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

Kazuya Saito

Yuka Akiyama

Abstract

The current study examined the impact of video-based conversational interaction on the longitudinal development (one academic semester) of second language (L2) 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 US via telecommunication tools, wherein the native speaking interlocutors were trained to provide interactional feedback in the form of recasts when the non-native 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 video-coded data showed that the experimental students incidentally worked on improving all linguistic domains of language, thanks to their native speaking interlocutors' interactional feedback (recasts, negotiation) during the treatment. The pre-/post-test data led to significant gains in their comprehensibility, fluency and lexicogrammar, but not in the accentedness and pronunciation dimensions of their spontaneous production abilities. No significant improvement was found for the control group. The findings in turn support the acquisitional value of interaction, especially with regards to certain aspects of linguistic competence related to the early phase of L2 speech learning.

Key words: Interaction, Second language speech, Listening; Pronunciation; Fluency; Vocabulary; Grammar

Title:

Video-based Interaction, Negotiation for Comprehensibility, and Second Language Speech

Learning: A Longitudinal Study

Running Head:

ROLE OF INTERACTION IN L2 SPEECH LEARNING

Authors

Kazuya Saito (Birkbeck, University of London)

Yuka Akiyama (Georgetown University)

Corresponding Author:

Kazuya Saito

Birkbeck, University of London

The Department of Applied Linguistics and Communication

26 Russell Square, London WC1B 5DT, UK.

Email: k.saito@bbk.ac.uk

Acknowledgement

Abstract

The current study examined the impact of video-based conversational interaction on the longitudinal development (one academic semester) of second language (L2) 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 US via telecommunication tools, wherein the native speaking interlocutors were trained to provide interactional feedback in the form of recasts when the non-native 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 video-coded data showed that the experimental students incidentally worked on improving all linguistic domains of language, thanks to their native speaking interlocutors' interactional feedback (recasts, negotiation) during the treatment. The pre-/post-test data led to significant gains in their comprehensibility, fluency and lexicogrammar, but not in the accentedness and pronunciation dimensions of their spontaneous production abilities.

Key words: Interaction, Second language speech, Listening; Pronunciation; Fluency; Vocabulary; Grammar

Video-based Interaction, Negotiation for Comprehensibility, and Second Language Speech

Learning: A Longitudinal Study

Whereas adult second language acquisition (SLA) is a complex phenomenon whose

process and ultimate attainment is affected by a range of individual factors, most theoretical

accounts have at least agreed with the fundamental idea that L2 learners can improve their oral

ability through increased conversational experience with other native (NSs) and non-native

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.

The current paper reports on an experimental study which examined the longitudinal

development (i.e., 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 the field of 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. This is because it 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 via 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 (i.e., comprehensible input) (Long, 1983).

Whereas NSs aim to retrieve meaning from NNSs' incomprehensible speech by using several

negotiation strategies, such as repetition, confirmation checks, and clarification requests in the

case of communication breakdown, they may signal comprehended yet erroneous speech through

recasting of NNSs' erroneous forms (i.e., interactional feedback) (Lyster & Saito, 2010).

Through this, NNSs may be induced to notice and understand the gap between their own

interlanguage system and the incoming input, and are subsequently pushed to repair their

production (i.e., comprehensible output) (Swain, 2005). Consequently, the entire conversational

move via incidental 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

series of experimental studies with a pre/post-test design. Such a design allows researchers to

control various features of L2 interaction as independent variables and test their impact on L2

development (for a review, see Mackey, 2012; 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 the 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 vary according to various

affecting 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 towards 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 overall, L2 interaction is beneficial for acquisition with a medium effect size. Though revealing, the findings pointed out several methodological issues worthy of further investigation. First, their claims regarding the effectiveness of interaction were derived from only a brief amount of treatment in the individual studies; the length typically lasted less than one hour due to the paucity of longitudinal work in the field of SLA. However, it is important to mention that the effect size of L2 interaction became larger at the time of the delayed post-tests (e.g., 1 month after interaction) compared to immediate post-tests, suggesting that the benefit of interaction needs to be assessed via a long-term framework. In her review, Ortega (2009) claimed that "a long-term view on the benefits of interaction may be particularly important" as a future direction for interaction researchers.

Second, most L2 interaction studies included in the meta-analysis exclusively focused on the effectiveness of interaction on lexical and morphosyntactic development. According to several descriptive studies, negotiation for meaning episodes happen in all linguistic domains

(pronunciation, vocabulary, grammar), and NNS learners are in general aware of pronunciation-focused feedback compared to grammar-focused feedback (e.g., Mackey, Gass, & McDonough, 2000). However, few studies have ever 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, the current study took a first step towards 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 from a longitudinal perspective.

Developing L2 Oral Ability

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

For instance, Derwing and Munro (2013) conducted a longitudinal investigation on how late immigrants could improve their oral abilities during seven years of immersion in Canada. The results showed that motivated and regular L2 users in particular progressively enhanced their overall comprehensibility, but that their foreign accentedness demonstrated little change

over time. Furthermore, a growing amount 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 (relatively

related to comprehensibility) is characteristic of the early phase of L2 speech learning (LOR = 1-

3 years), the gradual refinement of pronunciation (strongly tied with accentedness) is concerned

with the long-term delineation of L2 speech learning (LOR = 5+ years). For further discussion

on the relationship between L2 experience and learning in naturalistic SLA, see Fkege (2009)

and Saito (2015).

Notably, the aforementioned studies have exclusively focused on immigrants and ESL

students in naturalistic (rather than controlled) settings, where interaction with native and non-

native interlocutors with various backgrounds in a range of social contexts is common. Although

these studies typically asked L2 learners to self-report how much they used L2 with native and

non-native speakers at work and home in a retrospective manner, it remains unclear how such

self-report measures can reliably reflect their actual conversation experience which is subject to a

huge 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).

Motivation for Current Study

From one perspective, 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 highlighted 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 alike 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 based on the ACTFL guidelines.

With respect to NS-NNS interaction, Akiyama (in press) descriptively documented how focus on form practices longitudinally occurred 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 native speakers 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, the current study was designed to examine the effect of L2 interaction on the longitudinal development (one academic semester) of Japanese L2 English learners' oral abilities with a pre- and post-test 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). Their interactional gains were scrutinized via a range of speech assessment instruments measuring global, phonological, temporal, lexical and grammatical dimensions of language.

