Perceptual Acquisitions of Non-native Syllable Structures by Native Listeners of Japanese

Kyoko Nagao, Byung-jin Lim and Kenneth de Jong
Department of Linguistics, Indiana University
E-mail: knagao@indiana.edu, bylim@indiana.edu, kdejong@indiana.edu

ABSTRACT

This study aimed to examine how native and non-native language experience affects the adults' perception of non-native contrasts. More specifically, this study examined the perceptual acquisition of English syllable structures and voicing contrasts by Japanese listeners. The results from thirty-four native listeners of Japanese showed that even the listeners with virtually no exposure to English speech could extract both onset and coda types of structure syllables from the stream of repeated syllables, and the good performance of non-native syllable structures is accounted by the listeners' bias toward a 'foreign' category as well as the native phonology.

1. INTRODUCTION

Native language experience strongly influences perception of non-native contrasts by adult second language learners. However, influence from native language is not uniformly applied to all the non-native contrasts. Adults' perception is also affected by non-native experiences such as length of residence or age of learning [1].

The Perceptual Assimilation Model (PAM) proposed by Best[2] predicts discriminability by non-native listeners. The predictions by the PAM are made on comparison between native and non-native phonological systems. Native listeners perceptually assimilate non-native phones to their native phonemes based on detection of commonalities in gestures. There are six types of assimilation in the PAM. Two Category assimilation (TC type) is an assimilation type that two non-native phonemes are assimilated to the two distinctive phones in the native phonology. Single Category assimilation (SC type) and Category Goodness difference (CG type) are for assimilation patterns in which two non-native contrastive phonemes fall in to the same native phonemes. When the non-native contrastive phonemes are perceived as a good native phoneme for one and a bad native phoneme for another, performance could be relatively good. Uncategorized-Categorized pairs (UC type) are expected to be well discriminated because the one phone is categorized as a native category and the other one is categorized as a phonetic sound but outside of native phonological system. Both Uncategorizable (UU type) is a type that both non-native phonemes are outside of native phonological space. The predicted performance depends on the similarity of the non-native contrasts. Nonassimilable (NA type) is a case that the non-native sound is perceived as non-speech sounds.

Although the PAM has been examined by a lot of cross-language studies on segmental contrast, it has not been well-examined for contrasts at suprasegmental level. Also, the PAM focuses on mainly one aspect of perceptual ability, i.e. discrimination of contrasts, but not identification. Identification of non-native contrasts can tell us more clearly whether the learners can acquire the contrasts or not.

The current paper examined the perceptual acquisition of non-native syllable structures and voicing contrasts by adult second language learners and how non-native language experience affects acquisition of non-native contrasts. More specifically, we examined how native listeners of Japanese with different degrees of English exposure identify the non-native VC structures /ib/ and /ip/. CV structures /bi/ and /pi/ were employed for the comparison of voicing contrasts.

English and Japanese were selected for the purpose of this study because Japanese does not allow some syllable structures that English allows. English exhibits both coda and onset syllable (VC and CV) structures, and multiple consonants are allowed to follow or precede a vowel. However, Japanese is a language in which the dominant syllable structures are open-syllables, and which allows only one consonant in either onset or coda position. Japanese coda syllables are extremely limited, both in terms of their content and their frequency of occurrence. The possible coda consonants are either the moraic nasal /N/ (e.g. *bin* /biN/ = [biN] 'a bottle') or the first segment of geminate consonants /Q/ appearing only in the middle of word, (e.g. *kitte* /kiQte/ = [kitte] 'a stamp'). Voiced consonants are not allowed to form a geminate consonant.

