Language Learning & Technology
http://llt.msu.edu/vol13num3/tannerlandon.pdf
October 2009, Volume 13, Number 3
pp. 51–65

THE EFFECTS OF COMPUTER-ASSISTED PRONUNCIATION READINGS ON ESL LEARNERS' USE OF PAUSING, STRESS, INTONATION, AND OVERALL COMPREHENSIBILITY

Mark W. Tanner and Melissa M. Landon
Brigham Young University

ABSTRACT

With research showing the benefits of pronunciation instruction aimed at suprasegmentals (Derwing, Munro, & Wiebe, 1997, 1998; Derwing & Rossiter, 2003; Hahn, 2004; McNerney and Mendelsohn, 1992), more materials are needed to provide learners opportunities for self-directed practice. A 13-week experimental study was performed with 75 ESL learners divided into control and treatment groups. The treatment group was exposed to 11 weeks of self-directed computer-assisted practice using Cued Pronunciation Readings (CPRs). In the quasi-experimental pre-test/post-test design, speech perception and production samples were collected at Time 1 (week one of the study) and Time 2 (week 13). Researchers analyzed the treatment's effect on the learners' perception and production of key suprasegmental features (pausing, word stress, and sentence-final intonation), and the learners' level of perceived comprehensibility. Results from the statistical tests revealed that the treatment had a significant effect on learners' perception of pausing and word stress and controlled production of stress, even with limited time spent practicing CPRs in a self-directed environment.

INTRODUCTION

Computer-Assisted Language Learning (CALL) is of interest to language teachers and learners because it can provide individualized instruction and immediate feedback on the correctness of a learner's response to computerized tasks (Nagata, 1993). In computer-aided pronunciation (CAP), technology has increased learners' access to their and others' pronunciation performance through visual displays such as spectrographic analyses of individual phonemes or amplitude waves showing levels of intensity for isolated words or phrases (Anderson-Hsieh, 1992, 1994; Hardison, 2004; Molholt, 1988). While such programs can provide learners with immediate feedback regarding the accuracy of an utterance compared to that of a native speaker (NS), they typically require teacher supervision and interpretation. Pennington (1999) cites another drawback of CAP, stating that nearly all CAP programs focus exclusively on segmentals. This focus implies that intelligibility is primarily impaired by the articulation of individual sounds and ignores the influence of prosody. If intelligibility is prioritized above accuracy, a focus on key words, stress, rhythm, and intonation rather than the articulation of individual sounds, may be needed.

Empirical research has begun to confirm the importance of prosodic features in learners' overall intelligibility and perceived comprehensibility. Blau (1990) found that appropriate pausing patterns in native English speech had a significantly greater effect on non-native listeners' comprehension than either syntactic complexity or speech rate. Fayer and Krasinski (1995) found that native speakers' intelligibility judgment of non-native English speech significantly correlated with pause length. Native speakers rated non-native speakers (NNSs) with longer individual pauses and greater total pause time as less intelligible than those with more appropriate pause length. Towell, Hawkins, and Bazergui (1996) found that as British learners of French improved their pausing patterns in French, their overall fluency improved, as measured by speaking rate.

Regarding stress, Field (2005) found that when native English speech was manipulated to include incorrect lexical stress, the ability of both NS and NNS listeners to locate words in connected speech was

seriously affected (p. 419). Sentence stress, also known as primary stress (Hahn, 2004), also has a crucial role in successful communication. Hahn (2004) reported that correct sentence-level stress by an ESL speaker, compared to misplaced or omitted stress, led to improved listener comprehension and recall of content.

Regarding the importance of intonation, Levis (1999, 2002), Levis and Pickering (2004), and Jenkins (2004) have emphasized the importance of teaching intonation in context, preferably at a discourse-level, rather than within isolated sentences. Wennerstrom (1998) analyzed brief lectures given by Chinese ESL speakers and found that those using appropriate intonation contours received higher ratings on a speaking skills test. Similarly, Pickering (2004) found that NS teachers consistently used discourse-level intonational cues to "emphasize relationships between semantically related sections of the discourse and highlight information structure" (p. 38) while international teaching assistants (ITAs) did not. She noted that the ITAs' weaknesses in discourse-level intonation were likely to impede students' understanding.

This research into the effects of pausing, stress, and intonation on the comprehensibility of non-native English speech has prompted teachers and materials developers to devise various techniques for incorporating suprasegmental practice into the classroom. The earliest such technique, jazz chanting (Graham, 1978), continues to be used and advocated by teachers who have students chant poems and songs to become more familiar with English rhythm, stress, and intonation (Richman, 2005).

Other oral techniques are advocated in pedagogical materials to enhance learners' use of prosodic features. Mirroring, tracking, and shadowing involve imitating native speaker discourse models (Celce-Murcia, Brinton, & Goodwin, 1996). A fourth imitative technique, imitative conversation (Goodwin, 2004), has ESL learners select, analyze, and then replicate a brief one- to two-minute clip of dialogic speech from a movie or television show.

While such techniques have been promoted as ways to provide contextualized classroom practice for suprasegmentals, empirical research regarding the specific benefits of such techniques has been slow in coming. As Celce-Murcia, et al (1996) put it:

> There is no consensus in the literature about which of these techniques is most effective; we advise teachers to experiment with them and get feedback from their students as the learners themselves are the ultimate judges of what they find most useful (p. 310).

