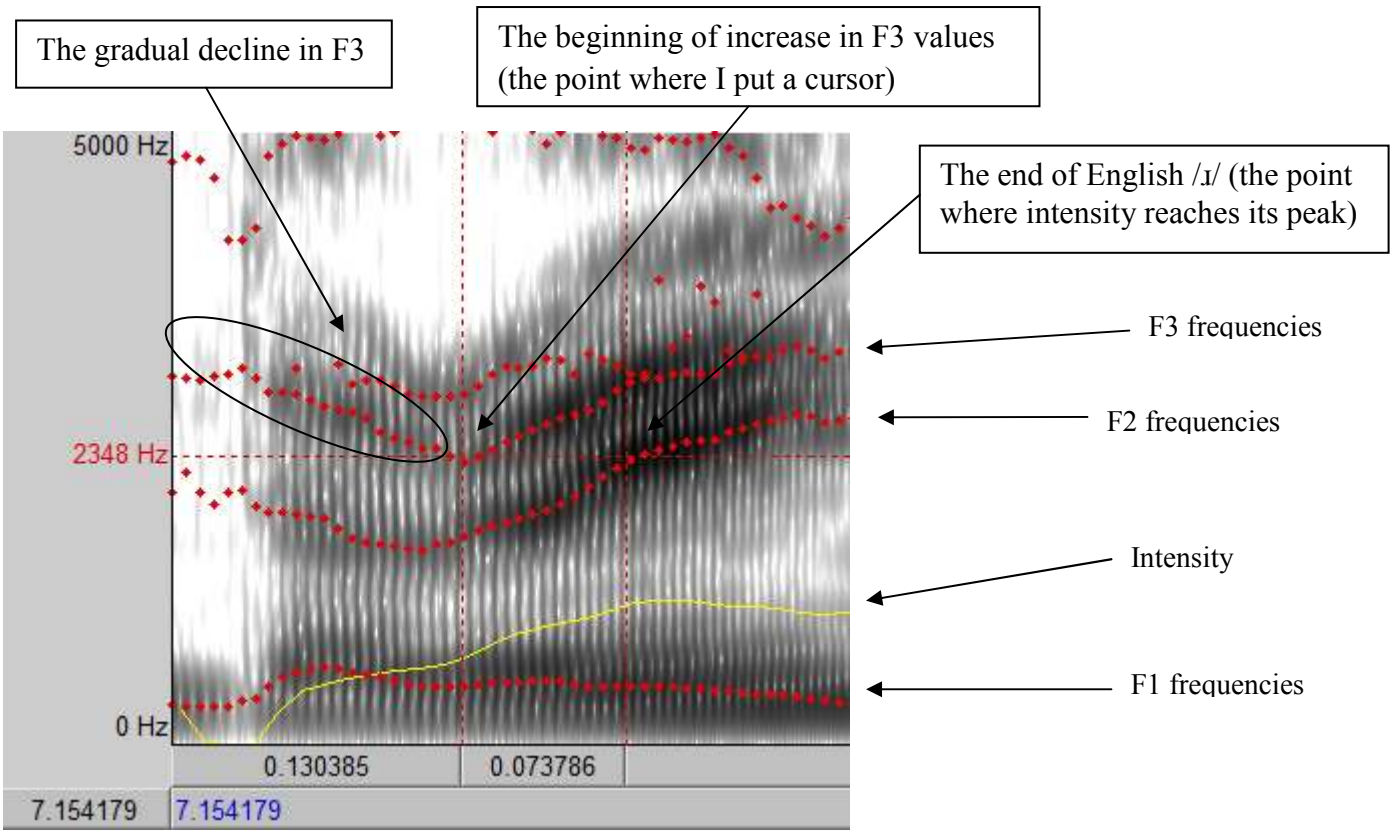
Acoustic Analyses

Acoustic analyses were conducted on the primary acoustic property of English /ɹ/ (i.e., F3 values) in all 2,700 tokens (2,450 words from 49 learners + 250 words from 10 NSs) in order to assess in depth to what degree Japanese learners of English exhibited gains resulting from FFI with and without EI in comparison with the NS baseline. As in Studies 1 and 2, Study 3 followed the Flege et al. (1995) procedure to acoustically analyze spectral aspects of natural English /ɹ/ tokens (i.e., F3 values) elicited from a variety of production tasks. All speech tokens were analyzed via the speech analysis software, *Praat* (Boersma & Weenik, 2009).

The beginning of word-initial English /ɹ/ was first identified via both the spectrographic representations and wave forms of the speech tokens in conjunction with the onset of the energy for all three formants. For English /ɹ/ embedded in continuous speech (i.e., the SP tokens), every effort was made to find the beginning of the word: As a reliable cue, the researcher carefully located the end of gradual decline in F3 values, given that English /ɹ/ exhibits relatively low F3

values compared to other vowel and consonant sounds in the English phonetic system (e.g., Bradlow, 2008; Hattori & Iverson, 2009; Iverson et al., 2003; for details of acoustic properties of English /ɹ/, see Espy-Wilson, 1992; Flege et al., 1995; Ladefoged, 2003). Finally, the cursor was put on the endpoint of the steady state of F3 values (i.e., F3 values again start to increase towards following vowels) and the auto-correlation method of linear predictive coding (LPC) analysis was used to measure F3 values of English at this point.[25] In order to capture the gender difference between 49 learners (7 male, 42 females), the formant analysis parameters were controlled by setting the maximum frequency to 5000 Hz for male voices and to 5500 Hz for female voices (Boersma & Weenik, 2009). The acoustic analysis procedure was visually summarized in Figure 13.

[25] The endpoint of English /ɹ/ was also determined by checking where F3 values as well as their intensity reach their peak (see Figure 14).



"the rain" from the SP task (spontaneous speech)

Figure 13. Procedure for the acoustic analysis for English /ɹ/

Post-hoc Analyses

Amount of Repair

Given that it still remains controversial whether immediate repair (i.e., repetition of CF) directly relates to acquisition in L2 morphosyntax studies (Leeman, 2003; Loewen & Philps, 2006; Lyster & Ranta, 1997; Lyster, 1998a, 1998b; Lyster & Izquierdo, 2009; McDonough, 2007; McDonough & Mackey, 2006), Study 3 explores this variable through post-hoc analysis. That is, the researcher carefully watched 24 hours of videotaped FFI lessons (3 FFI-only classes × 4 hours + 3 FFI+EI classes × 4 hours) in order to examine (a) whether the amount of repair was different or consistent between the FFI-only and FFI+EI groups; and (b) to what degree group differences, if any, influenced their pre- and post-test performance. I used the same coding schema developed in Study 2; three examples of the recast-repair sequence follow:

Example 8 (English Debate)

T: What kind of exercise do you like?

S: I don't like all sports.

T: I see. Not at all!

S: If I run /lʌn/*, I like running /lʌnɪŋ/* inside.

T: Running /ɹʌnɪŋ /. ← RECAST

S: Running /ɹʌnɪŋ/. ← REPAIR

Example 9 (Debate Activity)

S: If it rains, we can run /ɹʌn/ inside.

T: Rain /leɪn/. ← RECAST

S: Yes. ← NO REPAIR

Example 10 (Guessing Game)

S: We should eat bread /blɛd/...

