EFFECTS OF PRODUCTION TRAINING WITH VISUAL FEEDBACK ON THE ACQUISITION OF JAPANESE PITCH AND DURATIONAL CONTRASTS

Yukari Hirata

Department of East Asian Languages and Literatures
Colgate University
yhirata@mail.colgate.edu

ABSTRACT

This study examined the effects of production training with visual feedback on second language (L2) learners' production and perception development. Native speakers of English learning intermediate Japanese practiced producing words and sentences with the CSL-Pitch Program, during which time they tried to match the pitch and durational contours of their own production to those of the models on the computer screen. Their abilities to produce and perceive novel words and sentences including both pitch and durational contrasts were tested in citation and sentence contexts, and were compared to those of a control group. The results indicated that (1) this production training helped the trained subjects to both produce and perceive the pitch and durational contrasts, and that (2) the trained subjects' abilities in the two domains were greatly enhanced in the sentence context as well as in the citation context.

1. INTRODUCTION

A variety of computer-assisted perceptual training programs for L2 learners’ acquisition of difficult nonnative contrasts have been developed and tested for their efficacy (see [1] for a review). A series of studies, e.g., [2] and [3], have shown that perceptual training with stimuli varying in talkers and phonetic contexts has positive effects on the Japanese learner’s acquisition of English liquid contrasts. These findings were further extended in [4] to show that providing variations in length and rate of speech was also effective in the acquisition of difficult L2 contrasts. However, most training programs have examined the effects of perceptual training, except for a few studies, e.g., [5], and little is known as to whether similar effects of stimulus variability can be obtained with production training. The present study, therefore, examined the effects of production training that incorporated these previous findings on stimulus variability.

The present experiment utilized Kay Elemetrics’ CSL-Pitch Program, which enables one to see real-time representations of spoken utterances on a computer screen (Figure 1), and to compare a learner’s speech with that of a model. It immediately exhibits the utterances’ fundamental frequency contours, which correspond to our perception of pitch. However, the acoustics exhibited onscreen are not easily interpretable by non-phoneticians, e.g., the word boundaries do not correspond to breaks in the fundamental frequency lines. It is also not clear to...
the learner how steep a rise or a fall is significant or ignorable. To address this interpretation problem, prosody graphs were created as supplementary materials throughout training (Figure 1, see also 2.2).

A trained and a control group’s abilities were measured by a pre-test and a post-test in both production and perception domains. The testing methods were different from the training method, i.e., testing without CSL-Pitch Program (see 2.3). Furthermore, in each test of production and perception, their abilities were measured in citation and sentence contexts. It was found in [6] that, for the acquisition of Japanese durational contrasts, perceptual word training was effective only in the citation context, but that perceptual sentence training was effective in both the citation and the sentence contexts. The present study examined the efficacy of the production training that provided model speech in both word and sentence contexts, and examined how this training affected the trained subjects’ ability in the two contexts.

The questions addressed were: (1) Is the production training using visual feedback effective overall in acquiring Japanese pitch and duration contrasts even though the testing methods are different from the training task? (2) Does this production training have effects on both production and perception? (3) Given this production training that places emphasis on sentence-level practice, do the trained subjects improve their ability in both word and sentence contexts?

2. METHODS

2.1 Subjects

Subjects were 8 native speakers of English who were taking the intermediate Japanese course at the University of Chicago. Their age ranged from 18 to 21 years old. The subjects had exposure to spoken Japanese by listening to their instructors and audio tapes that were required for the present and the past (introductory) courses, and had experience in traveling in Japan for no more than two months. Half of the subjects participated in the pre-test, training, and the post-test, and the other half only participated in the pre- and the post-tests, serving as a control.

2.2 Training Procedure

Training materials included pairs and triplets of words contrasting in pitch or duration, or in both, e.g., ame vs. ame, tori vs. toori, and isho vs. isho vs. issho. Out of 10 training sessions, only sessions 1-3 provided words in isolation. Session 4 provided short phrases, and the rest of the 6 sessions provided sentences of different length. These utterances were recorded by 4 native speakers of Japanese, 2 males and 2 females. Each of them spoke at a rate comfortable for him or her, which resulted in variations in speech rates. This added difficulty for the learners in interpreting the fundamental frequency contours, but this treatment was intentional to test whether subjects effectively learn from the exposure to variations in voice, length, and rate of speech. To compensate for this difficulty, however, supplementary handouts were created.

A handout was made for each of the 10 sessions, explaining how to pronounce the utterances with proper pitch control using prosody graphs (Figure 1). A prosody graph is a schematic description of how we perceive fundamental frequency contours, and corresponds to our knowledge of pitch height, words, and morae or rhythmic beats [7]. Circles represent each mora (or beat), and ovals represent bimoraic syllables such as long vowels. The length of the vertical lines indicates perceived relative pitch height, and the horizontal lines at the bottom correspond to words or phrases.

