

EMPIRICAL STUDY

Video-Based Interaction, Negotiation for Comprehensibility, and Second Language Speech Learning: A Longitudinal Study

Kazuya Saito^a and Yuka Akiyama^b

^aBirkbeck, University of London and ^bGeorgetown University

This study examined the impact of video-based conversational interaction on the longitudinal development (one academic semester) of second language production by college-level Japanese English-as-a-foreign-language learners. Students in the experimental group engaged in weekly dyadic conversation exchanges with native speakers in the United States via telecommunication tools. The native speaker interlocutors were trained to provide interactional feedback (recasts) when the nonnative speakers' utterances hindered successful understanding (i.e., negotiation for comprehensibility). The students in the comparison group received regular foreign language instruction without any interaction with native speakers. The coded video data showed that the experimental students worked on improving all linguistic domains of language, likely in response to their native speaker interlocutors' interactional feedback (recasts, negotiation) during the treatment. The pretest–posttest data of the students' spontaneous production showed

We are grateful to *Language Learning* reviewers as well as the journal editor, Pavel Trofimovich, for their constructive feedback on earlier versions of the manuscript. We also acknowledge Shungo Suzuki, Masaki Eguchi, George Smith, Ze Shan Yao, and Ethan Beaman for their help in data collection and analyses. The project was funded by a Grant-in-Aid for Scientific Research in Japan (No. 26770202) to the first author and the ITEL Teletandem Project at Georgetown University to the second author.



This article has been awarded an Open Materials badge. All materials are publicly accessible in the IRIS digital repository at <http://www.iris-database.org>. Learn more about the Open Practices badges from the Center for Open Science: <https://osf.io/tvyxz/wiki>.

Correspondence concerning this article should be addressed to Kazuya Saito, Birkbeck, University of London, Department of Applied Linguistics and Communication, 26 Russell Square, London, WC1B 5DT, United Kingdom. E-mail: k.saito@bbk.ac.uk

that they made significant gains in the dimensions of comprehensibility, fluency, and lexicogrammar but not in those of accentedness and pronunciation.

Keywords interaction; second language speech; listening; pronunciation; fluency; vocabulary; grammar

Introduction

Adult second language acquisition (SLA) is a complex phenomenon affected by a range of individual factors. However, most theoretical accounts assume that second language (L2) learners can improve their oral ability through increased conversational experience with other native (NSs) and nonnative speakers (NNSs). Although many empirical attempts have been made to describe the features of L2 interaction as well as investigate the effects of L2 interaction on acquisition in controlled laboratory settings, most existing studies have only involved a brief amount of interaction treatment (< 1 hour) and have been exclusively concerned with L2 lexicogrammar development. This article reports on an experimental study that examined the longitudinal development (over one academic semester) of Japanese college students' spontaneous English production abilities via weekly dyadic conversation exchanges with NSs. Using a video conferencing tool, participants collaboratively worked on improving the comprehensibility of L2 speech with a primary focus on meaning (i.e., negotiation for comprehensibility).

Background

Conversational Interaction and SLA

Over the past 40 years, one of the most extensively researched topics in SLA has been the role of conversational interaction in language acquisition. This line of research has been generally motivated by several versions of the interaction hypothesis (e.g., Gass, 1997; Long, 1983, 1996; Pica, Holliday, Lewis, & Morgenthaler, 1989). The main tenet of the hypothesis states that adult SLA can be facilitated and promoted through conversational interaction with other NSs and NNSs because such interaction provides many opportunities to impact various aspects of SLA processes, especially when interlocutors encounter and work together on solving communication breakdown attributable to language.

To facilitate NNSs' comprehension, for example, NSs modify their speech (thus providing comprehensible input) through the repetition of utterances with emphasis of key words at a slower speech rate and the rephrasing of utterances with more frequent and simple words (Long, 1983). NSs try to retrieve meaning from NNSs' speech by using several negotiation strategies, such as repetition,

confirmation checks, and clarification requests in the case of communication breakdown, and they may signal comprehended, yet erroneous, speech through recasting of NNSs' erroneous forms, thereby providing interactional feedback (Lyster & Saito, 2010). NNSs may thus be induced to notice and understand the gap between their own interlanguage system and the input and subsequently be pushed to repair their production through comprehensible output (Swain, 2005). Consequently, the entire process of incidental, conversation-driven focus on form is believed to help NNSs improve their L2 performance in the most optimal manner (Goo & Mackey, 2013).

A number of researchers have probed the interaction–acquisition link by conducting experimental studies using a pretest–posttest design. Such a design allows researchers to control various features of L2 interaction as independent variables and test their impact on L2 development (for reviews, see Mackey, 2012, and Plonsky & Gass, 2011). Earlier intervention studies revealed that L2 learners were able to improve their grammatical and lexical performance when given opportunities to negotiate meaning through interaction rather than through mere exposure to simplified input (e.g., Ellis & He, 1999; Mackey, 1999). Subsequently, researchers further examined the extent to which such gains resulting from L2 interaction varied according to various factors. For instance, it has been shown that the efficacy of interaction can be increased when NNSs have sufficient proficiency with the target structures (e.g., Mackey & Philp, 1998, for developmentally ready learners) and/or relatively high aptitude (e.g., Goo, 2012, for working memory). Much research attention has been given to examining the facilitative role of more explicit and pedagogically elaborated feedback (e.g., Sheen, 2007, for metalinguistic correction). Other researchers have also shown that L2 interaction can be equally beneficial for interlanguage development when the treatment is delivered in various contexts (e.g., Ziegler, 2015, for face-to-face vs. teletandem communication; Gass, Mackey, & Ross-Feldman, 2005, for classroom vs. lab settings).

Mackey and Goo's (2007) meta-analysis of 26 intervention studies showed that L2 interaction was overall beneficial for acquisition with a medium effect size. Though revealing, these findings pointed out several methodological issues worthy of further investigation. First, claims regarding the effectiveness of interaction were derived from only a brief amount of treatment, as treatment length typically lasted less than 1 hour, reflecting the paucity of longitudinal work in SLA. In contrast, the effect size for benefits of L2 interaction was shown to be larger for delayed posttests (e.g., 1 month after interaction), compared to immediate posttests, suggesting that the benefit of interaction needs to be assessed via a long-term framework (Ortega, 2014).

Second, most L2 interaction studies included in the meta-analysis exclusively focused on the effectiveness of interaction for lexical and morphosyntactic development. According to several descriptive studies, episodes of negotiation for meaning happen in all linguistic domains (pronunciation, vocabulary, grammar), and L2 learners are generally more aware of pronunciation-focused feedback, compared to grammar-focused feedback (e.g., Mackey, Gass, & McDonough, 2000). However, few studies have experimentally examined how L2 learners can improve their pronunciation and fluency skills through negotiation for meaning during natural conversation experience with NSs. To this end, this study took a first step toward providing a longitudinal look at the extent to which interaction can impact not only lexicogrammar but also pronunciation and fluency aspects of adult L2 learners' spontaneous speech.

Developing L2 Oral Ability

L2 oral ability has been traditionally defined as a composite phenomenon and has been analyzed via global measures (comprehensibility, accentedness) and specific measures of pronunciation (segmentals, prosody), fluency (speech rate), vocabulary (appropriateness, richness), and grammar (accuracy, complexity; e.g., Trofimovich & Isaacs, 2012; Saito, Trofimovich, & Isaacs, 2015). Although it is often difficult for adult L2 learners to attain nativelike proficiency in all domains of language (Abrahamsson & Hyltenstam, 2009), recent L2 speech studies have suggested that L2 learners can enhance the overall comprehensibility of their speech, regardless of foreign accent, through increased interaction with other NSs and NNSs, typically operationalized as length of residence (LOR) in a L2 speaking environment.

For instance, Derwing and Munro (2013) conducted a longitudinal investigation into how late immigrants could improve their oral abilities during 7 years of immersion in Canada. The results showed that motivated and regular L2 users enhanced their overall comprehensibility but that their foreign accentedness demonstrated little change over time. Furthermore, a growing body of research on study-abroad learning has shown that adult learners indeed tend to show quick development of fluency (e.g., Mora & Valls-Ferrer, 2012) and lexicogrammar accuracy (Vercellotti, 2015), given a short-to-medium amount of immersion (e.g., 1 year of immersion). However, it requires a great deal of L2 experience (several years of immersion) to attain refined segmental and prosodic aspects of L2 speech (e.g., Flege, 2009).