T: Bread /bɹɛd/ ← RECAST

S: Eat bread /bɹɛd/ and fruit /flut/ ← REPAIR

T: Fruit/fɹut/ ← RECAST

S: Every day because many doctors recommend this kind of diet. ← NO REPAIR

Final Interview

In order to investigate the role that students' perceptions of FFI might play in determining its effectiveness, the researcher individually interviewed the 35 learners in the experimental groups (n = 17 for the FFI+EI group and n = 18 for the FFI-only group) after they completed the post-test. After receiving an explanation about the primary purpose of the current project (i.e., teaching how to produce and perceive English /ɹ/ in meaning-oriented classrooms), students were asked to estimate in a retrospective manner what proportion of their focus had been on meaning (i.e., English argumentative skills) and on form (i.e., English /ɹ/) during the FFI treatment.

Results

I will first present (a) to what extent the learners improved in their pronunciation of familiar and unfamiliar tokens, and then (b) whether the two individual variables (i.e., the amount of feedback and repair and the learners' perceptions of their focus on form or meaning) interacted to affect FFI and EI effectiveness, if any.

Pre- and Post- Test Data

First, the pre- and post-test scores were sorted out according to two lexical contexts: (a) familiar items and (b) unfamiliar items. Second, each participant's F3 values were averaged to obtain a group mean for each of the three variables: (a) test types, (b) following vowel backness

and (c) following vowel height. Last, separate three-factor ANOVAs (Group × Time × Test, Group × Time × Backness, Group × Time × Height) were performed in order to find any statistically significant differences between pre- and post-test sessions (within-group comparison) and among the three groups at the post-test sessions (between-group comparison). Third, any significant change in F3 values was interpreted in accordance with the benchmark of NS perceptions of English /ɹ/ which I established in Studies 1 and 2:

- Native-like English /ɹ/ (F3 = 1600-1700 Hz)
- Good English /ɹ/ (F3 = 2000-2280 Hz)
- Poor English /ɹ/ (F3 = 2280-2550 Hz)
- Hybrid exemplars (F3 = 2550-2800 Hz)
- English /l/ (F3 > 2800 Hz)

For the 10 NSs who took the same tests once, the descriptive and inferential statistics of their scores will be also reported for comparison reasons. An alpha level was set at a $p < .05$ level for all statistical analyses. Cohen's *d* was also calculated in order to measure the magnitude of instructional effectiveness between two contrast groups of means.[26]

Pre-test Data

In all of the contexts (familiar and unfamiliar items according to tests, backness, and height), a set of two-tailed *t*-tests on the pre-test data found no significant difference between the three groups at the time of the pre-test sessions nor the performance of the control group over

[26] 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 ($0.80 \leq d$). In all cases, whereas control group means were used to calculate between-group contrasts, pre-/post-test scores were used to calculate within-group contrasts.

time (between pre- and post-test sessions). This indicates that any changes in the experimental groups were not attributable to any group discrepancy at the onset of the study nor test-retest effects.

Familiar Items

Task types. A three-factor repeated-measures ANOVA, with Test (the CP test, the SP test) and Time (pre- and post-tests) as the repeated measure and Group (FFI-only, FFI+EI, control) as the between-group factor, found overall main effects for Test, $F(2,28) = 13.851$, $p < .001$. The result indicated that the participants in Study 3 had slightly more difficulty in the SP test ($M = 2527$ Hz, $SD = 313$ Hz) than in the CP test ($M = 2478$ Hz, $SD = 319$ Hz).

The ANOVA also found overall main interaction effects for Group and Time, $F(2, 46) = 24.510$, $p < .001$. Tukey's post hoc pairwise comparisons were employed to find the source of the significance (an alpha level of .05 was set for these and all subsequent post hoc analyses), revealing several interesting patterns. First, the FFI+EI group exhibited significant improvement in their performance of English /ɹ/ from hybrid exemplars ($M = 2632$ Hz, $SD = 288$ Hz) to good English /ɹ/ exemplars ($M = 2218$ Hz, $SD = 313$ Hz) ($p < .001$) with large effects ($d = 1.39$). They also outperformed the control group ($M = 2625$ Hz, $SD = 200$ Hz) at the time of the post-test sessions ($p < .001$) with large effects ($d = 1.54$). The post-test score difference between the FFI+EI group and the FFI-only group was close to statistical significance ($p =. 059$). Second, the FFI-only group also showed significant change over time from hybrid/poor exemplars ($M = 2632$ Hz, $SD = 288$ Hz) to poor English /ɹ/ exemplars ($M = 2455$, $SD = 333$) and outperformed the control group ($M = 2625$ Hz, $SD = 200$Hz), but with small-to-medium effects ($d = 0.56$ for the within-group difference and $d = 0.61$ for the between-group difference).

Vowel backness. A three-way ANOVA was conducted: Group (the FFI-only, FFI+EI, control) × Time (pre- and post-test) × Backness (singletons with front, central, and back vowels). The ANOVA found significant overall main effects for Backness, F (1, 46) = 104.107, p < .001. The learners produced English /ɹ/ with different F3 values according to its following vowel backness in the following order: (a) front vowels (M = 2602 Hz, SD = 319 Hz) > (b) central vowels (M = 2501 Hz, SD = 359 Hz) > (c) back vowels (M = 2393 Hz, SD = 308 Hz). The ANOVA revealed overall main interaction effects for Group and Time, F (2, 46) = 24.578, p < .001.According to Tukey's post hoc pairwise comparisons, the FFI+EI group significantly changed their performance of English /ɹ/ from hybrid exemplars (M = 2620 Hz, SD = 300 Hz) to good exemplars (M = 2185 Hz, SD = 309 Hz) (p < .001) with large effects (d = 1.42). Again, they outperformed both the control group (p < .001) with large effects (d = 1.58) and the FFI-only group (p = .030) with medium effects (d = 0.63) at the time of the post-test sessions. Similarly, the FFI-only group also showed significant change over time from hybrid exemplars (M = 2629 Hz, SD = 313 Hz) to poor exemplars (M = 2396 Hz, SD = 333 Hz) and outperformed the control group at the time of the post-test sessions. Yet, the magnitude of effectiveness was medium (d = 0.69 for the within-group difference and d = 0.72 for the between-group difference).

Vowel height. A three-way ANOVA was conducted: Group (the FFI-only, FFI+EI, control) × Time (pre- and post-test) × Height (singletons with high, mid, and low vowels). The results of the ANOVA confirmed significant overall main effects for vowel height. According to the post-hoc comparisons, the learners tended to generate significantly higher F3 values in (a) singletons with front vowels (M = 2556 Hz, SD = 311 Hz) than (b) singletons with central vowels (M = 2430 Hz, SD = 312 Hz) and (c) singletons with back vowels (M = 2496 Hz, SD = 358 Hz). The ANOVA revealed overall main interaction effects for Group and Time, F (2, 46) =

24.703, $p < .001$. The post-hoc comparisons revealed that there was a significant improvement over time for the FFI+EI group from hybrid exemplars ($M = 2619$ Hz, $SD = 276$ Hz) to good exemplars ($M = 2184$ Hz, $SD = 299$ Hz) ($p < .001$) with large effects ($d = 1.51$). In addition, they exhibited superior performance in comparison to the control group with large effects ($d = 1.65$) and the FFI-only group with medium effects ($d = 0.65$) at the time of the post-test sessions. A significant change was also found for the FFI-only group in within-group comparisons from hybrid exemplars ($M = 2629$ Hz, $SD = 312$ Hz) to poor exemplars ($M = 2396$ Hz, $SD = 349$ Hz) ($p < .001$) and in between-group comparisons (vs. the control group) at the time of the post-test ($p = .030$). The magnitude of effectiveness proved to be medium for both contexts ($d = 0.70$ and 0.74 respectively).

