Training Native English Speakers to Identify Japanese Vowel Length with Fast Rate Sentences

Yukari Hirata,[1] Elizabeth Whitehurst,[2] Emily Cullings,[3] Jacob Whiton,[4] and Carol Glenn[5]

[1, 2, 3, 4] Department of East Asian Languages and Literatures
Colgate University, Hamilton, NY, U.S.A.
[5] Department of Linguistics
Cornell University, Ithaca, NY, U.S.A.
yhirata@mail.colgate.edu

ABSTRACT

Native English speakers were trained to identify Japanese vowel length with sentences spoken at a fast rate, and their test scores before and after training were compared with those of a control group. Results indicated that the present training with little variability in speaking rate did not yield a robust perceptual improvement for the trained group. However, among the three rates tested (slow, normal, fast), the trained group showed significant improvement specifically for the slow-rate stimuli, distinct from the control group's.

Index Terms: Japanese, vowel length, speaking rate, perception, training

1. INTRODUCTION

This study examined the effect of speaking rate of sentences on native English (NE) speakers' ability to identify Japanese vowel length. Vowel length (short versus long) is phonemic in Japanese and known to be difficult for NE speakers to perceive [1, 2, 3]. Long vowels are 2.2-3.2 times longer in duration than short vowels within a given rate [4, 5, 6]. Only small differences have been observed between the formant frequencies of short and long vowels [5, 6, 7]. A long vowel is longer in duration than a short vowel by definition, but a long vowel spoken quickly can be shorter than a short vowel spoken slowly [8]. This durational overlap across speaking rates does not cause a problem for native Japanese speakers' perception as they use the surrounding speech context to normalize speaking rate and identify vowel length accurately [9].

Recent studies show that NE speakers' identification of Japanese vowel length is affected by the speaking rate of stimuli [10, 11, 12]. These studies showed that NE speakers' identification accuracy was lower when stimuli were spoken at a fast rate. The error analyses in [11, 12] showed that the NE speakers' identification accuracy was biased by the absolute duration of vowels: they showed a tendency to identify vowels with longer absolute duration as long, and vowels with shorter absolute duration as short, without regard to the surrounding speaking rates. It is not well known how learners of Japanese can cope with speaking rate variations and learn to distinguish vowel length accurately.

In training nonnative speakers to perceive difficult second language contrasts, Pisoni and Lively [13] showed that stimulus variability in speakers and phonetic contexts during training has a positive effect on nonnative speakers' perceptual learning. Hirata [14] showed that training with varied stimuli including different carrier sentences, words, and speakers improved NE speakers' perception of Japanese length contrasts. However, the specific factors in sentence training that influence the perceptual improvement in [14] are still unclear. Given that NE speakers perceived speech materials spoken at a fast rate with low accuracy [10, 11, 12], we questioned how their perceptual ability improves with fast rate sentence training.

Thus, this study focused on the effect of training with fast rate sentences on the identification of Japanese vowel length across different rates. Three questions were addressed. First, how do NE speakers perform in identifying Japanese vowel length with three speaking rates (slow, normal, fast)? Second, is training with only a fast rate effective in identifying stimuli across three speaking rates? Following Pisoni and Lively's [13] variability theory, little rate variability in the present training stimuli only with a fast speaking rate might not improve subjects' overall perceptual learning. However, if single rate training is sufficient, a group trained with fast rate sentences would improve distinctly from a control group in overall perceptual ability. Third, does fast rate sentence training improve subjects' performance only on fast rate test stimuli, and/or does the effect of fast rate training generalize to other speaking rates?

2. METHOD

Participants were assigned to one of two groups, training or control. The training group participated in a pretest, four sessions of perceptual training at a fast rate, and a post-test. Over the same time interval, the control group took only a pretest and a post-test. Each participant listened to a series of Japanese sentences through a computer using headphones, and was asked to identify whether the second vowel of a disyllable in the middle of a carrier sentence was short or long. The carrier sentences were written in Romanized letters with a blank for the location of the spoken target word, e.g., "Soko wa ___ to kaite arimasu."

