Perception of phonemic length contrasts in Japanese by native and non-native listeners

Keiichi Tajima[*], **Hiroaki Kato**[*], **Amanda Rothwell**[*], and **Kevin G. Munhall**[†]

[*] ATR Human Information Science Laboratories, Kyoto, Japan
ktajima@atr.co.jp, kato@atr.co.jp

[†] Queen's University, Kingston, Ontario, Canada
rothwell@atr.co.jp, munhallk@psyc.queensu.ca

ABSTRACT

Native English listeners' perception of Japanese phonemic length contrasts was investigated using a minimal-pair identification task with Japanese words that differed by the length of a vowel or consonant, e.g., "haken"–"hakken". Identification accuracy was found to be higher for listeners with Japanese exposure than for unexposed listeners, and higher after five days of perceptual identification training than before. Even though listeners were trained to identify words containing vowel length contrasts only, accuracy improved for other length contrasts also. However, accuracy was relatively poor when the contrasts occurred in certain positions within the target word. In sum, while training improved English listeners' perceptual ability, a significant gap remained before they reached native-like performance.

1 INTRODUCTION

The length of vowels and consonants can be used contrastively in Japanese, captured in Japanese phonology as a difference in mora count. The primary acoustic correlate of such length contrasts is said to be the relative duration of the vowel or consonant. Since languages such as English do not convey lexical contrasts by segment duration alone, perception of phonemic length is predicted to be difficult for native listeners of these languages. Past studies on second-language (L2) speech perception have indeed found evidence supporting this prediction (e.g., [1]). Meanwhile, studies have shown that perception of L2 phonetic contrasts is affected by listener-related factors such as amount of experience with L2 [2]. Likewise, the ability to perceive L2 phonetic contrasts has been shown to improve with perceptual identification training [3].

The main goal of the present study is to investigate native and non-native listeners' perception of various types of Japanese phonemic length contrasts occurring in naturally spoken words and sentences (see [4] for a related study using resynthesized stimuli). Specifically, this study first examined how non-native listeners' identification performance is influenced by exposure to Japanese and by perceptual training. Second, given that Japanese has several distinct types of phonemic length contrasts, this study addressed whether training with one type of contrast generalizes to other types of contrast or whether its effect is restricted to that particular contrast type. Third, the study explored whether listeners' perception of phonemic length varies depending on word position (whether the length distinction appeared word-initially, medially, etc).

2 METHODS

Three groups of normal-hearing listeners took part in the study: (1) seven English speakers aged 18–30 with no Japanese experience (Eng) who participated in the study at Queen's University in Canada, (2) seven English speakers aged 18–32 who had spent 1–6 months in Japan (JRE = Japan-Residing English), and (3) seven native Japanese speakers aged 19–22 (Jpn). Self-reports by listeners in group JRE indicated only minimal daily Japanese usage. Groups JRE and Jpn took part in the study at ATR Laboratories in Japan.

The study consisted of three phases: pretest, training, and posttest. Group JRE took part in all three phases, while groups Eng and Jpn took just the pretest.

Stimuli for the pretest and posttest consisted of Japanese word pairs that were segmentally and accentually matched except for a difference in one of the following four types of length contrasts in Japanese: (1) length of one of the five Japanese vowels "i e a o u" (18 pairs), e.g., "kado" (corner) – "ka:do" (card), (2) length of one of the obstruents "p t k s sh ch j" (15 pairs), e.g., "haken" – "hakken", (3) length of one of two nasals, "m" or "n" (18 pairs), e.g., "tanin" (stranger) – "tannin" (person in charge), or (4) length of the palatal sound "y" (18 pairs), e.g., "kyaku" (visitor) – "kiyaku" (statute)[1]. The contrast appeared in different word positions, e.g., in the initial syllable, e.g., "ka:do" – "ka::do", or final syllable "jinja" (shrine)

[1]The four contrast types are referred to in Japanese as choh'on, sokuon, hatsuon, and yoh'on, respectively.

ISBN 1-876346-48-5

– "jinja:" (ginger). For the nasal pairs, the contrast occurred either in the initial syllable[2], e.g., "tanin" – "tannin" or in a medial position, e.g., "kiri no e" (picture of a fog) – "kirin no e" (picture of a giraffe)[3]. The stimuli also included six nonword triplets of the form "ereCe" – "ere:Ce" – "ereCCe" where "C" is one of "p t k s m n" (see [4] for details).