Each session had a particular goal for subjects to focus on. Session 1 stated criteria for judging the goodness of the subjects’ productions. Variations resulting naturally from different speakers, speech rates and sentence lengths were systematically introduced and explained.

The subjects came individually to the Language Labs of the University of Chicago to practice with the CSL-Pitch Program. In each of 10 training sessions, subjects did the following: 1) read the handout, 2) opened each model audio file, 3) listened to the model tokens and watched their pitch patterns in real-time, 4) produced them on another empty window, 5) overlaid the model pitch contour onto their own pitch contour, and 6) repeated the procedure, producing the same words until their pitch contours matched those of the models.

Each session took about 30 minutes. The 10 training sessions were completed in three and a half weeks.

2.3 Testing Procedure

In order to assess the subjects’ changing abilities, production and perception tests were conducted before and after the training. For production tests, 21 words including pitch and duration contrasts were chosen (9 words practiced in training and 12 novel words), e.g., toki, toku, tokku, and tooku (‘to solve,’ ‘advantage,’ ‘a long time before,’ and ‘far away’). The same sets of words were used in the pre- and the post-tests. They were written on index cards using Japanese orthography, and the cards were randomized. A sentence frame was also written on a card. Subjects first repeated each word in the stack after hearing it spoken by the examiner. Then, their production of the 21 words was recorded three times in isolation (word context), and three times in the sentence frame (sentence context). Watashi wa ___ ga ienakatta ‘I could not say ___.’ In each repetition, the words were presented in a different random order.

For the evaluation of the subjects’ production, two native speakers identified all the utterances produced by subjects. For example, in the word set, toki, toku, tokku, and tooku, which were randomly ordered among the 21 test words and recorded by subjects, the native speakers were asked which of the four words they thought the subjects intended to say. The number of the evaluators’ identifications that matched the subjects’ intent was then calculated.

For perceptual tests, there were 30 words, which had not appeared in training, and 30 sentences using those words in each of the pre- and the post-tests. The materials recorded by two female native speakers of Japanese were digitized at 48 kHz on an SGI computer.

For the word perceptual tests, subjects first clicked the play button on the computer screen to listen to the word, and then clicked one of the 9 pitch patterns (HL, LH, HLL, LHL, LHH, HLLL, LHLH, LHHH; H=high, L=low) of prosody graphs that matched the word they had just heard. This method required the ability to identify not only the pitch but also the duration of the words. For example, for the word katee (3 morae) ‘family,’ the correct answer is LHH (representing 3 morae), and not LH, (representing 2 morae).

For the sentence perceptual tests, a frame sentence was written on the computer screen, e.g., Sokowa ___ to yonde kudasai, ‘Please read that as ___.’ When subjects clicked the
play button, they heard a sentence that included the target word in the underlined location, e.g., Sokowa chizu to yonde kudasai. Then the subjects were asked to choose the pitch pattern that matched the target word, e.g., chizu. The program in Tcl/Tk used in [6] was modified to execute this procedure.

2.4 Analyses
A four-factor analysis of variance (ANOVA) was conducted with the overall data. Factors were Group (trained vs. control), Test (pre-test vs. post-test), Context (word vs. sentence), and Domain (production vs. perception). Test, Context, and Domain were repeated measures factors. Regarding the first question addressed in the introduction, if the training had effects on the subjects’ overall abilities to produce and perceive pitch and durational contrasts, we would expect a Group x Test interaction. For the second question, if the training had effects on both production and perception, we would expect a Group x Test interaction but no Group x Test x Domain interaction. That is, the same pattern of a Group x Test interaction should be found in both production and perception domains. Finally, to examine the third question on the effects of training in the word and the sentence contexts, the subjects’ production and perception performance was examined separately in the two contexts.

3. RESULTS

3.1 Overall effects of training
A four-factor ANOVA with Group, Test, Context, and Domain revealed a significant Group x Test interaction [F(1, 6)=11.972, p=.013], indicating that the trained and the control groups showed different amounts of improvement from the pre-test to the post-test. The average percent correct pre-test scores were 52.4% for the trained and 52.3% for the control groups, and the difference between these scores was not significant [t(6)=0.008, p=.190]. However, their post-test scores were 69.5% for the trained and 57.3% for the control groups, and the difference between these scores was significant [t(6)=1.364, p=.003]. These results indicate that the training program was effective for improving the trained subjects’ overall ability to produce and perceive Japanese pitch and durational contrasts.

A main effect of Test was found [F(1, 6)=39.575, p=.001]. The average percent correct scores were 52.4% for the pre-test and 63.4% for the post-test. A main effect of Context was also found [F(1, 6)=23.422, p=.003]. The average percent correct score in the word context was 61.9%, higher than 53.8% in the sentence context. This indicates that even if one can perceive or produce a word in isolation correctly, it does not guarantee that the same word would be perceived or produced in the sentence context. A main effect was also found for Domain [F(1, 6)=143.190, p<.001]. The average percent correct scores were 44.0% for the perception domain, and 71.8% for the production domain. This difference was due to the different methods used to obtain the data in the perception and the production tests.