Clearly, empirical research is needed to investigate the actual effects of these pedagogical techniques on learners so teachers can make sound decisions about their use.

Empirical studies conducted in recent years have focused on the effects of various types of pronunciation instruction on learners' overall levels of intelligibility and comprehensibility. Intelligibility is "the extent to which a listener actually understands an utterance" (Derwing & Munro, 2005, p. 385) and is often evaluated through transcription or listening comprehension tasks performed by a listener. Comprehensibility is "a listener's perception of how difficult it is to understand an utterance" (Derwing & Munro, 2005, p. 385). Comprehensibility (also referred to as perceived comprehensibility) is often measured using a Likert scale to rate speech samples based on listeners' perceptions of how easily the speaker or speech stimuli can be understood.

Derwing, Munro, and Wiebe (1997) investigated how a 12-week course focusing on prosodic features might impact ESL learners considered to be fossilized in their English pronunciation skills. Learners were recorded at Time 1 and Time 2 reading sentences and telling a story based on picture prompts. Results from 57 NS listeners showed significant changes at Time 2 in the ESL speakers' intelligibility, comprehensibility, and accent. The researchers concluded that pronunciation instruction deemphasizing the importance of segmental units, combined with a greater focus on prosody and general speaking characteristics, can effectively change fossilized pronunciation patterns in individuals who have spent years in an English speaking environment.

Considering the 1997 study's positive finding, Derwing, Munro and Wiebe (1998) conducted a second study comparing the effects of different types of instruction on NNSs' comprehensibility, accentedness, and fluency. Forty-eight intermediate ESL students were exposed to one of three types of instruction (segmental accuracy, global features (e.g., stress, intonation, rhythm) and no specific program of instruction) for 12 weeks. The results indicated that in the sentence-reading tasks, the segmental and global treatment groups improved from Time 1 to Time 2 in comprehensibility and accentedness. In the narrative task, only the treatment group improved significantly. The researchers concluded "attention to both global and segmental concerns benefits ESL students. The global instruction, however, seems to provide the learner with skills that can be applied in extemporaneous speech production" (p. 407).

Derwing and Rossiter (2003) reanalyzed the data from the Derwing, et al. (1998) study in an effort to determine how the improvements in comprehensibility and fluency reported in the results for the narrative task were manifested in students' oral productions before and after pronunciation instruction. They concluded that if the goal of pronunciation teaching is to help students become more understandable, then instruction should include a stronger emphasis on prosody (p. 14).

While teachers and learners experiment with different techniques to promote measurable gains, one challenge facing learners is the number of teachers lacking formal preparation and training to teach pronunciation. Breitkreutz, Derwing, and Rossiter (2002) surveyed ESL teachers in Canada and found that 67% reported having no training in pronunciation instruction. Derwing and Munro (2005) cite additional studies indicating the lack of prepared pronunciation teachers in other English-speaking countries such as Britain and Australia. This lack of qualified teachers results in a lack of quality pronunciation instruction, suggesting a need for materials that enable learners to direct their own pronunciation learning outside the classroom.

Studies by Hardison (2004) and Pennington and Ellis (2000) have shown that computer technology can help second language learners learn prosodic patterns if the computer tasks focus learners' attention on how prosody works within a piece of discourse. As ESL learners become more aware of how these prosodic features function, they can begin to predict where pauses should occur, which syllables and words should be stressed, and whether intonation should rise or fall at the end of an utterance. This process of perception and prediction, followed by production of prosodic features, is advocated by researchers promoting the use of pronunciation strategies (Hahn & Hahn, 2007; Sardegna & Molle, 2008). Embedding these strategies into computer-assisted materials would benefit learners by allowing them to take control of their own learning and by providing discourse-length contexts in which to practice those prosodic features that improve intelligibility and comprehensibility.

Current Study

This study is intended to empirically evaluate a self-directed, computer-assisted technique that uses oral readings to improve students' perception and production of pausing, word stress, and sentence-final intonation. Oral reading techniques have been recommended by several researchers (Anderson-Hsieh, 1990; Firth, 1992; Ricard, 1986; Walker, 2005) to raise ESL learners' awareness of individual prosodic features. In the techniques they advocate, ESL learners listen to (150-300-word) passages recorded by NSs, mark the location of an individual suprasegmental feature (e.g., pausing or syllable stress), practice the reading orally with the appropriate feature marked, and then record themselves reading the passage.

The oral reading technique studied empirically here is referred to as CPR. This cued pronunciation reading technique differs from previous oral reading techniques discussed or advocated in the pedagogical literature in several key ways. First, the CPR tasks are almost entirely self-directed. Students are given an overview of pausing, word stress, and sentence-final intonation patterns in English. This overview raises their awareness of prosody and gives them the skills necessary to begin predicting the occurrence of prosodic features. Teachers' involvement is minimal because students complete the tasks on their own and teachers provide no feedback on students' recordings. A second difference is that students practice

perceiving and producing multiple suprasegmental features within a single passage instead of separate passages for each feature. A third difference between CPR and other oral reading techniques is that learners see an answer key after each perception task so they can check their perception and prediction of prosodic features before practicing the passage orally with the recorded NS model and recording it themselves.

