

PRODUCTION AND PERCEPTION OF THE RHYTHMIC PATTERN  
OF ENGLISH BY JAPANESE LEARNERS

Michiko Mochizuki-Sudo and Shigeru Kiritani

1. Introduction

Traditionally, languages are classified into various rhythmic types according to the difference in their rhythmic units: stress-timed, syllable-timed or mora-timed. In the case of stress-timed languages, the ISI (interstress interval) is considered to tend to keep equal durations, while the syllable- or mora-timed languages maintain equal syllable or mora durations.

Several experimental studies report the measurements of ISIs in English. In these studies, the duration of ISIs shows a considerable variability, and perfect isochrony on the level of production does not exist. Therefore, on the basis of these acoustic measurements, some researchers are opposed to the dichotomy between stressed-timed and syllable- or mora-timed languages. In a stress-timed language, however, there is a phenomenon which contributes to the lessening of the increase in the duration of ISIs due to the addition of unstressed syllables. A compensatory shortening of a stressed vowel with the addition of following unstressed syllables has been reported in English. The studies of Huggins (1975) and Fowler (1977) report that there exists what is called the "compression effect" in English. Huggins and Fowler's data reveal that "a stressed vowel is compressed in anticipation of a subsequent unstressed syllable or syllables in the ISI" (Fowler, p34). A compensatory effect is also reported for perception. Huggins (1972) reports that synthetic speech tends to remain temporally fluent when a change made in the duration of one segment is compensated for in an adjacent segment across a syllable or word boundary. These studies suggest that ISI is a temporal unit in English.

The purpose of this study is to compare the durational control of English in the English performance of native speakers of American English, a stress-timed language, to that of Japanese learners of English whose native language is mora-timed. The present paper consists of two parts, experiments on production and the perception of durational control in English by native speakers of American English and Japanese learners. The production experiment deals with the duration of an ISI and the compression effect of a stressed vowel. Specifically, the compression effect in the sentences produced by Japanese learners of English (good and poor learners) will be compared to that of American speakers. What we are interested in here is how learners of different proficiencies deal with the compression of stressed vowels as compared to native speakers. The second part of the paper deals with the perception of duration in English by native speakers and Japanese learners. The perceptual experiments are designed to study the discrimination of durational differences in vowels, and the influence of vowel duration on the perception of naturalness of temporal patterns.

## 2. Production experiment

### 2.1. Subjects

Three groups of speakers served as subjects: native speakers of American English, Japanese good learners of English(JGL) and Japanese poor learners of English(JPL). All five American subjects were undergraduate students at the University of Pennsylvania. Two groups of Japanese subjects (five subjects in each group) were selected on the basis of their speaking proficiency in English. The JPL were selected from a class at the lowest speaking level in the language program for non-native speakers at the University of Pennsylvania. Their stay in the U.S. ranged from one month to ten months at the time of our recording. The JGL were selected from among graduate students in the TESOL (Teaching English to Speakers of Other Languages) program of the Department of Education or from the Department of Linguistics at the University of Pennsylvania. Their recordings of the linguistic materials for the production experiment were presented to an instructor of the language program. Each was evaluated as being a much better speaker than the students of the highest level in the language program. The stay of the JGL subjects in the U.S. ranged from two to four years.

### 2.2. Linguistic materials

Four groups of English sentences (a total of 22) were devised for the purpose of this production experiment. The sentences in each group differed in the number of unstressed syllables that intervened between a target stressed syllable and the next stressed syllable. The number of unstressed syllables in all the groups ranged from zero to three. In other words, the number of the syllables in the target ISI ranged from one to four. The four groups of sentences are given below.

#### Group 1

1. They bought books by the carload.
2. They bought the books by the carload.
3. They bought up books by the carload.
4. They bought me books by the carload.
5. They bought up the books by the carload.
6. They bought up my books by the carload.
7. They bought me the books by the carload.
8. They bought up all my books by the carload.

