Effect of Dynamic Information of Formants on Discrimination of English Vowels in Consonantal Contexts by Japanese Listeners

Akiyo Joto

Department of Intercultural Studies
Prefectural University of Hiroshima, Hiroshima, Japan
joto@pu-hiroshima.ac.jp

Abstract

This study examined how differently native speakers of Japanese discriminated between the American English vowels /e/ and /ae/ in /CVp/ syllables with 20 different initial consonants, and how the differing discrimination was related to the formant changes throughout the vowels in comparison with the Japanese vowels /e/ and /a/. A perceptual test and formant analyses of the English and Japanese vowels were conducted. The results showed that there were significant differences in discrimination across the consonantal contexts: the discrimination of /e/ was significantly poorer when the initial consonant was /dg/, /g/, /ch/, /m/ or /th/ (voiceless), and that of /ae/, when it was /sh/, /h/ or /t/. It was found that the poorer discrimination was more related to the smaller formant ratios (F2/F1) and the higher F1 frequency in the latter part of /e/, and to the larger formant ratios and the lower F1 frequency in the latter part of /ae/. The changing of formant patterns throughout the two English vowels could be attributed to the poorer discrimination in the particular consonantal contexts.

Index Terms: vowel discrimination, cross lingual, Japanese

1. Introduction

The recent studies of the cross-language perception of vowels have been more concerned with vowels in consonantal contexts. Strange et al. (2001) investigated the perceptual assimilation of American English vowels in six consonantal contexts by Japanese listeners [10]. Bohn and Steinlen (2003) examined the identification of British English vowels in three consonantal contexts by native Danish listeners [2]. The results of both studies suggested that perceptual difficulties of non-native vowels would significantly vary with the phonetic contexts in which vowels were presented. More than ten years before their investigations, Joto (1988) found that the discrimination of the American English (AE) vowel /i/ from /e/ by Japanese listeners significantly differed across various consonantal contexts [4]. She also reported the discrimination between /e/ and /ae/ by Japanese listeners differed across the different initial consonants in the monosyllabic consonantal contexts /CVt/ and /CVd/ and that the differing discrimination was related with the changing formant patterns throughout the vowel nuclei [7], [8].

The English language has more than four times as many vowel phonemes as the Japanese language with the five vowel system. Some of the English vowel pairs are, therefore, difficult for native speakers of Japanese to discriminate. The AE mid-low front vowel /e/ is sometimes articulated very close to the low front vowel /ae/ even under the condition without the effect of on-going regional sound changes, the Northern Cities Shift. Strange et al. (2001), based on their experiment on perceptual assimilation, stated that the AE vowels /e/ and /ae/ were considered 'poor instances of any one Japanese category' and could be one of the pairs constituting 'Category Goodness' or 'Uncategorized versus Categorized' assimilation patterns [10]. According PAM (the perceptual assimilation model) proposed by Best (1995), the discrimination of this vowel pair would be expected to present intermediate difficulties [1].

This paper is concerned with whether and how the Japanese listeners discriminate between these two AE vowels in the monosyllabic consonantal context /CVp/ and how the discrimination is related with the formant patterns throughout the vowel duration. It also examined the formant patterns of the Japanese vowels /e/ and /a/ for comparison with the AE vowels to find out whether the language interference affects the discrimination.

Two kinds of experiments were carried out to clarify the problems: the first experiment was a perceptual test to examine the differences in the discrimination between the English vowels in different consonantal contexts by native speakers of Japanese, and the second one was an acoustical analysis of the English and Japanese vowels to investigate the relation between the vowel discrimination and the formant patterns of the vowels in various consonantal contexts. The changes of the formant patterns throughout the vowels were taken into consideration in the analysis because they can be considered as one of the important acoustic factors for vowel perception.

2. Experiment 1

A perceptual test was performed with native speakers of Japanese to investigate how differently they discriminated the English vowels /e/ and /ae/ in different consonantal contexts where the final consonant was /p/.