Unfamiliar Tokens

Task type. A three-way ANOVA (Group × Time × Task) found significant overall interaction effects for Group and Time, $F(2, 46) = 28.033$, $p < .001$. According to Tukey's post hoc pairwise comparisons, the FFI+EI group exhibited significant improvement in their performance of English /ɹ/ over time from hybrid exemplars ($M = 2622$ Hz, $SD = 286$ Hz) to good exemplars ($M = 2224$ Hz, $SD = 313$ Hz) ($p < .001$) with large effects ($d = 1.26$). They also outperformed not only the control group ($p < .001$) with large effects ($d = 1.38$) but also the FFI-only group ($p = .046$) with medium-large effects ($d = 0.79$) at the time of the post-test sessions. In contrast, the FFI-only group demonstrated significant change over time from hybrid exemplars ($M = 2600$ Hz, $SD = 270$ Hz) to poor exemplars ($M = 2472$ Hz, $SD = 314$ Hz) ($p < .001$) with small effects ($d = 0.43$), but did not outperform the control group at the time of the post-test sessions ($p = .067$).

Vowel backness. A three-way ANOVA (Group × Time × Backness) found significant overall main effects for Backness: (a) singletons with front vowels (M = 2602 Hz, SD = 313 Hz) > (b) singletons with central vowels (M = 2484 Hz, SD = 349 Hz) > (c) singletons with back vowels (M = 2402 Hz, SD = 312 Hz). The ANOVA also found overall interaction effects for Group and Time, F (2, 46) = 28.033, p < .001. According to Tukey's post hoc pairwise comparisons, the FFI+EI group revealed significant improvement in their performance of English /ɹ/ over time from hybrid exemplars (M = 2624 Hz, SD = 299 Hz) to good /ɹ/ exemplars (M = 2218 Hz, SD = 334 Hz) (p < .001) with large effects (d = 1.28), and outperformed the control group (p < .001) with large effects (d = 1.33) at the time of the post-test sessions. The post-test score difference between the FFI+EI group and the FFI-only group was close to statistical significance (p =. 051). The FFI-only group demonstrated significant change over time from hybrid exemplars (M = 2618 Hz, SD = 263 Hz) to poor English /ɹ/ exemplars (M = 2414 Hz, SD = 346 Hz) (p < .001) with small-to-medium effects (d = 0.58), but did not outperform the control group at the time of the post-test sessions (p =.087).

Vowel height. Because there were no singletons with low vowels, two vowel height factors (high and mid vowels) were considered. A three-way ANOVA (Group × Time × Height) revealed that there was a significant effect for Height, F (1, 46) = 19.871, p < .001, indicating that the learners tended to generate relatively higher F3 values in (a) singletons with high vowels (M = 2549 Hz, SD = 279 Hz) than (b) singletons with mid vowels (M = 2465 Hz, SD = 324 Hz). The ANOVA revealed significant overall main interaction effects for Group and Time, F (2, 46) = 29.693, p < .001. The post-hoc comparisons revealed that the FFI+EI group revealed significant improvement over time from hybrid exemplars (M = 2624 Hz, SD = 274 Hz) to good English /ɹ/ exemplars (M = 2237 Hz, SD = 310 Hz) (p < .001) with large effects (d = 1.32), and

outperformed both the control group ($p < .001$) with large effects ($d = 1.38$) and the FFI-only group ($p = .041$) with medium effects ($d = 0.67$) at the time of the post-test sessions. Although the FFI-only group also showed significant improvement over time from hybrid exemplars (M = 2609 Hz, SD = 256 Hz) to poor English /ɹ/ exemplars (M = 2434 Hz, SD = 312 Hz) ($p < .001$) with medium effects ($d = 0.60$), no significant difference between the FT+CF and control groups were found at the time of the post-test sessions ($p = .065$). A summary of all relevant results appears in Table 12.

NS Baseline

The descriptive statistics showed that 10 NSs generally produced English /ɹ/ with relatively low F3 values for both familiar items (M = 1620 Hz, SD = 136 Hz) and unfamiliar items (M = 1633 Hz, SD = 148 Hz), and a matched-paired t-test found no significant difference between the two lexical contexts, $p = .326$. As for three affecting variables, a set of one-factor repeated ANOVAs found that, although no difference was found in their performance between the CP tests and SP tests, NSs tended to produce lower F3 values for (a) English /ɹ/ preceding central and back vowels (M = 1673 Hz, SD = 159 Hz)[27] than front vowels (M = 1605, SD = 148), F (2, 18) = 4.923, $p = .019$, and (b) English /r/ following mid vowels (M = 1604 Hz, SD = 151 Hz) than high vowels (M = 1633 Hz, SD = 153 Hz), F (1, 9) = 9.186, $p = .0142$.

[27] The Tukey test did not find any significant difference between singletons with mid vowels and back vowels.

Table 12. *Summary of significant changes in F3 values*

Group	Lexical contexts	Variables	Within-group comparisons (pre- → post-tests)	Between-group comparisons (at post-tests)
FFI+EI group (n = 17)	Familiar items	Test types	**Large** effects (M = 2620 → 2218 Hz, d = 1.39)	**Large** effects (vs. control, d = 1.54)
		Vowel backness	**Large** effects (M = 2629 → 2185 Hz, d = 1.42)	**Large** effects (vs. control, d = 1.58) **Medium** effects (vs. FFI-only, d = 0.63)
		Vowel height	**Large** effects (M = 2619 → 2184 Hz, d = 1.39)	**Large** effects (vs. control, d = 1.65) **Medium** effects (vs. FFI-only, d = 0.65)
	Unfamiliar items	Test types	**Large** effects (M = 2622 → 2224 Hz, d = 1.26)	**Large** effects (vs. control, d = 1.38) **Medium-large** effects (vs. FFI-only, d = 0.79)
		Vowel backness	**Large** effects (M = 2624 → 2218 Hz, d = 1.28)	**Large** effects (vs. control, d = 1.33)
		Vowel height	**Large** effects (M = 2624 → 2237 Hz, d = 1.32)	**Large** effects (vs. control, d = 1.65) **Medium** effects (vs. FFI-only, d = 0.65)
FFI-only group (n = 18)	Familiar items	Test types	**Small-medium** effects (M = 2632 → 2455 Hz, d = 0.56)	**Medium** effects (vs. control, d = 0.61)
		Vowel backness	**Medium** effects (M = 2629 → 2396 Hz, d = 0.69)	**Medium** effects (vs. control, d = 0.72)
		Vowel height	**Medium** effects (M = 2629 → 2396 Hz, d = 0.70)	**Medium** effects (vs. control, d = 0.74)
	Unfamiliar items	Test types	**Small** effects (M = 2600 → 2472 Hz, d = 0.43)	*n.s.*
		Vowel backness	**Small-medium** effects (M = 2602 → 2414 Hz, d = 0.58)	*n.s.*
		Vowel height	**Medium** effects (M = 2609 → 2434 Hz, d = 0.61)	*n.s.*