Since this study's focus was the influence of self-directed readings on learners' perception and production skills, participants voluntarily completed the CPR tasks as extra-credit tasks outside the normal class time. To facilitate the use of CPR tasks in a self-directed context, each CPR was set up as a series of PowerPoint slides. Audio recordings from a NS model were embedded in the PowerPoint slides to let participants listen to them as many times as necessary to complete the listening (perception) activities. The 11 CPRs used in this study covered topics from telescope types to the Empire State Building, to similarities between Abraham Lincoln and John F. Kennedy. Study participants used the audio recording feature in Microsoft Word to record themselves reading the passage and save the recordings as an MP3 file for later analysis.

Research Questions

The following research questions were addressed:

To what extent do cued pronunciation readings practiced in a self-directed context affect intermediate ESL learners'…

1. perception of pausing, word stress, and sentence-final intonation?
2. use of pausing, word stress, and sentence-final intonation in controlled production?
3. perceived comprehensibility in spontaneous speech tasks?

METHOD

Participants

ESL Students. Seventy-five ESL learners enrolled full-time in a university ESL program participated in the study. They were all of intermediate-level proficiency and ranged in age from 17 to 54 (mean age = 25 years). All had spent between one week and two years in the US, with the median length of stay being four months. Participants reported having previously studied English for between two months and 17 years, with the median being four years. The participants' native language backgrounds were organized into three categories: Asian language speakers (e.g., Japanese, Chinese, Korean, n = 36), Romance language speakers (e.g., Spanish Italian, Romanian, n = 34), and other languages (e.g., Haitian Creole, Russian, Armenian, n = 5). Before this study began, the ESL program assigned the ESL participants to one of six intermediate-level classes, balanced as closely as possible for English language proficiency, L1, gender, and length of time in the US. Intermediate-level ESL learners were selected for this study because their normal curriculum involved the use of self-directed computer-based language learning tasks, and this prior experience would enable them to complete the cued pronunciation readings independently.

Native Speaking (NS) Informants. Previous studies researching pronunciation errors (Anderson-Hsieh, Johnson, & Koehler, 1992) have used NS informants to establish baselines for analyzing phonological errors in passages read aloud by NNSs. For this study, the researchers determined that NNS performance on the Time 1 and Time 2 perception and controlled production tasks could only be appropriately measured against NSs' performance of the same tasks. Ten NSs, 5 males and 5 females, all enrolled in a graduate TESOL program, served as informants. All 10 NSs were from the western United States and were recruited because of their knowledge of linguistics and their ability to speak Standard American English.

Teachers. Six intermediate listening/speaking classes were randomly selected for the study, each taught by a different teacher. Three classes were randomly selected as the control group, with the other three classes serving as the treatment group. The six teachers had one to three years of formal teaching experience. Each participating teacher agreed to neither address the specifics of the study in their curriculum, nor provide feedback to their students regarding the CPR tasks.

Listeners. Two listener groups were used. The first group consisted of ten novice NSs (i.e., people not accustomed to listening to or working with NNSs) recruited to evaluate spontaneous speech samples of ESL participants. These ten listeners (five men and five women) ranged in age from 21 to 52 (mean age = 29 years) and reported normal hearing acuity. The first group of listeners provided comprehensibility ratings for all 75 ESL participants. The second group of listeners included two expert judges who were both native English speakers and had advanced degrees in linguistics as well as extensive experience transcribing speech samples from native and non-native speakers. Previous studies (Anderson-Hsieh, et al., 1992; Derwing & Rossiter, 2003) investigating learners' errors in controlled speech tasks have used expert judges to classify segmental or suprasegmental errors to determine whether these errors interfered with comprehensibility. The expert judges in this study independently listened to the perception and production tasks, scoring passages for errors in the suprasegmental features being evaluated (pausing, word stress, sentence-final intonation). The expert judges' inter-rater reliability scores are reported in the results section.

PROCEDURE

Collection of Speech Perception and Production Data

Speech perception and production data at Time 1 and Time 2 were collected from all NNS participants in the ESL program's computer lab. The total test time was 20 minutes and data collection consisted of a sequence of seven computerized tasks: five spontaneous speech tasks, one perception task, and one controlled production task, in that order. The five spontaneous speech elicitation tasks were similar to those used in the Educational Testing Service (ETS) institutional version of the SPEAK (Spoken Proficiency English Assessment Kit) test (1999). In the spontaneous speech tasks, participants told a story based on a sequence of pictures, suggested potential solutions to a problem illustrated in the picture sequence, discussed the advantages and disadvantages of an issue, expressed their opinion on a debatable topic, and explained changes made in a schedule of events. For the perception task, participants listened to a passage recorded by a NS and marked pauses, stressed words, and sentence-final intonation (rising or falling) on a transcript. The controlled production task required participants to simply read a passage aloud. They were given one minute to read the passage silently before reading it aloud. The test was conducted during the participants' listening/speaking class period to avoid placing additional stress on the students. The spontaneous speech tasks used across Time 1 and Time 2 were of a similar type but included different pictures, topics, etcetera to avoid participants remembering the task prompts and practicing over the time period of the study (Derwing & Munro, 1994). Perception and controlled production tasks remained the same so that changes in perception and performance errors could be noted across Time 1 and Time 2. Neither teachers nor students could access testing stimuli during the treatment period.