It is important to emphasize that the NS interlocutors in this study were trained to provide interactional feedback in the form of recasts—the most preferred type of feedback in task-based telecollaboration, according to our precursor research (Akiyama, in press). During the training, we explicitly asked the NSs to provide feedback only when they perceived the NNS as incomprehensible (for details, see the subsection of recast training). Our intention here was to create interactional contexts which induced the Japanese learners to work *selectively* on certain linguistic errors directly relevant to comprehensibility (but not necessarily accentedness) in collaboration with their NS partners, while the primary focus of the entire interaction lay in meaning.[1] This decision reflects the widely-accepted view that comprehensibility and accentedness are two interrelated yet essentially different constructs, and that not all linguistic errors linked to accent hinder comprehensibility (e.g., Derwing & Munro, 1997).

The nature of this specific interaction—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 towards comprehensibility-related linguistic features can also be distinguishable 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," which is typical of teacher-student interactions (Lyster & Ranta, 1997, p. 42). In the latter contexts, feedback providers (teachers) *equally* attend to all linguistic errors in order to push receivers (students) to fill in 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 to what extent 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 the modeling of naturalistic L2 speech learning (Derwing & Munro, 2013; Saito, 2015), we formulated the following hypotheses. First, we predicted that 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), as they are found to be susceptible to quick changes according to study-abroad research (Mora & Valls-Ferrer, 2012). Second, it was also hypothesized that little improvement would appear in the attainment of enhanced accentedness and pronunciation. This is because adult L2 learners' successful acquisition of relatively difficult features 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).

Method

In total, 30 Japanese EFL students participated as NNS learners, and 15 American college students as NS interlocutors. Two kinds of flyers (one on conversational activities and the other on vocabulary/grammar activities) were prepared and distributed to Japanese EFL students. We divided the students interested in the former to the experimental group ($n = 15$), and those interested in the latter to the comparison group ($n = 15$). After they took the pre-tests 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 US by way of a

video-conferencing tool (Google Hangout). One week after the last session, the participants took

the post-tests, in Week 12. The timeline of the study is visually summarized in Figure 1.

FIGURE 1

Participants

NNS learners (Japanese EFL learners). A total of 30 freshman and sophomore

Japanese students majoring in business at a university in Japan participated as volunteers. At the

time of the project, all of them were registered in approximately three hours of EFL lessons per

week taught by Japanese instructors as a requirement for the school. According to the department

syllabus and our casual classroom observations, the content of the EFL classes mainly consisted

of a number of listening and reading activities without many opportunities to produce language.

The analysis of the language background questionnaire showed that all of the participants

had studied English in EFL classrooms for six years prior to the project, typically through

grammar-translation methods. They had little experience abroad except for family and school

trips, wherein they mainly used their L1 (Japanese) 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). A total of 15 NSs of English (9

males, 6 females; M_{age} = 21.2) learning Japanese at US universities participated in the study.

Some of them participated in this project as a one-credit course, while others were volunteer

exchange partners who expressed interest in practicing Japanese outside the classroom. Whereas

they had a wide variety of L2 Japanese learning experience (including intermediate to advanced

proficiency levels), they reported equally high familiarity with Japanese-accented English.

Experimental Group

The 15 Japanese NNS learners in the experimental group (10 males, 5 females)

participated in dyadic interaction with NS interlocutors in the U.S. via a video-conferencing tool

(i.e., *Google Hangout*) over one academic semester (9 sessions in total). For each session (60

min), the participants interacted with each other in English for the first 30 minutes, and switched

to Japanese for the second half. The details of the methodology used for the first half of the

interaction in English (Japanese students as NNSs; American students as NSs) are reported in the

current paper.[3] At the onset of the project (Week 2), both the NNS and NS participants received

explanation on the two-fold 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 how to promote 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. Our decision here corresponds to

Develotte, Guichon, and Vincent's (2010) concern that the potentially different amount of L2

learners' technological knowledge with video-conferencing tools and familiarity with video-

mediated conversations may influence its effectiveness. Due to the time difference between

Japan and the US, the participants were allowed to have the sessions with much flexibility

according to their schedules outside the classroom using their own computers. The NNSs were

required to report to the researcher the date/time of each session. All of the sessions were video-

recorded and stored as digital data for the purpose of analyses.

We decided to employ a type of information exchange task called *visual-based*

conversation, following the suggestion by Lee (2002), who found that two-way exchange of

information on real-life topics that are theme-based and minimally structured helped students

recycle ideas and reinforce language skills. Each week, NNSs were asked to find two visuals

(one for Japan and the other for the US) 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 may 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 may 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, 2015). 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

ROLE OF INTERACTION IN L2 SPEECH LEARNING

communicative flow of conversational moves (Long, 1996). Examples of recast episodes

retrieved from the current dataset are as follows:

Example 1: Pronunciation

      NNS: This is a picture of my **[kɒrɪdʒ]**.

      NS: Oh that's your **[kɒlɪdʒ]** (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 conveyance)

by drawing on recasts when their NS talkers' 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 essentially different constructs of L2 speech learning—
    improving comprehensibility (easier to understand) vs. reducing accentedness (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 on
    example recast episodes. For each episode, they discussed with the researcher on what
    kinds of errors the NNS talker made (i.e., trigger), and how the NS interlocutor helped
    the NNS retrieve their impaired comprehensibility (i.e., i.e., recasts, repair). Such recasts
    could occur as a 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 A).

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 onset (T1: the second session) and endpoint
(T2: the eighth session) of the project (30 min × 2 sessions × 15 dyads). Following the norm in

the previous research (e.g., Lyster & Ranta, 1997), the data was analyzed according to the three

crucial components of L2 interaction:

1. *Trigger* refers to the number of linguistic errors that NNS learners made in the domains

    of pronunciation (i.e., mispronunciation of segmentals and prosody), vocabulary (wrong

    word choice), and grammar (morphosyntactic errors).

2. *Feedback* refers to the number of recasts and negotiation strategies (confirmation checks,

    repetition, clarification requests) that NS interlocutors adopted in response to NNSs'

    errors.

3. *Uptake* refers to NNS learners' reaction to the feedback move, including repaired

    (successfully producing more targetlike production), needs repair (failing in self-

    correcting errors despite any sort of attempt), and no uptake (showing no reaction to NS

    interlocutors' feedback).