Taken together, the aforementioned studies lend some evidence to the claim that adult L2 learners tend to improve their L2 oral proficiency by selectively working on linguistic domains with high communicative value with the goal of

achieving successful social interaction and communication. Whereas the quick development of fluency and lexicogrammar (related to comprehensibility) is characteristic of the early phase of L2 speech learning (LOR = 1–3 years), the gradual refinement of pronunciation (strongly tied to accentedness) is typical of long-term L2 speech learning (LOR = 5+ years). For further discussion of the relationship between L2 experience and learning in naturalistic SLA, see Flege (2009) and Saito (2015).

Notably, most previous research has exclusively focused on immigrants and English-as-a-second-language (ESL) students in naturalistic (rather than controlled) settings, where interaction with NS and NNS interlocutors with various backgrounds in a range of social contexts is common. In prior research, L2 learners were often asked to self-report how much they used the L2 with NSs and NNSs at work and home in a retrospective manner. Nevertheless, it remains unclear how such self-report measures can reliably reflect actual conversation experience because it is likely subject to a large amount of individual variability on a daily basis. As Flege (2009) pointed out, it is still methodologically difficult to keep track of the exact amount and nature of L2 interaction for a prolonged period of time from a longitudinal perspective (cf. Ranta & Meckelborg, 2013).

The Current Study

One ideal testing ground for a longitudinal analysis of L2 interaction is the foreign language setting, where L2 learners' target language use is highly limited outside of classrooms. This allows researchers to control the quality/quantity of interaction as an independent variable when it comes to conducting experimental studies. To create communicatively authentic conversation opportunities in such foreign language settings, some SLA studies have targeted interaction between NNSs and NSs by way of synchronous computer-mediated communication tools. This kind of online-based telecollaborative interaction is considered as "one of the main technologically enhanced activities in modern foreign language education" (O'Dowd, 2011, p. 368), enabling all kinds of L2 learners to connect with NSs and NNSs all over the world (Belz, 2003). Similar to research findings in face-to-face interaction, it has been found that L2 learners who interact with NSs via video-conferencing tools also tend to have opportunities for negotiation for meaning, pushed output, and focus on form (Wang, 2006) and can improve their oral abilities (Monteiro, 2014).

To date, there have been few attempts to provide a longitudinal analysis of the role of conversational interaction in L2 speech learning in foreign language settings. One such example is Payne and Whitney's (2002) project, where

American learners of Spanish were divided into small groups (four to six students) and engaged in peer interaction activities in the target language (e.g., discussion of cultural texts and video) in either a face-to-face or text-chat mode over the course of a semester. The results showed that all participants significantly improved their oral ability, which was measured via an oral proficiency interview and assessed using the American Council on the Teaching of Foreign Languages guidelines.

With respect to NS–NNS interaction, Akiyama (in press) descriptively documented how focus on form practices occurred longitudinally in the context of video-based dyadic interaction. In this study, American learners of Japanese had biweekly opportunities to engage in task-based telecollaborative interaction with NSs of Japanese via a video-conferencing tool over one academic semester (15 weeks). To elicit incidental focus on form in an optimal fashion, the NS interlocutors were trained to provide various types of interactional feedback (e.g., recasts, prompts, explicit correction). According to the results of the survey analyses, the participants identified recasts as the most preferred type of interactional feedback. This is arguably because recasts were considered by the learners to be immediate, time saving, unobtrusive, and easy to provide (Loewen & Philp, 2006). From the interaction data, it was found that the majority of focus-on-form episodes (e.g., learners' successful repair) indeed resulted from their most preferred type of interactional feedback—recasts for a majority of the participants.

Building on this line of L2 interaction research, this study was designed to examine the effect of L2 interaction on the longitudinal development (one academic semester) of native Japanese L2 English learners' oral abilities with a pretest–posttest design. The quality and quantity of L2 interaction moves were coded and analyzed according to the type and number of triggers (pronunciation, vocabulary, grammar errors), interactional feedback (recasts, negotiation), and uptake (repair, needs repair, no uptake; Lyster & Ranta, 1997). The learners' interactional gains were calculated using a range of speech assessment instruments measuring global, phonological, temporal, lexical, and grammatical dimensions of language.

The NS interlocutors in this study were trained to provide interactional feedback in the form of recasts, which was the most preferred type of feedback in task-based telecollaboration in our precursor research (Akiyama, in press). During the training, we explicitly asked the NSs to provide feedback only when they perceived the NNSs as incomprehensible (see below). Our intention was to create interactional contexts that induced the Japanese learners to work selectively on certain linguistic errors directly relevant to comprehensibility

(rather than accentedness) in collaboration with their NS partners, while the primary focus of the entire interaction was on meaning.¹ This decision reflects the widely accepted view that comprehensibility and accentedness are two interrelated yet essentially distinct constructs and that not all linguistic errors linked to accent hinder comprehensibility (e.g., Derwing & Munro, 1997).

The nature of this specific interaction—termed here negotiation for *comprehensibility*—differs from the broad idea of negotiation for *meaning*, wherein any interactional feedback move is purely incidental as a means to “resolve communication breakdowns and to work toward mutual comprehension” (Pica et al., 1989, p. 65). At the same time, our feedback orientation toward comprehensibility-related linguistic features can also be distinguished from the underlying notion of negotiation of *form*, defined as “the provision of corrective feedback that encourages self-repair involving accuracy and precision and not merely comprehensibility” (Lyster & Ranta, 1997, p. 42), which is typical of teacher–student interactions. In the latter contexts, feedback providers (teachers) equally attend to all linguistic errors in order to push receivers (students) to linguistic nativelikeness, as many L2 learners would otherwise stay plateaued, especially after their linguistic performance has become sufficiently comprehensible (Swain, 2005).

The primary goal of the study was to explore the extent to which video-based negotiation for comprehensibility could help inexperienced Japanese learners with little experience abroad improve various dimensions of their L2 oral ability over one academic semester. In light of the extensive literature on naturalistic L2 speech learning (Derwing & Munro, 2013; Saito, 2015), the following hypotheses were formulated.

1. The effect of the interaction treatment would be clearly observed for those linguistic features related to the early phase of L2 speech learning (comprehensibility, fluency, vocabulary, grammar) because they have been found to be susceptible to quick changes according to study-abroad research (Mora & Valls-Ferrer, 2012).
2. Attainment of decreased accentedness and improved pronunciation would be limited because adult L2 learners’ successful acquisition of relatively difficult features of L2 segments and prosody requires a great amount of interactional experience (Trofimovich & Baker, 2006) and/or unique individual difference profiles, such as early age of acquisition (Abrahamsson & Hylstenstam, 2009) and high aptitude (Granena & Long, 2013).

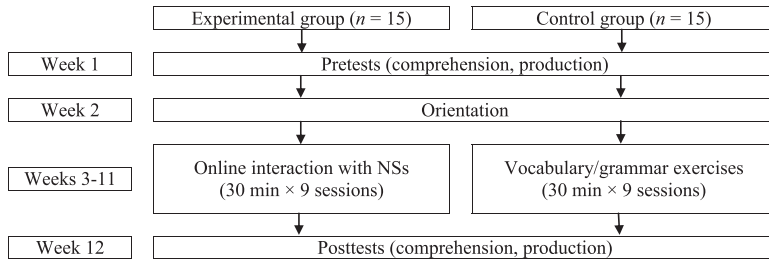


Figure 1 Project timeline.

Method

Participants

L2 Learners (Japanese English-as-a-Foreign-Language [EFL] Learners)

Thirty first- and second-year Japanese undergraduate students majoring in business at a university in Japan participated as volunteers. We recruited participants by distributing two different flyers, one offering conversational activities and the other offering vocabulary/grammar activities. We assigned the students interested in the former to the experimental group (10 males, 5 females; $M_{age} = 19.2$ years) and those interested in the latter to the comparison group (6 males, 9 females; $M_{age} = 18.9$ years). At the time of the project, all students were registered (as a university requirement) in approximately 3 hours of EFL lessons per week taught by Japanese instructors. According to the department syllabus and our casual classroom observations, the content of the EFL classes mainly consisted of listening and reading activities without many opportunities to produce language. After the participants took the pretests in Week 1, they joined weekly 30-minute extracurricular L2 activities outside of their regular EFL syllabus (i.e., 3 hours of language arts lessons) between Weeks 2 and 11. While the Japanese students in the comparison group did vocabulary/grammar exercise activities, those in the experimental group engaged in task-based conversation activities with their NS partners in the United States by way of a video-conferencing tool (Google Hangout). One week after the last session (in Week 12), the participants took the posttests. The timeline of the study is depicted in Figure 1.