English Instruction and CPR Use

All six classes in the study were taught using the ESL program's established syllabus and materials. The overall communicative curriculum was skills-based with attention to listening, speaking, reading, writing, and grammar. Because the study was intended to investigate the effects of self-directed computer-assisted pronunciation tasks, teachers neither graded recordings nor provided feedback. Teachers simply reminded students to complete the weekly CPR tasks, which counted as extra-credit for treatment-group classes.

During week two of the study, the treatment group classes received one 65-minute instructional period from one of the researchers. During this session, terms used in the CPRs (e.g., suprasegmental, stress, intonation) were defined and students were taken through the stages necessary to complete a sample CPR. The brief period of instruction was conducted in the computer lab so that students would understand the procedures for marking the task sheet, recording the passage, and saving the recorded file. Students in the treatment group were told to submit to the researchers one completed CPR recording and task sheet each week for the 11 weeks that followed that session. The CPR tasks were organized so that students could complete one portion of each reading on their own during the ten minutes they were allotted each day to work on the task after class hours. Participants were required to complete the tasks in the computer lab to ensure that only the students in the treatment group had access to the readings, with only one reading accessible per week, and so that a lab attendant was available to students if any difficulties arose.

SCORING PROCEDURES

Perception Task

Ten NS informants listened to the same passage presented to the NNS participants and marked the location of pauses, stressed words, and the direction (up or down) for sentence-final intonation. Their markings were pooled to determine agreed-upon locations for these suprasegmental features. Nine pause locations were identified by the NS informants. Thirty-six words were identified as having definite stress. Another 13 words were selected variably by the NSs as being stressed within the discourse; these 13 were labeled in the answer key as optionally stressed, so that whether or not an ESL participant selected one of these words as a stressed word, no error would be counted. NS informants agreed 100% on sentence-final intonation (pitch direction at the end of each of the seven sentences).

Once the answer key was created, the expert judges used the key to independently score the Time 1 and Time 2 perception tasks for each of the ESL participants. The following types of error were counted. An error was counted for pausing and word stress if a feature was *missing* (meaning the participant should have marked the feature but did not) or *incorrect* (the participant did mark the feature but should not have). These two error types constituted the *total* number of errors identified by the participant for each suprasegmental feature. Because the location of intonation markings was fixed, an intonation error was identified if the pitch direction at the end of the sentence was incorrectly marked.

Controlled Production Task

The same 10 NSs who had been used as informants for the perception task were recorded reading aloud the passage created for the controlled-production task. They followed the same procedure as the ESL participants in completing the task. The NSs were allowed to read the passage through silently before recording the passage to let them see the flow of the text, identify any unfamiliar words, and practice the pronunciation of the passage. Once the speech samples were recorded, the two expert judges independently evaluated them for the three suprasegmental features being studied (pausing, word stress, and sentence-final intonation). An answer key was constructed based on the pooled responses of the NSs.

For pausing, there was considerable agreement among the NSs as to locations where pauses should occur in the oral reading. Locations where 8 or more of the NSs paused collectively were labeled as required pauses. Places where 3 to 7 of the native speakers paused collectively were labeled as optional pauses. No pausing error was recorded if the ESL participant failed to pause in the optional locations. NNS pauses in any other location were counted as errors.

Word stress scoring was similar to that of pausing. Seventy-three syllables in the passage were stressed by 8 or more of the NS informants. Any other syllables stressed by the ESL participants were counted as errors. Syllables which participants should have stressed but did not were also counted as errors. There was 100% agreement among the native speakers on pitch direction for the 11 examples of sentence-final

intonation. When an ESL participant's sentence-final intonation rose instead of falling, fell instead of rising, or remained flat, it was counted as an error.

Spontaneous Speech Tasks

The five spontaneous speech tasks were included to evaluate whether the treatment significantly improved participants' level of perceived comprehensibility from Time 1 to Time 2. Ten novice NSs (people not accustomed to listening to or working with NNSs) rated these speech stimuli. Each was assigned to rate 30 students, randomly selected from each class and each time period (Time 1 or Time 2). The NSs listened to at least a 45-second speech sample from each of the five spontaneous speech tasks for each student before assigning a perceived comprehensibility score. The speech samples were recorded onto CDs in order to facilitate the rating process. Each set of spontaneous speech tasks was rated by two different listeners, resulting in four perceived comprehensibility scores for each examinee: two for Time 1 and two for Time 2, each from a different listener. The novice NS listeners used the five-point Likert scale shown in Table 1 in assigning a single perceived comprehensibility score to the spontaneous speech samples.

Table 1. *Descriptors for Perceived Comprehensibility Ratings*

4	**Speaker is very easy to understand.** Little (if any) listener effort is required. Errors (if any) are not distracting.
3	**Speaker is mostly comprehensible.** Listeners can understand with some effort. Errors are occasionally distracting.
2	**Speaker is sometimes comprehensible.** Significant listener effort is required. Errors are often distracting. Words and individual sentence meaning are usually comprehensible. Meaning of the overall recording is incomprehensible.
1	**Speaker is very difficult to understand.** Great listener effort is required. Errors are very distracting. Most words are intelligible, but sentence meaning is often unclear.
0	**Speaker is basically incomprehensible.** Only an occasional word is intelligible.
NR	**Not ratable**

RESULTS

Quantitative Findings

Perception tasks. After the Time 1 and Time 2 perception tasks were scored, the number of each error type for each examinee was entered into a spreadsheet for further analysis. As described previously, errors were counted for pausing and word stress if a feature was *missing* (meaning the participant should have marked the feature but did not) or *incorrect* (the participant did mark the feature but should not have). These error types constituted the *total* number of participants' errors for each suprasegmental feature. Pearson product-moment correlations showed no significant correlations between or within the three suprasegmental categories ($r = 0.43$ at $p > .05$). Therefore, each type of pausing, stress, and intonation error was analyzed separately.