2.1. Stimulus materials

The stimulus corpus consisted of both meaningful and nonsense monosyllabic English words which were composed of 20 sets of word pairs containing the vowel /e/ or /ae/ in the context of 20 different initial consonants with the final consonant of /p/. The initial consonants were /b/, /k/, /ch/, /d/, /f/, /g/, /h/, /dg/, /l/, /m/, /n/, /p/, /r/, /s/, /sh/, /t/, /th/

(voiceless), /th/ (voiced), /v/ and /z/. They were produced in random order by each of three male speakers of American English (General American). All the speakers were English teachers at a university in Japan. Two of them were in their thirties and one was in his forties. Their productions were recorded in an anechoic chamber, using an electret condenser microphone connected to a DAT recorder. The words of each speaker's production were sampled at 11kHz and randomly ordered with each word produced by each speaker presented twice on a computer for the experiment.

2.2. Listeners

Thirty-three female native speakers of Japanese participated in the perceptual test. They were all university students, whose age ranged from 19 to 20 years old, and had received standard English classes in junior and senior high schools in Japan. None of them had any experience of living in another country for six months or more, or had received education at any schools where instruction was done only in English. All the listeners were phonetically naïve. None of the participants reported having any hearing disability.

2.3. Procedures

Prior to the test, the listeners received instructions in Japanese. Each word was presented twice per each speaker's production, but not in serial order. They listened to 240 words in total (2 vowels x 20 consonants x 3 speakers x 2 times) through loudspeakers in a language laboratory. The stimuli were presented to the listeners once in random order at 2 second interstimulus intervals. They circled the word they heard on the sheets where 120 sets of word pairs were printed in alphabet letters. Each of the two test vowels was spelled with the same symbol in each of the words: the letter 'a' in the listed words represented the vowel /ae/ as in 'cat,' and the letter 'e,' the vowel /e/ as in 'bed.' This explanation was given to the listeners before the test.

2.4. Results

Table 1 and 2 show the error rates of the discrimination between /e/ and /ae/ in different initial consonantal contexts by the Japanese listeners. The overall error rate of /e/ (34.3%) was significantly higher than that of /ae/ (11.1%) (chi square test: $p<0.001$). The Japanese listeners, therefore, more often misheard /e/ for /ae/. In the previous studies by the author, the same results were found in the contexts /CVt/ and /CVd/ [7], [8]: the error rate of /e/ was significantly higher than that of /ae/.

The discrimination between the vowels, however, differed across the consonantal contexts. The error rate was significantly higher ($p<0.01$) when the initial consonant was /dg/, /g/, /ch/, /m/ or /th/ (voiceless) for /e/ and when it was /h/, /sh/ or /t/ for /ae/, while it was significantly lower ($p<0.01$) when the initial consonant was /b/, /d/, /h/ or /t/ for /e/ and when it was /z/ or /m/ for /ae/. The initial consonantal contexts in which the discrimination was better or poorer were different from the results in the contexts of /CVt/ and /CVd/. Thus it indicates the discrimination depends on both initial and final consonantal contexts.

Table 1: *Error rates of /ae/ in different consonantal contexts*

initial consonant	error rate (%)
sh	22.2
h	21.7
t	20.2
p	15.7
f	15.2
th (voiceless)	14.6
k	14.1
l	13.6
th (voiced)	13.1
g	11.6
b	10.6
s	9.1
ch	8.6
dg	8.6
d	8.1
n	7.6
r	7.6
v	7.6
z	5.1
m	4.0
mean	11.1

Table 2: *Error rates of /e/ in different consonantal contexts*

initial consonant	error rate (%)
dg	61.1
g	58.6
ch	51.5
th (voiceless)	46.0
m	45.5
k	41.4
th (voiced)	39.4
r	38.9
v	35.4
s	32.8
l	32.3
n	30.8
sh	28.8
z	27.3
f	26.3
p	26.3
d	22.2
b	21.7
h	21.2
t	21.2
mean	34.3

3. Experiment 2

Acoustical analyses were conducted on the English vowels /e/ and /ae/ and also on the Japanese vowels /e/ and /a/ in different consonantal contexts. The interference of the Japanese sounds was examined by comparing acoustic characteristics of the English vowels with those of the Japanese ones. The changes of formant patterns throughout the vowels were analyzed in order to find out which part of the vowel duration was most strongly related to the discrimination.