Based on a total of 15 hours of the coded data at two different time points (T1, T2), we

aimed to provide suggestive patterns regarding (a) how frequently the NNS learners in our study

received feedback according to different linguistic categories (pronunciation, vocabulary,

grammar); and (b) to what degree they successfully noticed/repaired their phonological, lexical

and grammatical errors.

Comparison Group

The remaining 15 Japanese NNS learners in the comparison group (6 males, 9 females)

were weekly involved with individual vocabulary/grammar exercise activities instead of task-

based interaction activities with NSs. The reasons for including the comparison group in the

study were two-fold. First, because we used identical materials in the pre/post-test sessions (see below), the analysis of the comparison group's performance allowed for an examination of any test-retest effect. 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 education (a few hours of Language and Arts lessons per week) without any opportunities for interaction. In so doing, we aimed to separate the patterns of the comparison group (i.e., effects of test-retest and one semester EFL education) from the L2 interaction gains in the experimental group, if any.

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 the part 5 in TOEIC (i.e., production practice). As piloted prior to the project, the weekly assignment typically took 30 minutes to complete at home. Their weekly assignment submissions were graded and recorded by the researcher.

Production Test

Traditionally, L2 speech has been measured via highly controlled production tasks, such as reading aloud written and audio prompts, so that researchers can analyze the linguistic structures of interest in a consistent and regulated manner (for 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 current 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 conveyance) under time pressure conditions (Spada & Tomita, 2010).

Materials. In the current study, a timed picture description was adopted to measure the participants' pronunciation, fluency, vocabulary and grammar performance during spontaneous speech. In our precursor research (e.g., Saito, 2015), the task was designed to allow L2 learners with various proficiency levels (including even low beginners) to contribute certain lengths of spontaneous speech without much dysfluency (filled and unfilled pauses, repetitions). The test could be also considered suitable especially for this specific project, as the nature of the task simulated what the participants actually did during dyadic conversation exchanges (i.e., picture descriptions followed by discussion).

Building on the picture narrative task used in the previous L2 speech literature (e.g., Munro & Mann, 2005), our participants were asked to describe seven individual pictures with only 5 sec of planning time per photo. 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 (for a review, Saito, 2014). For example, Japanese learners tend to neutralize the English /r/-/l/ contrast ("rain, rock, brew, crowd" vs.

"lane, lock, blue, cloud") and use Japanese borrowed words by inserting epenthetic vowels between consecutive consonants (/dəraɪvə/ for "drive," /θəri/ for "three," /səkaɪ/ for "sky") and after word-final consonants (/teɪbələ/ for "table," /myuzɪkə/ for "music," /roʊdə/ for "road"). The three pictures featured: (a) a table left on the drive way in the heavy rain" (key words: *rain, table, drive way*); (b) three guys who were playing rock music with one guy singing a song and the other two guys 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 sec of each picture description was 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 sec ($M = 10.3$ sec). The three picture descriptions were combined and stored as a single WAV file for each talker at the pre- and post-test sessions, resulting in 60 speech samples (30 NNSs × pre-/post-tests). The mean length of each file was 30.4 sec (ranging from 27.2 to 35.1), which can be considered as sufficiently long to provide the linguistic information required for L2 speech analyses, in line with previous research standards (e.g., Derwing & Munro, 1997 for 10-15 sec; Hopp & Schmid, 2013 for 10-20 sec; Trofimovich & Isaacs, 2012 for 30 sec). The mean number of words for each file was 40.3, ranging from 25-62 words.

Procedure. At Weeks 1 and 12, their picture descriptions were recorded individually in a quiet room at the university via 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

ROLE OF INTERACTION IN L2 SPEECH LEARNING

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

Novice raters. Five native speaking raters (2 males, 3 females) were recruited at an English speaking 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 native speaker 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 (they had never taken any linguistics courses nor taught English in ESL/EFL settings) and their low familiarity with Japanese-accented English (their mean self-report scores was 2.2, ranging from 1 to 3 on a 6-point scale: *6 = Very much, 1 = Not at all*). None of them 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 a custom software, Z-Lab (Yao, Saito, Trofimovich & Isaacs, 2013), developed using the commercial software package MATLAB 8.1 (The MathWorks Inc., Natick, MA, 2013). Then, they used a free-moving slider on a computer screen based on a 1000-point scale to evaluate comprehensibility (*0 = hard to understand, 1000 = easy to understand*) and accentedness (*0 = heavily accented, 1000 = little accent*), with the leftmost corner labeled with a frowning face and

the rightmost corner with a smiling face. To ensure that their ratings captured their intuitions as native speakers, they were allowed to play each sample only once. The raters first practiced the procedure with three trial samples (not included in the main dataset), and then proceeded with the assessment of all 60 samples. The entire session lasted for 1 hour with a 5-minute break halfway through. For training scripts and onscreen labels, see Appendix B.

Rater consistency. According to Cronbach's alpha, the inter-rater agreement was relatively high among the five novice raters for comprehensibility (α = .91) and accentedness (α = .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 traditions in L2 speech research, we used linguistically-trained judges' impressionistic analyses 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 native speaking raters (2 males, 3 females) were recruited at an English speaking 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 ($M_{years\ of\ teaching}$ = 5.7 years ranging from 3 to 13.5 years). They reported relatively high familiarity with Japanese-accented English (M = 4.8 ranging from 4 to 6: $6 = Very\ much,\ 1 = Not\ at\ all$).

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 the same moving slider 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

wanted to until they felt satisfied with their judgements.

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 dataset). 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 judgements. Finally, they proceeded with the main rating sessions, which

lasted for approximately 2 hours (with a 10-min intermission halfway). For onscreen labels and

training scripts, see Appendix C.

Rater consistency. Given the relatively high Cronbach's alpha among the five expert

raters for segmentals ($\alpha = .90$), word stress ($\alpha = .85$), intonation ($\alpha = .81$), and speech rate ($\alpha = .91$), their audio judgement scores were considered sufficiently consistent, and then averaged

to derive a single mean score per talker.