Other comprehensibility studies involving Time 1 and Time 2 speech samples (Derwing et al., 1997, 1998; Derwing & Rossiter, 2003) have used repeated-measures ANOVAs to analyze group performance with treatment as the between factor and time as the within factor. An extension of ANOVA that provides

a statistical means for eliminating the linear effects of a particular variable is called an analysis of covariance (ANCOVA) (Vogt, 1993). Following Vogt (1993), ANCOVAs were run using participants' scores at Time 1 as the covariate to control for learners' varying abilities at the outset of the study. To account for possible Type I Error with running multiple ANCOVAs, the alpha level was adjusted to .01.

ANCOVAs were performed on perception task data using treatment as the independent variable and total scores for three categories of prosodic error (pausing, word stress, and sentence-final intonation) as the dependent variables. Results are given in Table 2. These ANCOVAs indicated that the overall effect of treatment was significant ($p < .01$) for perception of pausing and word stress, but not intonation.

Table 2. *Analyses of Covariance for Perception Task: Summary*

Error Category	Source	*df*	*F*	*p*
Perception of Pausing	Treatment	1, 71	9.07	.004
Perception of Word Stress	Treatment	1, 71	21.63	< .001
Perception of Sentence-final Intonation	Treatment	1, 71	5.14	.027

Additional ANCOVAs were run to identify the effects of treatment on the specific types of pausing and word stress errors (*missing* vs. *incorrectly* placed). Results showed that the treatment group made a significantly greater reduction in *missing* pause marks than the control group ($F(1,71) = 10.07$, $p < .01$), but not in *incorrect* pause marks ($F(1,72) = 0.04$, $p = .843$). For errors in perception of word stress, ANCOVA results showed that *missing* word stress marks were also significantly reduced for the treatment group ($F(1,71) = 33.10$, $p < .01$), but *incorrect* word stress marks were not ($F(1,72) = 0.58$, $p = .450$). These results indicate that the treatment enabled participants to significantly reduce the number of instances where they were unable to perceive pauses and stressed syllables.

Controlled Production Task. Once the expert judges finished independently marking each of the 75 ESL participants' Time 1 and Time 2 controlled production passages, inter-rater reliability scores were obtained. Inter-rater reliability Pearson coefficients (r) were as follows: pausing, .84; stress, .88; and intonation, .98. The few discrepancies that occurred were resolved by listening to each questionable segment while referring to the NS recordings as a baseline. As in the perception task, errors were counted for pausing and word stress in the controlled production task if a feature was *missing* (meaning the participant should have produced the feature but did not) or *incorrect* (the participant did produce the feature but should not have). The total numbers of each error type for each student, as well as their total sentence-final intonation errors, were added to the spreadsheet.

ANCOVAs were performed on the controlled production task data using treatment as the independent variable and scores for the three categories of prosodic error (pausing, word stress, and sentence-final intonation) as the dependent variables. Participants' Time 1 scores were used as a covariate. Results of these analyses, shown in Table 3, indicate that the overall effect of treatment was significant ($p < .01$) for controlled production of word stress. The results for controlled production of pausing and sentence-final intonation, however, were not significant.

Table 3. *Analyses of Covariance for the Controlled Speech Production Task: Summary*

Error Category	Source	*df*	*F*	*p*
Production of Pausing	Treatment	1, 67	2.22	.141
Production of Word Stress	Treatment	1, 67	7.73	.007
Production of Sentence-final Intonation	Treatment	1, 67	0.33	.570

An additional ANCOVA was used to analyze the mean gains in production of word stress to determine which type(s) of error (*missing*, *incorrect*, or both) the treatment had affected. The results revealed that the treatment group participants produced fewer *incorrectly* stressed syllables at Time 2 ($F(1,67) = 7.48$, $p < .01$), indicating that the CPR treatment had improved their ability to use word stress appropriately.

Spontaneous Speech Tasks. A final ANCOVA was run to determine whether treatment significantly affected participants' perceived comprehensibility ratings. In the ANCOVA, the independent variable was treatment, with mean gains in perceived comprehensibility ratings as the dependent variable and Time 1 scores as the covariate. The results of the analysis showed no significant effect of treatment ($F(1,69) = 0.06$, $p = .802$).

To account for possible differences in rater severity between the NS listeners, a FACETS analysis was performed. A FACETS analysis looks at the interaction of multiple facets involved in the rating procedures, including examinee ability, test difficulty, and rater severity. To determine rater consistency, FACETS provides an expected rating for each examinee and compares it to the observed rating. Based on this comparison, it then calculates an infit mean square value for each rater, as well as a mean and standard deviation of these infit values for the group of raters. Two standard deviations from the mean (in both directions) constitute the range of raters who are acceptably consistent. Four facets were included in the analysis for this study: group (treatment vs. control), time (Time 1 vs. Time 2), students, and raters. The FACETS analysis showed that the infit scores for all ten NS listeners were within the allowable limits of 2 standard deviations. (Infit mean values = 0.88, $SD = 0.29$, Range = 0.30 to 1.46). These results indicate that none of the NS listeners were inconsistently severe or lenient in their ratings.