2.1 Participants

There were 15 participants in the training group and 14 participants in the control group. The participants were monolingual native speakers of American English (ages: 18-23), and had little to no prior exposure to spoken Japanese. One had traveled to Japan for three weeks, but had not studied the language. One had watched a few anime movies in Japanese. One had participated in another perceptual study of Japanese speech for less than an hour. All participants had studied at least one foreign language other than Japanese, and had an average of 6 years of study in that language, although the time span ranged from eight weeks to twelve years. The most common languages studied were Spanish, French and Latin. None of the participants had traveled in a non-English speaking country for more than two months prior to the age of ten. All participants reported normal hearing. Participants were paid for completing the experiment.

2.2 Training procedure

Participants assigned to the fast-only training participated in four training sessions in addition to the pretest and post-test. Participants completed the experiment over a minimum period of 11 days and a maximum of 17 days. Participants had any two consecutive training sessions at least 24 hours apart, and no more than four days apart.

The target words in training were nonsense Japanese words in the form of /mVmV/ and /mVmVV/ (V=/a, e, i, o, u/, e.g., /mama/ vs. /mama:/), with the pitch accent always on the first vowel. These words were embedded in a single carrier sentence spoken by four native Japanese speakers (Male 1, Female 1, Male 2, and Female 2; Table 1) who were instructed to speak as fast as possible.

In training, each of the four sessions contained 160 trials (5 vowels x 2 lengths x 2 repetitions x 8 blocks). Each session was broken into eight blocks and each block consisted of 20 trials. At the beginning of each block, the participant heard three examples, and then saw the answer, either "short" or "long," written on the screen. Then they could continue into training. The participants were asked to take a three-minute break after the fourth block.

Participants received feedback on each response during training. The feedback consisted of a screen with the correct answer written as "long" or "short" and the target word spelled out in Romanized letters. Words with short vowels (containing two moras) had two dots over them, and words with a long vowel (containing three moras) had three dots. When participants answered correctly, the sign "Correct" appeared on the screen and it enabled the participants to go on to the next trial. When participants answered incorrectly, the sign "Sorry…" appeared, and they were made to click "Play again" to hear the stimulus repeated three additional times.

2.3 Testing procedure

For the pretest and post-test, there were a total of 180 stimuli each. Five pairs of real Japanese words were used in testing, i.e., /rubi/-/rubi:/, /ise/-/ise:/, /rika/-/rika:/, /kato/ vs. /kato:/ and /saju/-/saju:/, with the pitch accent always on the first vowel. The stimuli were recorded by two native Japanese speakers (Male 3 and Female 3; Table 1). Both speakers recorded the words in three different carrier sentences for both pretest and post-test. Each sentence was spoken at slow, normal and fast rates. When the stimuli were recorded, speakers were given the following definition of the speaking rates [15]: "tempo that is relaxed and comfortable" for the normal rate, "slowest tempo possible while keeping the sentence flowing together without obviously inserting breaks between words" for the slow rate, and "fastest tempo possible without making errors" for the fast rate. In summary, the pretest and post-test each included 5 vowels x 2 lengths x 3 rates x 3 carrier sentences x 2 speakers for a total of 180 trials.

Each test was broken into six blocks and each block consisted of 30 trials. In each block, there was only one carrier sentence spoken by only one speaker (Table 1). The stimuli were presented in a random order across rates. No words, sentences, or speakers from training were used in tests. The pretest and the post-test had identical words but different carrier sentences. At the beginning of each block, the participant were given two examples with the same nonsense words used in