3.2 Training effects on production and perception
The above four-factor ANOVA with Group, Test, Context, and Domain revealed no interaction among Group, Test, and Domain [F(1, 6)=.034, p=.860]. This assures that the Group x Test interactions in the two domains were comparable to the Group x Test interaction reported in the previous section. Figure 2 shows the percent correct scores of the trained vs. the control groups in each domain. For both perception and production, the trained group showed greater improvement from the pre- to the post-tests than the control group, reflecting the Group x Test interaction reported in 3.1. For perception, the difference between the pre-test scores of the two groups was 0.9%, and this difference was not significant [t(6)=.084, p=.101]. However, the difference between the post-test scores of the two groups was 12.1%, and this difference was significant [t(6)=1.453, p=.004]. Comparable results were found for production. The difference between the pre-test scores of the two groups was 1.0%, and this difference was not significant [t(6)=.091, p=.580]. However, the difference between the post-test scores of the two groups was 12.4%, and this difference was significant [t(6)=1.209, p=.044].

In summary, the trained subjects significantly improved their ability not only to produce but also to perceive the pitch and durational contrasts.

3.3 Training effects on word vs. sentence contexts
Next is the examination of whether the trained subjects improved their abilities in both the word and the sentence contexts. The four-factor ANOVA with Group, Test, Context, and Domain revealed a marginal interaction among Group,
than in the sentence contexts. In this figure, the subjects’ production and perception scores in each context were combined and averaged. In the sentence context, the trained group showed a significant improvement: the difference between the pre- and the post-tests for the trained group was 21.4%, and this difference was significant \( t(3)=4.949, p=.016 \). In contrast, the control group showed no significant improvement \( t(3)=.006, p=.996 \): the difference between the pre- and the post-tests for the trained group was 0%. In the word context, although the improvement of the trained and the control groups looks parallel in Figure 3, the trained group’s improvement (12.8% from the pre- and the post-tests) reached significance \( t(3)=3.326, p=.045 \) while the control group’s improvement (10%) did not reach significance \( t(3)=2.132, p=.123 \).

In summary, the trained groups showed significant improvement in the two contexts, but they improved more drastically in the sentence than in the word contexts. In contrast, the control group’s improvement was significant in neither context, although they showed greater improvement in the word than in the sentence contexts.

4. CONCLUSIONS

There were three major findings in the present experiment. First, the production training using visual feedback was found to be effective for the acquisition of Japanese pitch and durational contrasts (3.1). The training provided stimuli that varied in voice, length, and rate. The focus of the training was for subjects to match the fundamental frequency contours of these training materials to those of the models, and there was no explicit comparison of two halves of minimal pairs in 6 out of 10 sessions. In those 6 sessions minimal pairs were only implicitly used in sentences, e.g., kawai vs. kawaii and kau vs. kau in Figure 1. This training resulted in the increased intelligibility of the subjects’ production as evaluated by Japanese native speakers. Furthermore, this training resulted in the increased ability to identify the pitch and durational patterns of the words that they had not practiced in training.

The trained subjects’ significant improvement indicates that stimulus variability has positive effects not only in perceptual training as previously found [2] and [3], but also in production training as in the present experiment.

The second finding was that the production training had effects on both production and perception (3.2). It is likely that subjects necessarily learned to perceive the training materials correctly in order to match their production to the model, and that this production training, in fact, involves implicit perceptual training. There is controversy about whether perceptual training leads to production improvement rather than the other way around [8]. The answer to this question might not be one or the other, and perception and production trainings might rather be in a mutually enhancing relationship in the course of L2 speech development.

Finally, the present training method using both word and sentences was found to have an effect on the trained subjects’ acquisition of the durational and pitch contrasts in both citation and sentence contexts (3.3), extending the finding by [6]. The control group showed greater improvement in the word than in the sentence contexts, although neither one was significant. In contrast, the trained subjects made robust improvement particularly in the sentence context while they also made significant but less improvement in the word context. It will be of interest to examine in the future which specific aspects of sentence training contribute to these positive effects, e.g., uncontrolled speech rates, or different lengths of stimuli.

5. ACKNOWLEDGMENT

This experiment was conducted using the facilities of the Language Laboratories and Archives of the University of Chicago, supported by Consortium for Language Teaching and Learning. I especially thank Karen L. Landahl and Barbara Need of the University of Chicago and Jack Dovidio of Colgate University for their assistance and encouragement. Thanks also go to Jon Bernard for creating and implementing the Tcl/TK program for the experiment.

6. REFERENCES