#### Group 2

1. Pete plays the piano.
2. Pete can play the piano.
3. Peter plays the piano.
4. Peter can play the piano.
5. Peterson plays the piano.
6. Peterson can play the piano.

#### Group 3

1. A dog house is a good idea.
2. A doggy house is a good idea.
3. A dog in the house is a good idea.
4. A doggy in the house is a good idea.

#### Group 4

1. We need a dog house.
2. We need a doggy house.
3. We need a dog in the house.
4. We need a doggy in the house.

As shown in this list, the four groups of sentences contained different target stressed vowels: [ɔ], [i] and [ʌ]. In addition, some sentences were devised to be different regarding the position of the target ISI or in containing ISIs which differed in the presence or absence of a word boundary.

### 2.3. Procedures

The utterance list of twenty-two sentences was arranged in a pseudo-random order. Stress placements were marked in the text in order to obtain the same sentence stress patterns for all subjects. They were instructed to produce the sentences at a comfortable speaking rate and to read each sentence through without pausing. When they inserted a pause or made a mistake, the subjects were asked to repeat the sentence from the very beginning. They repeated the text six times. Before the recording, they were provided sufficient time for familiarization and practice.

For each subject, five repetitions of each sentence (a total of 110) were selected from the total of six repetitions uttered. Broad-band spectrograms were made from the tape recordings. Then we measured the durations of the target stressed vowels, [ɔ] in the sentences of Group 1, [i] in those of Group 2, and [ʌ] in Groups 3 and 4. The onset of the vowels was defined as the instant a sharp rise appeared in the power of the first formant. Since all the target vowels were followed by a stop, [t] or [g], the offset of the vowels was defined as the abrupt decrease in power corresponding to the stop closure. Furthermore, we measured the duration of the target ISI, i.e., the interval between the onset of the target stressed vowel and that of the next stressed vowel.

### 2.4. Results

Figure 1(a) shows the average durations of the ISIs for the sentences of Group 1 produced by the American subjects, the Japanese good learners and the Japanese poor learners. As shown in this figure, the duration of the 1-syllable ISI was shortest for the Americans, longer for the Japanese poor learners (JPL) and the longest for the Japanese good learners (JGL). For the 2-, 3- and 4-syllable ISIs, the shortest ISIs were also produced by the Americans, yet the order of the JPL and the JGL was reversed, i.e., shorter ISIs for the JGL than for the JPL.

Turning to the rate of increase in the durations of the ISIs

with the addition of unstressed syllables for the sentences of Group 1, we observed a difference among these three groups of subjects. The average rates of increase in duration from a 1-syllable ISI to a 4-syllable ISI were 2.5 for the Americans, 2.8 for the JGL and 3.3 for the JPL. The durations of the ISIs in Group 1 become longer at the highest rate for the JPL, at a lower rate for the JGL and at the lowest rate for the Americans. These rates of increase can be considered to reflect the generally-claimed longer durations of unstressed syllables by the JPL.

As we have seen in Group 1, the rates of increase in the duration of an ISI for Groups 2, 3 and 4 are lower for the native speakers and the JGL than the JPL. The average rates of increase in duration from a 1-syllable ISI to a 4-syllable ISI for these groups of sentences are 2.2 for the Americans and the JGL and 2.7 for the JPL.

In addition to the rates of increase in the durations of the ISIs, we found a difference in the durational control of stressed vowels among the subject groups. Figure 2(a) shows the average durations of target stressed vowels by the Americans, the JGL and the JPL for the sentences of Group 1. As shown in the figure, the compression effect of a stressed vowel is much more conspicuous in the native speakers and the JGL compared to the JPL. The percentages of compression for stressed vowels from a 1-syllable ISI to a 2-syllable ISI in Group 1 are 9% for the Americans, 19% for the JGL and 1% for the JPL. Here, we can observe an exaggeration of the compression for the good learners. Thus, it is confirmed that not only the longer durations of unstressed syllables but also the small percentage of compression in the duration of the stressed vowels contributes to the high rate of increase in the duration of ISIs for the JPL.