Table 1: Training and Testing Stimuli

Block	Carrier Sentence	Rate
Training		
Session 1 (a)-(h) Male 1	Soko wa ____ to kaite arimasu.	fff*
Session 2 (a)-(h) Female 1	Soko wa ____ to kaite arimasu.	fff
Session 3 (a)-(h) Male 2	Soko wa ____ to kaite arimasu.	fff
Session 4 (a)-(h) Female 2	Soko wa ____ to kaite arimasu.	fff
Pretest		
(a) Male 3	Sore ga ____ da to omoimasu.	snf**
(b) Female 3	Asoko ni ____ to arimasu.	snf
(c) Male 3	Koko wa ____ ja arimasen.	snf
(d) Female 3	Hontoo ni ____ wa kaitenai.	snf
(e) Male 3	Soko de ____ to wa iwanakatta.	snf
(f) Female 3	Kitto ____ de wa nai deshoo.	snf
Post-test		
(a) Male 3	Kore ga ____ da to kikimashita.	snf
(b) Female 3	Koko ni ____ to arimasu ne.	snf
(c) Male 3	Are wa ____ ja nai desu yo.	snf
(d) Female 3	Zettai ni ____ wa kakareteta.	snf
(e) Male 3	Soshite ____ to itte kudasai.	snf
(f) Female 3	Tabun ____ de wa arimasen.	snf

* "fff" indicates that only fast rate stimuli are presented within the blocks.
** "snf" indicates that stimuli of slow, normal, and fast rates are presented randomly within the block.
(a)-(h) indicate blocks.

training (e.g., /mama/ and /meme:/), and then given the answer, either "short" or "long," written on the screen. The participants were asked to take a three-minute break after the first three blocks.

The procedure of testing was the same as that of training, except that participants did not receive any feedback on their responses.

2.4 Analysis

A three-way analysis of variance was conducted with the percent correct test scores. Factors were group (trained vs. control), test (pretest vs. post-test), and rate (slow vs. normal vs. fast). Group was a between-subjects factor, and Test and Rate were within-subjects factors. *Post hoc* tests were conducted with Bonferroni correction. Regarding the first question addressed in the introduction, if the participants' perceptual ability differs across different speaking rates, we would expect a main effect of rate. For the second question, if the training with only a fast rate is effective for the overall test scores, we would expect a significant group x test interaction. Finally, the third question was whether the fast-rate training improves only on fast rate stimuli or generalizes to other speaking rates. If the improvement due to fast-rate training depends on the speaking rate of test stimuli, we would expect a significant group x test x rate interaction.

3. RESULTS

3.1 Overall effect of speaking rates

With regard to our first question, a significant effect of rate was found [$F(2,54) = 23.36$, $p < 0.001$]: the percent correct test score averaged across test and group was higher for slower rates. The average score for the slow rate score (72.0%) was significantly higher than the score for the normal rate (68.3%) [$p = 0.017$], and the normal rate score was significantly higher than the fast rate score (63.4%) [$p = 0.001$]. A test x rate interaction was also significant [$F(2,54) = 3.18$, $p = 0.049$], indicating that scores averaged across group improved more for slower rates. The scores significantly improved from the pretest (68.5%) to the post-test (75.7%) for the slow rate (7.5% improvement) [$p < 0.001$]. The improvement for the normal rate (pre vs. post = 64.1 vs. 72.6%; improvement of 8.5%) was also significant [$p < 0.001$]. However, the scores did not significantly improve for the fast rate (pre vs. post = 61.7 vs. 65.1%; improvement of 3.4%) [$p = 0.089$].

3.2 Overall effect of training

As for the second question, an effect of fast rate training was not found overall. On average, participants improved their test scores (pre vs. post = 64.7 vs. 71.0%) [$F(1,27) = 28.66$, $p < 0.001$], but the amount of improvement from the pretest to the post-test did not significantly differ between the two groups [no group x test interaction: $F(1,27) = 1.60$, $p = 0.216$]. The test scores did not significantly differ for the pretest (trained: 65.4%; control: 64.1%) [$p = 0.77$] or for the post-test (trained: 73.1%; control: 68.9%) [$p = 0.35$]. Both groups showed significant improvement [trained: 7.8% improvement, $p < 0.001$; control: 4.8% improvement, $p = 0.015$]. Thus, the improvement of both groups' overall scores was attributed simply to the fact that they took the test twice. This result supports Pisoni and Lively's variability theory in that little variability of training stimuli (single rate, single carrier sentence, and single consonant context) did not help trained subjects' overall perceptual learning.