The analysis of the language background questionnaire showed that all participants had studied English in EFL classrooms for 6 years prior to the project, typically through grammar-translation methods. They had little experience abroad except for family and school trips, when they mainly used Japanese, their first language (L1), throughout their brief stays in L2 speaking

environments (<1 month). No participants reported any experience at private language schools to practice conversational English during the project, indicating that their L2 use with NSs was limited outside of classrooms. Given their homogeneous L2 learning backgrounds, typical of many other EFL instructional settings (i.e., a significant lack of L2 conversational experience), these participants could be considered as inexperienced speakers.²

NS Interlocutors (American Learners of Japanese)

Fifteen NSs of English (nine males, six females; $M_{age} = 21.2$ years) learning Japanese at U.S. universities also participated in the study. Some participated in this project as part of a one-credit course, while others were volunteer exchange partners who expressed interest in practicing Japanese outside the classroom. They had a wide variety of L2 Japanese learning experience (spanning intermediate to advanced proficiency levels) and reported high familiarity with Japanese-accented English.

Experimental Group

Fifteen of the Japanese L2 learners constituted the experimental group and participated in dyadic interaction with NS interlocutors in the United States via the video-conferencing tool Google Hangout over one academic semester (nine sessions in total). In each 60-minute session, the participants interacted with each other in English for the first half of the session and switched to Japanese for the second half. The methodology and results for the 30 minutes of interaction in English (Japanese students as NNSs, American students as NSs) are reported here.³ At the onset of the project (Week 2), both the NNS and NS participants received explanation on the twofold intention of the telecollaboration activities: (a) promoting NNSs' experience in meaningful conversation with NSs as the main goal and (b) improving the linguistic quality of NNSs' speech through negotiation for meaning as the secondary goal. To this end, the participants received training not only on how to proceed with the task-based interaction, but also on how to negotiate for comprehensibility (see below).

Task-Based Interaction

During the interaction activities (Weeks 3–11), communication mainly occurred via video, with minimal use of the multimodal features of Google Hangout (e.g., text chat, screen sharing). This was done to ensure that the video-conferencing environment resembled face-to-face interaction as much as possible. This decision addresses Develotte, Guichon, and Vincent's (2010) concern that the

potentially different amount of L2 learners' technological knowledge of video-conferencing tools and familiarity with video-mediated conversations may influence their effectiveness. Due to the time difference between Japan and the United States, the participants were allowed to conduct the sessions with much flexibility to accommodate their schedule outside the classroom using their own computers. The NNSs were required to report to the researcher the date and time of each session. All of the sessions were video-recorded and stored as digital data for the analyses.

We employed a type of information exchange task called visual-based conversation, following the suggestion of Lee (2002), who found that the two-way exchange of information on real-life topics that are theme-based and minimally structured helped students recycle ideas and reinforced language skills. Each week, NNSs were asked to find two visuals (one for Japan and the other for the United States) that represented the theme of the week and to prepare two discussion questions for each visual image. For instance, if the theme of the week was pop culture, NNSs might have chosen a visual of a Japanese idol group for the Japanese visual and Hollywood movies for the American visual. This type of open-ended yet authentic task requires various functional skills such as describing, narrating, and expressing opinions (Lee, 2002) and prompts negotiation for meaning (Doughty & Pica, 1986).

Recast Training

To promote the acquisitional value of L2 interaction, NS interlocutors were explicitly asked to provide conversational modifications—interactional feedback in the form of recasts—in response to NNSs' linguistic errors that might hinder successful comprehension, as was the case in previous L2 interaction studies (e.g., Mackey et al., 2000). Recasts are defined as the reformulation of erroneous L2 speech and have been identified as the most frequent type of interactional feedback in NS–NNS dyadic interaction as well as student–teacher classroom interactions (Goo & Mackey, 2013). From a theoretical perspective, recasts are believed to play an important role in adult SLA because they provide both positive (modeling correct form) and negative (signaling errors) evidence to NNSs without interrupting the communicative flow of conversational moves (Long, 1996). Examples 1–3 illustrate typical recast episodes in our data set.

Example 1. Pronunciation

NNS: This is a picture of my [kɔɾɪdʒ].

NS: Oh that's your [kɔɾɪdʒ] (college).

Example 2. Vocabulary

NNS: I do not like terror attraction.

NS: Ah, you don't like scary rides.

Example 3. Grammar

NNS: Ah. Do you enjoy party very well?

NS: Do I enjoy partying? Yeah I do. I do like partying.

During the orientation in Week 2, NSs received training from the researcher on how to negotiate for comprehensibility (i.e., selectively focus on errors related to message delivery) by drawing on recasts when their NNS interlocutors' linguistic errors interrupted the communicative flow of L2 interaction. In keeping with similar L2 interaction studies (Mackey, 1999; Mackey et al., 2000), the training procedure was elaborated and operationalized as follows.

1. NSs were clearly told that the primary focus of the project was to complete conversational tasks successfully, providing interactional feedback only where natural and appropriate.
2. NSs were introduced to two different constructs of L2 speech learning—improving comprehensibility (becoming easier to understand) versus reducing accentedness (becoming more nativelike).
3. NSs were explicitly instructed to provide recasts when they found their NNS partners' linguistic errors to impair the comprehensibility (but not necessarily accentedness) of their speech.
4. To familiarize themselves with the procedure, NSs watched a series of video clips of examples of recast episodes. For each episode, they discussed with the researcher the kinds of errors the NNS interlocutor made (i.e., trigger) and how the NS interlocutor helped the NNS improve in comprehensibility (i.e., recasts, repair). Such recasts could occur as part of negotiation strategies (e.g., confirmation requests, clarification requests) after communication breakdown and/or when NSs perceived NNSs' errors as having the potential to pose communication problems in the future (for error correction scripts, see Appendix S1 in the Supporting Information online).

Coding of Interaction Patterns

To explore the nature of communicative focus on form during the semester-long L2 interaction activities, a linguistically trained coder watched the video-recorded interactions of the 15 dyads at the pretest (the second session: Week 4)

and posttest (the eighth session: Week 10) of the project (30 minutes \times 2 sessions \times 15 dyads). Following the norm established in previous research (e.g., Lyster & Ranta, 1997), the data were analyzed according to the three crucial components of L2 interaction:

1. Trigger, defined as the number of linguistic errors that L2 learners make in the domains of pronunciation (i.e., mispronunciation of segmentals and prosody), vocabulary (wrong word choice), and grammar (morphosyntactic errors).
2. Feedback, defined as the number of recasts and negotiation strategies (confirmation checks, repetition, clarification requests) that NS interlocutors adopt in response to NNSs' errors.
3. Uptake, defined as L2 learners' reaction to the feedback move, including repair (successfully producing more targetlike production), needs repair (failing in self-correcting errors despite any attempt), and no uptake (showing no reaction to NS interlocutors' feedback).

Based on the total of 15 hours of coded data at two different time points (pretest, posttest), we aimed to provide suggestive patterns regarding (a) how frequently the L2 learners received feedback according to different linguistic categories (pronunciation, vocabulary, grammar) and (b) to what degree they successfully noticed/repared their phonological, lexical, and grammatical errors.

Comparison Group

The remaining 15 Japanese NNS learners formed the comparison group and participated in weekly individual vocabulary/grammar activities instead of task-based interaction activities with NSs. The reasons for including the comparison group in the study were twofold. First, because we used identical materials in the pretest and posttest sessions, the analysis of the comparison group's performance allowed for an examination of any test-retest effects. Second, the comparison group's performance could also be considered as an index of how much the Japanese students could improve their oral abilities over one academic semester via EFL instruction (a few hours of language arts lessons per week) without any opportunities for interaction. In so doing, we aimed to separate the learning patterns in the comparison group (i.e., effects of test-retest and one semester of EFL instruction) from any L2 interaction gains in the experimental group.