A similar tendency was also observed for other sentence groups. The percentages of compression for a stressed vowel from a 1-syllable ISI to a 2-syllable ISI in Group 2 are 21% for the Americans, 24% for the JGL and 7% for the JPL. Likewise, 22%, 18% and 8% compressions were observed in Group 3 for the Americans, the JGL and the JPL, respectively. The stressed vowels in Group 4 also showed a 22%, 18% and 12% shortening from a 1-syllable ISI to a 2-syllable ISI. Therefore, the percentage of compression in a stressed vowel was higher for the native speakers and the JGL than the JPL in all sentence groups.

However, an exception to this compression effect was observed for the vowels of 3- and 4-syllable ISIs in Groups 3 & 4. The vowel durations of 3-syllable and 4-syllable ISIs were longer than the vowels of 2-syllable ISIs for all three groups. When there was a phrase boundary between the target vowel and the following unstressed syllable, the duration of the stressed vowel was not shortened.

When we compare the Americans and the Japanese, the Japanese produced the target vowel of Groups 1 & 2 with a longer duration than the Americans, while the Americans produced the target vowel of Groups 3 & 4 with a longer duration. This result can be considered to stem from the difference in the target vowels. The Japanese produced a longer [ɔ] and [i] than the Americans, while the former produced a shorter [ʌ] than the latter. In other words, the Japanese learners are relying heavily on durational

difference as a distinction between the so-called long and short vowels in American English.

### 3. Perceptual experiment--influence of the lengthening and shortening of a stressed vowel on the perception of the naturalness of temporal patterns

#### 3.1. Procedure

The preliminary experiment confirmed that Japanese listeners are not less sensitive than the Americans in discriminating the durations of stressed vowels. In this perceptual experiment, the influence of the lengthening and shortening of a stressed vowel on the perception of the naturalness of temporal patterns was measured using LPC analysis-synthesis speech. The seven original sentences used are shown below.

1. A dog house is a good idea.
2. A doggy house is a good idea.
3. We need a dog house.
4. We need a doggy house.
5. Pete plays the piano.
6. Peter plays the piano.
7. Peterson plays the piano.

These sentences were selected on two bases: use of different vowels and the different positions of the target words. The duration of a stressed vowel in each of the seven English sentences, [ʌ] in "dog" and "doggy" and [i] in "Pete", "Peter" and "Peterson", was varied in 7 steps from -47% to +79% of the original duration. In each of the seven groups of sentences, the subjects heard 49 sentences (7 repetitions of different versions of one original sentence). The task of the subjects was to judge the naturalness of the temporal pattern in each sentence.

The percentages of lengthening and shortening were decided based on the results of a preliminary experiment in which the vowel duration in the sentence, "A dog house is a good idea", was varied over a range of -120 ms and +200 ms. In the preliminary experiment, +100 ms (+53%) and -60 ms (-32%) changes in the vowel duration produced a lower than 50% naturalness in responses. Since the original vowel durations in the seven sentences differed, the percentage, and not the absolute duration, was used to manipulate the duration (otherwise, some shortenings would produce a below 0 duration).

There were seven groups of sentences in this experiment, therefore, The subjects heard 343 sentences in total. After every fifth sentence, a marker sound was inserted. The sentences were separated by an interval of 2 seconds. The subjects were given a 10-minute rest after the fourth group and a short break in the other cases.

#### 3.2. Subjects

Five Americans and ten Japanese served as subjects. Two sessions were held on different days for each of the subjects in

an anechoic room at the University of Tokyo. The American subjects were either college students or college graduates in their twenties. All of the Japanese subjects were freshmen at Tamagawa University who had neither lived nor studied abroad before. Before the perceptual experiments, the Japanese subjects were asked to read the text of the seven English sentences used for the perceptual experiments. They repeated the text five times. Broad-band spectrograms were made from the tape recordings. Measurements were then obtained for the durations of the target stressed vowels and those of the ISIs. Their results show a similarity to the JPL in the production experiment. The high rate of increase in the duration of the ISIs and the small percentage of compression for stressed vowels, with the addition of unstressed syllables, are nearly alike.