3.3 Specific effects of training across three rates

Although there was no overall effect of training, an effect of training was found for one of the three rates when they were examined separately. There was a significant group x test x rate interaction [$F(2,54) = 3.54$, $p = 0.036$]. Figure 1 shows how each group's improvement from the pretest to the post-test differed according to speaking rate of test stimuli. For the fast rate test stimuli, neither group improved their test scores significantly (trained: 1.8%; control: 5%). For the normal rate stimuli, both groups improved their scores significantly (trained: 10.9%; control: 5.9%). Since both groups improved, however, this improvement must be regarded as an effect of

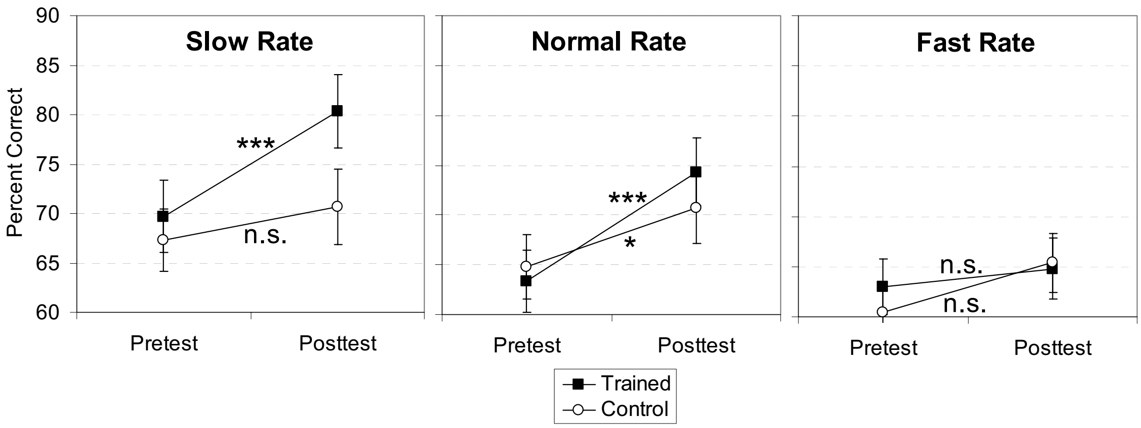

Figure 1: Pretest and post-test scores of trained and control groups for each rate. Significant difference between pretest and post-test [*** p =< 0.001; ** p =< 0.01; * p =< 0.05], non-significant difference between pretest and post-test [n.s. p > 0.05].

taking the tests twice, and not as a training effect. What distinguished the two groups, i.e., the genuine effect of training, was their scores on the slow rate stimuli: the trained group showed significant improvement (10.7%) [$p = 0.001$], while the control group did not (3.5% improvement) [$p = 0.118$]. Thus, with regard to the third question, we found that the training effect was not rate specific: the fast rate training improved scores for slow rate stimuli.

4. DISCUSSION AND CONCLUSIONS

The present experiment supported previous studies [10, 11, 12] in that speaking rate is an important factor for assessing the NE speakers' ability to identify Japanese vowel length. The participants' identification accuracy was higher for the slow than normal rate, and higher for the normal than fast rate. The degree to which nonnative listeners developed perceptual ability over two weeks also depended on the speaking rate of stimuli: both groups improved their scores most for the slow rate, but showed little improvement for the fast rate.

Second, this study revealed that perceptual training with sentences spoken at the fast rate was not effective overall in improving the ability of the trained group to accurately identify Japanese vowel length. This contrasts with the robust improvement that participants made after a comparable amount of training with more variable sentence stimuli in [14]. The present result is consistent with the variability theory [13] in that little stimulus variability in training does not yield nonnative listeners' robust perceptual learning. To further investigate the issue of rate variability, we are currently examining whether nonnative listeners would show robust perceptual improvement with training with mixed rates (both slow and fast) and little improvement with training only with a slow rate.

