Language Awareness
Vol. 20, No. 1, February 2011, 45–59

Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology

Kazuya Saito*

Department of Integrated Studies in Education, McGill University, 3700 McTavish, Montreal, Quebec H3A1Y2, Canada

(*Received 20 October 2009; final version received 4 October 2010*)

This paper reports on an instructed second language acquisition study that investigated the effects of explicit phonetic instruction on second language pronunciation by adopting two different outcome measurements (i.e. a rubric of accentedness as well as comprehensibility). Twenty native Japanese learners of English in ESL (English as a second language) settings participated in the current study and were randomly assigned to the experimental group and the control group. After they received four-hour instruction with the target pronunciation features of English-specific segmentals /æ,f,v,θ,ð,w,l,ɹ/, the comprehensibility and perceived foreign accent of the participants' oral production in English were evaluated by four native English listeners. Results suggested that explicit instruction had a significant effect on comprehensibility especially in the sentence-reading task, although a significant reduction in foreign accent was not obtained in any contexts.

Keywords: pronunciation; second language learning; intervention; intelligibility

Comprehensibility and accentedness

Whereas second language (L2) speech research has convincingly shown that it is tremendously difficult to attain native-like L2 pronunciation skills and foreign accent is a normal aspect of L2 speech (e.g. Flege, 2003; Piske, MacKay, & Flege, 2001), Derwing and Munro (2005) claimed that L2 speech should be considered from two different perspectives: accentedness and comprehensibility. According to them, accentedness is 'the degree to which the pronunciation of an utterance sounds differ from an expected pronunciation pattern', and comprehensibility is 'listeners' estimation of difficulty in understanding an utterance' (p. 385). In order to test the interrelationship between the two criteria, Derwing and Munro (1997) conducted an experiment in which 26 native English (NE) listeners rated 48 ESL (English as a second language) learners' speech productions. Essentially, they found that accentedness did not necessarily interfere with comprehensibility: even heavily accented speech can be highly comprehensible. Although some L2 learners who strongly strive for nativeness should not be discouraged to pursue their ambitions, researchers as well as teachers should set realistic goals for L2 learners based on empirical research evidence, such as comprehensibility rather than accentedness (for similar discussion, see also Levis, 2005; Setter & Jenkins, 2005).

*Email: kazuya.saito@mail.mcgill.ca

ISSN 0965-8416 print / ISSN 1747-7565 online

DOI: 10.1080/09658416.2010.540326
http://www.informaworld.com

Although the importance of explicit phonetic instruction (i.e. explicitly teaching segmental and suprasegmental elements of the target language) has been extensively discussed in the field of experimental phonetics as well as second language education (see Derwing, 2008 for a comprehensive overview on pronunciation teaching studies), its pedagogical implications could be limited, arguably because its impacts on comprehensible pronunciation remain unclear. That is, these veins of L2 speech studies did not clearly state whether their research goals were: (1) to eliminate students' accentedness with a view to native-like pronunciation; or (2) to aim at improving their comprehensible pronunciation to exceed minimal requirement for successful L2 communication (e.g. Elliot, 1997; Macdonald, Yule, & Powers, 1994; Missaglia, 1999). In order to assess the *communicative* value of pronunciation teaching, researchers should focus on students who make efforts to learn L2 pronunciation rules and become comprehensible enough. Therefore, the current study investigates the relative effects of instruction via two different evaluation methods: accentedness and comprehensibility.

Pronunciation teaching

Many second language acquisition (SLA) theorists argued that the first step for effective instruction is to have learners to become consciously aware of formal rules of the target language features, because the metalinguistic awareness is central to interlanguage development (Schmidt, 2001). In this respect, DeKeyser (2003) emphasises the importance of explicit instruction option defined as follows: 'An instructional treatment is explicit if rule explanation forms part of the instruction (deduction) or if learners are asked to attend to particular forms and try to find the rules themselves (induction)' (p. 321). In the context of pronunciation teaching, Derwing and Munro (2005) claimed that 'students learning L2 pronunciation benefit from being explicitly taught phonological form to help them notice the difference between their own productions and those of proficient speakers in the L2 community' (p. 388). Similarly, Venkatagiri and Levis (2007) also maintained that explicit instruction can help learners develop 'phonological awareness' (i.e. conscious knowledge of segmentals and suprasegmentals), which might play a key role in L2 speech intelligibility (see also Field, 2005).

Although very few, SLA researchers have recently begun to pay ever more attention to research-based pronunciation teaching studies (Derwing, Munro, & Wiebe, 1998; Elliot, 1997; Macdonald et al., 1994; Missaglia, 1999; Neri, Mich, Gerosa, & Giuliani, 2008). For example, Macdonald et al. (1994) specifically tested four different pedagogical activities in the speech of 23 adult Chinese learners of English: traditional drilling activities, self-study with tape recording, interactive activities, and a no-intervention control condition. Results showed that 120 NE listeners noticed improvement, although there was no preference between non-control groups. On the other hand, Missaglia (1999) examined the details of the effects of two different strategies, *suprasegmental* training and *segmental* training: students given segmental and suprasegmental feedback performed generally well, but suprasegmental-centred instruction enabled students to do better than those who got segmental-centred instruction. As mentioned earlier, these studies, however, adopted only the accentedness criteria to assess students' improvements; it still remains unclear the extent to which their instruction treatment impacted learners' comprehensible pronunciation (Derwing, 2008; Derwing & Munro, 2005; Levis, 2005).