Qualitative Results

Following the completion of the Time 2 testing, all 39 treatment group participants completed a brief follow-up survey regarding their experience using the CPRs. In the survey, students indicated the amount of effort they had put into completing the readings and the level of pronunciation progress they perceived in their speech. They also provided feedback on positive aspects of the readings, difficulties encountered, and recommendations for improving the CPRs. The results from this survey are discussed below. It should be noted that the ESL program administrators recommended that treatment group participants limit their work on the CPRs to ten minutes per day to allow sufficient time to complete their regular homework.

Level of Effort Expended. Of the 39 students who participated in the treatment group, 37 participants (94.8%) agreed or strongly agreed that they could have put more effort into completing the pronunciation readings. The treatment group's performance was assessed collectively, but the overall number of CPRs completed by the treatment participants varied. Nineteen students (48.7%) completed nine or more of the 11 CPRs. The remaining students completed eight or fewer of the readings. Students' reasons for not completing all of the CPRs ranged from simply forgetting to do them to not having time to complete the tasks in the computer lab.

Level of Perceived Pronunciation Progress. The follow-up survey included nine statements designed to assess the level of pronunciation progress treatment participants felt they had made by using the CPRs. Students responded to the statements using a four-point Likert scale that ranged from 1 (*strongly disagree*) to 4 (*strongly agree*). In Table 4, the statements are rank ordered by the percentage of students who agreed or strongly agreed with each statement.

As a whole, treatment group participants were positive about their use of the cued pronunciation readings. While 34 out of the 39 students (87%) recognized that they had more work to do on their pronunciation, they felt they had improved. Thirty-two (82%) of the students felt that as a result of the treatment, they could understand English conversations more easily and they had increased their knowledge of English

pronunciation. Thirty-one participants (79%) felt that because of the CPRs, they could communicate more effectively in situations previously difficult for them, they had more confidence when speaking English in public, and they could speak more fluently and correctly in English. Most treatment participants (64%) also noted that because of the readings, they could more easily recognize their own pronunciation errors.

Positive Aspects of CPRs. Treatment students' qualitative comments were positive about the specific ways in which they benefited from the readings. Six students (15.4%) commented that they liked seeing where the pausing and word stress occurred. For example: "I likes the fact that I had to study the reading with stress and pauses before reading. It helped me to understand better the pronunciation reading. I also liked the PowerPoint presentation. It was cool." Nine students commented that they felt the readings helped them speak clearly and more accurately: "I learned that [it] is more important to speak clear and using pauses than faster."; "I learned how to stress syllables and where to [pause]." and "I learned a lot about correct English pronunciation and I tried to copy it every time I did those exercises. I loved it!" Other positive comments included learning new vocabulary from the readings, developing a greater awareness of suprasegmental features, and utilizing the computer's recording and playback features to identify individual pronunciation errors.

Table 4. *Students' Responses Showing Perceived Level of Progress* (*n* = 39)

Statement	Percentage who Agreed and Strongly Agreed
I feel that I still have more work to do on my pronunciation, but I can see that I have made progress	87% (*n*=34)
I can understand English conversations more easily	85% (*n*=33)
By doing the pronunciation readings, I learned a lot about English pronunciation	82% (*n*=32)
By doing the pronunciation readings, I learned how to speak more fluently and correctly in English	79% (*n*=31)
I have more confidence now when speaking English in public	79% (*n*=31)
I can now communicate more effectively in situations that were difficult for me before	77% (*n*=30)
I feel that people can now understand my speech more easily	77% (*n*=30)
Because of the pronunciation readings, I can work on correcting my pronunciation errors	77% (*n*=30)
Because of the pronunciation readings, I can more easily recognize my own pronunciation errors	64% (*n*=25)

Difficulties with CPRs. Students made 31 comments regarding problems they had experienced with the CPRs. These largely focused on the difficulty of the perception tasks (identifying stressed syllables and understanding the native English speakers' pronunciation of the words in the passages) and having to imitate the pronunciation patterns of the native English speakers (talking fluently, linking words and phrases together with limited pausing, and producing multi-syllabic words). Example comments:

> "Most difficult was to understand where the stresses in the sentences were"
>
> "It's difficult sometimes when the native speaker speak very fast and I couldn't understand the pronunciation of each words"
>
> "Long words and difficult pronunciation vocabularies"

"Linked some words or phrases as native speakers do"

Recommendations for Change. Fifteen students (38.5%) identified time constraints as a concern for them in completing the readings, since they were already spending several hours a day studying English. Two of them suggested that the readings be made available for students to do at home, where it was more convenient for them to complete the tasks, rather than being required to do the tasks in the computer lab. Four students (10.3%) recommended receiving specific feedback on their final recordings to let them know what suprasegmental mistakes they continued to make. Four students (10.3%) recommended selecting more interesting topics for the readings, but they did not provide any suggestions for alternative topics.