Post-hoc Analyses

Amount of Recasts and Repair

Eighteen learners in the FFI-only group received 579 recasts and repaired 474 of them, yielding a relatively high repair rate (81.19%). On average, each learner in the FFI-only group received 32.17 recasts (M = 8.04 per lesson). Intriguingly, fewer recasts (n = 303) were directed to 17 learners in the FFI+EI group but with a high repair rate (277 repairs = 91.42%). On average, each learner in the FFI+EI group received 17.82 recasts (M = 4.46 per lesson).The results showed that, although both the FFI-only and FFI+EI groups demonstrated high awareness towards focus on form (81% and 91% of repair rates), the former group processed twice as many recasts and repairs as the latter group did.

Endpoint Interview

According to the final interview, the learners in both groups were split quite evenly between those who claimed to focus on form and those claiming to have focused on meaning. The analysis of individual self-report scores identified (a) 18 learners with relative focus on meaning (n = 10 for the FFI+EI group, n = 8 for the FFI-only group), (b) six learners with equal focus on meaning and form (n = 2 for the FFI+EI group, n = 4 for the FFI-only group), and (c) 11 learners with relative focus on form (n = 5 for the FFI+EI group, n = 6 for the FFI-only group). The mean of their self-reported portion scores were as follows: (a) the FFI+EI group (M = 56.17% for meaning vs. M = 43.82% for form) and (b) the FFI-only group (M = 57.77% for meaning vs. M = 42.22% for form)

Discussion

L2 learners decode L2 phonological information both at a lexical level (e.g., “read” “reader” “reading”) and a phonetic level (e.g., sensitivity to “/ɹ/” in “read” distinguishes it from

"lead") (Baker & Trofimovich, 2008; Flege et al., 1996). From a theoretical perspective, the latter procedure is particularly important for restructuring existing representations into new phonetic categories represented in long-term memory and, ultimately, for generalizing the newly-acquired phonetic knowledge to new lexical contexts (Flege, 1995, 2003, 2007, 2009). In Studies 1 and 2, the effects of FFI were tested inducing learners to notice and practice English /ɹ/ during meaningful discourse at a lexical level. The results did show some evidence of acquisition but with several limitations (i.e., the moderate improvement only within familiar lexical contexts). In this regard, Study 3 investigated the role of EI in FFI as a way to help learners attend to the segmental aspect of English /ɹ/ at a phonetic level.

Given that Japanese learners of English generally produced hybrid exemplars (F3 values = 2550-2800 Hz) at the onset of the project compared to NS baseline (F3 values = 1600-1700 Hz), in this section, I will first re-examine to what degree FFI alone is facilitative of their L2 pronunciation development of English /ɹ/ in both familiar and unfamiliar lexical contexts as assessed by measures that had been improved since Studies 1 and 2 (e.g., by including familiar and unfamiliar items in pre- and post-test materials and designing a new picture description task). Subsequently, I will turn my discussion to examining to what degree adding EI to FFI lessons enhances the acquisitional value of the overall instructional treatment.

Re-examining FFI Effectiveness

For the first research question concerning the generalizability and magnitude of FFI effectiveness, the results of Study 3 found that the FFI-only group generally changed their F3 values from 2600-2700 Hz to 2400-2500 Hz with small-to-medium effects (d = 0.40-0.70) between pre- and post-test sessions. Importantly, their post-test scores were greater than those of the control group with medium effects (d = 0.60-0.70) in familiar items but not in unfamiliar

items. That is, the FFI-only group enhanced their performance from hybrid exemplars to poor English /ɹ/ exemplars with small-to-medium effects in the case of familiar lexical items, but not in unfamiliar lexical contexts. This improvement pattern was similar across test types (CP vs. SP), vowel backness (singletons with front, central, and back vowels) and vowel height (singletons with high, mid and low vowels). The results of Study 3 with beginner/intermediate Japanese EFL learners in Japan echoed those of the original study with intermediate Japanese ESL learners in Canada.

Different from traditional decontextualized instruction methods (e.g., audio-lingual methods), FFI in Study 3 was integrated into meaning-oriented lessons, encouraging learners to focus on both form and meaning simultaneously (out of 18 learners, 12 self-reported either more focus on meaning than form or equal focus on both form and meaning during FFI lessons). In line with similar FFI studies in L2 morphosyntax development, this type of FFI could impact learners' developing L2 phonological system both at a controlled- and spontaneous-speech level irrespective of the following vowel contexts. Noteworthy, however, is that the magnitude of FFI effectiveness was small-to-medium and its generalizability to unfamiliar lexical contexts was limited. This could indicate that FFI delivered at a lexical level (i.e., FT+CF) might not be sufficient to trigger L2 learners' phonemic and phonetic noticing and awareness, especially in the case of the relatively difficult segmental acquisition of /ɹ/ by Japanese learners of English. That is, these learners might need either more exposure to exemplars or other kinds of modified input beyond FFI in order to establish a new phonetic category for English /ɹ/ in their long-term memory representation.

Adding EI to FFI Lessons

With respect to the second research question which asked whether and to what degree providing EI enhances the generalizability and magnitude of FFI effectiveness, Study 3 found that the FFI+EI group generally changed their F3 values over time from 2600-2700 Hz to 2100-2200 Hz with large effects in both familiar and unfamiliar items (d = 1.30-1.40). In other words, the FFI+EI group not only demonstrated considerable improvement from hybrid exemplars to good English /ɹ/ exemplars, but also transferred the instructional gain from familiar lexical contexts to unfamiliar lexical contexts. Importantly, out of six possible contexts (2 lexical factors [familiar and unfamiliar items] ×3 affecting variables [test types, vowel backness, vowel height]), the FFI+EI group outperformed the control group with large effects (d = 1.50-1.65) in all cases and the FFI-only group with medium-large effects (d = 0.60-0.80) in four contexts. According to the interview data, EI did not interfere with the learners' simultaneous focus on form and meaning (out of 17 learners, 12 self-reported either more focus on meaning than form or equal focus on both form and meaning during FFI lessons).

As I predicted earlier, the advantage of adding EI to FFI treatment could be due mainly to several factors which will be discussed in detail in the following subsection especially in relation to (a) noticing, (b) proceduralization and generalization, and (c) the amount of reapir.

Noticing

First, the instructor's exaggerated model pronunciation of only English /ɹ/ (and English /l/) could directly help the learners notice the perceptual difference between the new sound (English /ɹ/) and its L1 counterpart (Japanese tap). In addition, the learners were explicitly taught the primary articulatory gestures for English /ɹ/ (i.e., lip rounding and tongue movement) (Bradlow, 2008; Campbell et al., 2010), which might have also promoted their increased

awareness of the new sound form in a complementary fashion (Catford & Pisoni, 1970). Importantly, EI was implemented before FFI lessons so that the beginner/intermediate learners in Study 3 could fully use their limited attentional resources to attend to the new sound form in L2 input under no communicative pressure. The results could support the view that EI might be necessary with respect to linguistic features which learners would otherwise have tremendous difficulties in noticing through mere exposure to L2 input (Ellis, 2002; VanPatten, 2002, 2004; Spada & Lightbown, 2008).