### 3.3. Results

Figure 3 shows the results for the influence of the duration of a stressed vowel [i] on the perception of the naturalness of the temporal pattern in the sentence "Pete plays the piano". The result for the American subjects is indicated by a solid line and the Japanese by a dotted line. The average percentages of "natural" judgments for each group of subjects are plotted against the percentages of the shortening or lengthening of the target stressed vowel.

We can observe in Figure 3 that the temporal naturalness of the sentences deteriorates as the vowel duration undergoes a change. The more change the vowel undergoes, the more the deterioration of naturalness. When the target vowel is lengthened by 53%, the "natural" judgment for the Americans decreases to 31%, while that for the Japanese remains as high as 86%. In this sentence, where the target vowel is [i], the native speakers react more sensitively to the lengthening of the target vowel than the Japanese, while the latter react more sensitively to shortening than the former. When the vowel is shortened by 32%, the "natural" judgment for the Japanese is as low as 33%, while that for the Americans is still 77%. In addition, the peak of the "natural" judgments is different for the Americans and the Japanese. The peak of the "natural" judgments for the native speakers is at the original vowel duration, while that for the Japanese is at a lengthening of 26%.

The estimates for a 50% threshold of "natural" judgments are shown in Table 1. The data for the natural judgments were approximated by the cumulative normal distribution, and the threshold values were determined by the method of least squares. As shown in the table, the Americans respond more sensitively to the lengthening of a stressed vowel in their judgment of naturalness for all the sentences: the naturalness of the temporal pattern degrades with less of a lengthening of the target vowel, compared to the Japanese subjects. In the case of vowel shortening, we find a difference in the relative sensitivity of the Americans and the Japanese between the target vowels [ʌ] and [i]. The Japanese need less shortening than the Americans for the vowel [i] to reach the 50% threshold. On the other hand, the Americans need less shortening to reach the 50%

threshold than the Japanese for the vowel [ʌ]. This difference is related to the difference in the peak of "natural" judgments for [i] and [ʌ]. The natural duration for the Japanese shifted to a longer duration in the case of [i], while it shifted to a shorter duration for [ʌ].

It should be noted that the range of naturalness is narrower for the Americans in all the sentences but one, even though the Japanese are more sensitive to the shortening of the vowel [i] than the Americans. The range of the natural-sounding stressed vowel is much narrower for the native speakers than the Japanese.

#### 4. Discussion

It has been generally stated, and confirmed by some studies that Japanese learners of English can neither produce short enough unstressed syllables nor reduce vowels. These longer durations of unstressed syllables by Japanese learners most likely contribute to their higher rate of increase in the durations of ISIs. Actually, the results of the production experiment in this study confirmed that the rate of increase in the durations of ISIs due to the addition of unstressed syllables is highest for the JPL, lower for the JGL and lowest for the native speakers of American English.

However, the longer durations of unstressed syllables are not the sole factor contributing to a higher rate of increase in the durations of ISIs. The present study revealed that Americans and Japanese good learners show a compression effect in the duration of a stressed vowel within the ISI with the addition of unstressed syllables, while the Japanese poor learners demonstrate very little compression effect. It has been noted in some studies that Americans control the duration of an ISI using relatively short durations of unstressed syllables and the compression of stressed vowels in the direction of minimal lengthening when unstressed syllables are added. The results of this study show that good learners control the duration of an ISI, though to a lesser degree, in a similar way as native speakers.

On the perceptual side, it was noted in the preliminary experiment that Japanese are not less sensitive than Americans in discriminating the durations of stressed vowels. In spite of this ability to discriminate, the Japanese range for the natural duration of a stressed vowel was found to be broader than for Americans. When Americans decide the naturalness of temporal patterns, they are more sensitive than Japanese to the change of the vowel duration and a resulting change in the duration of ISIs. Thus, this study points out that a different durational pattern for Japanese learners, compared to native speakers, results not only from difficulty in controlling articulatory movements, but also from the inability of the Japanese to make perceptual judgments of natural durational patterns. Further studies are now being conducted on the perceptual effect of compensatory changes in the durations within ISIs.