3.1. Speech materials

The same speech materials as in experiment 1 were used for experiment 2 for English. The Japanese materials were the disyllabic/tri-moraic meaningful and nonsense words containing the vowel /e/ or /a/ with a final syllable of /to/. The vowels /e/ and /a/ were chosen as Japanese vowels for this experiment because they are, according to Strange et al., the sounds to which the English vowels /e/ and /ae/ are perceptually most similar respectively [9]. The words to be analyzed were of three moras/two syllables (/C_1V_1 $C_2C_2V_2$/) which were presumed to be most equivalent to the English materials because an English closed syllable sounds like geminate consonants plus a vowel to Japanese ears: for

example, /bet/ is phonetically translated as /betto/ in Japanese. Some initial consonants do not have parallel sounds in the two languages, or do not generally appear before /e/ or /a/ in a Japanese mora/syllable. Therefore, the number of initial consonants in the Japanese words was 16 for /e/ (/k/, /g/, /sh/, /dg/, /s/, /dz/, /ch/, /t/, /d/, /n/, /hw/, /h/, /b/, /p/, /m/, /r/) and 18 for /a/ (/k^{j}/, /g/, /s/, /dz/, /sh/, /dg/, /t/, /d/, /ch/, /n/, /h/, /b/, /p/, /hw/, /m/, /j/, /r/, /w/).

The Japanese words were produced by six male native speakers of the standard Japanese (Kanto dialect). Their ages ranged from 18 to 26. For the recording, they were given the instruction to produce each word with an accent on the first syllable, so that pitch fell on the last mora/syllable.

3.2. Method

The English and Japanese vowels in each consonantal context were acoustically analyzed by computer, using speech analysis software. In order to examine the temporal changes of formant frequencies, F1 and F2 frequencies were measured at the temporal locations 25%, 50% and 75% throughout the vowel portion of each word on the display of LPC formant tracks with a 25 ms hamming window employed. The ratios of the F2 frequency to the F1 frequency (henceforth, formant ratios) at each temporal location were calculated on the basis of the measurements of formant values.

3.3. Results

Table 3 shows the changes of the formant patterns throughout each English and Japanese vowel. In each English vowel, F1 rose and F2 fell throughout the vowel duration, hence, the formant ratios decreased in the latter part of the vowel portion. In the Japanese vowel /e/, both F1 and F2 slightly went up at the midway of the vowel and went down at the 75% temporal location. Thus the formant ratio slightly increased in the latter half of the vowel nucleus.

Table 3: *Mean formant frequencies in Hz and formant ratios of the English and Japanese vowels at three temporal locations.*

temporal locations		English		Japanese	
		/e/	/ae/	/e/	/a/
25%	F1	594.3	720.5	468.8	628.7
	F2	1674.1	1594.7	1837.8	1369.5
	F ratio	2.82	2.21	3.92	2.18
50%	F1	663.8	789.3	482.8	663.8
	F2	1611.3	1555.7	1845.1	1288.8
	F ratio	2.43	1.97	3.82	1.94
75%	F1	694.6	832.8	447.7	630.2
	F2	1535.8	1491.3	1803.8	1275.1
	F ratio	2.21	1.79	4.03	2.02

As for /a/, F1 rose at the midway and fell at the 75% temporal location, and F2 decreased throughout the vowel. The formant ratio increased a little at the 75% temporal location. The magnitude of rising and falling of Japanese vowel formants was not so great as that of English ones: both F1 and F2 formant frequencies of English vowels changes more than 100Hz throughout the vowel, whereas the changes of Japanese vowel formants were less than 100 Hz, and the formant patterns did not seem to change much throughout the Japanese vowels.