The stimuli were produced by two professionally trained native Japanese talkers of various ages (one male aged 55 and one female aged 44). Each talker produced each item above in two contexts: (1) in isolation, and (2) embedded in one of ten carrier sentences of roughly equal length, e.g., "ima kara ___ to iimasu" (I will say ___ now). Each item was embedded in different sentences for different talkers. A subset of the vowel length pairs and the obstruent length pairs, as well as the nonwords, were produced at three self-selected speaking rates, slow, normal, and fast, in both contexts. The nasal length and palatal "y" pairs were produced only at a normal rate. Professional talkers were used because they were thought to be more likely than naive talkers to produce the length contrasts reliably at various speaking rates.

Listeners participated in a single-stimulus, forced-choice identification task. On each trial, listeners heard a stimulus through headphones and identified it by choosing among items presented on the computer screen. On trials using real words as stimuli, two alternatives comprising a minimal pair were displayed. On trials using nonwords, three alternatives comprising a triplet as described above were displayed. Alternatives were presented using romanizations of the Japanese words as illustrated in the examples above. Long vowels were transcribed as "i: e: a: o: u:" to avoid undesirable grapheme-to-phoneme correspondences for "ee" and "oo". These conventions were explained to the listeners, and the length contrasts were illustrated using examples, prior to the test.

The pretest and posttest each consisted of 1296 trials, lasting roughly 2.5 hours. The trials were blocked by each available combination of talker, rate, context, and contrast type. The blocks appeared in a fixed order, but trials within each block were randomized.

After the pretest, group JRE underwent five days of identification training. Three training sessions were conducted on each day, with no more than three free days intervening consecutive training days. Each training session had 240 trials, consisting of 60 vowel length minimal pairs that were different from the test stimuli, presented two times each in a random order. Unlike the pretest and posttest, training stimuli consisted solely of words contrasting in vowel length produced in isolation at a normal rate. Items in each session were spoken by one of five professionally trained Japanese talkers of various age groups (three females and two males, aged 35–65). One of the talkers appeared during the tests. The five talkers appeared in a fixed order across the 15 sessions.

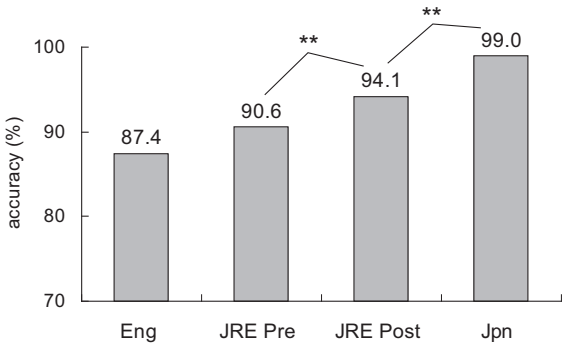

Figure 1: Identification accuracy as a function of exposure to Japanese and perceptual training ($n = 7$ for each bar). $^{**}p < .01$.

Training trials were identical to the test trials except that subjects were given immediate feedback concerning their responses, and repeated a trial until they selected the correct response. Following the training, group JRE took part in the posttest, which was identical to the pretest.

3 RESULTS AND DISCUSSION

3.1 Effect of training and exposure

Figure 1 examines how Japanese exposure and identification training affect listeners' performance, by plotting mean identification accuracy for group Eng, group JRE's pretest and posttest, and group Jpn. The figure is based on data from all rates, contexts, and contrasts. Results of pairwise comparisons between adjacent group means are also indicated[4].

Figure 1 shows that mean accuracy increases from left to right. That is, starting from the left, comparison of group Eng's score and group JRE's pretest score indicates that English listeners who had some exposure to Japanese performed slightly better on average than listeners who did not. The difference, however, did not reach significance, due to considerable within-group variability. Next, comparison of group JRE's pretest and posttest scores reveals that training improved listeners' performance. This difference was small but significant ($p < .01$); in fact, all seven listeners improved from pretest to posttest, although several listeners who scored above 95% in the pretest improved only minutely in the posttest. Finally, comparison of group JRE's posttest score and group Jpn's score indi-

[2]Or following the initial syllable's vowel, to be precise.

[3]Short noun phrases were used instead of single words because there were no word pairs that contained a nasal length contrast in non-initial positions.

[4]All statistical tests in this study were conducted on arcsine-transformed values of the percentage scores.