During the orientation (Week 2), the 15 Japanese students in the comparison group were explicitly told that the purpose of the project was to improve their L2 vocabulary and grammar ability with the goal of attaining higher scores on

the Test of English for International Communication (TOEIC). They were also instructed on how to complete the take-home assignment every week. Between Weeks 3 and 11, the NNSs were asked to practice using a variety of vocabulary and grammar activities, which consisted of vocabulary recall tests based on JACET 8000 (Aizawa, Ishikikawa, Murata, & Iso, 2005; i.e., comprehension practice) and fill-in-the-blank grammar questions in Part 5 of the TOEIC (i.e., production practice). The weekly assignments, which typically took 30 minutes to complete at home, were graded and recorded by the researcher.

Production Test

Traditionally, L2 speech has been measured via highly controlled production tasks, such as reading aloud of written and audio prompts, so that researchers can analyze the linguistic structures of interest in a consistent and regulated manner (for a review, see Piske, MacKay, & Flege, 2001). Yet, other researchers have pointed out that such tasks allow adult L2 learners to carefully monitor their correct production, potentially drawing on their metalinguistic knowledge (Abrahamsson & Hyltenstam, 2009). To measure the actual state of L2 representation and processing abilities, many scholars have emphasized the importance of eliciting spontaneous speech via free-constructed tasks (e.g., picture narratives) by inducing L2 learners to pay equal attention not only to pronunciation, fluency, vocabulary, and grammar (i.e., linguistic form) but also to successful task completion (i.e., meaning) under time pressure conditions (Spada & Tomita, 2010).

In this study, a timed picture description was adopted from Saito (2015) to measure the participants' pronunciation, fluency, vocabulary, and grammar performance during spontaneous speech. The task was designed to allow L2 learners with various proficiency levels (including even low beginners) to contribute spontaneous speech without much dysfluency (filled and unfilled pauses, repetitions). The test also could be considered especially suitable for this project because it simulated the interaction that the participants engaged in during dyadic conversation exchanges (i.e., picture descriptions followed by discussion).

As part of the production test, the participants were asked to describe seven individual pictures with only 5 seconds of planning time per picture. To reduce the effect of task familiarity on their performance, the first four pictures were used for practice, and the remaining three pictures were used for the final analyses. Each picture contained three word cues as hints to facilitate task completion. These key words were carefully chosen to represent a range of pronunciation problems typical for Japanese learners of English (see

Saito, 2014). For example, Japanese learners tend to neutralize the English /r/-/l/ contrast (*rain, rock, brew, crowd* vs. *lane, lock, blue, cloud*) and insert epenthetic vowels between consecutive consonants (/dərɑrvə/ for *drive*, /θəri/ for *three*, /səkaɪ/ for *sky*) and after word-final consonants (/teɪbəl/ for *table*, /myuzɪkə/ for *music*, /rəʊdə/ for *road*). The three pictures featured: (a) a table left on the driveway in the heavy rain (key words: *rain, table, driveway*), (b) three men playing rock music with one man singing a song and the other two men playing guitars (key words: *three guys, guitar, rock music*), and (c) a long road under a blue sky with a lot of clouds (key words: *blue sky, road, cloud*).

The first 10 seconds of each picture description were extracted for each participant. We carefully edited and cut full phrases to keep the samples as natural as possible, therefore, the length of the speech samples varied from 8.5 to 12.3 seconds ($M = 10.3$). The three picture descriptions were combined and stored as a single wav file for each speaker at the pretest and posttest sessions, resulting in 60 speech samples (30 NNSs \times 2 tests). The mean length of each file was 30.4 seconds (27.2–35.1), which can be considered as sufficiently long to provide the linguistic information required for L2 speech analysis, in line with previous research standards (e.g., Derwing & Munro, 1997, for 10–15 seconds; Hopp & Schmid, 2013, for 10–20 seconds; Trofimovich & Isaacs, 2012, for 30 seconds). The mean number of words for each file was 40.3 (25–62).

At Weeks 1 and 12, the picture descriptions were recorded individually in a quiet room at the university using a Roland-05 audio recorder (set at 44.1-kHz sampling rate and 16-bit quantization) and a unidirectional condenser microphone. To avoid any confusion and misunderstanding regarding the task procedure, all instructions were delivered in Japanese by trained L1 Japanese assistants.

Global Analyses

The global construct of L2 speech was assessed based on novice raters' intuitive judgment of how easily the L2 speech samples could be understood (i.e., comprehensibility), as well as how nativelike they were in comparison to NSs (i.e., accentedness; Derwing & Munro, 1997). Given that these two global domains were conceptualized as a reflection of how NSs react to speech in real-life communication with NNSs, novice participants were recruited for rating purposes (Trofimovich & Isaacs, 2012; Saito, Trofimovich et al., 2015; Saito, Webb, Trofimovich, & Isaacs, 2015).

Novice Raters

Five NS raters (two males, three females) were recruited at an English-medium university in Montreal, Canada ($M_{age} = 24.0$ years). According to the results of a language background questionnaire, they spoke English more than 90% of the time. Additionally, at least one of their parents was a NS of English. They were carefully selected according to the definition of novice rater in Isaacs and Thomson (2013) in terms of their lack of linguistic and pedagogical experience (i.e., they had not taken any linguistics courses nor taught English in ESL/EFL settings) and their low familiarity with Japanese-accented English (i.e., their mean self-report score was 2.2, ranging from 1 to 3 on a 6-point scale: 6 = *very much*, 1 = *not at all*). None reported any hearing problems.

Procedure

After receiving a brief amount of instruction on comprehensibility and accentedness, the raters listened to 60 speech samples delivered in a randomized order using Z-Lab (Yao, Saito, Trofimovich, & Isaacs, 2013), custom software developed using the commercial software package MATLAB (2013). Then they used a free-moving slider on a computer screen based on a 1,000-point scale to evaluate comprehensibility (0 = *hard to understand*, 1,000 = *easy to understand*) and accentedness (0 = *heavily accented*, 1,000 = *little accent*), with the leftmost endpoint labeled with a frowning face and the rightmost endpoint with a smiling face. To ensure that their ratings captured their NS intuitions, they were allowed to play each sample only once. The raters first practiced the procedure with three trial samples (not included in the main data set) and then proceeded with the assessment of all 60 samples. The entire session lasted 1 hour with a 5-minute break halfway through. For training scripts and onscreen labels, see Appendix S2 in the Supporting Information online.

Rater Consistency

Cronbach's alpha for the inter-rater agreement of the five novice raters was high for comprehensibility ($\alpha = .91$) and accentedness ($\alpha = .93$). The novice raters' scores were therefore considered sufficiently consistent and were averaged to derive a single score per rated category for each speaker.

Pronunciation and Fluency Analyses

Following the tradition in L2 speech research, we used linguistically trained judges' impressionistic evaluations of segmentals (e.g., Piske, Flege, MacKay, & Meador, 2011), prosody (e.g., Trofimovich & Isaacs, 2012), and fluency (e.g., Pinget, Bosker, Quené, & de Jong, 2014).

Expert Raters

Five NS raters (two males, three females) were recruited at an English-medium university in Montreal, Canada ($M_{age} = 29.0$ years). In line with the definition of expert raters (Isaacs & Thomson, 2013), they were not only graduate students in applied linguistics with extensive experience with linguistic analyses of L2 speech but also had taught English in ESL/EFL settings ($M = 5.7$ years; 3–13.5). They reported relatively high familiarity with Japanese-accented English ($M = 4.8$, ranging from 4 to 6: 1 = *not at all*, 6 = *very much*).

Audio Measures

For the pronunciation and fluency analyses, the raters listened to the 60 speech samples in a randomized order via the MATLAB software (Yao et al., 2013) and used a moving slider like that used by the novice raters to evaluate: (a) segmentals (substitution, omission, or insertion of individual consonant and vowel sounds), (b) word stress (misplaced or missing primary stress), (c) intonation (appropriate, varied use of pitch moves), and (d) speech rate (speed of utterance delivery).⁴ Given the demanding nature of the rating task (the simultaneous judgement of four domains of L2 speech), the raters were allowed to replay each sample as many times as they wished until they felt satisfied with their judgments.