An exception is the often-cited study conducted by Derwing et al. (1998). In their quasi-experimental classroom study, 48 ESL participants with various L1 backgrounds

were divided into three groups (a segmental group, a suprasegmental group, and a no-instruction group) and the training lasted for 20 hours per week for 10 weeks in total. Those 48 ESL students were asked to do a pre-test and a post-test that consisted of a sentence-reading task (controlled speech level) and a picture-description task (spontaneous speech level), with 48 NE listeners evaluating all speech stimuli. The study produced significant results: (1) ESL participants both in the segmental and suprasegmental groups improved whereas those in the no-instruction group did not; and (2) a close examination revealed that suprasegmental training enhanced performance in a picture-description task while segmental training enhanced performance in a sentence-reading task. This study provided an elaborate research framework as a cornerstone for future pronunciation teaching studies: in order to clearly observe learners' gains from pronunciation teaching, experiments need sufficient amount of data, an adequate length of instructional period, and valid measurement methods in order to validate learners' gains from pronunciation teaching.

Given a growing attention to the importance of instruction with research-based evidence, we now need to pursue a fine-grained analysis of how pronunciation teaching can be facilitative of L2 pronunciation development. Importantly, few studies precisely mentioned what linguistic features were targeted in their treatment and how the intervention was implemented (e.g. Derwing et al. [1998] focused on all of segmental and suprasegmental aspects of English; Neri et al. [2008] targeted 28 words). So, the extent to which instruction was effective to enhance the targeted pronunciation features could not be accurately explained in the previous studies. This might be due to several reasons. First, their main research goals were arguably to test just the overall potential of their methods by including many general pronunciation rules and using composite instructional options. Second, in most studies, participants were ESL students from various countries, making it difficult to individualise the teaching syllabus. If teachers have students with homogeneous characteristics (e.g. L1 background, age, and motivation), they could prioritise targets by making instructional treatment tailored to their students' potential problems (Derwing & Munro, 2005).

Thus, the current study narrows down the scope of the learners, highlighting a specific L1/L2 speech contrast known to be difficult for adult native Japanese speakers (NJs) learning English in ESL settings. That is, the current study carefully follows the framework of instructed SLA studies: instructed SLA research has to: (1) specify the linguistic target of the instruction; (2) document the detailed nature of the treatment; and (3) adopt appropriate outcome measurements (Doughty, 2003; Ellis, 2006; Norris & Ortega, 2000).

Current study

Segmental-based instruction

L2 learners such as native Japanese learners of English (NJs) are likely to have L1–L2 transfer problems at segmental levels, which, in turn, negatively affect NE listeners' perception (Flege, 2003; Piske et al., 2001). One of the implications of NJs' relative difficulties in segmental production could be that NJs might benefit from segmental-based instruction that focuses especially on these English-specific segmentals (see also Riney, Takada, & Ota, 2000). While effectiveness of suprasegmental-based instruction is still controversial (e.g. Levis, 2005), one of the advantages of segmental-based instruction is its *teachability* (Jenkins, 2000; Setter & Jenkins, 2005): all English segmental features can be clearly explained on the basis of three main phonetic characteristics: (1) *articulator organs*; (2) *place of articulation*; and (3) *manner of articulation* (Ladefoged, 2003). In addition, especially when L2 learners share the same L1, their problems could be predicted, to a

great extent, based on cross-linguistic differences between L1 and L2 phonetic systems (Flege, 2003), and Lambacher (1999) reported that many Japanese learners of English have difficulties especially in pronouncing eight *English-specific segmentals* /æ,f,v,θ,ð w,l,ɹ/, because they do not exist in Japanese. In sum, the current study mainly examines the efficacy of segmental-based instruction for adult NJs learning English, especially focusing on these English-specific sounds.[1]

Perception activities

The assumption underlying the current study is that if learners become explicitly aware of the English-specific segmentals /æ,f,v,θ,ð,w,l,ɹ/ and establish a strong mental representation of these sounds with their fundamental phonetic traits, learners will actually pronounce them, making an impact on the overall intelligibility of their English speech so that NE listeners can comprehend them without much effort. Explicit instructional options for pronunciation teaching comprise two stages of perception activities: *identification* and *discrimination.*

First, at the *identification* stage, participants are given a clear account of formal properties of English-specific sounds one by one in a sequence, focusing on the three fundamental phonetic characteristics of speech sounds (articulator organs, place of articulation, and manner of articulation). Then, participants are asked to produce individual sounds according to what they were taught. Second, at the *discrimination* stage, given cross-linguistic comparisons of English and Japanese phonetic systems, learners are alerted to which Japanese sounds they might confuse with English sounds and asked to discriminate the target English sounds from the closest Japanese counterparts (e.g. /æ/ vs /a/, /f/ vs. /Φ/, /v/ vs. /b/, /θ/ vs. /s/, /ð/ vs. /z/; Lambacher, 1999). Then, participants are asked to produce both a target English sound and a close Japanese sound to consciously make a contrast between them.

Production activities and feedback techniques

Swain (1985) maintains that not only input but also output plays a critical role in second language learning to push learners to be more conscious and careful about the accuracy of targeted language forms. In pronunciation teaching, production activities through output of the target language forms could enable learners to test their learned metalinguistic knowledge and strengthen their physical domains by actually sensing articulatory events and modifying their productions with respect to pronunciation.