In addition, there are differences in the segmental makup of the two languages as well. In general, that stop consonants in English and Japanese are similar but not identical. The most notable difference is that English voiceless stops usually have strong aspiration. Measured as voice onset time (VOT), category boundary locations between voiced and voiceless stops are not identical in both languages. Homma [3] reported that the average VOT of Japanese intervocalic /p/ and /b/ are 77 and 55 milliseconds, respectively, while Lisker [4] showed that the average VOT of intervocalic American English /p/ and /b/ are 120 and 75 milliseconds, respectively. Thus, English voiceless stops are pronounced with longer VOT

ISBN 1-876346-48-5

than Japanese voiceless ones, while the VOT for English voiced stops overlapped with the ranges of VOT for Japanese voiceless ones.

If we assume that perceptual identifications are based on unsyllabified segments, we would expect good performance from the Japanese listeners, as long as the phonemes are roughly equivalent. When the syllabic organization is included in the categorization, coda structures such as /ib/ and /ip/are both uncategorized, but onset structures such as /bi/ and /pi/ are categorized as native structures. Therefore, the coda and onset structures could be treated as an UC type assimilation, which predicts good discrimination between coda and onset structures. If native listeners apply the phonological restriction on geminate consonants, /ib/ and /ip/ could be also an UC type. Otherwise, a pair of coda structures could be an UU type. A pair of onset structures /bi/ and /pi/ could be TC type or CG type because phonetic boundaries for voicing contrasts are different between the two languages. A pair of /ib/ and /bi/ could be an UC type.

Certain exposure to a particular non-native language could enable the learners of the language to recategorize the once uncategorized phoneme or structure as a categorized one. This then could produce a difference in perceptions of coda structures between the learners with a different degree of exposure to the non-native system.

2. METHODS

Subjects

Thirty-four native listeners of Japanese participated in the study. They were grouped into three subgroups in terms of their degree of English exposure. The first group, JA (= Japanese Advanced) were fourteen Japanese listeners who had resided in the United States more than three months. All of them were students of Indiana University, Bloomington ranging in age from 21 to 31 years old, with a mean of 24.4 years old.

Another two groups of Japanese listeners were recruited and run in Japan. We will consider these two subgroups as monolingual Japanese listeners. One subgroup, JY (= Japanese Younger), consisted of twelve younger listeners. They were undergraduate students of Tezukayama Gakuin University in Osaka, Japan, whose ages ranged from 18 to 21 (mean= 19.7) years old. The other subgroup, JO (= Japanese Older), consisted of eight older listeners whose ages ranged from 42 to 60 (Mean= 51.6) years old. All of the older listeners had been lived in the same area as the younger listeners for more than twenty years. Three of the JY and five of the JO had been to abroad, but only for leisure trips for a short period. None had taken any English classes outside of a regular school system. English education in Japan has placed little importance on listening and speaking skills; hence we consider taking English in Japanese regular schools not to count as exposure to English speech.

Eighteen native listeners of American English also participated as a control group (ENG). Their ages ranged from 18 to 23 (mean= 20) years old. None reported that they had been diagnosed with any hearing problems.

Stimuli

Four original utterances were spoken by four native speakers of American English. The speakers repeated either /ib/, /ip/, /bi/, or /pi in time with a rate-controlling metronome. The metronome period started from 450 ms/syllable (Slow) and shortened to a period of 200 ms/syllable (Fast). From the fastest 22 syllables, we spliced each three repeated syllables as one stimulus. Splicing was done from the vowel offset to vowel onset. Hence, the fastest stimulus contained the three fastest syllables. The last and the middle syllables of the second fastest stimulus were the middle and the first syllables of the fastest stimulus. Twenty-one stimuli were spliced from each original utterance. In total, 336 stimuli were prepared for the perception experiments.

Procedure

Listeners were asked to identify what the speaker is repeating by selecting one of the four syllables on a computer screen. Listeners were allowed to listen to each stimulus as many times as they wished. There was one difference between the JA and the two groups of monolingual Japanese. The labels used for the advanced learners were ‘eeb’, ‘eep’, ‘bee’, and ‘pea’, which were the same labels used for the English listeners, while the monolingual listeners were presented with ‘ib’, ‘ip’, ‘bi’, and ‘pi’, based on the Roman writing convention adopted in Japan.