Finally, the finding for our third question was that training with *fast-rate* sentences had an effect on nonnative listeners' improvement not in the *fast rate*, but in the *slow rate*. Previous training studies for nonnative listeners indicated that abilities acquired during training are limited to deal with kinds of stimuli they had received [13, 16, 17]. For example, training in the word-initial position had an effect on the improvement in the practiced context, but did not yield successful improvement in the word-medial context [16]. The present study is inconsistent with these previous findings, as the participants trained with the fast rate improved their perceptual scores mainly at the slow rate. It is possible that the role of speaking rate variability in training is distinct from the roles of other variability such as in voice or phonetic context. Further studies are necessary to investigate this issue.

5. ACKNOWLEDGMENTS

This study has been supported by the National Science Foundation (Award No. BCS-0418246). The authors would like to thank Jon Bernard for his assistance with the training program.

6. REFERENCES

[1] Landahl, K. & Ziolkowski, M., "Discovering phonetic units: Is a picture worth a thousand words?" *Papers from the 31st Regional Meeting of Chicago Ling. Soc.*, 1: 294-316, 1995.
[2] Toda, T., "Strategies for producing mora timing by non-native speakers of Japanese." In Daini Gengo Shuutoku Kenkyuukai (Ed.), *Acquisition of Japanese as a Second Language*, 157-197, Bonjinsha, Tokyo, 1997.
[3] Oguma, R., "Perception of Japanese long vowels and short vowels by English-speaking learners." *Japanese-Language Education Around the Globe*, 10: 43-55, 2000.
[4] Han, M., "The feature of duration in Japanese." *Onsei no Kenkyuu*, 10: 65–80, 1962.
[5] Tsukada, K., *An acoustic phonetic analysis of Japanese-accented English*. Doct. diss., Macquarie University, 1999.
[6] Ueyama, M., *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese*. Doct. diss., University of California at Los Angeles, 2000.
[7] Kondo, Y., *Production of schwa by Japanese speakers of English: A crosslinguistic study of coarticulatory strategies*. Doct. diss., University of Edinburgh, UK, 1995.
[8] Hirata, Y., "Effects of speaking rate on the vowel length distinction in Japanese." *J. Phonetics,* 32(4): 565-589, 2004.
[9] Hirata, Y. and Lambacher, S. G., "Role of word-external contexts in native speakers' identification of vowel length in Japanese." *Phonetica,* 61: 177-200, 2004.
[10] Tajima, K., Rothwell, A., & Munhall, K. G., "Native and non-native perception of phonemic length contrasts in Japanese: Effect of identification training and exposure." *J. Acoust. Soc. Amer., Vol. 112(5): 2002, p 2387(A).*
[11] Hirata, Y., "Effects of speaking rate on native English speakers' identification of Japanese vowel length." *ISCA Workshop on Plasticity in Sp. Percept., 2005, p 36(A).*
[12] Hirata, Y., "Factors affecting native English speakers' identification of Japanese vowel length." *First ASA Workshop on Sec. Lang. Speech Learning, 2005, p 13(A).*
[13] Pisoni, D. B. and Lively, S. E., "Variability and invariance in speech perception: A new look at some old problems in perceptual learning." In W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, 433-459, York Press, Timonium, MD, 1995.
[14] Hirata, Y., "Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts." *J. Acoust. Soc. Amer.,* Vol. 116(4): 2384-2394, 2004.
[15] Port, R. F. *The influence of speaking tempo on the duration of stressed vowel and medial stop in English Trochee words.* Doct. diss., Indiana University, 1977.
[16] Morosan, D. and Jamieson, D. G., "Evaluation of a technique for training new speech contrasts: Generalization across voices, but not word-position or task." *J. Speech and Hearing Research*, 32: 501-511, 1989.
[17] Strange, W. & Dittmann, S., "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English." *Percept. Psychophysics*, 36: 131-145, 1984.