Vocabulary and Grammar Analyses

All speech samples (each of which included one talkers' 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, ehh). As is the case with the previous L2

vocabulary and grammar literature (e.g., Saito, Webb, Trofimovich, & Isaacs, 2015; Crossley,

Salsbury , & McNamara, 2015), the transcripts were submitted to the following lexicogrammar

analyses conducted by a linguistically-trained coder: (a) lexical appropriateness (i.e., the ratio of

non-nativelike word choice), (b) lexical variation (i.e., the number of different words per

minute), and (c) morphosyntactic accuracy in verbs (i.e., the ratio of tense, aspect, modality, and

subject-verb agreement errors), nouns (the ratio of plural usage errors related to countable and

uncountable nouns), and articles (the error ratio of article usage in terms of definite, indefinite

and null).

Results

In this section, we first descriptively analyzed the experimental students' interactional

behaviours (the number of feedback moves and uptake individual talkers processed in each

dyad), and then examined the participants' global, phonological, temporal, lexical and

grammatical gain scores in relation to those of the comparison group.

Details of Interaction Moves

The first aim of the analysis was to estimate how negotiation for comprehensibility

generally took place among the experimental group which was explicitly encouraged to work on

errors relevant with successful communication (but not necessarily with linguistic

nativelikeness). 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 onset (second session: T1) and endpoint (eighth session: T2) of the project.

TABLE 1

Overall patterns. The descriptive results of the interactional features of the 15 dyads at

T1 and T2 are summarized in Table 1. NS interlocutors only occasionally provided interactional

feedback on NNSs' errors (targeting only 22.9% [T1], 13.4% [T2] of errors). It is possible that,

consistent with the instructions we gave them, the NSs generally prioritized communicative flow over the correction of linguistic errors, except when errors were perceived 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. Interestingly, 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 a feedback move, the NNS learners demonstrated a good amount of effort to modify their output over time. Indeed, the talkers showed relatively high uptake (both repair and needs repair) towards approximately half of these feedback moves (36.7%, 64.7% for recasts; 48.5%, 69.6% for negotiation), indicating that the NNS learners looked to the NS interlocutors' feedback in order to find which linguistic errors they needed to work on to be successfully understood.

TABLE 2

Linguistic dimensions. Next, we investigated the quality of interactional treatment according to different linguistic (pronunciation, vocabulary, grammar) dimensions. As summarized in Table 2, although the NNS learners made a relatively large number of pronunciation (56.3%, 51.8%) and grammar (36.9%, 41.0%) errors, the NS interlocutors directed a relatively small amount of feedback in response to these errors (18.1%, 10.1% for pronunciation; 23.1%, 12.5% for grammar Although the NNSs made far fewer vocabulary errors (6.0%, 7.2%), the NS interlocutors seemed to focus more than half of their feedback on lexical issues (56.3%, 51.1%).

At the beginning of the project, the NNS 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%) vs. vocabulary (44.4%) vs. grammar (36.5%). In the end, the NNS learners showed

relatively high sensitivity (uptake) to pronunciation (76.6%) and vocabulary (65.2%) feedback

compared to grammar feedback (54.0%).

To sum up, the descriptive results indicated that (a) the raw number of NNS errors

(pronunciation and grammar > vocabulary) and the NS feedback ratio patterns (vocabulary >

pronunciation and grammar) remained static throughout the project; (b) the NNSs became

increasingly sensitized to, in particular, pronunciation and vocabulary feedback (compared to

grammar feedback) around the end of the project.

Effects of Interaction on the Development of L2 Oral Ability

The second aim of the statistical analyses was to examine the extent to which the

participating students in the experimental group, who were engaged with L2 interaction over one

academic semester, improved their L2 production skills. Their performance was compared to that

of the comparison group, who only practiced vocabulary and grammar exercise activities.

To check for any pre-existing differences in the experimental and comparison groups'

oral ability, their performance at the time of the pre-test sessions was analyzed according to the

rated domains of the production test (comprehensibility, accentedness, pronunciation, fluency,

lexicogrammar). Since 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 $p < .05$

and adjusted to $p < .025$ via Bonferroni correction.

According to the results, prior to the project, the two groups were comparable for their

global production ($z = -1.43$, $p = .161$ for comprehensibility; $z = -.0.04$, $p = .967$ for

accentedness), pronunciation ($z = -0.24$, $p = .806$ for segmentals; $z = -0.83$, $p = .412$ for word

stress; z = -1.30, p = .191 for intonation), fluency (z = -0.85, p = .412 for speech rate), vocabulary

(z = -1.49, p = .135 for appropriateness; z = -1.22, p = .233 for variation), and grammar domains

(z = -.166, p = .870 for verb error ratio; z = -.293, p = .806 for noun error ratio, z = -.687, p

= .512 for article error ratio).

To examine the presence/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 pre- and post-tests

were used as a dependent variable with an alpha level set at $p < .025$ level (Bonferroni

corrected). The magnitude of their improvement over time (pre → post) was measured by

Cohen's d analysis.[5] The descriptive and inferential statistics of the pre- and post-test production

scores are summarized in Table 3.

TABLE 3

The experimental group significantly improved their overall comprehensibility (but not

accentedness) scores (p = .012) with medium effects (d = 0.58). In terms of specific domains of

L2 speech, the experimental group significantly enhanced their perceived fluency (speech rate),

vocabulary variation (the number of different words per minute), and the accurate use of

grammar (verb and article error ratio) with large effects (d = 1.17, 0.82, 0.90, and 1.00,

respectively). In contrast, the comparison group did not note any significant improvement in any

contexts ($p > .025$).

Discussion

Given the lack of longitudinal evidence on the acquisitional value of negotiated

interaction in adult SLA, the present study took an exploratory approach by investigating the

effect of weekly video-based interaction sessions on the development of Japanese EFL learners'

ROLE OF INTERACTION IN L2 SPEECH LEARNING

oral abilities. Our 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 scrutinized by way of global (comprehensibility, accentedness) as well as

linguistic (pronunciation, fluency, vocabulary, grammar) production analyses.

Nature of Interactional Treatment

With respect to the content of the interactional treatment, we performed a detailed

examination of the video-coded data, which brought to light how interactional feedabck 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 which negatively affected message conveyance (10.1-23.1% of

pronunciation and grammar errors; 51.1-56.3% of vocabulary errors). The descriptive results

presented here concurred with what we explicitly asked the NS interlocutors to do during the

recast training: Selectively attend to those linguistic forms with particularly high communicative

value while ignoring other linguistic errors which may not have had an immediate influence on

the flow of the ongoing communication. At best, the data suggest 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- (but not form-) oriented

interaction.