From L2 instructional perspectives, the transition from rule-based performance to memory-based performance implies two different types of production activities, *controlled* practice and *communicative* practice (Lyster, 2007). In the continuum of *controlled* activities, participants in the current study are encouraged to actively practice three levels of reading tasks: (1) *segmental-level reading task*; (2) *word-level reading task*; and (3) *sentence-level reading task* (see, for details of each activity, Derwing, Munro, & Thomson, 2004). In the continuum of *communicative* activities, participants are given picture-description activities in which they are asked to describe two pictures as if they were explaining to somebody who had never seen them.

In this type of activity, learners are required to pay attention to their accuracy in syntax and word choices as well as their pronunciation, which could simulate spontaneous speech settings.

In the course of output, corrective feedback is offered to encourage L2 students to produce more output, notice their errors, and self-repair errors in phonetic forms

(articulator organs, place of articulation, and manner of articulation). In previous classroom observational studies, teachers frequently used *recasts* (i.e. showing model pronunciations) in response to students' mispronunciation or unclear pronunciation (e.g., see, for their analysis of pronunciation-focused feedback in adult ESL classrooms, Ellis, Basturkmen, & Loewen, 2001). In addition, the current study also adopts more explicit feedback to increase the amount of learners' uptake and practice. This type of output-prompting feedback is called *prompts* that 'offer learners an opportunity to self-repair by generating their own modified response' such as clarification requests, repetition of error with intonation, elicitation, and metalinguistic clues (Lyster, 2007, p. 108). When necessary, the instructor provided objective feedback with the aid of the acoustic analysis software Praat (Lambacher, 1999).

Methods

Participants

Twenty adult NJs of intermediate proficiency participated in the present study (mean = 27.6 years). They had just arrived in the upstate NY to study at the university level, with the exception of four participants who already had stayed for more than six months (time in the USA: mean = 2.3 months). They studied in ESL settings, taking regular academic courses in English as well as two or three ESL classes per week. When the researcher conducted the personal interview before the experiment, all of them reported that they were highly motivated to learn English and had many opportunities to use English academically and socially on a daily basis. All of them had learned English for more than 10 years since their entrance to seventh grade in junior high school in Japan. They were divided into two groups, an experimental group (Participants 1–10) and a control group (Participants 11–20) as equally as possible, based on their age, gender, TOEFL scores, and length of time in the USA.

For the 10 participants in the experimental group, the intent of the study was explained and the participants clearly showed their interest in participating in the study. As for the other 10 participants in the control group, they were not informed of the existence of an experimental group in order to avoid compromising the internal validity of the experiment: it has been shown that participants tend to become highly unmotivated or motivated if they know that they are in a control group (Parker, 1990). Ten participants in the control group were not given any instructional treatment; they were given free choice regarding whatever they wanted to do while the other participants in an experimental group received instruction. According to their self-reports after the experiment, some participants in the control group studied at the library while others took ESL classes during the experiment period. Importantly, examining how the control group performed is assumed to reveal: (1) test–retest effects (i.e. learners might have benefited because the same test was used twice); and (2) any kind of learners' improvement due to their stay in an English-speaking country for about a month (i.e. simple exposure to the target language might have unconsciously improved learners' pronunciation such as the other segmentals, supra-segmentals, and speech rate which are not the focus of the current study). Taken together, the only difference between participants in the experimental and control group was that the former students received four-hour explicit phonetic instruction whereas the latter students did not. The current study aims at measuring the true effects of phonetic instruction, by comparing *instructed* learners in the experimental group with *uninstructed* learners in the control group.

Instructional settings

All instruction was offered in a laboratory setting by the researcher to one or two students as tutoring sessions,[2] lasting one hour per week for a total of four weeks. An instructor gave instructional treatment to one or two students as tutoring sessions. The researcher/instructor is a non-native speaking teacher (L1, Japanese) and graduated from an MA programme in linguistics in the USA with a concentration in TESOL. Instruction was presented both in Japanese and English depending on the extent of learners' understanding.

Listeners

Four NE listeners participated in the present study (one male and three females). They were recruited from X University (in the USA). All had grown up in the United States and reported normal hearing. All listeners were experienced instructors of either phonetics or ESL classes at X University. Moreover, they were categorised as 'trained NE listeners' because they also reported their regular contact with a wide variety of ESL learners and their familiarity with L2 speech.

Speech samples

The current study used a sentence-reading task in order to measure participants' performance at the controlled level, and a picture-description task in order to measure their performance at the spontaneous level (e.g. Derwing & Munro, 1997; Derwing et al., 1998; Munro, Derwing, & Morton, 2006). With respect to the sentence-reading task, for instance, Derwing et al. (1998) used three sentences that have 'high-frequency lexical items' in a general sense. Since this syllabus targeted the phones /æ,f,v,θ,ð,w,l,ɹ/ the current study deliberately composed four loaded sentences that can equally assess participants' performance on each of these phones (see the Appendix). In total, four sentences had 39 loaded words that included a number of problematic phones for NJs out of 50 words (see Table 1). When participants were asked to read, these four sentences were randomly presented to participants together with three non-loaded sentences as distracters that included few problematic sounds in order to avoid overattention to what was being tested.

In the picture-description task that was designed to evaluate participants' spontaneous speech, the current study randomly selected one set of two pictures. Participants were asked to describe the image freely as if they were talking to somebody who had never seen the pictures before. Twenty seconds of their speech were randomly selected in the middle and used as speech stimuli. The order of the two tasks was fixed for all participants. First,

Table 1. Contents of loaded sentences out of 50 words.