Besides the identification task, the listeners were asked to estimate how sure they were to their answer. Preliminary analyses of these confidence ratings indicate that the ratings were not very sensitive to the category structures evident in the identification data. We will not discuss listeners’ confidence level of their answers in current paper. In general, it took less than one hour to complete the entire experiment.

There are important differences between current study and previous studies related with the PAM. First, our collected data is based on listeners’ identification performance not discrimination abilities as seen in other PAM studies. We assume that identification results are equivalent to discrimination results in order to examine the native adults’ perceptual performance on non-native contrasts. Second, to probe the variability which the perceptual systems have to deal with, we employed natural speech elicited in a rate-variation paradigm. Rate-varied productions have previously been found to affect listeners’ perception of both syllable affiliation and stop voicing [5].

Analysis

Four types of syllables were employed in this study. Based on their syllable structures, /pi/ and /bi/ were grouped as CV types, and /ip/ and /ib/ as VC types. When a VC-type syllable is repeated with accelerated rate in English, the listeners’ perceived syllable structure becomes CV from VC[5, 6]. This is called perceptual

resyllabification. Perceptual resyllabification has proven to be a robust and cross-language phenomenon [7, 8]. Since native listeners of English tend to misperceive VC as CV at fast rate, accuracy is not an appropriate measurement for the stimuli we used. In order to see whether or not the non-native listeners could recognize VC structures, mean percent CV responses to the VC stimuli were computed at each speech rate for each speaker. These CV responses collected from the three groups of non-native listeners were compared with the responses from the native listeners. Since perceptual shift is rate sensitive, separate statistical tests were conducted for slow and fast rates. The six stimuli which appeared earliest in the utterance are grouped as Slow stimuli, whereas the last six stimuli were grouped as Fast stimuli.

3. RESULTS

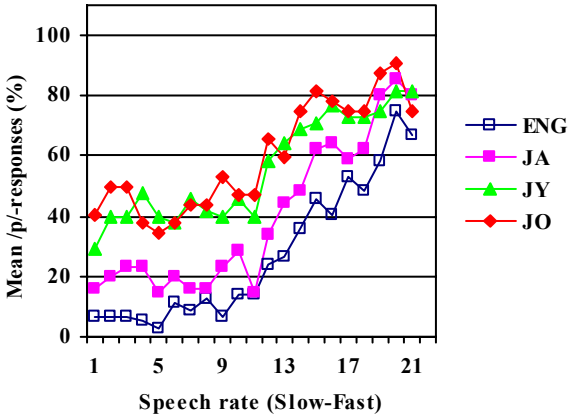

FIG. 1 Percent /p/ identification of /bi/ tokens by four groups of listeners.

Let us first look at the CV-stimuli. JA and JO perceive about 97 % of the CV-stimuli as CV and their results are very similar to the English listeners. JY shows relatively poor perception (81 %), but overall patterns are similar to the English listeners. Voicing identification for /pi/ are good for all listener groups. The mean /p/-responses for the /pi/ for each group are above 94 % for Slow, and above 85 % for Fast. FIG. 1 displays mean /p/-responses to /bi/-stimuli from the four groups of listeners. JA listeners show slight preference to /p/ compared with the English listeners. Less English exposed JY and JO also prefer /p/ for English /bi/. These results suggest that the Japanese listeners can identify non-native VC structures, but their perception of VC can be influenced by the most dominant native structure CV as speech rate increases. However, this CV bias at fast rate is not found in the JY listeners.

Next, we will look at the perceptions of non-native syllable structures. All three groups of Japanese identified the VC structure in a very similar fashion to the English control group. When the rate is slow, listeners' identifications of VC are accurate and consistent among the listeners. For the Slow /ib/-stimuli, the mean CV-responses from JA, JY, JO, ENG groups are 4.4, 6.2, 2.6, and 3.8 %, respectively. For the Slow /ip/-stimuli, the mean CV-responses from JA, JY, JO, and ENG groups are 3.2, 1.1, 3.4, and 1.9 %, respectively.