Overall, the descriptive results indicated that the nature of the interactional treatment in

the current study could be considered pertinent to the fundamental idea of L2 conversation as

promoting opportunities to negotiate for meaning—not only can NNS 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 NNS learners in the current 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 other L2 interaction research which has shown some L2 learners' high noticing/awareness of pronunciation-focused feedback relative to grammar-focused feedback (Mackey et al., 2000), the NNS learners in the study were also engaged in modifying pronunciation and vocabulary errors with a higher uptake ratio (76.6%, 65.2%) than grammar errors (54.0%).

Effects of Interaction

To answer our research question (the acquisitional value of the semester-long interaction), significant improvement was found in the experimental group (participating in interaction sessions with NSs), but not the comparison group (involving grammar- and vocabulary-focused exercise). 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 native and non-native speakers (e.g., Flege, 2009). At the same time, our results regarding 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 oral ability at a spontaneous level (Spada & Tomita, 2010).

Importantly, the results of our comprehensive L2 speech measures further revealed that the extensive interaction activities were relatively 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, the findings shed some light on the complex

mechanisms underlying the acquisitional value of interaction. It appears that NNSs are driven to

enhance those linguistic dimensions of speech which are related to native speakers'

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 et al., 2015). In contrast, such communicatively-oriented L2

learning may not clearly relate to other areas of language strongly tied with native speakers'

nativelikeness judgements, 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 the current study (i.e., one academic semester).

Combining the analyses of the video-coded data of the treatment as well as the pre/post-

tests allowed us to derive tentative conclusions regarding the potential and limits of L2

interaction from a longitudinal perspective. With respect to the potential of the treatment, NNS

learners can communicatively work on improving pronunciation, fluency, vocabulary and

grammar with primary attention to maintaining successful communication with NS interlocutors.

Consequently, NNS learners can improve the overall comprehensibility of their L2 speech to

improve lexical variation and morphosyntactic accuracy to 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 significant impacts on

nativelikeness-related features entailing a great deal of learning difficulty (foreign accentedness,

prosody, segmentals).

In comparison with the longitudinal developmental patterns of L2 adults immersed in naturalistic settings, the differential effects of interaction on determining comprehensibility versus accentedness indicates the 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 NNS learners did successfully repair their mispronounced and unclear utterances following NS interlocutors' feedback during the interactional treatment (repair ratio = 10-20%), such interaction-based learning did not appear to ultimately change the present state of their L2 phonetic knowledge and performance, which was, in theory, measured via the pre-/post-test materials. In fact, previous L2 speech research has provided extensive evidence that L2 pronunciation learning is a slow, gradual, and extended process which 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 L2 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 suprasegmental-based instruction (consisting of explicit instruction as well as focused practice activities on L2 prosody and fluency) led even experienced ESL learners (with approximately 20 years of LOR) to improve specific pronunciation features as well as overall comprehensibility. Similarly, Saito (Saito, 2013; Saito & Lyster, 2012) examined the acquisitional value of teachers' provision of recasts to Japanese learners' mispronunciation of English /r/ in the context of meaning-oriented

classrooms. They found that a communicative focus on form could lead to gains not only at a controlled but also at a spontaneous speech level. Further research is needed which examines how to integrate such an intentional approach into communicatively-oriented NS-NNS interaction through focused tasks designed to elicit learners' use of specific phonological features, and/or through feedback techniques targeting certain phonological errors.

Limitations

Due to the exploratory nature of the project, several topics worthy of further investigation need to be addressed for future L2 speech studies of this kind. First of all, the findings were based on a relatively small sample size ($N = 30$). Future work should replicate the current study with a larger number of L2 learners with different L1 backgrounds. In particular, the multifaceted effect of L2 interaction should be further analyzed via various production measures, test materials, and analysis methods.

In the current dataset, although the experimental group significantly enhanced most of their 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). The findings here contradict those of the previous studies that found the lexicogrammatical accuracy to be important for native speakers' speaking proficiency judgements (Crossley et al., 2015) and subject to immediate improvement within a short amount of immersion (Schmitt, 1997). Whereas the differential amount of interaction benefits was found in the study according to different grammatical morphemes (verbs, articles > nouns), the 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]

According to methodological standards in L2 vocabulary research, however, what length of speech sample is required for robust lexical analyses has still remained controversial and open to further validation. Whereas spontaneous speech samples in our study could be considered as relatively short ($M = 40.5$ words), the length of speech samples has been up to two to three minutes in previous studies (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 intriguing to further examine 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).

Next, it is important to remember that our discussion related to the potential/limitations of L2 interaction was exclusively limited to a specific context of L2 learners—inexperienced Japanese EFL students in the early phase of L2 speech learning. Thus, it would be intriguing to examine whether, to what degree and how L2 interaction can help L2 learners with various proficiency levels (including not only beginner but also advanced learners) *continue* to improve their production ability with not only communicatively salient features (e.g., fluency, lexicogrammar), but also those linguistic structures without much communicative value or learnability (e.g., segmentals, prosody), especially beyond one academic semester.

Third, the current dataset 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 re-examine this topic while controlling for a number of common traits of interaction found in the

previous research, such as different types of interactional feedback (Lyster & Saito, 2010), task structures (Gass, Mackey, Alvarez-Torres, & Fernandez-Garcia, 2009), interlocutors (Storch, 2002), and the presence/absence of opportunities for repair (Leeman, 2003).

Last, we would like to emphasize the importance of conducting not only quantitative, but also qualitative analyses on 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 over longer periods of time, as in the current study. Such 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 through incidental focus on form (without any explicit instruction) (Lyster & Saito, 2010).

Conclusion

In the context of inexperienced EFL speakers, the current study examined the impact of negotiation for comprehensibility via video-based interaction on the longitudinal development of L2 oral production. Results showed that the experimental group significantly improved the global, temporal and grammatical qualities of their L2 speech. On the one hand, the findings support the acquisitional value of interaction, especially with regard to those features of L2 speech competence directly linked with 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 have much learning difficulty (e.g., pronunciation) and are highly relevant to perceived nativelikeness. In conclusion, task-based

interaction can greatly help inexperienced learners become successful communicators in the

early phases of L2 speech learning because it results in a communicative focus on linguistic form

that is crucial for meaning conveyance. At the same time, however, it may require an extensive

amount of L2 interaction (> one academic semester) and/or intentional focus on form (e.g.,

explicit instruction) for learners to attend to and practice non-salient and difficult features as a

way to refine the linguistic nativelikeness of their speech and attain more advanced L2 oral

ability.