Proceduralization and Generalization

Subsequently, these learners with high sensitivities to sound-sized units of L2 phonological information could make the best of the subsequent FFI activities where the new sound form was embedded at a lexical level. Namely, they practiced the target sound feature via a number of communicative activities in order to (a) proceduralize more targetlike representations (i.e., the large gain was found across different tasks and following vowel conditions) and (b) generalize the newly-acquired phonetic knowledge to unfamiliar lexical contexts (the large effects were apparent not only in familiar items but also in unfamiliar items).

Amount of Repair

Although both FFI+EI and FFI-only groups received CF as a part of the FFI treatment, the post-hoc analyses showed that the learners in the FFI-only group generated twice as much repair (8.04 repairs for one student per lesson) as those in the FFI+EI group did (4.64 repairs for one student per lesson); the pre- and post-test data revealed, however, that the FFI+EI group showed robust and generalizable gains, but the instructional gain for the FFI-only group was restricted to lexical items appearing in the instructional materials. Despite considerable debate as to the role of repair in L2 morphosyntax development (Leeman, 2003; Loewen & Philp, 2006;

Lyster & Izquierdo, 2009; McDonough, 2007; McDonough & Mackey, 2006), the results of Study 3 at least suggest that the amount of repair is not necessarily related to transforming the nature of the representation (a transition from the lexical- to the phonetic-level abstractions). Instead, regardless of the amount of self-modified output, FFI might positively affect learners' processing abilities to retrieve information based on the present state of their mental representation (the learners in both the FFI+EI and FFI-only groups could equally produce English /ɹ/ without any variance across different tasks and following vowel conditions). In other words, FFI can be an important variable to ameliorate L2 learners' retrieval process in production especially after they develop more targetlike representations with the aid of EI. However, this suggestion deserves future research, because the study was not originally designed to control the amount of repair nor to investigate the timing of CF (and EI) in relation to learners' representations and their processing abilities to access these representations; this point will be further discussed in the next section.

Conclusion and Future Directions

Whereas the SLM posits that learners' noticing of *perceived* difference between a new sound and its L1 counterpart leads them to establish a new phonetic category which is a crucial component in later online production stages (Flege, 1995, 2003, 2007, 2009), the current study further examined the role of EI as a way to promote learners' phonetic-level noticing in the context of L2 pronunciation development of /ɹ/ by Japanese learners of English. The results revealed that the learners who received FFI (FT + CF) with and without EI at the beginning of FFI lessons demonstrated two different types of L2 phonological development.

On the one hand, the learners in the FFI-only group demonstrated small-to-medium improvement particularly in familiar lexical contexts; their gains, however, were not transferred

to unfamiliar lexical contexts. This finding implies that, without remedial techniques such as EI, L2 learners tend to continue to decode L2 phonological information at a lexical level as their *default interlanguage strategies*, and their access to the relevant mental representations is restricted to lexical items occurring in the instructional materials and, thus, *context-specific*. On the other hand, the learners in the FFI+EI group who started with the noticing of perceptual aspects of English /ɹ/ via EI before FFI showed (a) large gains (change from hybrid exemplars to good English /ɹ/ exemplars) across different tasks and following vowel conditions; and (b) ability to generalize their gains to new lexical contexts beyond the instructional materials. This result indicates that these learners established strong *context-invariant* representations in the underlying system with various levels of processing abilities (controlled vs. spontaneous production; various following vowel conditions). In short, the results of Study 3 showed the relative impact of phonetically-driven L2 phonological learning (FFI+EI) over lexically-driven L2 phonological learning (FFI-only), especially in the case of the relatively difficult segmental acquisition of /ɹ/ by Japanese learners of English.

Finally, the findings of Study 3 provides several pedagogical recommendations for teaching L2 segmental sounds in relation to the FFI model developed by Lyster in the area of L2 morphosyntax studies (Lyster, 2004b, 2007; Ranta & Lyster, 2007): Noticing → Awareness → Practice. First, at the noticing phase, some isolated intervention such as EI might be necessary in order to push L2 learners (a) to attend to sound-sized units of L2 phonological information and (b) to notice the perceptual difference between a new sound and its L1 counterpart. This noticing could be a first step towards restructuring existing representations and establishing a new phonetic category in a learner's representational system (Flege, 1995, 2003, 2007, 2009). Next, at the awareness stage, L2 learners should be given communicative tasks where they can further

process the target sound either receptively or productively in meaningful lexical contexts (i.e., FT), which further help learners to develop and internalize the phonetic representation.[28] Last, at the practice phase, L2 learners should be gradually encouraged to practice the target sound feature in a productive mode in order to proceduralize more targetlike phonetic representations (i.e., a transition from effortful to automatic use of phonetic knowledge). Given that the corrective force of pronunciation-focused recasts is quite salient to L2 learners (L2 learners' repair rate following pronunciation-focused recasts is reported at around 80% in a wide range of classrooms; Ellis et al., 2001; Kim & Han, 2007; Lyster, 1998b; Sheen, 2006), teachers could use this technique to push their students to repair their nontargetlike production of target sounds at a lexical level. The proposed model for instructed L2 phonological development is visually summarized in Figure 14.

I conclude this paper with several future directions for L2 speech acquisition research of this kind. First, future examination is warranted to investigate the timing of of introducing EI, FT and CF to L2 learners according to the differential level of their developing representation and processing abilities (as proposed in the model in Figure 14). In this regard, future research needs to adopt not only *production* measures but also *perception* measures, because change in the perception phase entails change in a learners' representational system and precedes change in the production phase (Bradlow et al., 1997, 1999; Flege, 1995, 2003, 2007, 2009). Therefore, given that Studies 1, 2, and 3 focused only on change in the production phase, despite some efforts made at each phase (i.e., adopting not only CP but also SP), it is still possible that the learners could have consciously and carefully produced English /ɹ/ drawing on their explicit articulatory

[28] Some SLA researchers strongly argue that the learning at this stage should be *receptive* rather than *productive* especially in order to impact L2 developing system (VanPatten, 2002, 2004; VanPatten & Cardierno, 1993).

