PERCEPTUAL LEARNING OF JAPANESE MORA SYLLABLES BY NATIVE SPEAKERS OF AMERICAN ENGLISH: EFFECTS OF TRAINING STIMULUS SETS AND INITIAL STATES

Tsuneo Yamada
Osaka University
Suita, Osaka 565, Japan

Reiko A. Yamada
ATR Human Information Processing Research Laboratories
Seika, Kyoto 619-02, Japan

Winifred Strange
University of South Florida
Tampa, Florida 33620, USA

ABSTRACT

Native speakers of American English were trained to identify Japanese short vowels, long vowels and special phoneme /Q/. This laboratory training using natural tokens was effective for generalization into unfamiliar phonemic contexts and unfamiliar talker. Different sets of training stimuli had a significant effect on the early stage of learning processes, but not on the final stage.

1. INTRODUCTION

Native speakers of American English (AE) have difficulty in perceiving some of the sound contrasts in Japanese. For example: /i/ plus /j/ versus palatalization of the preceding consonant (e.g., /ki-ja/ vs. /kja/), long vowel versus short vowel (e.g., /sa-a/ vs. /sa/), presence versus absence of the phoneme /Q/, which makes a geminata with the following consonant by prolonging the following consonant (e.g., /sa-ta/ vs. /sa-Q-ta/ or /sa-t-ta/), and so on. Suprasegmental contour and moraic unit play essential roles in distinguishing these contrasts. In the above examples, hyphens show the moraic boundary. Note that members in each contrast differ in the number of morae.

Difficulties arise for adult English speakers learning Japanese due to the differences in the phonological systems of the English and Japanese. It had been believed that adult learners hardly learn novel phonetic contrasts that do not occur in their own language. Recently, however, it has been experimentally demonstrated that even adult learners can modify their phonetic categories, if adequate training methods are used (cf. [1,2,3]). Segmental phonetic contrasts, such as English /ð/ vs. /θ/ for native speakers of French, and English /r/ vs. /l/ for native speakers of Japanese, were used as training items in those studies.

In this paper, we focused on the suprasegmental contrasts rather than segmental contrasts, and examined AE speakers' perception of Japanese short vowel vs. long vowel, and short vowel with /Q/ vs. short vowel without /Q/. In the experiment, these two contrasts were examined together, using "short vowel", "long vowel" and "short vowel plus /Q/" triplets. We trained adult AE speakers to identify Japanese /CV-CV/, /CV-V-CV/ and /CV-Q-CV/ syllables in order to determine whether adult learners can improve their ability to distinguish these suprasegmental contrasts. Processes of learning and generalization, and the effects of variability in the training stimuli on those processes will be discussed.

2. METHODS

2.1. Subjects

Fifteen native speakers of American English served as subjects. They were monolinguals whose dialect was General American, and had no Japanese language experience.

2.2. Stimuli

Stimulus materials were Japanese syllables which constituted minimal triplets, /CV-CV/, /CV-V-CV/, /CV-Q-CV/, where the first CV and the final CV were identical among the members (Table 1). Triplets with various consonants and vowels were produced by three female and one male native speakers of Japanese (Table 2). There were two sets of training materials, the C2-variable and V1-variable sets. In the C2-variable set, C2 varied among 10 consonants whereas V1 was fixed as /o/. In contrast, in the V1-variable set, V1 varied among 5 vowels whereas C2 was fixed as /k/. In order to have the same number of triplets in these sets, two different combinations of C1 and V2 were used in the V1-variable set, and only a single combination was in the C2-variable set. Thus, each set consisted of 10 triplets. The training stimuli were produced by the three female talkers.

Table 1. Examples of the minimal triplets contrasting short vowel, long vowel and /Q/ used as stimulus materials.

Response Category	Structure	Examples
short V	C1V1-C2V2	/kato/
long V	C1V1-V1-C2V2	/kaato/
Q	C1V1-Q-C2V2	/kaQto/

In addition, there were two sets of test materials, the full-test set and mini-test set. In the full-test set, all possible combinations of five vowels and twelve consonants, i.e., 60 triplets, produced by all four talkers, were used. Three of these talkers were those used in the training, and the fourth was the male talker not used in the training. The mini-test set was a subset of the full-test set (see Table 2), and consisted of only the tokens produced by the three talkers used in the training. Syllables in the test sets were selected among the ones which were not used in the training sets.