DISCUSSION

Results from this quasi-experimental study showed that treatment had a significant effect ($p < .01$) on perception of pausing, perception of word stress, and controlled production of word stress. Further investigation of improvement in pause perception revealed that *missing* pause marks decreased significantly for the treatment group participants at Time 2, indicating that their awareness of appropriate pausing had increased and they had become better able to hear and correctly identify NS pause locations. The cued pronunciation readings also had a significant effect on the participants' ability to correctly perceive stressed syllables, specifically shown by a significant reduction in *missing* stress marks. In the controlled production task, the treatment participants also had significantly fewer instances of *incorrectly* placed stress.

The final question addressed in this study was whether treatment significantly affected ESL participants' level of perceived comprehensibility. This question is important because previous research (Derwing et al., 1997, 1998; Derwing & Rossiter, 2003) has shown that treatment can have a short-term effect on performance directly related to the type of treatment participants received. Results from the current study showed that there was no significant change in the learners' level of perceived comprehensibility at Time 2. The lack of such a finding may be due to the length of the treatment (11 weeks), the level of sustained effort put forth by the treatment group participants (a maximum of 10 minutes per day over the 11 week study), or the students' level of motivation to participate. Given these factors and others that the researchers may not be aware of, it is important to note that there still were areas in which treatment group participants improved significantly more than students in the control group.

Qualitative data gathered from treatment group participants provided additional insights into the learners' use of CPRs. They noted that the readings helped them learn a great deal about English pronunciation, learn to speak English more fluently and correctly, have more self-confidence when speaking in public or in situations difficult for them, and feel that people could understand their speech more easily.

Implications

The findings from this study provide further empirical evidence strengthening claims about the pedagogical use of oral reading techniques for pronunciation improvement. NNSs in the treatment group significantly improved their perception of pausing and word stress and their production of word stress with limited exposure to the available treatment (11 weeks, 10 min. per day) by using self-directed, computer-assisted cued pronunciation readings. It might be argued that participants in the treatment group merely improved their familiarity with the procedure and format for marking prosodic features. However, the fact that their markings were more correctly placed on the post-test indicates that not only were treatment group participants better able to perceive the requisite features, but their ability to predict where those features should occur also improved. It is important to remember that the treatment participants received no feedback other than the answer keys provided in the PowerPoint slides. For language instructors who do not feel comfortable teaching pronunciation or who cannot fit it into their curriculum,

self-directed, computer-assisted cued pronunciation readings can provide an effective way to help students improve their ability to perceive, predict, and produce prosodic features outside of class.

A second implication is that as in other areas of language acquisition, for many learners, the ability to perceive suprasegmental features may precede the ability to correctly produce them. This implication stems from the fact that the CPR treatment had a significant effect on two areas of prosodic perception (pausing and stress), but only on one area of controlled production (stress), and no significant effect on perceived comprehensibility. This implication supports previous pedagogical advice suggesting that students should be empowered with predictive and perceptive skills necessary to improve production (Dickerson, 1995). In turn, perception and controlled production may precede improvements in learners' spontaneous speech.

A third implication is based on recommendations for change given by the treatment participants. Students may be more motivated to complete self-directed, computer-assisted pronunciation tasks if the tasks are available for them to do at a time and in a location of their choosing. In this study, participants were required to complete the tasks in the ESL program's computer lab to control access to the readings and so that a lab attendant was available for assistance.

Limitations

Although this study was carefully planned and implemented, some important limitations should be noted. First, the ESL program placed heavy time constraints on the study, as mentioned earlier. A longer, more intensive treatment might have had a more significant impact on the learners' performance.

A second limitation involves the issue of feedback. Since the readings were self-directed, the participants received no teacher feedback on their weekly markings or recordings. Four students expressed that they might have benefited more from the readings if they had received specific feedback on their progress (or lack thereof).

A third limitation is the varying level of effort put forth by the treatment group participants. As discussed previously, only half of the treatment group participants completed nine or more readings. Participants who completed fewer readings (particularly the five students who did fewer than three CPRs) may have contributed to the limited number of significant differences that occurred between the treatment and control groups.

A final limitation involves the length of the tasks used at Time 1 and Time 2. To avoid learner fatigue while providing sufficient discourse for scoring prosodic errors, the perception and controlled-production paragraphs were fairly short (seven and 11 sentences respectively), particularly with regard to the number of tokens available for testing sentence-final intonation. A passage with more sentences and more varied pitch direction would provide additional tokens for assessing participants' perception and production of suprasegmentals.

CONCLUSION

The quantitative analysis of the pre-test and post-test results revealed that treatment group participants made significant gains ($p < .01$) in three areas: perception of pausing, perception of word stress, and controlled production of word stress. The treatment had a significant positive effect on the participants' total pause marking errors, specifically on the number of missing pause marks. For perception of stressed words, the treatment had a significant positive effect on the total number of stress marking errors, specifically in the area of missing stress marks. Finally, treatment had a significant effect on the total number of stress placement errors produced in the controlled production task. More specifically, participants had fewer instances of stressing syllables that should not receive stress.

Suggestions for Future Research

With several popular pronunciation textbooks (Grant, 2001; Meyers & Holt, 1998; Miller, 2006) advocating the use of imitative oral readings as a pronunciation improvement technique, further empirical research needs to done. This research will help identify what learner variables (e.g., students' proficiency level, level of motivation, attitude toward the host culture) and pedagogical variables (e.g., time on task, duration of practice, teacher feedback) most influence second language learners' ability to effectively use these types of pronunciation practice activities both inside and outside the classroom.