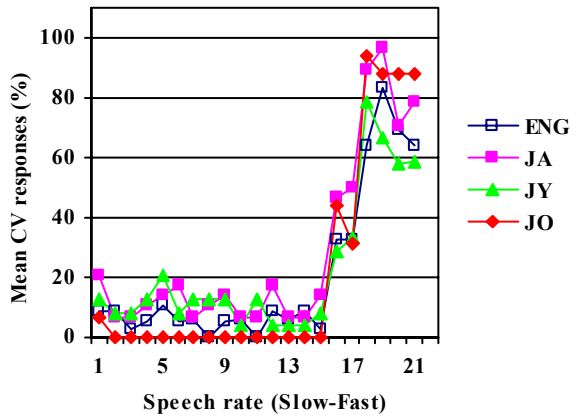

FIG. 2 Percent CV identification by the four groups of listeners for the VC-stimuli produced by Speaker CD.

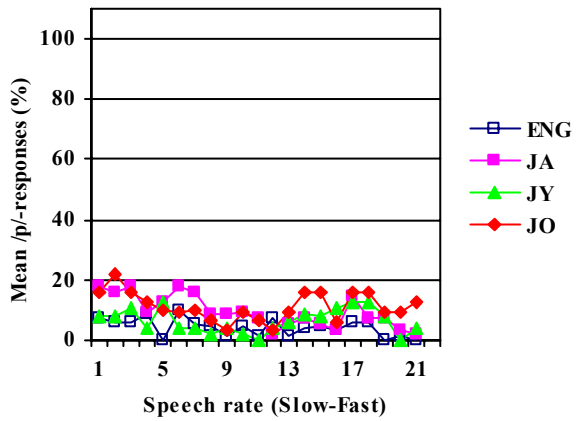

FIG. 3 Percent /p/ identification to /ib/-stimuli by four groups of listeners.

However, there is a difference among the four groups at fast rate. FIG. 2 shows the mean proportions of CV responses by the four groups of listeners to the VC-stimuli originally produced by CD as a function of speech rate. Taking the ENG results as a reference, we can say that as rate increases, JA and JO listeners tend to perceive VC as CV, whereas JY listeners exhibited a tendency against perceptual resyllabification. To test whether or not the four groups of listeners are different at fast rate, a two-way ANOVA on the listener groups and the speakers were performed, using CV responses to the Fast VC-stimuli. The main effect of the groups was significant, $F(3, 176)= 9.44$, $p < .05$. There was no significant interaction effect, $F(9, 176)= 1.08$, $p > .05$. Sheffé's post-hoc test results showed that only the two pairs of JY and ENG and of JA and JO are not significantly different. The same post-hoc test for the Slow stimuli revealed that the groups of the listeners are significantly different from one other except a

ISBN 1-876346-48-5 © 2003 UAB

pair of JY and JO. These results suggest that the Japanese listeners can identify non-native VC structures, but their perception of VC can be influenced by the most dominant native structure CV as speech rate increases. However, this CV bias at fast rate is not found in the JY listeners.

FIG. 3 shows the mean proportions of /p/-responses by the four groups of listeners to the /ib/-stimuli as a function of speech rate. The consonants in the /ib/ were identified accurately, while the listener's identification of /p/ for /ip/ was not very good. JA and JO exhibited a tendency toward /p/ for the /ib/ as seen in FIG. 1. However the young, less experienced listeners did not follow this tendency, especially at slow rate.