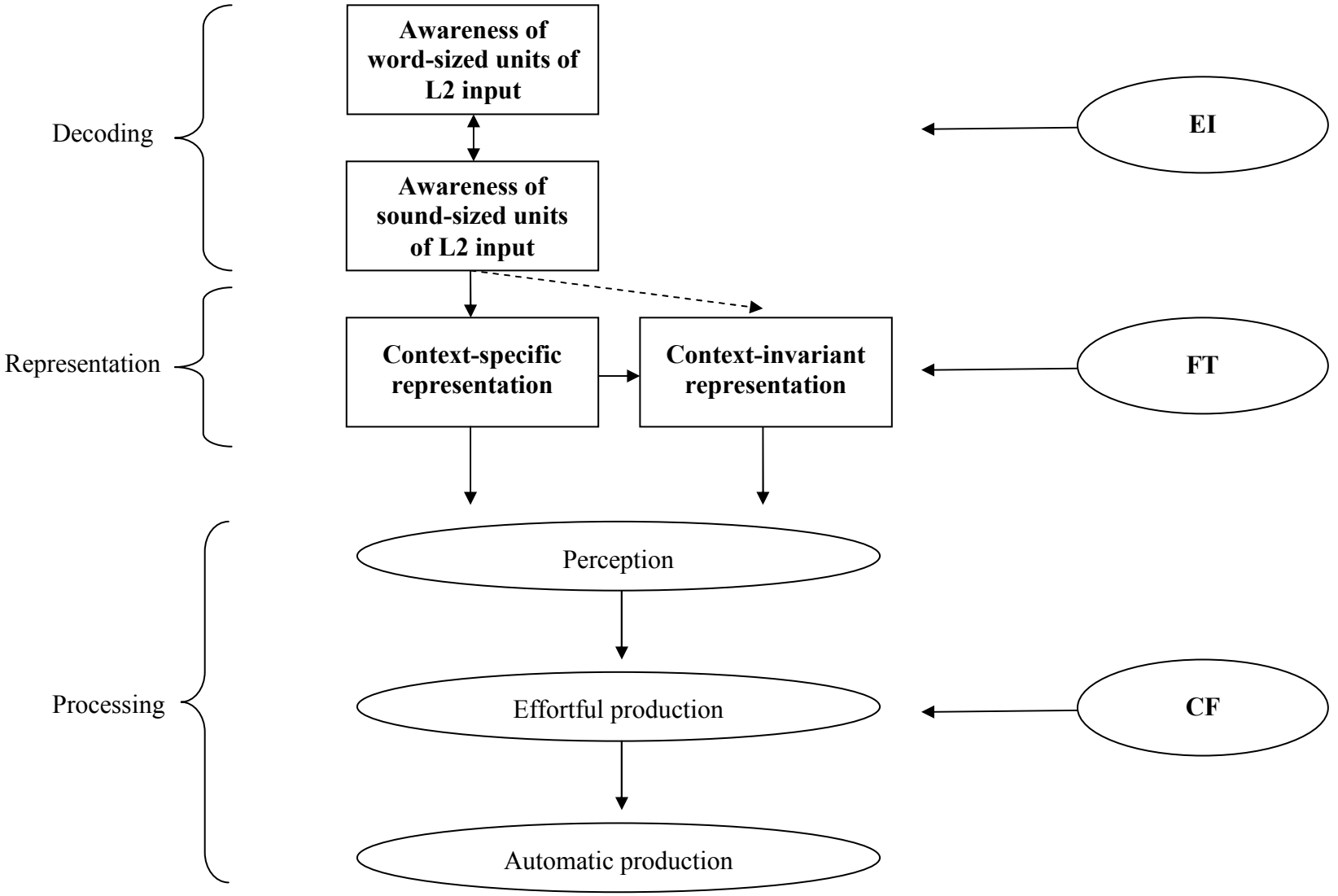


Figure 14. Pedagogic model for instructed L2 phonological development

knowledge (i.e., *monitoring*) even without establishing nor developing new a representation. In other words, it is necessary to implement a wide range of perception and production tests simultaneously, in order to examine (a) which combination of FFI techniques actually impact the L2 developing system at the initial stages of SLA (which could be mainly measured via perception tests) and (b) to what degree they proceduralize the newly-acquired phonetic knowledge at later stages of SLA (which could be mainly measured via production tests).

Adopting these outcome measures will shed light on the multifaceted aspects of instructed L2 phonology development. For example, although Studies 1 and 2 found that the FT-only group did not show any significant improvement in the production measures, there is some possibility that the gain might have begun to appear in the perception phase (but not in the production phase), arguably because FT-only could be hypothesized to be more effective than FT+CF at a stage when learners are beginning to notice the perceptual aspects of English /ɹ/ and developing a new phonetic category in a receptive mode (Lyster, 2004b; Lyster, 2007; Ranta & Lyster, 2007). Another example is found in the results of the FFI-only group (FT+CF) in Studies 1, 2, and 3. From the perspective of information processing theory, pronunciation-focused recasts tend to push L2 learners to generate self-modified output under communicative pressure, which again requires a great deal of attentional resources; such output-prompting FFI technique could be *too early* for students with less targetlike representations (VanPatten, 2002, 2004; VanPatten & Cardierno, 1993; VanPatten & Okiennon, 1996). In fact, the results of the FT+CF group (i.e., the small-to-medium improvement in their production of English /ɹ/ in familiar lexical contexts) might signal that introducing FT and CF without EI (to beginner-intermediate learners) yields a limited amount of acquisition.

Second, future research needs to tease apart several FFI variables which were conflated in Study 3. For instance, although the FFI+EI group showed robust and generalizable gains, it raises another question—whether it is necessary to include CF (i.e., output-based instruction[29]) in FFI treatments. That is, given that providing CF to students with incomplete representations of the target form might even distract their attention from noticing and developing awareness of new sounds in L2 input, one could argue that it is not CF but EI which explains the gains made by the FFI+EI group. In order to isolate the effects of EI on L2 pronunciation development, it would be interesting to compare the FT+CF+EI group not only with the FT+CF group but also with the FT+EI group.

[29] With respect to pronunciation errors, it may be the case that all CF types (i.e., recasts, prompts, explicit correction) have sufficient illocutionary force to convey their corrective message and to elicit modified output (Carpenter et al., 2007; Ellis et al., 2001; Kim & Han, 2007; Han, 2008; Lyster, 1998b; Lyster & Saito, 2010b; Mackey et al., 2000; Sheen, 2006). In this respect, providing CF (and pushing students to modify their non-nativelike production) could be considered an output-based instructional option.

REFERENCES

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*, 543-574.

Anderson-Hsieh, J. (1992) Using electronic visual feedback to teach suprasegmentals. *System*, *20*, 51-62.

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. (2008). Lexical and segmental influences on child and adult learners' production of second language vowels. *Concordia Working Papers in Applied Linguistics*, *1*, 30-54.

Benati, A. (2004). The effects of structured input activities and explicit information on the acquisition of the Italian future tense . In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 207 – 225). Mahwah, NJ: Erlbaum.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistics experience: Theoretical and mythological issues in cross-language speech research* (pp. 171–204). Timonium, MD: York Press.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775–794.

Best, C., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Bialystok, E. (1978). A theoretical model of second language learning.. *Language Learning, 28,* 69-83.

Boersma, P., & Weenik, D. (2009). *Praat: Doing phonetics by computer*. Retrieved from http://www.praat.org (Last viewed September, 2009).

Bradlow, A. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In Hansen Edwards, J. G. and Zampini, M. (Eds.), *Phonology and Second Language Acquisition* (pp. 287-308): Amsterdam: John Benjamins.

Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Teaching Japanese listeners to identify English /r/ and /l/: Long term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977-985.

Bradlow, A. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106, 707-729.

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 /r/ and /l/. *Journal of the Acoustical Society of America, 101*, 2299–2310.

Breitkreutz, J., Derwing, T., & Rossiter, M. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19*, 51-61.