Targeted phones	Total number of loaded phones	Examples
/æ/	6	married, happy, bad
/f, v/	6	office, food, of, visa
/θ, ð/	6	think, things, they, the
/w/	4	when, woman, with, workers
/l, ɹ/*	17	read, letter, recently, limited

*/r/ sounds in the word initial and medial positions were counted.

they were given the sentence-reading task. Participants read four loaded sentences together with three non-loaded sentences as distracters. Next, participants performed the picture-description task. After participants fully understood the task, each practiced by using one set of two pictures to familiarise themselves with the task. Immediately afterwards, they moved onto another set of two pictures as the real test. Participants were allowed sufficient time to think about what to say and ask the instructor any questions about the vocabulary and expressions they intended to use. In the first picture, a man and a woman are on stage acting in a play. In the second picture, both of them are at a restaurant and he is giving her a ring as a surprise present. Several participants reported that the two pictures did not seem to have one coherent story, so they opted to explain each of them independently. The same test procedure was conducted for both pre- and post-tests. All speech stimuli were recorded by the computer software, Praat, and 200 speech samples per stimuli were collected in total (4 sentences $\times$ 20 participants $\times$ 2 pre/post-tests $=$ 160 sentences; 1 picture-description $\times$ 20 participants $\times$ 2 pre/post-tests $=$ 40 descriptions), excluding all practicing.

Controlled stimuli

As in Derwing et al.'s (1998) study, the present study made controlled recordings of two NEs (one male, one female). Those two NEs were asked to do the same tasks, and an additional 10 samples/stimuli were added (4 sentences $\times$ 2 participants $+$ 1 picture-description $\times$ 2 participants $=$ 10 samples/stimuli). It was expected that all of their stimuli would be rated as very proficient compared to the 20 NJs. Thus, if raters failed to detect two NEs' stimuli among samples of NJs, they had to be eliminated because their rating abilities were not reliable. In sum, 210 samples per stimuli (200 NJs' samples/stimuli $+$ 10 NEs' samples/stimuli $=$ 210 samples/stimuli) were collected. They were randomised on a data CD.

Overall procedure

First, at the entry point (Time 1), all 20 participants were asked to do the pre-test (sentence-reading task/picture-description task), and a brief personal interview was conducted to determine their backgrounds in terms of their knowledge of pronunciation and personal impression towards pronunciation teaching. Next, one week after T1, the 10 participants in the experimental group took part in one-hour instructional session per week for four weeks while the other 10 participants studied in the library. Two weeks after the last lesson (seven weeks after the pre-test [T1]), all participants were asked to do the same task as a post-test to measure their pronunciation levels at the exit point (Time 2). The procedure is summarised in Figure 1.

Human rating method

For the purpose of measuring learners' improvement in segmental accuracy and their mutual intelligibility, the current study adopted a human rating method in the same manner as similar previous studies (e.g. Derwing et al., 1998). Four trained NE listeners were asked to listen to one data CD that contained 210 randomised stimuli and rate them on the basis of the 9-point scale on a rubric of accentedness (from 1 $=$ native-like to 9 $=$ heavily accented) and comprehensibility (from 1 $=$ no effort to understand to 9 $=$ very hard to understand).

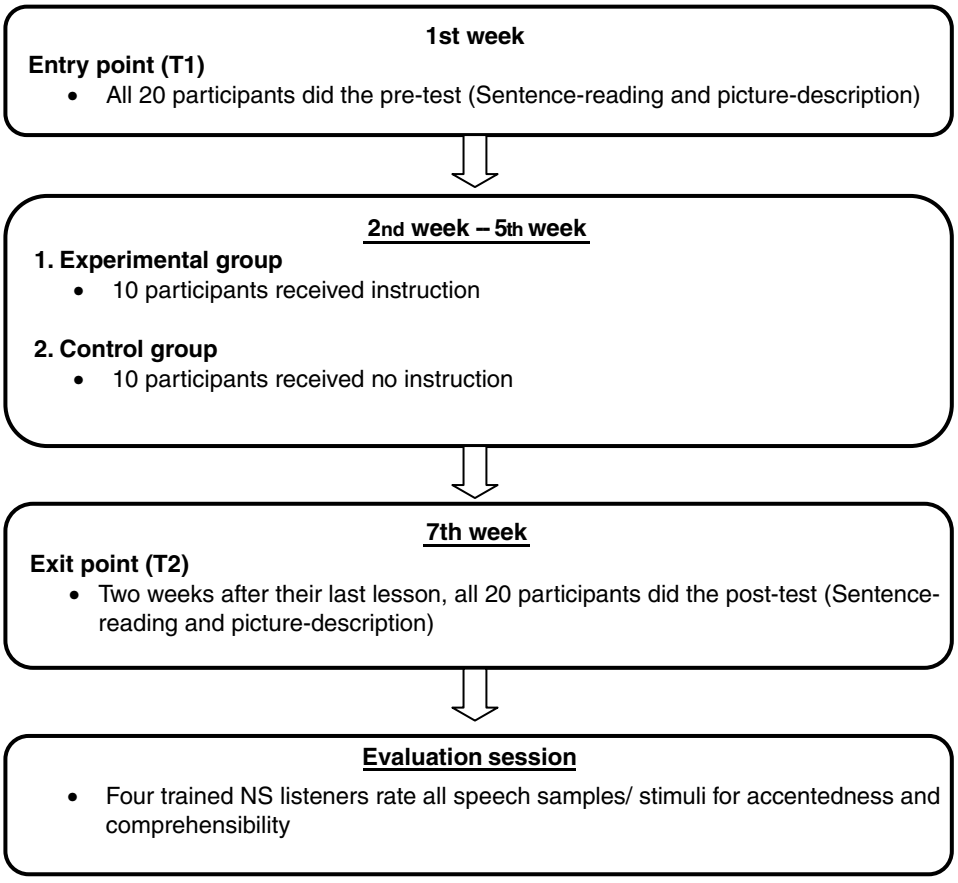

Figure 1. Summary of the measurement procedure.