4. DISCUSSION

Perception of onset consonants by Japanese listeners clearly shows the effect from their native voicing categorization. As expected from the VOT differences between Japanese and English, Japanese listeners tend to identify the non-native /b/ as /p/. Examining the VOT values of the second syllable in each stimulus, we found that Japanese listeners show more /p/ responses than English to the stimuli with the same ranges of VOT. We also found that a negative linear relationship between the strength toward the native category and the degree of exposure to English. This can be a CG type assimilation. It can be said that non-native experience aids to adjust the gap between native and non-native systems.

The good Japanese identifications for the non-native syllable structures tell us that segmental analyses can be undertaken when they judge the syllable structures. Otherwise, the results for /ib/ and /ip/ should show a good performance for syllable perception but chance level performance for voicing perception. The differences seen in the results for /ib/ and /ip/ suggest that structural analysis is also employed by Japanese listeners. If the Japanese listeners apply their native phonological knowledge for identifying the non-native syllable structures, the results for the possible coda sequence /ip/ should be better than the ones for /ib/, which is ungrammatical coda sequence. However there are no differences in structure perceptions for /ib/ and /ip/ within each listener group. Voicing perception for /ib/ is better than ones for /ip/. This is contrary to the PAM prediction that /ip/ is expected to be better than /ib/. Therefore, we can say that the non-native pair of VC and CV are assimilated to a UC type.

The results from the three groups of Japanese listeners are against the expectation that the more experienced listeners can perform well. Instead, the least and the most experienced listeners perceive the VC stimuli in the same way, that is, they showed a bias toward the CV perception for VC stimuli. On the other hand, the less experienced Japanese younger listeners do not exhibit the tendency toward CV. Although it does not reach statistical significance, Japanese younger listeners were more likely to respond as a VC when the stimuli became difficult to judge. It can be considered that this is due to a response bias toward the more 'foreign' or 'odd' category by the non-native listeners. The bias toward odd categories is also found in Japanese /l/-/r/ perceptions. A bias toward /r/ in /l/ -/r/ perception [9] can be due to a special attention toward /r/-pronunciation common among Japanese learners of English. Sometimes in the limited context, inexperienced listeners are 'better than' the experienced due to external biases toward 'foreign' category.

5. CONCLUSION

Perception of non-native syllable structures can be acquired without an extensive exposure to the non-native language. However, perception of a native language influences non-native syllable structures even for the experienced listeners. Non-native categories initially categorized as 'foreign' or uncategorized can be re-categorized or assimilated to a close native category.

ACKNOWLEDGMENTS

This work is supported by the NIDCD (grant # R03DC04095-01A2) and the NSF (grant # BCS-9910701). The authors thank Minoru Fukui for helping us to conduct the experiment in Japan.

REFERENCES

[1] 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*, W. Strange, Ed. Baltimore: York Press, 1995, pp. 305-320.

[2] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience*, W. Strange, Ed. Baltimore: York Press, 1995.

[3] Y. Homma, "Durational relationship between Japanese stops and vowels," *Journal of Phonetics*, vol. 9, pp. 273-281, 1981.

[4] L. Lisker, "Closure duration and the intervocalic voiced-voiceless distinction in English," *Language*, vol. 33, pp. 42-49, 1957.

[5] K. de Jong, B.-j. Lim, and K. Nagao, "The Interaction of Syllabification and Voicing Perception in American English," *IULC Working Papers Online*, vol. 1, pp. 1-12, 2001.

[6] R. H. Stetson, *Motor Phonetics: A Study of Speech Movements in Action*. Amsterdam: North-Holland, 1950.

[7] B.-j. Lim, K. de Jong, and K. Nagao, "Cross-language perception of syllable affiliation: effects of voicing and language background," *BLS*, vol. 27, 2001.

[8] K. Nagao, B.-j. Lim, and K. de Jong, "Perception of rate induced resyllabification: Cross-language comparison," *Indiana University Working Papers in Linguistics*, vol. 3, 2002.

[9] A. R. Bradlow, D. B. Pisoni, R. Akahane-Yamada, and Y. i. Tohkura, "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," *JASA*, vol. 101, pp. 2229-2310, 1997.