Procedure

The raters first received thorough instruction from a trained research assistant on the four pronunciation (segmentals, word stress, intonation) and fluency (speech rate) categories. They then practiced the rating procedure with five speech samples (not included in the main data set). For each practice file, the raters were asked to explain their decisions, with the assistant providing feedback to ensure that they had correctly understood the linguistic categories during their audio judgments. Finally, they proceeded with the main rating sessions, which lasted for approximately 2 hours (with a 10-minute break halfway through). For onscreen labels and training scripts, see Appendix S3 in the Supporting Information online.

Rater Consistency

Given the relatively high Cronbach's alpha achieved by the five expert raters for segmentals ($\alpha = .90$), word stress ($\alpha = .85$), intonation ($\alpha = .81$), and speech rate ($\alpha = .91$), their judgement scores were considered sufficiently consistent and then averaged to derive a single mean score per speaker for each pronunciation and fluency category.

Vocabulary and Grammar Analyses

All speech samples, each of which included one speakers' three timed picture descriptions, were orthographically transcribed and cleaned by removing obvious mispronunciations based on contextual information available in the pictures (e.g., *ought side* was transcribed as *outside*) and orthographic markings of pausing (e.g., uh, um, oh, eh). As was the case in previous L2 vocabulary and grammar literature (e.g., Saito, Webb et al., 2015; Crossley, Salsbury, & McNamara, 2015), the transcripts were submitted to the following lexicogrammar analyses conducted by a linguistically trained coder: (a) lexical appropriateness (ratio of nonnative word choices to total number of words); (b) lexical variation (number of different words per minute); and (c) morphosyntactic accuracy in verbs (ratio of tense, aspect, modality, and subject–verb agreement errors to total number of words), nouns (ratio of plural usage errors related to countable and uncountable nouns to total number of words), and articles (error ratio of article usage in terms of definite, indefinite, and null to total number of words).

Results

We first analyzed descriptively the experimental students' interactional behaviors (the number of feedback and uptake moves for individual NNSs in each dyad) and then examined the participants' global, phonological, temporal, lexical, and grammatical gain scores in relation to those of the comparison group.

Interaction Behavior

The first aim of the analysis was to estimate the extent to which negotiation for comprehensibility took place within the experimental group, which was explicitly encouraged to work on errors relevant to successful communication but not necessarily relevant to linguistic natelikeness. To this end, we analyzed communicative focus-on-form episodes (trigger, feedback, uptake) during the NS–NNS interaction activities by way of the coded data of 15 dyads at the beginning (second session, T1) and end (eighth session, T2) of the project.

Overall Patterns

The results of descriptive statistics are summarized in Table 1. The NS interlocutors only occasionally provided interactional feedback on NNSs' errors, targeting only 22.9% (T1) and 13.4% (T2) of errors. It is possible that, consistent with the instructions given, the NSs generally prioritized communicative flow over the correction of linguistic errors except when errors were perceived

Table 1 Overall interaction patterns of trigger, feedback, and uptake at T1 (Week 4) and T2 (Week 10)

Interaction pattern	T1		T2	
	<i>n</i>	%	<i>n</i>	%
Error triggers: All linguistic domains	664		795	
Feedback ^a				
No feedback	512	77.1	688	86.5
Recasts	117	17.6	74	9.3
Negotiation	35	5.3	33	4.1
Uptake after recasts ^b				
Repair	21	17.9	21	28.3
Needs repair	22	18.8	27	36.4
No uptake	74	63.2	26	35.1
Uptake after negotiation ^c				
Repair	3	8.5	6	18.1
Needs repair	14	40.0	17	51.5
No uptake	18	51.4	10	30.3

Notes. ^aPercentage was calculated from the total number of error triggers.

^bPercentage was calculated from the total number of errors triggering recasts.

^cPercentage was calculated from the total number of errors triggering negotiation.

to be detrimental to the flow of L2 communication. In order to retrieve impaired meaning, the NS interlocutors primarily used recasts, just as they were trained to do. Additionally, they also used negotiation (repetition, confirmation checks, clarification requests) as a secondary strategy, especially when they encountered communication breakdowns as a result of serious linguistic problems. In response to such NS feedback moves, the L2 learners exerted considerable effort to modify their output over time. Indeed, the interlocutors showed relatively high uptake (both repair and needs repair) toward approximately half of these feedback moves (T1 = 36.7%, T2 = 64.7% for recasts; T1 = 48.5%, T2 = 69.6% for negotiation), indicating that the L2 learners looked to the NS interlocutors' feedback in order to find which linguistic errors they needed to work on to be successfully understood.

Linguistic Dimensions

Next, we investigated the quality of interactional treatment according to different linguistic dimensions (pronunciation, vocabulary, grammar). As summarized in Table 2, although the L2 learners made a relatively large number of pronunciation (T1 = 56.3%, T2 = 51.8%) and grammar (T1 = 36.9%, T2 = 41.0%)

Table 2 Pronunciation-, vocabulary-, and grammar-related patterns for trigger, feedback, and uptake at T1 (Week 4) and T2 (Week 10)

Interaction pattern	Pronunciation				Vocabulary				Grammar			
	T1		T2		T1		T2		T1		T2	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Error triggers	334	56.3	453	51.8	48	6.0	45	7.2	272	36.9	297	41.0
Feedback ^a												
No feedback	282	82.0	406	89.6	21	43.8	22	48.9	209	76.8	260	87.5
Recasts	47	13.7	31	6.8	18	37.5	19	42.2	52	19.1	24	8.1
Negotiation	15	4.4	16	3.5	9	18.8	4	8.9	11	4.0	13	4.4
Uptake after recasts ^b												
Repair	6	12.8	8	25.8	8	44.4	6	31.6	7	13.5	7	29.2
Needs repair	12	25.5	15	48.4	0	0.0	7	36.8	10	19.2	5	20.8
No uptake	29	61.7	8	25.8	10	55.6	6	31.6	35	67.3	12	50.0
Uptake after negotiation ^c												
Repair	1	6.2	3	18.8	2	22.2	2	50.0	0	0.0	1	7.7
Needs repair	6	37.5	10	62.5	2	22.2	0	0.0	6	54.6	7	53.9
No uptake	8	50.0	3	18.8	5	55.5	2	50.0	5	45.5	5	38.5

Notes. ^aPercentage was calculated from the total number of error triggers within each linguistic domain.

^bPercentage was calculated from the total number of errors triggering recasts.

^cPercentage was calculated from the total number of errors triggering negotiation.

errors, the NS interlocutors directed a relatively small amount of feedback in response to these errors (T1 = 18.1%, T2 = 10.1% for pronunciation; T1 = 23.1%, T2 = 12.5% for grammar). Although the L2 learners made far fewer vocabulary errors (T1 = 6.0%, T2 = 7.2%), the NS interlocutors seemed to focus more than half of their feedback on lexical issues (T1 = 56.3%, T2 = 51.1%). At T1, the L2 learners demonstrated a similar amount of uptake (including repair and needs repair) in response to approximately 40% of the NSs' feedback (including both recasts and negotiation), regardless of linguistic focus: pronunciation (40.3%), vocabulary (44.4%), or grammar (36.5%). At T2, the L2 learners showed relatively high sensitivity (uptake) to pronunciation (76.6%) and vocabulary (65.2%) feedback, compared to grammar feedback (54.0%).

In summary, the descriptive results indicated that (a) the raw number of NNS errors (pronunciation and grammar > vocabulary) and NS feedback patterns (vocabulary > pronunciation and grammar) remained static throughout the

project and (b) the NNSs became increasingly sensitized to pronunciation and vocabulary feedback (compared to grammar feedback) by the end of the project.

Effects of Interaction on the Development of L2 Oral Ability

The second aim of the analyses was to examine the extent to which the students participating in the experimental group (i.e., those who were engaged with L2 interaction over one academic semester) improved their L2 production skills. Their performance was compared statistically to that of the comparison group, who only practiced vocabulary and grammar exercise activities.