Brown, A. (1995). Minimal pairs: minimal importance? *ELT Journal, 49*(2), 169–175.

Campbell, F., Gick, B., Wilson, I., & Vatikiotis-Bateson. (2010). Spatial and temporal properties of gestures in North American English /ɹ/. *Language and Speech, 53,* 49-69.

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

Catford, J., & Pisoni, D. (1970). Auditory vs. articulatory training in exotic sounds. *Modern Language Journal, 54,* 447-81.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching pronunciation: A reference for learners of English to speakers of other languages*. Cambridge: Cambridge University Press.

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

Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27*, 49–64.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21,* 46-66.

de Bot, K. (1983). Visual feedback of intonation: I. Effectiveness and induced practice behaviour. *Language and Speech, 26,* 331-350.

de Bot, K. (1996). The Psycholinguistics of the Output Hypothesis. *Language learning, 46*, 529-555.

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom application. *TESOL Quarterly, 16*, 71-77.

de Bot, K., Paribakht, T. S., & Wesche, M. B. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition. *Studies in Second Language Acquisition, 19*, 309-329.

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

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-51). Cambridge, UK: Cambridge University Press.

DeKeyser, R. (2003). Implicit and explicit learning. In M. Long & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 313-348). Malden, MA: Blackwell.

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

DeKeyser R, Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's processing instruction: an update. *Language Learning,* 52, 480-523.

DeKeyser, R., & Sokalski, K. (1996). The differential role of comprehension and production practice. *Language Learning, 46*, 613-642.

Derwing, T. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and second language acquisition,* pp. 347-369. Amsterdam: John Benjamins.

Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition, 12*, 1–16.

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

Derwing, T. M. & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*, 476-490.

Derwing, T. M., & Munro, M. J. (2010). Accentuating the Positive: Directions in Pronunciation Research. *Language Teaching, 43,* 366-368.

Derwing, T., Munro, M., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly*, *34*, 592–603.

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

Derwing, T.M., Rossiter, M.J., Munro, M.J. & Thomson, R.I. (2004). L2 fluency: Judgments on different tasks. *Language Learning, 54,* 655-679.

Diehl, R.., Lotto, A., & Holt, L. (2004). Speech perception. Annual *Review of Psychology, 55,* 149–179.

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. (2003). Instructed SLA: Constraints, compensation, and enhancement. In M. Long & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 257–310). Malden, MA: Blackwell.

Doughty, C., & Williams, J. (Eds.). (1998). *Focus on forms in classroom second language acquisition.* Cambridge: 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: Cambridge University Press.

Elliott, R. E. (1995) Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language, *The Modern Language Journal,* 79, 356-371.

Elliott, A. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania, 80*, 95-108.

Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24,* 143–188

Ellis, R. (2001). Investigating form-focused instruction. *Language Learning, 51*, 1–46.

Ellis, R. (2002). Does form-focused instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition,* 24, 223–236.

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-focussed instruction on L2 acquisition. *AILA Review, 19*, 18-41.

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407-452). Oxford: Oxford University Press

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

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

Ellis, R., Rosszell, H., & Takashima, H. (1994). Down the garden path: another look at negative feedback. *JALT Journal, 16,* 9-24.

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

Erlam, R.M. (2003). Evaluating the relative effectiveness of structured-input and output-based instruction in foreign language learning: results from an experimental study. *Studies in Second Language Acquisition, 25*, 559-583.

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

Fernandez, C. (2008). Reexamining the role of explicit information in processing instruction. *Studies in Second Language Acquisition, 30*, 277-305.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399–423.

Flege, J. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America, 86*, 1684-1697.

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: Mouton de Gruyter.

Flege, J. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353-382). Berlin: Mouton de Gruyter.

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

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. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527-552.

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

Flege, J., Munro, M, & MacKay, I. R. A. (1995b). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*, 1-26.

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

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

Gass, S. (1997). *Input, interaction, and the second language learner.* Mahwah, NJ: Erlbaum.

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.

Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics, 22,* 123-134.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds L and R. *Neuropsychologia*, *9,* 317–323.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38*, 201–223.

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

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24,* 495-522.

Hardison, D. (2010). Visual and auditory input in second-language speech processing. *Language Teaching, 43*, 84-95.

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America, 125*(1), 469-479.

Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA. *Studies in Second Language Acquisition, 31*, 559-575.

Herron, C. (1991). The garden path correction strategy in the foreign language classroom. *The French Review, 64,* 966-977.

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.

Ingram, J. C., & Park, S. (1998). Language, context, and speaker effects in the identification and discrimination of English /ɹ/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America, 103*, 1161–1174.

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

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

Iverson, P., & Kuhl, P. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/ *Journal of the Acoustical Society of America, 92*, 1130-1140.

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

Izumi, S. (2002). Output, input enhancement, and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition*, *24*, 541-577.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception and Psychophysics, 40,* 205-215.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology, 43*, 88-96.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, *23*, 83–103.

Jiang, N. (2000). Lexical development and representation in a second language. *Applied Linguistics, 21*, 47-77.

Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition, 24,* 617-637.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics. 25*, 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language acquisition. *Language Learning*, 57, 1-33.

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

Jusczyk, P.W. (1997). *The discovery of spoken language.* Cambridge, MA: Bradford Books, MIT Press.

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

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness, 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*. Oxford: Oxford University Press.

Krashen, S. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon.

Kuhl, P. K. (1991). Human adults and human infants show a ”perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50*, 93–107.

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

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5,* 831-843.

Kuhl, P., Wiliams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistics experience alters phonetic perception in infants by 6 month age. *Science, 255*, 606–608.

Ladefoged, P. (2003). *Phonetic data analysis: an introduction to instrumental phonetic fieldwork*. Oxford: Blackwell.

Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer-Assisted Language Learning, 12*, 137–156.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*, 227-247.

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition, 25*, 37-63.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press

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

Lin, Y. (2001). Syllable simplification strategies: A stylistic perspective. *Language Learning, 51,* 681-718.

Lin, Y. (2003). Interphonology variability: sociolinguistic factors affecting L2 simplification strategies. *Applied Linguistics, 24*, 439–464.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychology Review, 74,* 431–461.

Lively, S., Logan J., & Pisoni, D. (1993). Training Japanese listeners to identify English /ɹ/ and /l/. II: The role of phonetic environments and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242–1255.

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.

Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 361-377). Oxford: Oxford University Press.

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

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

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition*. *Vol. 2: Second language acquisition* (pp. 413-468). New York: Academic Press.

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

Long, M., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge, UK: Cambridge University Press.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and Language Learning: Integrating Theory and Practice* (pp. 123-167). Clevedon, England: Multilingual Matters Ltd.

Lyster, R. (1998a). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition, 20*, 51-81.

Lyster, R. (1998b). 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. (2002). Negotiation in immersion teacher-student interaction. *International Journal of Educational Research, 37,* 237-253.

Lyster, R. (2004a). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399-432.

Lyster, R. (2004b). Research on form-focused instruction in immersion classrooms: Implications for theory and practice. *Journal of French Language Studies, 14,* 321-341.

Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam/Philadelphia: John Benjamins.

Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning, 59,* 453-498.