The results of this study provide empirical evidence supporting other researchers' claims (Hardison, 2004; Pennington & Ellis, 2000) that computer technology can help second language learners more accurately perceive and produce prosodic features. In the current study, self-directed computer-assisted CPRs provided treatment group participants an opportunity to develop their perception abilities (in pausing and word stress) and production abilities (in word stress). These findings, however, suggest that further work is needed to explore the effects that CPRs can have on language learners' perception and production of specific suprasegmental features. What impact might CPRs have if the readings were integrated into the ESL curriculum? How might periodic feedback from teachers, as called for by participants in this study's treatment group, affect learners' production of suprasegmental features? Could regular teacher feedback focused on specific errors produced by ESL learners in the CPR tasks result in significant changes in the learners' level of perceived comprehensibility? Could learners' use of CPRs over an extended time period (longer than 11 weeks) result in greater gains in the perception and production of suprasegmentals as well as improvement in ratings of perceived comprehensibility? Clearly, several issues still need to be addressed.

ACKNOWLEDGMENTS

We are grateful to the ESL learners and native English speaker raters whose participation made this study possible. We are also grateful to the three anonymous reviewers and journal editors for their helpful feedback on earlier versions of this manuscript.

ABOUT THE AUTHORS

Melissa Landon holds an MA in TESOL from Brigham Young University. She is currently a full-time mother and part-time researcher. Her research interests include assessment and pronunciation instruction.

E-mail: melisma37@yahoo.com

Mark Tanner is an Assistant Professor in the Linguistics and English Language Department at Brigham Young University. His research interests include pronunciation pedagogy and comprehensibility research, self-directed learning, and second language teacher education.

E-mail: mark_tanner@byu.edu

REFERENCES

Anderson-Hsieh, J. (1990). Teaching suprasegmentals to international teaching assistants using field-specific materials. *English for Specific Purposes, 9*, 195-214.

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

Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal, 11*(4), 5-22.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42,* 529-555.

Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly, 24*, 746-752.

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2002). Pronunciation teaching practices in Canada. *TESL Canada Journal, 19,* 51-61.

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

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

Derwing, T. M., Munro, M. J., & Weibe, G. E. (1997). Pronunciation instruction for fossilized learners: Can it help? *Applied Language Learning, 8,* 217-235.

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

Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning, 13*, 1-17.

Dickerson, L. (1995). Autonomy and motivation: A literature review. *System, 23,* 165-174.

Fayer, J., & Krasinski, E. (1995). Perception of hesitation in nonnative speech. *Bilingual Review, 20*, 114-121.

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

Firth, S. (1992). Pronunciation syllabus design: A question of focus. In P. Avery & S. Ehrlich (Eds.), *Teaching American English Pronunciation* (pp. 173-183). Oxford: Oxford University Press.

Goodwin, J. M. (2004). *Imitative conversations*. Demonstration presented at the international TESOL Conference, Miami, FL.

Graham, C. (1978). *Jazz chants: Rhythms of American English for students of English as a second language*. New York: Oxford University Press.

Grant, L. (2001). *Well said: Pronunciation for clear communication (2nd Edition)*. Boston, MA: Heinle & Heinle.

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

Hahn, L. D., & Hahn, K. (2007). *Phrase stress essentials*. Paper presented at the international TESOL conference, Seattle, WA.

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quanitative and qualitative findings. *Language Learning & Technology, 8*, 34-52. Retrieved from http://llt.msu.edu

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics, 24*, 109-125.

Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly, 33*, 37-63.

Levis, J. M. (2002). Reconsidering low-rising intonation in American English. *Applied Linguistics 23*, 56-82.

Levis, J. M., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System, 32*, 505-524.

McNerney, M., & Mendelsohn, D. (1992). Suprasegmentals in the pronunciation class: Setting priorities. In P. Avery & S. Ehrlich (Eds.), *Teaching American English pronunciation* (pp. 185-196). Oxford: Oxford University Press.

Meyers, C., & Holt, S. (1998). *Pronunciation for success: Student workbook*. Burnsville, MN: Aspen Productions.

Miller, S. F. (2006). *Targeting pronunciation: Communicating clearly in English (2nd Edition)*. New York: Houghton Mifflin.

Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly, 22*, 91-111.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal, 77*, 330-339.

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning, 12*, 427-440.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic clues. *The Modern Language Journal, 84*, 372-389.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes, 23*, 19-43.

Ricard, E. (1986). Beyond fossilization: A course in strategies and techniques in pronunciation for advanced adult learners. *TESL Canada Journal, Special Issue 1*, 243-253.

Richman, B. (2005). *Songs for Speaking: Lyrics for Pronunciation*. Paper presented at the Rocky-Mountain TESOL Conference, Salt Lake City, UT.

Sardegna, V., & Molle, D. (2008). *Empowering students with pronunciation learning strategies*. Demonstration given at the international TESOL conference, New York, NY.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*, 84-119.

Vogt, W. P. (1993). *Dictionary of statistics and metholodogy: A non-technical guide for the social sciences.* Newbury Park, CA: Sage Publications.

Walker, R. (2005). Using student-produced recordings with monolingual groups to provide effective individualized pronunciation practice. *TESOL Quarterly, 39*, 550-58.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition, 20*, 1-25.