To check for any preexisting differences in the experimental and comparison groups' oral ability, their performance at the time of the pretest sessions was analyzed according to the domains rated in the production test (comprehensibility, accentedness, pronunciation, fluency, vocabulary, lexicogrammar). Because the number of participants in each group ($n = 15$) was relatively low, a series of nonparametric Mann-Whitney tests were conducted, with the alpha level set at .05 and adjusted to .025 using the Bonferroni correction. Prior to the project, the two groups were comparable in their global production ($z = -1.43, p = .16$, for comprehensibility; $z = -.04, p = .97$, for accentedness), pronunciation ($z = -.24, p = .81$, for segmentals; $z = -.83, p = .41$, for word stress; $z = -1.30, p = .19$, for intonation), fluency ($z = -.85, p = .41$, for speech rate), vocabulary ($z = -1.49, p = .14$, for appropriateness; $z = -1.22, p = .23$, for variation), and grammar domains ($z = -.17, p = .87$, for verb error ratio; $z = -.29, p = .81$, for noun error ratio; $z = -.69, p = .51$, for article error ratio).

To examine the presence or absence of any significant improvement over time, a set of nonparametric Wilcoxon Signed Ranks tests were performed for the experimental and comparison groups, respectively. The participants' production scores at the pretest and posttest were used as a dependent variable with the alpha level set at .025 (Bonferroni corrected). The magnitude of their improvement over time was measured through Cohen's d .⁵ The descriptive and inferential statistics of the pretest and posttest production scores are summarized in Table 3.

The experimental group significantly improved its overall comprehensibility (but not accentedness) scores ($p = .01$) with a medium effect size ($d = .58$). In terms of specific domains of L2 speech, the experimental group significantly enhanced its perceived fluency (speech rate), lexical variation (number of different words per minute), and the accurate use of grammar (verb and article error ratio), with large effect sizes ($d = 1.17, .82, .90$, and 1.00 , respectively). In contrast, the comparison group did not make any significant improvement in any context ($p > .025$).

Table 3 Descriptive and inferential statistics for the production test scores over time

Production skill	Group	Pretest		Posttest		Pretest-posttest comparison		
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>z</i>	<i>p</i>	<i>d</i>
Global impression ^a								
Comprehensibility	E	315	165	420	191	2.49	.012*	0.58
	C	316	140	375	126	0.48	.629	0.44
Accentedness	E	292	161	320	175	1.07	.281	0.16
	C	284	136	266	101	0.45	.650	0.15
Pronunciation ^b								
Segmentals	E	343	149	365	147	1.07	.281	0.14
	C	329	114	367	100	1.93	.053	0.35
Word stress	E	416	118	452	128	1.70	.088	0.29
	C	371	99	386	123	0.51	.609	0.13
Intonation	E	361	146	418	168	2.01	.044	0.36
	C	307	113	347	122	2.10	.036	0.34
Fluency ^b								
Speech rate	E	329	157	514	159	3.40	.001*	1.17
	C	378	167	431	137	1.59	.112	0.34
Vocabulary								
Appropriateness ^c	E	2.8%	2.6%	1.1%	1.6%	-2.00	.045	0.78
	C	4.9%	3.8%	4.8%	4.1%	-0.18	.851	0.02
Variation ^d	E	38.4	11.9	48.6	12.9	-3.23	.001*	0.82
	C	43.4	6.7	46.0	17.1	-0.34	.733	0.20
Grammar								
Verb error ratio	E	7.6%	3.7%	4.4%	3.0%	-2.55	.011*	0.90
	C	8.5%	4.9%	8.0%	3.2%	-0.11	.909	0.32
Noun error ratio	E	2.6%	3.2%	3.4%	3.5%	-0.86	.386	0.23
	C	3.1%	3.8%	4.6%	4.0%	-0.73	.462	0.38
Article error ratio	E	7.5%	5.8%	2.4%	4.2%	-2.76	.006*	1.00
	C	6.7%	5.2%	6.9%	5.0%	-0.62	.532	0.03

Notes. E = experimental group, C = control group.

^bMaximum score = 1,000 points.

^cRatio of nativelike word choices to all words.

^dAverage number of words per minute.

* $p < .025$.

Discussion

Given the lack of longitudinal evidence for the acquisitional value of negotiated interaction in adult SLA, this study took an exploratory approach by

investigating the effect of weekly video-based interaction sessions on the development of Japanese EFL learners' oral abilities. This study is a novel contribution in that (a) the nature of interactional treatment was analyzed according to the type/amount of trigger, feedback, and uptake and (b) oral development was assessed through analyses of global (comprehensibility, accentedness) as well as specific linguistic (pronunciation, fluency, vocabulary, grammar) production.

Nature of Interactional Treatment

With respect to the content of the interactional treatment, we performed a detailed examination of the coded video data, which brought to light how interactional feedback actually took place across all linguistic domains (pronunciation, vocabulary, grammar). The NS interlocutors initiated negotiation for comprehensibility via recasts or other negotiation strategies upon encountering errors that negatively affected message delivery (10.1–23.1% of pronunciation and grammar errors; 51.1–56.3% of vocabulary errors). The descriptive results presented here confirmed what the NS interlocutors were explicitly asked to do during the recast training, that is, to selectively attend to those linguistic forms with particularly high communicative value while ignoring other linguistic errors that might not have an immediate influence on the flow of the ongoing communication. At best, the data suggested that Japanese EFL learners were given an adequate number of opportunities to work on their errors by drawing on negative and/or positive evidence available in the NSs' feedback in meaning-oriented interaction.

Overall, the descriptive results indicated that the nature of the interactional treatment in this study could be considered pertinent to the fundamental idea of L2 conversation as promoting opportunities to negotiate for meaning—not only can L2 learners receive more comprehensible input via interactional feedback from their NS interlocutors, but they are also pushed to modify their erroneous output to make themselves more easily understood (Long, 1996). As is the case with previous literature detailing the nature of L2 interaction, the L2 learners in this study demonstrated a similar distribution of linguistic errors (pronunciation, grammar > vocabulary) and received more intensive feedback on vocabulary than pronunciation and grammar errors (cf. Loewen & Philp, 2006). Similar to the L2 interaction research showing that L2 learners tend to be more aware of pronunciation-focused feedback relative to grammar-focused feedback (Mackey et al., 2000), the L2 learners in the study were also engaged in modifying pronunciation (76.6%) and vocabulary (65.2%) errors with a higher uptake ratio than for grammar (54.0%) errors.

Effects of Interaction

This study was designed to examine the acquisitional value of the semester-long interaction, and we did find significant improvement in the experimental group (participating in interaction sessions with NSs) but not in the comparison group (doing grammar- and vocabulary-focused activities). These results are consistent with the theoretical consensus in the field of L2 speech research that adult L2 learners improve their proficiency as a function of increased conversational experience with other NSs and NNSs (e.g., Flege, 2009). At the same time, our results for the comparison group echoed previous findings in instructed SLA that decontextualized L2 learning via drill (instead of interaction) activities may not be an optimal method for making any tangible changes in L2 spontaneous oral production ability (Spada & Tomita, 2010).

More importantly, the results of our comprehensive L2 speech measures further revealed that the extensive interaction activities were facilitative of certain (but not all) areas of L2 oral ability development. Interaction seemed to have a significant impact on comprehensibility, fluency, vocabulary, and grammar but not necessarily on accentedness and pronunciation. In accordance with the L2 speech literature, these findings shed some light on the complex mechanisms underlying the acquisitional value of interaction. It appears that NNSs can enhance some linguistic dimensions of speech that are related to NSs' understanding, such as comprehensibility (Derwing & Munro, 1997), optimal speech rate (Munro & Derwing, 2001), rich and complex vocabulary usage (Crossley et al., 2015), and grammatical accuracy (Saito, Webb et al., 2015). In contrast, communicatively oriented L2 learning may not clearly relate to other areas of language strongly tied to NSs' judgments of nativelikeness, such as accentedness (Derwing & Munro, 1997) and segmental and prosodic accuracy (Munro, Derwing, & Burgess, 2010; Trofimovich & Baker, 2006), at least within the time frame of this study (i.e., one academic semester).

Combining the analyses of the coded video data of the treatment as well as the analyses of the pretests and posttests allowed us to derive tentative conclusions regarding the potential and the limits of L2 interaction from a longitudinal perspective. With respect to the potential of the treatment, L2 learners can work communicatively on improving pronunciation, fluency, vocabulary, and grammar with primary attention devoted to maintaining successful communication with NS interlocutors. Consequently, L2 learners can enhance the overall comprehensibility of their L2 speech, improve lexical variation and morphosyntactic accuracy, and attain more targetlike production at an optimal tempo. With respect to the limits of the treatment, despite a number of instances of communicative focus on phonological form during the

interaction activities, negotiation for comprehensibility alone may not have been sufficient to make a significant impact on nativelikeness-related features entailing a great deal of learning difficulty (foreign accentedness, prosody, segmentals).