References

Aizawa, K., Ishikikawa, S., Murata, T., & Iso, T. (2005). JACET 8000 Eitango (JACET 8000 Word List). *Tokyo: Kirihara-shoten*.

Abrahamsson, N. & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in a second language – listener perception *vs.* linguistic scrutiny. *Language Learning, 59*, 249–306. doi: 10.1111/j.1467-9922.2009.00507.x

Belz, J. (2003). Linguistic perspectives on the development of intercultural competence in telecollaboration. *Language Learning & Technology, 7,* 68–117. Retrieved from http://llt.msu.edu/vol7num2/belz/default.html

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

Bardovi-Harlig, K., & Comajoan, L. (2008). Order of acquisition and developmental readiness. In B. Spolsky and F. M. Hult (Eds.). The Handbook of Educational Linguistics (pp. 383-397. Malden, MA: Blackwell.

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

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

Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility. *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

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., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language learning, 51*(2), 281-318. doi: 10.1111/1467-9922.00156

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.

Ellis, R., & Sheen, Y. (2006). Re-examining the role of recasts in L2 acquisition. *Studies in Second language Acquisition, 28*, 575-600. doi: 10.1017/S027226310606027X

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

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

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

Gass, S. M., Mackey, A., Alvarez-Torres, M., & Fernandez-Garcia, M. (1999). The effects of task repetition on linguistic output. *Language Learning, 49*, 549-580. doi: 10.1111/0023-8333.00102

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2

    learning. *Studies in Second Language Acquisition*, *34*, 445-474. doi:

    http://dx.doi.org/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

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: http://dx.doi.org/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: http://dx.doi.org/10.1017/S0272263103000020

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.

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

Lyster, R., & Saito, K. (2010a). 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., 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., & Philps, J. (1998). Conversational interaction and second language development:

   Recasts, responses and red herrings? *Modern Language Journal, 82,* 338-356.

   doi: 10.1111/j.1540-4781.1998.tb01211.x

Monteiro, K. (2014). An experimental study of corrective feedback during video-conferencing.

   Language *Learning & Technology, 18*(3), 56–79. Retrieved from

   http://llt.msu.edu/issues/october2014/monteiro.pdf

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

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of

   foreign language learners. *Applied Linguistics*. doi: 10.1093/applin/amu024

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. 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

Munro, M., & Mann, V. (2005). Age of Immersion as a predictor of foreign accent. *Applied

   Psycholinguistics, 26,* 311-341. doi: http://dx.doi.org/10.1017/S0142716405050198

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.

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:

   http://dx.doi.org/10.1017/S0261444810000194

Payne, J. S., & Whitney, P. J. (2002). Developing L2 oral proficiency through synchronous

CMC: Output, working memory, and interlanguage development. *CALICO Journal*,

20(1), 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., 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

Piske, T., Flege, J., MacKay, & Meador, D. (2011). Investigating native and non-

native vowels produced in conversational speech. In M. Wrembel, M. Kul &

Dziubalska-Kołaczyk, K. (Eds.), *Achievements and perspectives in the acquisition of*

second language speech: New Sounds 2010 (pp. 195-205). Switzerland: Peter Lang.

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. *The Canadian*

Modern Language Review, 69, 1-33. doi: http://dx.doi.org/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.

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*. 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*. 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:

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:

Lawrence 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: http://dx.doi.org/10.1017/S0272263106060013

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility.

Bilingualism: Language and Cognition, *15*, 905-916. doi:

http://dx.doi.org/10.1017/S1366728912000168

Vercellotti, M. L. (2015). The development of complexity, accuracy, and fluency in second

language performance: A longitudinal study. *Applied Linguistic.* doi:

10.1093/applin/amv002

Wang, Y. (2006). Negotiation of meaning in desktop videoconferencing-supported distance

language learning. *ReCALL, 18*, 122-145. doi:

http://dx.doi.org/10.1017/S0958344006000814

Yao, Z., Saito, K., Trofimovich, P., & Isaacs, T. (2013). *Z-Lab.* Retrieved August 15, 2013, from

    https://github.com/ZeshanYao/Z-Lab

Ziegler, N. (2015). Synchronous computer-mediated communication and interaction: A meta-

    analysis. *Studies in Second Language Acquisition*. Advance online publication. doi:

    10.1017/S027226311500025X

Endnotes

1. Our decision to train NSs to provide recasts on certain salient (but not all) linguistic errors during task-based interaction corresponds to the L2 interaction research standard which 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 (easier to understand) vs. reducing accentedness (more nativelike)—and were explicitly asked to use recasts in order to help their NNS partners to achieve the former (but not the latter) goal of L2 speech learning.

2. According to some empirical research (e.g., Jian, 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 competence in adult SLA, which is assumed to be tied with L2 learners' spontaneous speaking abilities. Thus, we speculate that our participants, who had learned English only through EFL education without many opportunities for conversational practice, could be at least considered as "inexperienced" learners.

3. In the other part of the project, their roles reversed (Japanese students as NSs; American students as NNSs) and they followed a similar training procedure with regards to how the NSs helped the NNSs improve their oral ability via recasts in the case of communication breakdowns. The details of the results will be reported in another paper.

4. In our validation study (Saito et al., 2015), the expert rating 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 in the field of instructed SLA are roughly classified as small ($d < 0.30$), medium ($0.30 \leq d < 0.80$), or large ($0.80 \leq d$).

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: 7.5%, 7.6%) compared to noun morphology (2.6%) at the time of the pre-tests.