After each stimulus, a 20-second pause was presented for listeners to grade. Listeners were advised to use the entire scale.

Listener training

The current study added two extra activities in order to enhance the reliability of raters. First, in order to enhance intra-rater reliability, they were given specific instruction to make sure that only segmental aspects of pronunciations were considered, because variables such as their own lexicogrammatical competences could alter the ratings (Derwing & Munro, 1997). Second, in order to enhance inter-rater reliability, listeners did a warm-up session together by using six example sounds randomly selected from the data set. In the session, four NE listeners discussed what score a speech stimulus deserved and checked how the other listeners rated the same speech tokens. The whole session was divided into two days (each session took 1 hour and 30 minutes including both training and listening). All four NE listeners were paid $30 at the end of the experiment.

Data computation

The four NE trained listeners are identified as Raters 1, 2, 3, and 4. Also, 10 NJ participants in the experimental group are labelled as Participants 1–10, and the 10 NJ participants in the control group are identified as Participants 11–20. For the sentence-reading task, the mean of each participant's reading of four sentences was obtained from each rater. On the pre-test, Participant 1 got '7', '8', '7', and '6' (four sentences) from Rater 1. Thus, the mean is '7', and this figure '7' is given to Participant 1 as his performance of the sentence-reading task on the pre-test. For the picture-description task, one score was given by each rater both at the pre-test and the post-test.

A set of two-way mixed-design analysis of variance (ANOVA) tests (experimental/control group × pre/post-test) were administered for four contexts: (1) *accentedness/sentence reading*; (2) *accentedness/picture description*; (3) *comprehensibility/sentence reading*; and (4) *comprehensibility/picture description*. The α level was set at $p < .01$. Follow-up statistic tests, effect size estimates based on Cohen's d-index and mean vector analysis, were conducted, especially with respect to observed findings in the context of comprehensibility which is more closely related to mutual intelligibility (e.g. Derwing & Munro, 1997). All scores from the four raters were used: 10 participants (either in the experimental or the control group) × 4 raters = 40 variables. The procedure is summarised in Figure 1.

Results

Inter-rater reliability

Each of four trained NE listeners was given 210 stimuli (200 NJs' speech stimuli + 10 NE's speech stimuli) and provided 420 responses (210 accentedness responses, 210 comprehensibility responses). The inter-class correlation (between all four raters) was calculated for accentedness ($r = .66$) and comprehensibility ($r = .53$) by following the manner of Ebel (1951). The r-values were statistically significant at $p < .001$, indicating adequate reliability among the four raters.

Pre-test

A two-tailed t-test was conducted on pre-test values which found no significant difference between participants in the experimental group and the control group in any cases.

Accentedness

In the context of accentedness, no significant differences were found in the sentence-reading tasks, $F(1, 18) = 0.000$, $p = .984$, $d = 0.01$ for Group (i.e. the mean difference between experimental and control groups at the time of post-tests), $F(1, 18) = 0.578$, $p = .457$, $d = 0.07$ for Time (i.e. the difference between pre- and post-test sessions), nor in the picture-description tasks, $F(1, 18) = 1.050$, $p = .319$, $d = 0.44$ for Group, $F(1, 18) = 3.334$, $p = .084$, $d = 0.18$ for Time. It is noteworthy that even within the experimental group who received instruction, some increased their received accentedness both in the sentence-reading task to some degree (e.g. Participant 5, mean = 6.25 → 7.68 points; Participant 7, mean = 5.63 → 7.5 points) as well as in the picture-reading task (Participant 8, mean = 6.56 → 6.81 points).

Comprehensibility

In contrast to accentedness, instruction yielded statistically significant gains with respect to comprehensibility. First, with respect to sentence reading, the ANOVA results revealed significant effects for the overall Group × Time interaction, $F(1, 18) = 19.945$, $p = .00003$. A simple main effect for Time was found significant for the experimental group (mean = 3.83 → 3.27 points), $F(1, 18) = 30.15$, $p = .00001$, i.e. $p < .01$, with a large effect size ($d = 0.96$), while those in the control group did not show any significant gains (mean = 3.82 → 3.73 points). In particular, three participants in the experimental group improved their mean value in the sentence-reading task (four sentences) to a great degree: Participant 1 (mean = 3.56 → 2.56 points), Participant 3 (mean = 3.38 → 2.31 points), and Participant 8 (mean = 4.13 → 3.15 points).