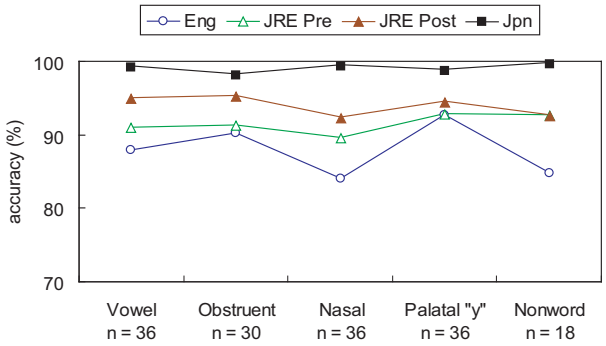

Figure 2: Identification accuracy for the four types of length contrasts and for the nonwords.

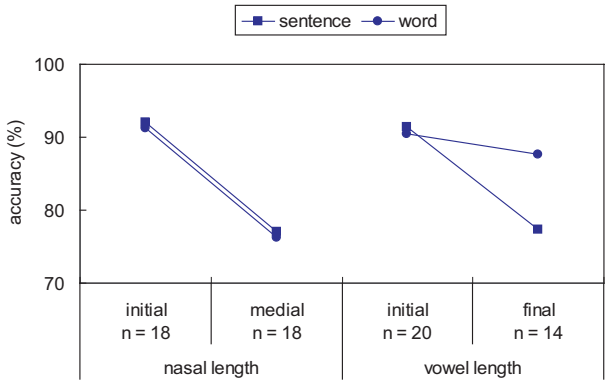

Figure 3: Group Eng's mean scores as a function of context (word, sentence) and position (initial, medial/final) for words involving nasal length (left half) and vowel length (right half) contrasts.

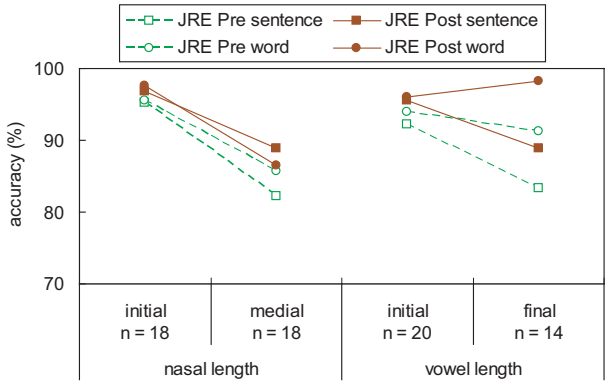

Figure 4: Group JRE's mean pretest and posttest scores as a function of context and position for words involving nasal or vowel length contrasts.

cates that even though training improved performance, non-native listeners still responded less accurately than native listeners did ($p < .01$) (see [4] for a similar result based on "ambiguous" speech stimuli).

3.2 Generalization to untrained contrasts

Figure 2 examines the effect of exposure and training separately for the four types of length contrasts and for the nonwords. The figure is based on stimuli spoken at a normal rate (see [5] for an analysis of how speaking rate influenced listeners' performance).

Comparison of group Eng's scores and group JRE's pretest scores suggests that English listeners with Japanese exposure scored somewhat higher than listeners with no exposure for some contrast types. Statistical comparisons between these groups' scores for each contrast type, however, did not show any significant differences. Turning to the effect of training, given that group JRE was trained with vowel length pairs only, it is possible that these listeners' performance might improve only for this particular type of pairs. Figure 2 shows that mean accuracy was higher after training than before not just for vowel length pairs but also for other length contrasts. A two-way ANOVA showed a significant effect of test (pre vs. post) ($p < .05$) but no significant effect of contrast type nor a significant interaction. Individual comparisons indicated that the differences were statistically significant for the vowel contrast ($p < .05$) and the obstruent contrast ($p < .01$), but not for the nasal and palatal "y" contrasts. These results suggest that perceptual training with vowel length contrasts generalizes to at least some other types of length contrasts in Japanese.

3.3 Effect of position within word

Figures 3–4 depict how non-native listeners' performance varied depending on where the length contrast appeared within the word and on whether the word appeared in isolation or in a sentence. The left half of each figure shows data involving nasal length contrasts, comparing two positions, initial vs. medial. The right half of each figure shows data from words involving vowel length contrasts, comparing two positions, initial vs. final. Figures 3 and 4 show results from groups Eng and JRE, respectively. Group Jpn's data are not plotted because a similar analysis indicated that scores were above 98% in all conditions, suggesting no effect of position or context[5].

Looking at group Eng's performance in Figure 3, the left half shows that accuracy was high when the nasal contrast appeared in initial position, but was substantially lower when it appeared in medial position. The presence or absence of a carrier seems to have no effect on listeners' performance for this contrast. In the right half of the figure, data from words containing vowel length contrasts indicate that accuracy was again relatively high when it was the initial position that the contrast appeared in. When the contrast appeared in

[5]For stimuli containing obstruent length contrasts, the effect of position was not analyzed because all obstruent contrasts occurred in word-initial position. For stimuli contrasting in the length of the palatal "y", the effect of position was not analyzed because there were too few stimuli available in each position to allow proper analysis.