2.3. Procedures

Subjects were assigned to the C2-variable training group, V1-variable training group or control group. In addition to the pretest and post-test, a mini-test on every training day and an interim test just in the middle of the entire training phase were administered. The experiment consisted of eleven daily sessions. Familiarization and pretest were administered on Day 1, and followed by four days of training (Day 2-5), an interim test (Day 6), four days of training (Day 7-10) and a post-test (Day 11). On every training day, while subjects in the C2- and V1-variable training groups took both one training session and one mini-test session, those in the control group took only the mini-test session.

In the pretest, interim test and post-test, the full-test set of 720 stimuli (60 triplets by 4 talkers) were tested without repetition (720 trials). In each training session, 10 triplets by 3 talkers (i.e., 90 stimuli) were presented randomly with three repetitions (270 trials). In the mini-test session, 14 triplets by 3 talkers (i.e., 126 stimuli) were presented without repetition (126 trials).

Table 2. Combinations of V1 and C2. Combinations for the V1-variable training were enclosed with a square, and those for the C2-variable training were enclosed with ellipses. Combinations marked with "f" were used in the full-test, and those marked with "m" were used in the mini-test.

C2 \ V1	a	i	u	e	o
k	f,m	f,m	f,m	f,m	(f,m)
s	f	f	f	f	(f,m)
t	f	f	f	f	(f,m)
h(F)	f	f	f	f	(f,m)
g	f	f	f	f	(f,m)
z	f	f	f	f	(f,m)
d	f	f	f	f	(f,m)
b	f	f	f	f	(f,m)
p	f	f	f	f	f
ʃ	f	f	f	f	(f,m)
tʃ	f	f	f	f	f
dʒ	f	f	f	f	(f,m)

A forced-choice task from three alternatives, "short-V," "long-V" and "short-V+Q", were used in all the tests and the training. Subjects were seated in a quiet room, and listened to the stimuli through a headphone (Stax, SR-Λ Signature). They responded by typing one of the keys, which corresponded to the alternatives. In the tests, any feedback was not given to the subjects. In the training, in contrast, feedbacks were given to them immediately after the responses; chime sound and graphical coin for correct response, and buzzer for incorrect response. Experimental events and data collection were controlled by workstation (NeXT Cube Turbo).

3. RESULTS & DISCUSSION

3.1. Effects of training: Comparisons in pretest, interim test and post-test

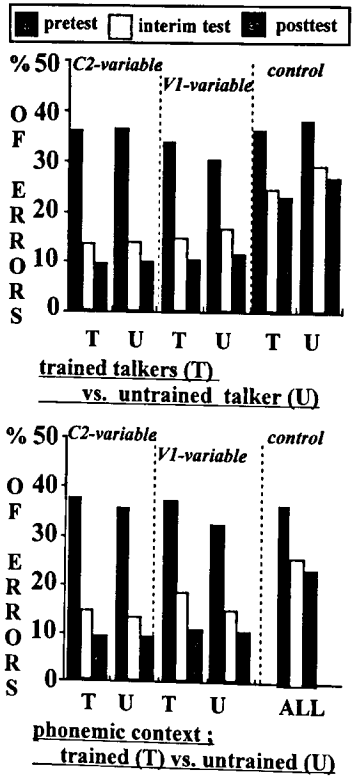

Figure 1. Generalization processes in interim test and posttest. The ordinate indicates mean percentage of errors in sub-categories, (1) talkers used in training vs. an unfamiliar talker (upper) , and (2) trained phonemic contexts vs. untrained contexts (lower).

Error rates in pretest, interim test and post-test are shown in Figure 1. Mean error rates in the pretest were around 30-40% in all three groups, which were lower than chance level of 67%. Such low error rates stemmed from the fact that AE listeners could distinguish short vowels from long vowels without training in some phonemic contexts, although they had difficulties when the contrasts occurred before /s/, /ʃ/, /tʃ/, or /k/.

Both the C2-variable and V1-variable training groups showed significant improvements in the interim test and post-test. These were true not only for the stimuli produced by trained female talkers but also for those produced by an untrained male talker (Figure 1, top panel), suggesting that training effects generalized into the unfamiliar talker. The equal amounts of improvements were also observed in trained phonemic contexts and untrained phonemic contexts in both training groups (Figure 1, bottom panel), suggesting that training effects generalized into unfamiliar phonemic contexts. These results showed that, through such training, adult AE speakers modified their phonetic categories, so that they can distinguish short vowel, long vowel and short vowel plus /Q/.