Second, in the context of picture description, there was no statistical significance for Time, $F(1, 18) = 1.88$, $p = .186$, $d = 0.20$ nor for Group, $F(1, 18) = 0.00$, $p = .986$, $d = 0.71$. Interestingly, a close examination of the results noted a wide range of individual performance both at the pre-test and the post-test, which might have obviated the overall statistical significance. For example, the mean vector analyses for the picture-description task in comprehensibility showed marked improvement for some participants if we categorise the mean value of a score change of more than –1.0 point as 'improvement'. Four participants showed remarkable improvement: Participant 1 (mean = 4.25 → 2.25 points), Participant 2 (mean = 4.25 → 3.25 points), Participant 4 (mean = 7 → 4.75 points), and Participant 9 (mean = 6.5 → 4.5 points). The details of the results are summarised in Table 2.

Discussion

The current study examined the efficacy of explicit phonetic instruction in the context of NJs and found noteworthy results. Most importantly, in order to boost teachability and learnability of comprehensible pronunciation to a great degree, this study clearly identified pronunciation targets (*L2-specific segmentals*), the nature of the instructional treatment (*explicit instruction/feedback techniques*), and evaluation methods (not in pursuit of *native-like* but *comprehensible* pronunciation) specifically for NJs. The experiment confirmed that explicit instruction benefited NJs' comprehensibility in the experimental group especially at the controlled speech level (*sentence-reading*). A more detailed examination of individual participants revealed that participants exhibited general improvement in comprehensibility even in the cases of the spontaneous speech (*picture-description*; Participants 1, 2, 4, and 9), but with considerable variance between participants; the lack of statistical significance in the ANOVA comparisons could be due to the fact that learners deliberately processed the phonetic knowledge from instruction and actively applied it in natural speech settings at different rates. Given more instruction time, perhaps more participants would begin to show improvements that might influence the overall statistical significance.

On the contrary, unlike other previous studies (Derwing et al., 1998; Elliott, 1997), a significant reduction in foreign accents was not clearly achieved in the sentence-reading nor in the picture-description tasks in the current study. Although it is difficult to draw any decisive conclusion about the effectiveness of instruction on accentedness, especially in conjunction with the relatively small number of participants ($n = 20$) and short length of instruction (four hours) in the current study, one of the possible interpretations could be that it is cognitively and physically *difficult* for adult L2 learners to decrease the degree of accentedness, but explicit phonetic instruction can facilitate their acquisition of comprehensible

Table 2. Summary of comprehensibility ratings.

	Sentence reading		Picture description	
	Pre-test	Post-test	Pre-test	Post-test
Experimental group				
Participant 1	2.94	2.56	4.25	2.25
Participant 2	3.38	2.31	2.5	3.75
Participant 3	3.75	3.31	7.00	4.75
Participant 4	4.00	3.81	3.75	4.00
Participant 5	4.31	3.81	5.00	5.00
Participant 6	3.75	3.75	3.75	4.50
Participant 7	4.75	4.12	4.50	7.25
Participant 8	3.75	3.06	6.50	4.50
Participant 9	4.13	3.50	6.00	6.25
Participant 10	3.56	2.56	3.00	3.25
Average	mean = 3.83* (SD = 0.47)	mean = 3.27* (SD = 0.59)	mean = 4.75 (SD = 1.31)	mean = 4.45 (SD = 1.47)
Control group				
Participant 11	2.19	2.00	2.19	2.00
Participant 12	4.75	4.81	4.75	4.81
Participant 13	3.44	3.62	3.44	3.62
Participant 14	4.50	3.62	4.50	3.62
Participant 15	4.06	3.93	4.06	3.93
Participant 16	3.69	3.62	3.69	3.62
Participant 17	4.19	4.35	4.19	4.35
Participant 18	3.63	3.62	3.63	3.62
Participant 19	4.13	4.25	4.13	4.25
Participant 20	3.63	3.56	3.63	3.56
Average	mean = 3.82 (SD = 0.67)	mean = 3.73 (SD = 0.70)	mean = 5.27 (SD = 1.33)	mean = 5.58 (SD = 1.53)

*Significant differences at $p < .01$.

L2 speech production more *immediately* (Derwing & Munro, 2005). As implications for practitioners in L2 classrooms, empirical findings in the current study suggest that teachers should assess students' performance on the basis of 'comprehensibility' rather than 'accentedness' in order to measure their true improvements. There is a great possibility that students whose L2 speech is sufficiently comprehensible could be underestimated as 'no-improvement' with respect to accentedness evaluations. For example, in the case of the sentence-reading task, Participants 6, 8, and 9 were judged as 'regressing' from T1 to T2 on a rubric of accentedness (mean = 7.00 → 7.25, 7.25 → 7.56, and 6.19 → 6.68 points, respectively), but they were considered as 'improving' on a rubric of comprehensibility (mean = 4.31 → 3.81, 4.75 → 4.12, and 3.75 → 3.06 points, respectively). This result concurs with that of Derwing and Munro in that 'although some features of accent may be highly salient, they do not necessarily interfere with intelligibility' (1997, p. 11).

Unfortunately, many practitioners as well as students in L2 classrooms still believe that it is highly possible and plausible to attain native-like pronunciation with no foreign accent (Derwing & Munro, 2005; Field, 2005; Jenkins, 2000). In fact, Levis pointed out that 'many teachers, especially those unfamiliar with pronunciation research, may see the rare learner who achieves a native-like accent as an achievable ideal, not an exception' (2005, p. 370). However, it has to be noted that L2 speech studies have never provided evidence that L2 instruction could eradicate foreign accent (Piske et al., 2001) and it is possible that

forcing such a goal might inhibit many L2 learners who have different motivations and aims (Field, 2005). In short, as Derwing and Munro claim, 'out of concerns for the welfare of L2 students, teachers should help them to set realistic goals on the basis of current research findings' (2005, p. 384).