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

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

Lyster, R., & Saito, K. (2010b). Interactional feedback as instructional input: A synthesis of classroom SLA research. *Language, Interaction and Acquisition, 1,* 276-297.

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning*, *44*, 75–100.

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

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language*

acquisition: A collection of empirical studies (pp. 407-452). Oxford: Oxford University Press

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

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

McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational interaction in second language acquisition,* (pp.323-338). New York: Oxford University Press.

McDonough, K. & Mackey, A. (2006). Responses to recasts: Repetitions, primed production and linguistic development. *Language Learning, 56*, 693-720.

Meador, D., Flege, J. E., & MacKay, I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition, 3,* 55-67.

Mochizuki, M. (1981). The identification of /l/ and /ɹ/ in natural and synthesized speech. *Journal of Phonetics, 9,* 283-303

Moyer, A. (1999). Ultimate Attainment in L2 Phonology: The Critical Factors of Age Motivation and Instruction. *Studies in Second Language Acquisition*, *21*, 81-108.

Munro, M. (2008). Foreign accent and speech intelligibility. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 193-218). Amsterdam: John Benjamins.

Munro, M., & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*, 289–306.

Munro, M., & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learnin*g, *49,* Supplement 1, 285-310.

Munro, M., & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*, 451–468.

Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*, 520–531.

Munro, M., Derwing, T., & Morton, S. (2006) The mutual intelligibility of L2 speech, *Studies in Second Language Acquisition*, *28*, 111–131.

Munro, M., Flege, J., & MacKay, I. (1996). The effect of age of second-language learning on the production of English vowels. *Applied Psycholinguistics, 17*, 313–334.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning, 50*, 617-673.

Nation, P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning, 21*, 393-408.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, *51*, 719-758.

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

Oi, K. (2005). Teaching argumentative writing to Japanese EFL students using the Toulmin model. *JACET Bulletin, 41*, 123-140.

Pennington, M. (1996). *Phonology in English language teaching*. London: Longman.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177–228). Cambridge: MIT Press.

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

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

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: 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.

Rivers, W., & Temperley, M. (1978). *A practical guide to the teaching of English as a second language*. Oxford: Oxford University Press.

Robinson, P. (1996). Learning simple and complex rules under implicit, incidental rule-search conditions, and instructed conditions. *Studies in Second Language Acquisition*, *18,* 27–67.

Robinson, P. (2011). *Task-based language learning*. Oxford: Blackwell.

Rose, K., & Kasper, G. (Eds). (2002). *Pragmatics in language teaching*. Cambridge: CUP.

Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation teaching in EFL settings: The case of English vowels and Japanese learners of English. *Linguistic Journal, 3*, 17-41.

Saito, K. (2009, March). *The impact of English-specific segmentals on mutual intelligibility: Suggestions for pronunciation teaching*. Paper presented at The 30th Annual NY TESOL Applied Linguistics Conference, Columbia University, New York City, NY.

Saito, K. (2010). Introduction to English pronunciation. In *Hearing Marathon*. Tokyo: ALC press.

Saito, K. (2011a). Re-examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness*., *20*(1).

Saito, K. (2011b). Differential effects of phonological and lexicogrammatical errors on NS and NNS listeners' perceptions of comprehensibility: An exploratory study. *The Journal of Asia TEFL.*

Saito, K. (under revision). Towards indentifying teaching and learning priorities for acquiring intelligible pronunciation. *The Canadian Modern Language Review*.

Saito, K. (under review a). Variables affecting the effects of recasts on L2 pronunciation development of /ɹ/ by Japanese learners of English.

Saito, K. (under review b). Re-examining effects of form-focused instruction on L2 pronunciation development of /ɹ/ by Japanese learners of English: The role of explicit phonetic instruction.

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

Sanz, C., & Morgan-Short, K. (2004). Positive evidence vs. explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54,* 35-78.

Schmitt, N. (2008). State of the art: Instructed second language vocabulary acquisition. *Language Teaching Research, 12,* 329-363.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). Cambridge, UK: Cambridge University Press.

Schwartz, B. (1993). On explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15*, 147-163.

Scovel, T. (1988). *A time to speak*. Newbury House, New York.

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

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

Setter, J. & Jenkins, J. (2005). Pronunciation. *Language Teaching, 38*, 1-17.

Sharwood Smith, M. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition, 15,* 165-179.

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

Sheen, Y. (2007). The effects of corrective feedback, language aptitude and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 301-322). Oxford: Oxford University Press.

Sheen, Y. (2008). Recasts, language anxiety, modified output and L2 learning. *Language Learning, 58*, 835-874.

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. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching, 29,* 73-87.

Spada, N., & Lightbown, P. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly, 42*, 181–207.

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

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning, *Language Learning, 26*, 233-253.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowkey, MA: Newbury House.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–484). Mahwah, NJ: Lawrence Erlbaum.

Swain, M., & Lapkin, S. (1982). *Evaluating bilingual education: A Canadian case study.* Clevedon, UK: Multilingual Matters.

Takagi, N. (1993). *Perception of American English /r/ and /l/ by adult Japanese learners of English: A unified view.* Ph.D. Dissertation, University of California, Irvine.

Takashima, H., & Ellis, R. (1999). Output enhancement and the acquisition of the past tense. In R. Ellis (Ed.), *Learning ad second language through interaction* (pp.173-188). Amsterdam: John Benjamins.

Tomasello, M., & Herron, C. (1989). Feedback for language transfer errors: the garden path technique. *Studies in Second Language Acquisition, 11,* 385-395.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning. 2nd Edition*. New York: Macmillan.

Trofimovich, P., & Baker, W. (2006). Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

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.

Trofimovich, P., Lightbown, P. M., Halter, R., & Song, H. (2009). Comprehension-based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition, 31*, 609-639.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning, 52*, 755–803.

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

VanPatten, B., & Cardierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15*, 225-243.

VanPatten, B., & Okiennon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition, 18,* 495-510.

Varonis, E., & Gass, S. (1982) The comprehensibility of non-native speech. *Studies in Second Language Acquisition, 4,* 114–136.

Venkatagiri, H., & Levis, J. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness, 16*(4), 263–277.

Walley, A. C., & Flege, J. E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics, 27*, 307-332.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology, 50*, 509-535.

White, J. (1998). Getting the learner's attention: a typographical input enhancement study In C. Doughty and J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 85-113). Cambridge: Cambridge University Press.

White, L. (1987). Against comprehensible input: the input hypothesis and the development of L2 competence. *Applied Linguistics, 8,* 95-110.

White, L., Spada, N., Lightbown, P., & Ranta, L. (1991). Input enhancement and L2 question formation. *Applied Linguistics, 12*, 416-432.

Yamada, R. A. (1993). Effect of extended training on /r/ and /l/ identification by native speakers of Japanese. *Journal of the Acoustical Society of America, 93,* 2391.

Yamada, R. A. (1995). Age and acquisition of second language speech sounds: Perception of American English /p/ and /l/ by native speakers of Japanese. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305–320). Timonium, MD: York Press.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past-tense forms. *Studies in Second Language Acquisition, 32*, 235-263.

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *Neuroimage, 46*, 226-240.