ISBN 1-876346-48-5

word-final position, performance depended on the context. That is, accuracy was high when the word appeared by itself but was low when it was embedded in a carrier. In Figure 4, group JRE's data exhibit a similar pattern of results in both the pretest and posttest.

Separate ANOVAs were carried out for the nasal and vowel length contrasts, and for data from group Eng, group JRE's pretest, and group JRE's posttest, with Context and Position as within-subjects factors. For the nasal length contrast, there was a significant main effect of Position ($p < .001$). For the vowel length contrast, a significant main effect of Position ($p < .05$) and a significant Position-by-Context interaction ($p < .05$) were found. This pattern of result was found for group Eng, and JRE's pretest and posttest. No significant effects were found for group Jpn.

One noteworthy characteristic of the result in Figures 3 and 4 is that native English listeners showed high accuracy for contrasts that appeared in the initial syllable, for both nasal and vowel length pairs. This may have arisen from the fact that Tokyo Japanese has a low-to-high or high-to-low lexical tone alternation between the first and second moras of a word or phrase. As such, a length contrast that appears in the initial syllable is accompanied not only by a difference in duration, but also by a difference in pitch movement. It is therefore possible that English listeners made use of both duration and pitch cues to identify phonemic length contrasts that occurred in initial syllables, while they relied primarily on duration for other positions.

Another salient pattern of result in Figures 3 and 4 is that accuracy is high when the length contrast is utterance-final (i.e., both word-final and not followed by a carrier sentence), but is low when it is followed by something else (either the remainder of the target stimulus or the carrier sentence). This pattern seems somewhat more difficult to account for unequivocally. On one hand, an opposite trend might be predicted if one assumes that judgment of segment duration would be easier when the length distinction is immediately followed by another vowel or consonant than when it is not followed by any phonetic material. In fact, past studies have found evidence supporting this prediction [6]. On the other hand, however, it is also possible that highly salient cues for utterance-final length contrasts were in fact available in the stimuli used in the present study,perhaps by virtue of the fact that the materials were produced by professionally trained talkers, who are likely to speak more clearly than non-professional talkers. Further perception tests as well as acoustic analysis of the stimuli are necessary to find plausible explanations of the results.

4 CONCLUSIONS

Results from this study suggest that exposure and training improve English listeners' perception of length contrasts in Japanese. However, differences in perceptual tendencies remain between native and non-native listeners even after training. The level of performance of the English listeners is generally high, showing a score above 87% even for listeners with little Japanese exposure. It remains to be seen whether this is because perception of phonemic length in Japanese is not very difficult for English listeners in the first place, or whether the specific procedures used in the present study, e.g., the repetitiveness and length of the tests, were problematic. Additional experiments with modified procedures are necessary to address this question.

ACKNOWLEDGMENT

This work was conducted in collaboration with Reiko Akahane-Yamada, ATR Laboratories. This research was supported in part by the Telecommunications Advancement Organization of Japan.

REFERENCES

[1] R. Oguma, "Perception of Japanese long vowels and short vowels by English-speaking learners," *Japanese-Lang. Educ. around the Globe*, vol. 10, pp. 43–54, 2000.

[2] R. A. Yamada, "Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, W. Strange, Ed. York Press, Timonium, MD, 1995.

[3] S. E. Lively, D. B. Pisoni, R. A. Yamada, Y. Tohkura, and T. Yamada, "Training Japanese listeners to identify English /r/ and /l/ III. Long-term retention of new phonetic categories," *J. Acoust. Soc. Am.*, vol. 96, pp. 2076–2087, 1994.

[4] H. Kato, K. Tajima, A. Rothwell, R. Akahane-Yamada, and K. G. Munhall, "Native and non-native perception of moraic phonemes in Japanese: Measurement of categorization," in *Proc. 2003 Spring Meeting Acoust. Soc. Japan*, Tokyo, Japan, 2003.

[5] K. Tajima, H. Kato, A. Rothwell, and K. G. Munhall, "Native and non-native perception of moraic phonemes in Japanese: Effect of identification training and exposure," in *Proc. 2003 Spring Meeting Acoust. Soc. of Japan*, Tokyo, Japan, 2003.

[6] H. Kubozono, "Temporal neutralization in Japanese," in *Papers in Laboratory Phonology VII*, C. Gussenhoven and N. Warner, Eds. Mouton, to appear.