Future direction

Over the past 20 years, instructed SLA has been extensively discussed within L2 syntax and L2 morphology but without much attention to L2 phonology (Norris & Ortega, 2000). However, as Ellis claimed, 'the linguistic target of the instruction can be *phonological*, lexical, grammatical or pragmatic' (2006, p. 157, emphasis added). Similarly, Doughty and Williams (1998) also assumed that empirical findings in instructed SLA studies could be applicable to all types of language features including L2 phonology. The current study, therefore, took a first step to further examine the mechanism of instructed SLA in L2 phonology by contrasting 'instructed learners' and 'uninstructed learners'. Although implications from the current study that adopted only one type of instruction (i.e. exclusively explicit instruction in a lab setting) with such small samples have to be interpreted with caution, it is still important to make pedagogical suggestions for L2 classrooms and to inform future directions for instructed L2 phonology studies. This section will mainly discuss two crucial research components – (1) instructional treatment and (2) outcome measures.

Instructional treatment

The explicit phonetic instruction in the current study could be categorised as one type of *form-focused instruction* (FFI): Spada defined FFI as 'any pedagogical effort which is used to draw the learners' attention to language form either implicitly or explicitly' (1997, p. 73). Given that the current study showed that learners should be given FFI on particular pronunciation problems such as L2-specific sounds, and encouraged to practice these basic forms through repetition in a rather decontextualised manner, a next step could be to investigate how L2 learners practice it through more communicative activities to improve not only accuracy but also fluency and automaticity of their pronunciation after sufficiently raising their awareness about rule-based knowledge (Trofimovich & Gatbonton, 2006).

Importantly, in the context of L2 grammar teaching, Lyster suggested that FFI can be most effective 'when implemented in communicative contexts, to ensure that learners will be able to transfer what they learn in the classroom to communicative interaction outside the classroom' (2007, p. 43). This instructional notion draws on *transfer appropriate processing* theory whereby learning language form through meaningful activities will help learners improve their retrieval in future communication settings rather than metalinguistic knowledge learned in decontextualised contexts such as mechanical drills (see also Doughty, 2003; Ellis, 2006). In fact, one could argue that the lack of communicative practice in the current study might have resulted in learners' improvements limited only in the controlled speech level (sentence-reading).

Taken together, future studies are now called for in order to examine the potential of different types of FFI and feedback in pronunciation teaching in depth. Among very few studies that have explored communicative activities for pronunciation teaching (Pica, 1984; Rogerson & Gilbert, 1990), Celce-Murcia presented a framework for designing such materials: after finding and prioritising students' problems, teachers should 'find lexical/grammatical contexts with many natural occurrences of the problem sound' and 'develop communicative tasks that incorporate the word' (1987, p. 10). Importantly, Brown

suggested that 'the same type of activities used to teach other language areas communicatively can also be used to teach pronunciation' (1995, p. 173).

Outcome measures

As one of the most obvious shortcomings, the current study drew on the small database (i.e. 20 NJs completed four sentence readings and one picture description for evaluation purposes). Thus, the results of the current study need to be considered as very tentative ones. Truly, one could emphasise the importance of conducting larger scale experiments with more participants and test tokens (e.g. Derwing et al. [1998] with 48 participants; Elliott [1997] with 66 participants). However, the current study actually revealed that the adoption of only human rating methods as the reliable outcome measure to assess improvement resulting from instruction could be problematic because it seriously limits the construct validity and feasibility of L2 pronunciation teaching research of this kind. Note that, although the current study asked NE listeners to rate a small data set (20 NJs $\times$ 5 tokens [four sentence readings + one picture description] $\times$ 2 test occasions [pre- and post-test sessions]), the listening sessions actually lasted three hours in total. In fact, the decision had to be made to divide the entire sessions into two separate days due to fatigue problems (i.e. all of the four listeners agreed that it would have been difficult to provide trustworthy ratings if the length of one session had exceeded more than 90 minutes). Although it could be still possible to ask NE listeners to rate a number of speech tokens produced by many participants, it would inevitably put a great deal of burden not only on listeners but also on researchers (e.g. Derwing et al., 1998 for six hours of listening; Elliot, 1997 for six hours of listening). Importantly, other previous studies tended to adopt a small number of participants (e.g. Macdonald et al., 1994, 23 participants; Neri et al., 2008, 28 participants).

Intriguingly, L2 speech studies have adopted acoustic analyses in order to measure acoustic properties of L2 speech sounds (e.g. formant frequencies, intensity, and duration) at a fine-grained level, and very few intervention studies have begun to use this technique as another possible outcome measure for future pronunciation teaching research (e.g. Saito, 2007; Saito & Lyster, in press). However, given that it still remains unclear to what degree any changes in such acoustic properties can immediately or ultimately affect NE listeners' overall comprehension which is arguably the main focus of pronunciation teaching (Ladefoged, 2003), it might be necessary to wait for future research to elaborate new outcome measures with an innovative research design, which could combine both human rating methods and acoustic analyses in a complementary fashion.

Acknowledgements

Parts of this study were presented at the annual meeting of the Interfaces of Intelligent Computer-Assisted Language Learning in Waterloo, Ontario, on December 8, 2007, and at Georgetown University Round Table in Washington DC on March 13, 2009. I gratefully acknowledge Roy Lyster, Kim van Poeteren, Talia Issac, and anonymous *Language Awareness* reviewers for their helpful comments on earlier versions of this article.