4. Discussion

Strange reported that the vowel pair of /e/ and /ae/ is considered to be "Category-Goodness" or "Uncategorizable vs Categorized" Type, according to Best's PAM model, and that it would be expected to be of intermediate perceptual difficulty [1], [10]. However, experiment 1 indicated that the two vowels were not discriminated equally (11.1% for the mean error rate of /ae/, and 34.3% for /e/). Moreover, each vowel was discriminated differently across the consonantal contexts. The error rates of /e/ across the consonantal contexts varied from the highest of 61.1 % to the lowest of 21.2 %; the discrimination ranged from 'poor' to 'good.' As for the discrimination of /ae/, the error rates were also dependent on the initial consonant, from the highest of 22.2% with the consonant of /sh/ to the lowest of 4.0% with /m/.

In order to clarify the relation of the differing discrimination to the formant changes, the correlation was measured between the error rates, the formant ratios of /e/ and /ae/ and F1 and F2 frequencies of /e/ and /ae/ at each temporal location.

Table 4: *Correlation coefficients of error rates with F1 and F2 frequencies and formant ratios of /e/ and /ae/ at three temporal locations.*

temporal locations		/e/	/ae/
25%	F1	0.12	-0.29
	F2	0.05	0.26
	F ratio	0.05	0.32
50%	F1	0.39	-0.51
	F2	0.1	0.07
	F ratio	-0.22	0.45
75%	F1	0.56	-0.58
	F2	0.25	-0.38
	F ratio	-0.39	0.38

As shown in Table 4, a stronger positive correlation between the formant ratios and the error rates and a stronger negative correlation between the error rates and the F1 frequencies were seen at the 50% and 75% temporal locations of the vowel /ae/. It indicates the discrimination of /ae/ becomes poorer when the formant ratios are larger and the F1 frequencies are lower in the latter part of the vowel portion, which implies /ae/ is closer to /e/ in the values of formant ratios and the F1 frequency. Therefore, the lower F1 frequencies were related to the changing of the formant patterns in the latter part of the vowel.

In fact, the formant ratio at the 75% temporal location was bigger (1.78) and the F1 was lower (831.9Hz) in the high

error group of /ae/ (/sh/, /h/, /t/) than in the low error group (/m/, /z/) (1.77 and 849.7Hz, respectively).

As for /e/, however, a negative correlation was found between error rates and the formant ratio at the 75% temporal location. This means that the discrimination of the vowel becomes poorer when the formant ratios are smaller in the latter half of the vowel portion. In addition, a stronger positive correlation was seen in the F1 frequency at the 75% temporal location. It demonstrates that the discrimination of /e/ becomes poorer when the formant ratio is smaller and also when the F1 frequency is higher in the latter part of the vowel.

In the high error group of /e/ (/dg/, /g/, /ch/, /m/, /th/ (voiceless)), the mean formant ratio actually was smaller (2.18) than in the low error group (/b/, /d/, /h/, /t/) (2.20). The mean F1 frequency was higher (705.5 Hz) in the high error group than in the low error group (691.2 Hz).

Thus, the poorer discrimination of /e/ in the particular consonantal contexts was correlated with the decrease of the formant ratios and the increase of F1 frequency in the latter part of the vowel which implies the lowering of the tongue height. On the other hand, the poorer discrimination of /ae/ was related to the bigger formant ratios and the decrease of F1 frequency in the latter half of the vowel duration which means the raise of the tongue height in articulation. In other words, the discrimination was poorer in the particular consonants when the two vowels became closer to each other in the latter part of each vowel portion.

From the viewpoint of the language interference, the poor discrimination does not seem to be attributed to the acoustic similarity or dissimilarity between the English vowels and the Japanese ones. It is suggested that compared with Japanese vowels which undergo little formant changes, the greater changes of formants throughout the English vowels could be one of the factors which make it more difficult for Japanese listeners to perceive the English vowels.