Table 1

Overall Interaction Patterns of Trigger, Feedback and Uptake

Errors			Feedback			Uptake		
	T1	T2		T1	T2		T1	T2
			No feedback	*n* = 512 (77.1%)	*n* = 688 (86.5%)			
All linguistic domains	*n* = 664	*n* = 795	Recasts	*n* = 117 (17.6%)	*n* = 74 (9.3%)	Repair	*n* = 21 (17.9%)	*n* = 21 (28.3%)
						Needs repair	*n* = 22 (18.8%)	*n* = 27 (36.4%)
						No uptake	*n* = 74 (63.2%)	*n* = 26 (35.1%)
			Negotiation	*n* = 35 (5.3%)	*n* = 33 (4.1%)	Repair	*n* = 3 (8.5%)	*n* = 6 (18.1%)
						Needs repair	*n* = 14 (40.0%)	*n* = 17 (51.5%)
						No uptake	*n* = 18 (51.4%)	*n* = 10 (30.3%)

Table 2
Pronunciation-, Vocabulary- and Grammar-Related Patterns of Trigger, Feedback and Uptake

Errors			Feedback			Uptake		
	T1	T2		T1	T2		T1	T2
			No feedback	*n* = 282 (82.0%)	*n* = 406 (89.6%)			
Pronunciation	*n* = 344 (56.3%)	*n* = 453 (51.8%)	Recasts	*n* = 47 (13.7%)	*n* = 31 (6.8%)	Repair	*n* = 6 (12.8%)	*n* = 8 (25.8%)
						Needs repair	*n* = 12 (25.5%)	*n* = 15 (48.4%)
						No uptake	*n* = 29 (61.7%)	*n* = 8 (25.8%)
			Negotiation	*n* = 15 (4.4%)	*n* = 16 (3.5%)	Repair	*n* = 1 (6.2%)	*n* = 3 (18.8%)
						Needs repair	*n* = 6 (37.5%)	*n* = 10 (62.5%)
						No uptake	*n* = 8 (50.0%)	*n* = 3 (18.8%)
	T1	T2		T1	T2		T1	T2
			No feedback	*n* = 21 (43.8%)	*n* = 22 (48.9%)			
Vocabulary	*n* = 48 (6.0%)	*n* = 45 (7.2%)	Recasts	*n* = 18 (37.5%)	*n* = 19 (42.2%)	Repair	*n* = 8 (44.4%)	*n* = 6 (31.6%)
						Needs repair	*n* = 0 (0%)	*n* = 7 (36.8%)
						No uptake	*n* = 10 (55.6%)	*n* = 6 (31.6%)
			Negotiation	*n* = 9 (18.8%)	*n* = 4 (8.9%)	Repair	*n* = 2 (22.2%)	*n* = 2 (50.0%)
						Needs repair	*n* = 2 (22.2%)	*n* = 0 (0%)
						No uptake	*n* = 5 (55.5%)	*n* = 2 (50.0%)

ROLE OF INTERACTION IN L2 SPEECH LEARNING

	T1	T2		T1	T2		T1	T2
			No feedback	n = 209 (76.8%)	n = 260 (87.5%)			
						Repair	n = 7 (13.5%)	n = 7 (29.2%)
			Recasts	n = 52 (19.1%)	n = 24 (8.1%)	Needs repair	n = 10 (19.2%)	n = 5 (20.8%)
Grammar	n = 272 (36.9%)	n = 297 (41.0%)				No uptake	n = 35 (67.3%)	n = 12 (50.0%)
						Repair	n = 0 (0%)	n = 1 (7.7%)
			Negotiation	n = 11 (4.0%)	n = 13 (4.4%)	Needs repair	n = 6 (54.6%)	n = 7 (53.9%)
						No uptake	n = 5 (45.5%)	n = 5 (38.5%)

ROLE OF INTERACTION IN L2 SPEECH LEARNING

Table 3
The Results of Descriptive and Inferential Statistics of the production Test Scores over Time

			Pre-test (1000 points)		Post-test (1000 points)		Improvement (pre →post)		
			M	*SD*	*M*	*SD*	*z*	*p*	*d*
Global impression	Comprehensibility	Experimental	315	165	420	191	2.49	.012*	0.58
		Control	316	140	375	126	0.48	.629	0.44
	Accentedness	Experimental	292	161	320	175	1.07	.281	0.16
		Control	284	136	266	101	0.45	.650	0.15
Pronunciation	Segmentals	Experimental	343	149	365	147	1.07	.281	0.14
		Control	329	114	367	100	1.93	.053	0.35
	Word stress	Experimental	416	118	452	128	1.70	.088	0.29
		Control	371	99	386	123	0.51	.609	0.13
	Intonation	Experimental	361	146	418	168	2.01	.044	0.36
		Control	307	113	347	122	2.10	.036	0.34
Fluency	Speech rate	Experimental	329	157	514	159	3.40	.001*	1.17
		Control	378	167	431	137	1.591	.112	0.34
Vocabulary	Appropriateness	Experimental	2.8%	2.6%	1.1%	1.6%	-2.00	.045	0.78
		Control	4.9%	3.8%	4.8%	4.1%	-0.18	.851	0.02
	Variation	Experimental	38.4	11.9	48.6	12.9	-3.23	.001*	0.82
		Control	43.4	6.7	46.0	17.1	-0.34	.733	0.20
Grammar	Verb error ratio	Experimental	7.6%	3.7%	4.4%	3.0%	-2.55	.011*	0.90
		Control	8.5%	4.9%	8.0%	3.2%	-0.11	.909	0.32
	Noun error ratio	Experimental	2.6%	3.2%	3.4%	3.5%	-0.86	.386	0.23
		Control	3.1%	3.8%	4.6%	4.0%	-0.73	.462	0.38
	Article error ratio	Experimental	7.5%	5.8%	2.4%	4.2%	-2.76	.006*	1.00
		Control	6.7%	5.2%	6.9%	5.0%	-0.62	.532	0.03

Note. * stands for a statistically significant improvement at a *p* < .025 level.
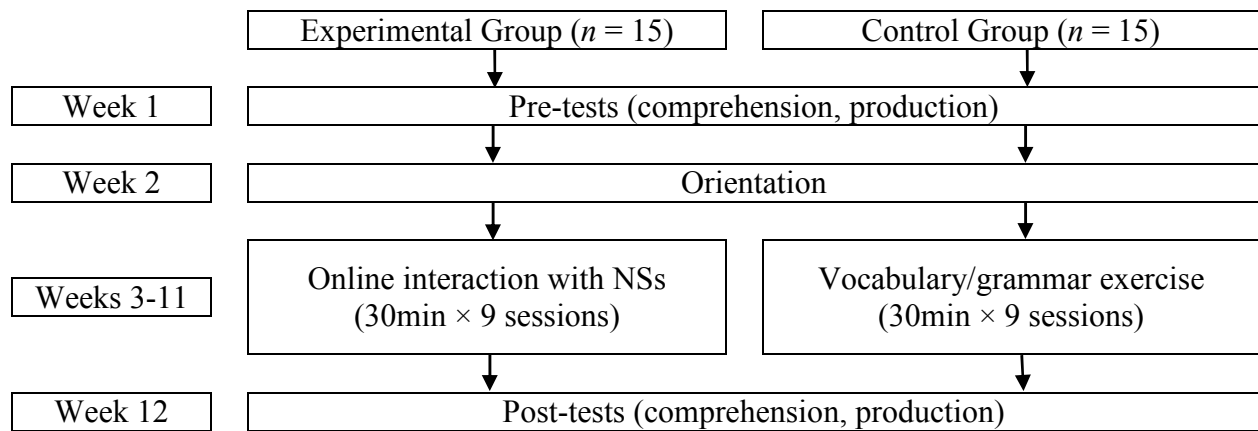
Figure 1. Time framework of the current project