In comparison to the developmental patterns of L2 adults immersed in naturalistic settings, the differential effects of interaction on comprehensibility versus accentedness suggest a relatively strong benefit of negotiated interaction during the early phase of L2 speech learning. That is, extensive exposure to NS–NNS interaction could lead beginner-to-intermediate L2 learners to become more successful communicators due to the quick development of fluency (Mora & Valls-Ferrer, 2012) and lexicogrammar (Vercellotti, 2015). On the other hand, although some L2 learners did successfully repair their mispronounced and unclear utterances following NS interlocutors' feedback during the interactional treatment (repair ratio of 10–20%), such interaction-based learning did not appear to ultimately change the state of their L2 phonetic knowledge and performance, which was, in theory, measured via the pretests and posttests. In fact, previous L2 speech research has provided extensive evidence that L2 pronunciation learning is a slow, gradual, and extended process that occurs over several years of immersion and that its ultimate attainment is greatly influenced by other individual difference factors, such as aptitude and age of acquisition (Flege, 2009).

One intriguing way to maximize the effect of interaction for adult L2 oral development, entailing not only improved fluency and lexicogrammar but also refined segmental and prosodic accuracy, concerns the provision of explicit phonetic instruction. For instance, Derwing, Munro, Foote, Waugh, and Fleming (2014) found that instruction consisting of explicit teaching as well as focused practice activities targeting L2 prosody and fluency led even experienced ESL learners (with approximately 20 years of LOR) to improve in specific pronunciation features as well as overall comprehensibility. Similarly, Saito examined the acquisitional value of teachers' recasts targeting Japanese learners' mispronunciation of English /r/ in the context of meaning-oriented classrooms, reporting that a communicative focus on form could lead to gains not only at the controlled but also at the spontaneous speech levels (Saito, 2013; Saito & Lyster, 2012). Further research needs to examine how to implement such an instructional approach within communicatively oriented NS–NNS interaction by focusing on tasks designed to elicit learners' use of specific phonological features and/or feedback techniques targeting specific errors.

Limitations

Due to the exploratory nature of this project, several topics worthy of further investigation need to be addressed in future research. First, the findings were based on a relatively small sample size. Future work should replicate this study with a larger number of L2 learners from different L1 backgrounds. In particular, the multifaceted effect of L2 interaction should be further analyzed via various production measures, test materials, and analysis methods.

Although the experimental group significantly enhanced its vocabulary (variation) and grammar (verb/article error ratio) performance in L2 speech, such improvement was not found in the other domains of lexicogrammar (i.e., lexical appropriateness, noun error ratio). These findings contradict those of previous studies that found lexicogrammar accuracy to be important for NSs' speaking proficiency judgments (Crossley et al., 2015) and subject to improvement within a short amount of immersion (Schmitt, 1998). While different amounts of interaction benefits were found in this study depending on the different grammatical morphemes involved (verbs, articles > nouns), previous morpheme studies have identified adult L2 learners' tendency to acquire noun plurality before third person plurality, tense, and article (Bardovi-Harlig & Comajoan, 2008).⁶

In addition, the issue of adequate length of speech samples required for robust lexical analyses remains controversial and open to further validation. Whereas the spontaneous speech samples in this study could be considered relatively short ($M = 40.5$ words), the length of samples analyzed in previous research has included up to 2–3 minutes of speech (e.g., Crossley et al., 2015). Recent L2 vocabulary research has shown that 100 words may be necessary for certain lexical analyses (e.g., Koizumi & In'nami, 2012, for diversity). Thus, it would be interesting to examine further the impact of L2 interaction on speaking ability by adopting different speaking tasks, especially more argumentative, formal, and complex ones, whereby L2 learners are induced to demonstrate their productive L2 lexical and grammatical knowledge to a greater degree (see Hulstijn, Schoonen, de Jong, Steinel, & Florijn, 2012).

It is also important to remember that our discussion concerning the potential or limitations of L2 interaction was exclusively limited to a specific group of L2 learners—inexperienced Japanese EFL students in the early phase of L2 speech learning. Thus, it would be important to examine whether, to what degree, and how interaction can help L2 learners at various proficiency levels (including not only beginner but also advanced learners) to continue to improve their production ability not only for communicatively salient features (e.g., fluency, lexicogrammar) but also for those linguistic structures without

much communicative value or learnability (e.g., specific segmental contrasts), especially beyond one academic semester.

Moreover, we adopted a quantitative approach, which allowed us to attribute the participants' L2 development to the interaction process as a whole. To understand which specific features of L2 interaction differentially affect SLA processes, future studies could reexamine this topic while controlling for a number of common characteristics of interaction found in previous research, such as different types of interactional feedback (Lyster & Saito, 2010), task structures (Gass, Mackey, Alvarez-Torres, & Fernandez-Garcia, 1999), interlocutors (Storch, 2002), and the presence or absence of opportunities for repair (Leeman, 2003).

Finally, we would like to emphasize the importance of conducting not only quantitative, but also qualitative analyses of the relationship between negotiation for meaning and the development of oral ability in the long run. One promising direction would be to probe the perceptions of participants via stimulated recall instruments (Mackey et al., 2000), especially when the interlocutors interact with the same partners (as in this study) over longer periods of time. These studies will directly answer several unstudied questions, such as which linguistic errors NSs perceive to be detrimental to their successful understanding of foreign-accented speech (Trofimovich & Isaacs, 2012) and which features NNSs have difficulty in noticing and repairing without any explicit instruction (Lyster & Saito, 2010).

Conclusion

This study examined the impact of negotiation for comprehensibility via video-based interaction on inexperienced EFL speakers' longitudinal development of L2 oral production. Results showed that the experimental group significantly improved in global, temporal, and grammatical qualities of their L2 speech. On the one hand, the findings supported the acquisitional value of interaction, especially with regard to those features of L2 speech competence directly linked to successful communication (e.g., comprehensibility, fluency, grammar). On the other hand, the findings also revealed the limited efficacy of interaction, particularly for certain linguistic structures which are subject to much learning difficulty (e.g., pronunciation) and are highly relevant to perceived native-likeness. In conclusion, task-based interaction can greatly help inexperienced learners become successful communicators in the early phases of L2 speech learning because it provides a communicative focus on linguistic form that is crucial for expression of meaning. At the same time, however, it may require an extensive amount of L2 interaction (longer than one academic semester) and/or

intentional focus on form (e.g., explicit instruction) for learners to attend to and practice nonsalient and difficult features as a way to refine the linguistic nativelikeness of their speech and attain more advanced L2 oral ability.

Final revised version accepted 8 January 2016

Notes

- 1 Our decision to train NSs to provide recasts for certain salient (but not all) linguistic errors during task-based interaction corresponds to the L2 interaction research standard that considers the completion of communicative tasks as a primary objective, providing feedback only where natural and appropriate (without causing dysfluencies and learner irritation; see Mackey, 1999; Mackey et al., 2000). In our project, however, NSs were highly aware of the two essentially different constructs of L2 speech learning—improving comprehensibility (becoming easier to understand) versus reducing accentedness (becoming more nativelike)—and were explicitly asked to use recasts in order to help their NNS partners to achieve the former (but not necessarily the latter) goal of L2 speech learning.
- 2 According to some empirical research (e.g., Jiang, 2007), spending an extensive amount of time in decontextualized instruction (e.g., grammar-translation method) does not necessarily help the development of integrated knowledge or automatic performance in adult SLA, which is assumed to be tied to L2 learners' spontaneous speaking abilities. Thus, we speculate that our participants, who had learned English only through EFL instruction without many opportunities for conversational practice, could be at least considered as inexperienced learners.
- 3 In the other part of the project, the participants reversed roles (Japanese students as NSs, American students as NNSs) and followed a similar training procedure with regards to how the NSs helped the NNSs improve their oral ability via recasts in cases of communication breakdown. The details of this project are reported in another manuscript.
- 4 In our validation study (Saito, Trofimovich et al., 2015), the expert rater scores were found to be correlated with the actual phonological and temporal properties of L2 speech objectively measured via Praat (Boersma & Weenink, 2012): the number of vowel and consonant errors for segmentals ($r = .64$); the number of prosodic errors for word stress ($r = .72$) and intonation ($r = .54$); and the mean length of run and the number of unfilled pauses for speech rate ($r = .79, .49$, respectively).
- 5 According to Cohen (1988), effect sizes are roughly classified as small ($d < .30$), medium ($.30 \leq d < .80$), or large ($d \geq .80$).
- 6 A reviewer pointed out that the relative effectiveness of interaction on the acquisition of verb and article morphology could be attributed to the fact that there was considerable room for improvement with these features (error ratio of 7.5% and 7.6%, respectively), compared to noun morphology (2.6%) at the time of the pretests.