In the teaching of English sounds in Japan, there has been little attention or even realization of the problem of the contextual dependent differences in the vowel discrimination. The results of our experiments, however, clearly indicate that vowels in consonantal contexts should be presented to learners for their better discrimination, which could lead to their better communicative ability in English. Practically, learners need more exposure to the training of vowel discrimination in the various consonantal contexts with an emphasis on the contexts with poorer discrimination.

The results also suggest that the studies of the cross-lingual perception of vowels, which have not extensively dealt with variation in consonantal contexts, should be more concerned with vowels in consonantal contexts for a more elaborate description of vowel perception.

5. Conclusions

It was found that there were significant differences in the discrimination between AE vowel /e/ and /ae/ across the consonantal contexts by Japanese listeners. It is suggested that this variability of discrimination could be attributed to the dynamic changes of formants throughout the vowels: the decrease of formant ratios and the increase of F1 frequency in the latter part of the vowel /e/, and the greater formant ratios and the decrease of F1 frequency in the latter part of the vowel /ae/ had more effect on the poorer discrimination between the two English vowels in the particular consonantal contexts.

6. Acknowledgements

This study was partially supported by Grant-in -Aid for Scientific Research on Priority Areas (A) from the Ministry of Education, Culture, Sports, Science and Technology in Japan.

7. References

[1] Best, C. T., "A Direct Realist View of Cross-Language Speech Perception," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. by W. Strange, York Press, Baltimore:171-204, 1995.

[2] Bohn, Ocke-Schwen and Steinlen, Anjak. "Consonantal Context affects Cross-Language Perception of Vowels," Proc. of 15th ICPhS, 2289-2292, 2003.

[3] Joto, A., "Nihonjin-gakushuusha no Shiin-myakuraku niyoru Eigo-boin /e/ no Shikibetsuryoku (Discrimination of the English Vowel /e/ in Consonantal Context by Japanese Learners)", *The Bulletin of the Phonetic Society of Japan*, no. 186: 15-18, 1987.

[4] Joto, A., "Nihonjin-gakushuusha no Shiin-myakuraku no Eikyoo: CVC Onsetsu niokeru /i/ to /e/ nitsuite (Discrimination of the English Vowel /i/ in Consonantal Context by Japanese Learners)", *Persica,* no.15: 1-10, 1988.

[5] Joto, A. "The Effect of Consonantal Context and Acoustic Characteristics on English Vowel Discrimination by Japanese Learners", *AILA '99 (12th World Congress of Applied Linguistics) Program Abstract*: 382-383, 1999.

[6] Joto, A. "The Effect of Consonantal Context and Acoustic Characteristics on the Discrimination between the English Vowel /i/ and /e/ by Japanese Learners", *Proc. of ICSLP 2000* Vol. III: 957-960, 2000.

[7] Joto, A., Imaishi, M., Nagase, Y., and Funatsu, S. "Discrimination of English Vowels in Consonantal Contexts by Native Speakers of Japanese and its Relations to Dynamic Information of Formants," Proc. of 7th ICSLP, vol. 2, 1133-1136, 2002.

[8] Joto, A., Imaishi, M., Nagase, Y., and Funatsu, S. "Effect of Consonantal Context on the Discrimination of the American English Vowel /ae/ by Japanese Listeners," English Phonetics, vol. 6, 259-273, 2003

[9] Strange, W., Akahane-Yamada, R., Kubo, R.,Trent, S. A., Nishi, K., and Jenkins, J. J. "Perceptual Assimilation of American English Vowels by Japanese Listeners," *Journal of Phonetics* 26: 311-344 , 1998.

[10] Strange, W., Akahane-Yamada, R., Kubo, R.,Trent, S. A. and Nishi, K. "Effects of Consonantal Context on Perceptual Assimilation of American English Vowels by Japanese Listeners," JASA, vol. 109, no.4, 1691-1704, 2001.