Appendix A: Error correction video script

Episode 1: Pronunciation
NS: How's the weather in Japan?
NNS: It's very sunny. I always use a **PARASOL** (pronounced with an equal emphasis
         on each syllable)
NS: What's that?
NNS: I use an umbrella.
NS: Oh, **PArasol**!
NNS: Yes, parasol. It protects you from the Sun.

Episode 2: Pronunciation
NS: It's very hot today… How is it in Tokyo?
NNS: Well, it is very **[mugi].**
NS: [mugi]? What do you mean?
NNS: It's like it's very wet. You know? The air is wet.
NS: Ah, it's muggy [mʌgi].
NNS: Oh sorry. Yes, it's muggy [mʌgi].

Episode 3: Vocabulary
NS: What's your plan for the weekend?
NNS: Oh, I am going to climb Mt. Fuji for the first time. So, I am I am **very exciting**!
NS: Yeah, you're very **excited**. I bet. Who are you going with?
NNS: I am **single**.
NS: Sorry?
NNS: I am one.
NS: Oh, you're **on your own**?
NNS: Yes, it's my adventure.

Episode 4: Vocabulary
NNS: What should I say to my professor when I want to **make a reservation**?
NS: What do you mean?
NNS: I mean, I want to see her in her office hours.
NS: Oh, so you want to **make an appointment**.
NNS: Yes, appointment.

Episode 5: Grammar
NS: Do you play sports?
NNS: My father **like** sports but I don't.
NS: Oh your father **likes** sports.
NNS: Yes. I **used playing** basketball. But now I don't.
NS: You **used to play** basketball.
NNS: Yes, I used to play basketball.

Episode 6: Grammar
NNS: How was your break?
NS: It was great. I spent a month in Italy with my relatives.

NNS: Nice. I **have gone** to Italy two times.
NS: What do you mean?
NNS: I went to Italy last year and as child.
NS: Oh, you **have been** to Italy two times.
NNS: Yes, I have been to Italy.

Appendix B: Training materials and onscreen labels for comprehensibility and accentedness measures

Comprehensibility	This term refers to how much effort it takes to understand what someone is saying. If you can understand with ease, then a speaker is highly comprehensible. However, if you struggle and must listen very carefully, or in fact cannot understand what is being said at all, then a speaker has low comprehensibility.
Accentedness	This refers to how much a speaker's speech is influenced by his/her native language and/or is colored by other non-native features.

1. Comprehensibility

Difficult to understand		Easy to understand

2. Accentedness

Heavily accented		Not accented at all
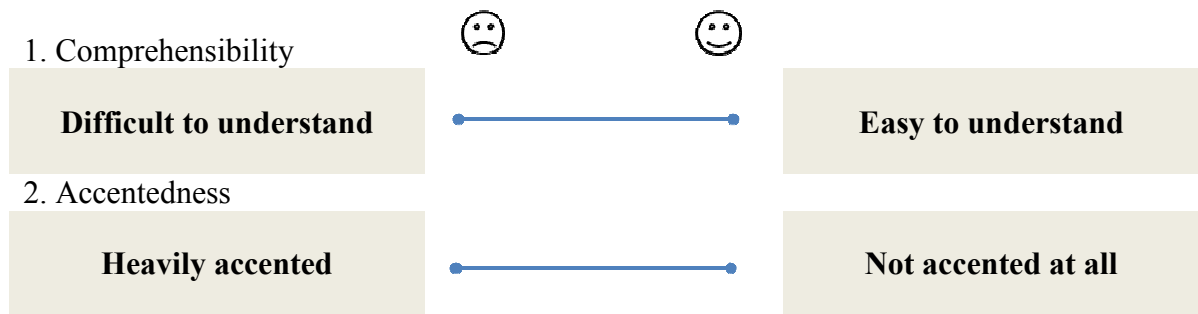
Appendix C: Training materials and onscreen labels for audio- and transcript-based measures

A. Pronunciation and fluency categories	
Segmental errors	This refers to errors in individual sounds. For example, perhaps somebody says "*road*" "*rain*" but you hear an "l" sound instead of an "r" sound. This would be a consonant error. If you hear someone say "*fan*" "*boat*" but you hear "*fun*" "*bought*," that is a vowel error. You may also hear sounds missing from words, or extra sounds added to words. These are also consonant and vowel errors.
Word stress	When an English word has more than one syllable, one of the syllables will be a little bit louder and longer than the others. For example, if you say the word "computer", you may notice that the second syllable has more stress (comPUter). If you hear stress being placed on the wrong syllable, or you hear equal stress on all of the syllables in a word, then there are word stress errors.
Intonation	Intonation can be thought of as the melody of English. It is the natural pitch changes that occur when we speak. For example, you may notice that when you ask a question with a yes/no answer, your pitch goes up at the end of the question. If someone sounds "flat" when they speak, it is likely because their intonation is not following English intonation patterns.
Speech rate	Speech rate is simply how quickly or slowly someone speaks. Speaking very quickly can make speech harder to follow, but speaking too slowly can as well. A good speech rate should sound natural and be comfortable to listen to.

☹   ☺

1. Vowel and/or consonant errors

Frequent	●————————●	**Infrequent or absent**

2. Word stress errors affecting stressed and unstressed syllables

Frequent	●————————●	**Infrequent or absent**

3. Intonation (i.e., pitch variation)

Too varied or not varied enough	●————————●	**Appropriate across stretches of speech**

4. Speech rate

Too slow or too fast	●————————●	**Optimal**