References

- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*, 249–306. doi:10.1111/j.1467-9922.2009.00507.x
- Aizawa, K., Ishikikawa, S., Murata, T., & Iso, T. (2005). *JACET 8000 Eitango* [JACET 8000 Word List]. Tokyo, Japan: Kirihara-shoten.
- Akiyama, Y. (in press). Learner beliefs and corrective feedback in telecollaboration: A longitudinal investigation. *System*.
- Bardovi-Harlig, K., & Comajoan, L. (2008). Order of acquisition and developmental readiness. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 383–397). Malden, MA: Blackwell.
- Belz, J. (2003). Linguistic perspectives on the development of intercultural competence in telecollaboration. *Language Learning & Technology, 7*, 68–117.
- Boersma, D., & Weenink, P. (2012). Praat: Doing phonetics by computer (Version 5.3.14) [Computer software]. Retrieved from <http://www.praat.org>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics, 36*, 570–590. doi:10.1093/applin/amt056
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 12*, 1–16.
- Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning, 63*, 163–185. doi:10.1111/lang.12000
- Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning, 64*, 526–548. doi:10.1111/lang.12053
- Develotte, C., Guichon, N., & Vincent, C. (2010). The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL, 22*, 293–312. doi:10.1017/S0958344010000170
- Doughty, C., & Pica, T. (1986). “Information gap” tasks: Do they facilitate second language acquisition? *TESOL Quarterly, 20*, 305–325. doi:10.2307/3586546
- Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition, 21*, 285–301.
- Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Clevedon, UK: Multilingual Matters.
- Gass, S. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Erlbaum.
- Gass, S., Mackey, A., Alvarez-Torres, M., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning, 49*, 549–580. doi:10.1111/0023-8333.00102

- Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning, 55*, 575–611. doi:10.1111/j.0023-8333.2005.00318.x
- Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition, 34*, 445–474. doi:10.1017/S0272263112000149
- Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition, 35*, 127–165. doi:10.1017/S0272263112000708
- Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29*, 311–343. doi:10.1177/0267658312461497
- Hopp, H., & Schmid, M. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics, 34*, 361–394. doi:10.1017/S0142716411000737
- Hulstijn, J. H., Schoonen, R., de Jong, N. H., Steinel, M. P., & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing, 29*, 203–221. doi:10.1177/0265532211419826
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly, 10*, 135–159. doi:10.1080/15434303.2013.769545
- Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language acquisition. *Language Learning, 57*, 1–33. doi:10.1111/j.1467-9922.2007.00397.x
- Koizumi, R., & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System, 40*, 554–564.
- Lee, L. (2002). Enhancing learners' communication skills through synchronous electronic interaction and task-based instruction. *Foreign Language Annals, 35*, 16–24. doi:10.1111/j.1944-9720.2002.tb01829.x
- Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition, 25*, 37–63. doi:10.1017/S0272263103000020
- Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *The Modern Language Journal, 90*, 536–556. doi:10.1111/j.1540-4781.2006.00465.x
- Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126–141. doi:10.1093/applin/4.2.126
- 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.

- Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19, 37–66.
- Lyster, R., & Saito, K. (2010). Corrective feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32, 265–302. doi:10.1017/S0272263109990520
- Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition*, 21, 557–587.
- Mackey, A. (2012). *Input, interaction, and corrective feedback in L2 learning*. Oxford, UK: Oxford University Press.
- 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, UK: Oxford University Press.
- Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses and red herrings? *The Modern Language Journal*, 82, 338–356. doi:10.1111/j.1540-4781.1998.tb01211.x
- Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471–497.
- MATLAB (Version 8.1) [Computer software]. (2013). Natick, MA: MathWorks.
- Monteiro, K. (2014). An experimental study of corrective feedback during video-conferencing. *Language Learning & Technology*, 18, 56–79.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46, 610–641. doi:10.1002/tesq.34
- Munro, M. J., & Derwing, T. M. (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. J., Derwing, T. M., & Burgess, C. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, 52, 626–637. doi:10.1016/j.specom.2010.02.013
- O'Dowd, R. (2011). Online foreign language interaction: Moving from the periphery to the core of foreign language education? *Language Teaching*, 44, 368–380. doi:10.1017/S0261444810000194
- Ortega, L. (2014). *Understanding second language acquisition*. New York: Routledge.
- Payne, J. S., & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal*, 20, 7–32. doi:10.1558/cj.v20i1.7-32
- Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, 11, 63–90. doi:10.1017/S027226310000783X

- Pinget, A. F., Bosker, H. R., Quené, H., & de Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing, 31*, 349–365. doi:10.1177/0265532214526177
- Piske, T., Flege, J., MacKay, I., & Meador, D. (2011). Investigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 195–205). Frankfurt am Main, Germany: Peter Lang.
- Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accents in an L2: A review. *Journal of Phonetics, 29*, 191–215. doi:10.1006/jpho.2001.0134
- Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning, 61*, 325–366. doi:10.1111/j.1467-9922.2011.00640.x
- Ranta, L., & Meckelborg, A. (2013). How much exposure to English do international graduate students really get? Measuring language use in a naturalistic setting. *Canadian Modern Language Review, 69*, 1–33. doi:10.3138/cmlr.987
- Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /r/ to adult Japanese learners. *Language Learning, 63*, 499–529. doi:10.1111/lang.12015
- Saito, K. (2014). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics, 24*, 250–277. doi:10.1111/ijal.12026
- Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning, 65*, 563–595. doi:10.1111/lang.12120
- Saito, K., & Lyster, R. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /r/ by Japanese learners of English. *Language Learning, 62*, 595–633. doi:10.1111/j.1467-9922.2011.00639.x
- Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*. Published online September 29, 2015. doi:10.1093/applin/amv047
- Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2015). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness and sense relations. *Studies in Second Language Acquisition*. Published online August 18, 2015. doi:10.1017/S0272263115000297
- Schmitt, N. (1998). Tracking the incremental acquisition of a second language vocabulary: A longitudinal study. *Language Learning, 48*, 281–317. doi:10.1111/1467-9922.00042
- 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, UK: Oxford University Press.
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263–308.
doi:10.1111/j.1467-9922.2010.00562.x
- Storch, N. (2002). Patterns of interaction in ESL pair work. *Language learning, 52*, 119–158. doi:10.1111/1467-9922.00179
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). Mahwah, NJ: Erlbaum.
- 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. doi:10.1017/S0272263106060013
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*, 905–916.
doi:10.1017/S1366728912000168
- Vercellotti, M. L. (2015). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistic*. Published online March 13, 2015. doi:10.1093/applin/amv002
- Wang, Y. (2006). Negotiation of meaning in desktop videoconferencing-supported distance language learning. *ReCALL, 18*, 122–145.
doi:10.1017/S0958344006000814
- Yao, Z., Saito, K., Trofimovich, P., & Isaacs, T. (2013). Z-Lab [Computer software]. Retrieved from <https://github.com/ZeshanYao/Z-Lab>
- Ziegler, N. (2015). Synchronous computer-mediated communication and interaction: A meta-analysis. *Studies in Second Language Acquisition*. Published online October 12, 2015. doi:10.1017/S027226311500025X

Supporting Information

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

Appendix S1. Error Correction Video Script.

Appendix S2. Training Materials and Onscreen Labels for Comprehensibility and Accentedness Measures.

Appendix S3. Training Materials and Onscreen Labels for Audio- and Transcript-Based Measures.