Notes

1. Note that these eight English-specific sounds were also identified as most important pronunciation features to teach by 48 experienced NJ English teachers belonging to a private language school located in Tokyo (Saito, 2009).

2. Out of the 10 participants in the experimental group, two students received instruction in pairs, because they were friends with each other and preferred to do so rather than in a one-on-one meeting. As an instructor, I did not observe any significant difference in the amount and quality of instruction they received compared with the other students who received tutored instruction individually.

Notes on contributor

Kazuya Saito is an instructor as well as a doctoral candidate in the Department of Integrated Studies in Education at McGill University. His research examines the pedagogical effectiveness of FFI and corrective feedback on L2 pronunciation development.

References

Brown, A. (1995). Minimal pairs: Minimal importance? *ELT Journal*, *49*(2), 169–175.
Celce-Murcia, M. (1987). Teaching pronunciation as communication. In D. Morley (Ed.), *Current perspectives in pronunciation: Practice anchored in theory* (pp. 5–12). Washington, DC: TESOL.
DeKeyser, R. (2003). Implicit and explicit learning. In M. Long & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 313–348). Malden, MA: Blackwell.
Derwing, T. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 347–369). Amsterdam: John Benjamins.
Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, *2*, 1–16.
Derwing, T., & Munro, M. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379–397.
Derwing, T., Munro, M., & Thomson, R. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*, 655–679.
Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393–410.
Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In M. Long & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 257–310). Malden, MA: Blackwell.
Doughty, C. , & Williams, J. (Eds.) (1998). *Focus on forms in classroom second language acquisition*. Cambridge: Cambridge University Press.
Ebel, R. (1951). Estimating the reliability of ratings. *Psychometrica*, *16*, 407–424.
Elliott, A. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, *80*, 95–108.
Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. *Applied Linguistics*, *19*, 18–41.
Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning*, *51*, 281–318.
Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, *39*, 399–423.
Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities* (pp. 319–355). Berlin: Mouton de Gruyter.
Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
Ladefoged, P. (2003). *Phonetic data analysis: An introduction to instrumental phonetic fieldwork*. Oxford: Blackwell.
Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer-Assisted Language Learning*, *12*, 137–156.
Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 367–377.
Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamin.

Macdonald, D., Yule, G., & Powers, M. (1994). Attempts to improve English L2 pronunciation: The variable effects of different types of instruction. *Language Learning*, *44*, 75–100.
Missaglia, F. (1999). Contrastive prosody in SLA: An empirical study with adult Italian learners of German. In J.J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A.C. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Science* (Vol. 1, pp. 551–554). Berkeley, CA: University of California.
Munro, M., Derwing, T., & Morton, S. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*, 111–131.
Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, *21*, 393–408.
Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528.
Parker, R. (1990). Power, control, and validity in research. *Journal of Learning Disabilities*, *23*, 613–620.
Pica, T. (1984). The selective impact of classroom instruction on second language acquisition. *Applied Linguistics*, *6*, 214–222.
Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accents in an L2: A review. *Journal of Phonetics*, *29*, 191–215.
Riney, T., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly*, *34*, 711–737.
Rogerson, P., & Gilbert, J. (1990). *Speaking clearly*. Cambridge: Cambridge University Press.
Saito, K. (2007). The influence of explicit phonetic instruction on pronunciation teaching in EFL settings: The case of English vowels and Japanese learners of English. *Linguistic Journal*, *3*, 17–41.
Saito, K. (2009, March). *The impact of English-specific segmentals on mutual intelligibility: Suggestions for pronunciation teaching*. Paper presented at the 30th Annual NY TESOL Applied Linguistics Conference, Columbia University, New York City, NY.
Saito, K., & Lyster, R. (in press). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of English /ɹ/ by Japanese learners of English. *Language Learning*.
Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 1–32). Cambridge: Cambridge University Press.
Setter, J., & Jenkins, J. (2005). Pronunciation. *Language Teaching*, *38*, 1–17.
Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching*, *29*, 73–87.
Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley, MA: Newbury House.
Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in L2 Spanish word processing: Implications for pronunciation instruction. *Modern Language Journal*, *90*, 519–535.
Venkatagiri, H., & Levis, J. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness*, *16*(4), 263–277.

Appendix. Sentences used for pre/post-tests

1. When do you think they are going to read letters?
   /wɛn du jʊ θɪŋk ðei əɹ ˈgoʊɪŋ tə ɹid ˈlɛtəɹz?/ (*6 out of 10 words are loaded*)
2. I guess a married woman is usually happy with her office life.
   /aɪ gɛs ə ˈmæɹid ˈ wʊmən ɪz ˈu ʒuəlɪ ˈhæpɪ wɪð həɹ ˈɔfɪs laɪf/ (*8 out of 12 words are loaded*)
3. He has at least nine things to complete on campus because of his visa.
   /hi hæz ət list naɪn θɪŋs tə kəmˈplit ən ˈkæmpəs bɪˈkɔz əv hɪz ˈvizə/ (*7 out of 14 words are loaded*)
4. Recently, the amount of food is very limited and that is bad for workers.
   /ˈɹi səntlɪ ðə əˈmaʊnt ə(v) fud ɪz ˈvɛɹɪ ˈlɪm ɪtɪd ənd ðæt ɪz bæd fəɹ wɜɹkəɹz/ (*10 out of 14 words are loaded*)