3.2. Acquisition process: Analysis of mini-tests.

Error rates in mini-tests are shown in Figure 2. While the C2-variable training group showed drastic improvements in the first three sessions, the V1-variable group needed larger number of training sessions to generalize into both trained and untrained phonemic contexts.

There observed a significant effect of consonant (C2) on the error rates: The voiceless fricatives (/s/, /h(F)/ and /ʃ/) made the distinction of the preceding target part, contrasting short vowel, long vowel and short vowel plus /Q/, difficult to distinguish. Since Japanese /Q/ is a geminata, in which the second consonant is preceded by the homorganic consonant, the /Q/ mora is filled with noise when the second consonant is a fricative, while it is filled with silence for the other consonants. The above result implies that the /Q/ filled with silence is easier for AE speakers to distinguish in the contrasts than the /Q/ filled with noise. In contrast to the effect of consonant, there was no effect of vowel (V1). These results pointed out that variation in the following consonant affects the perception more strongly than variation in the preceding vowel: Furthermore, the stimulus set with high variability in the following consonant is effective for generalizing in the early stage of learning on such contrasts.

4. CONCLUSIONS

The native speakers of AE learned the distinction of the Japanese moraic contrasts, among short vowel, long vowel and short vowel followed by /Q/. The training effects generalized not only into the untrained phonemic environments, but also into the untrained talker. The acquisition processes were clarified by analyzing performance in the mini-tests, which were administered after every training session. It was found that distinction in the contrasts is difficult when a fricative is the following consonant, and that training with high variability in the following consonants facilitates acquisition in the early stage of training. However, the enough amount of training trials promotes the equivalent improvements in the training with low variability in the following consonants.

Speech perception is originally to be *learned* in an individual linguistic environment. Therefore, the viewpoint that speech perception has dynamic aspects as an acquisition process is also necessary. In this paper, it was found that the acquisition process is affected by the structure of training stimuli. It was implicated that analysis of the learning process is essential for clarifying the mechanisms to acquire non-native contrasts.

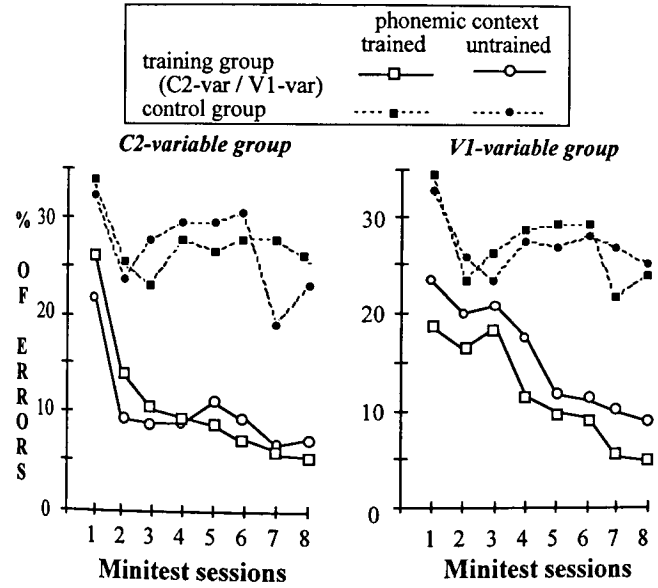

Figure 2. Generalization processes in the minitests. The ordinate indicates mean percentage of errors in each training condition, that is, (1) training group vs. control group and (2) trained phonemic contexts vs. untrained contexts. Constituents of each context condition in control group correspond to those in each training group.

ACKNOWLEDGMENTS

This research was partly supported by a fellowship from the Ministry of Education, Science and Culture, Japan to T. Yamada. We wish to thank Desiderio Saludes and Brett H. Fitzgerald for their assistance in conducting this experiment.

REFERENCES

[1] Jamieson, D.G., & Morosan, D.E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, **40**, 205-215.

[2] Logan, J.S., Lively, S.E., & Pisoni, D.B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, **89**, 874-886.

[3] Strange, W., & Dittmann, S. (1984). Effects of discrimination training in the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, **36**, 131-145.