## References

- Fowler, C. A. (1977) Timing control in speech production. Ph. D. thesis, University of Connecticut.
- Fujisaki, H. & Sugitou, M. (1977) Physical properties of speech. In Japanese 5: Phonemes (S. Ono & T. Shibata, editors), pp. 63-106. Tokyo: Iwanami.
- Huggins, A. W. F. (1972) On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America*, 51, 1279-1290.
- Huggins, A. W. F. (1975) On isochrony and syntax. In *Auditory analysis and perception of speech* (G. Fant & M. A. A. Tatham, editors), pp. 455-464. London: Academic Press.
- Lehiste, I. (1972) The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018-24.
- Lehiste, I. (1973) Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, 54, 1228-1234.
- O'Connor, J. D. (1965) The perception of time intervals. Progress Report, Phonetics Laboratory, University College, London, 2.11-15.
- Prator, Clifford H. (1957) *Manual of American English Pronunciation*. New York: Holt, Rinehart and Winston.
- Shen, Y. and G. G. Peterson (1962) Isochronism in English. University of Buffalo. *Studies in Linguistics, Occasional Papers*, 9, 1-36.
- Takefuta, Y. (1973) A study of rates in English. *Studies of Sounds*, 16, 231-247.
- Uldall, E. (1971) Isochronous stresses in R. P., In *Form and substance: Phonetic and linguistic papers presented to Eli Fischer-Jorgensen* (L. L. Hammerich, Roman Jakobson & Eberhard Zwirner, editors), pp. 205-210. Copenhagen: Akademisk Forlag.
- Yoshida, K. et al. (1984) *Listening in English*. Tokyo: Taisyukan.



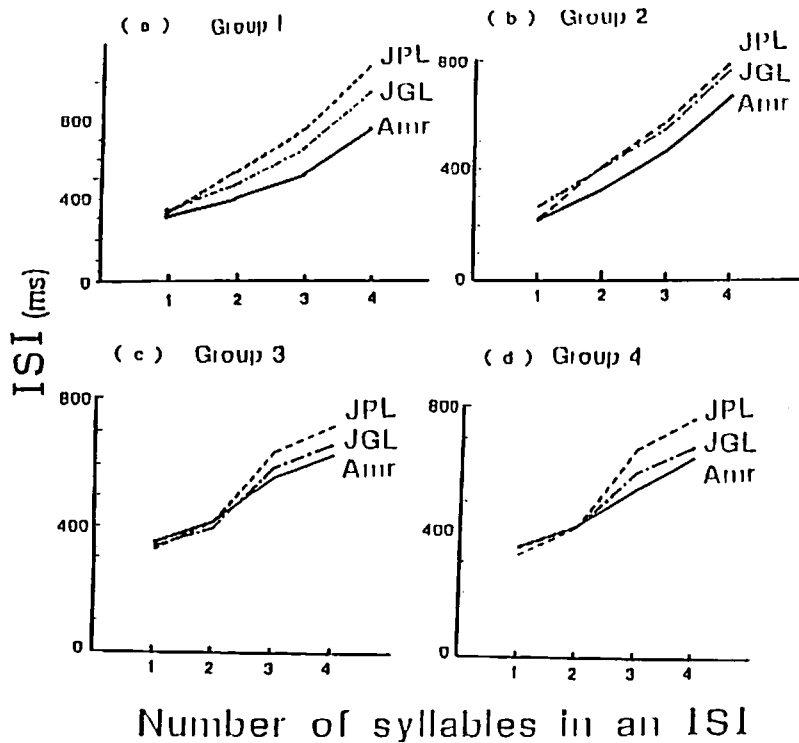


Figure 1. Durations of ISI (durations between the target vowel and the next stressed vowel).

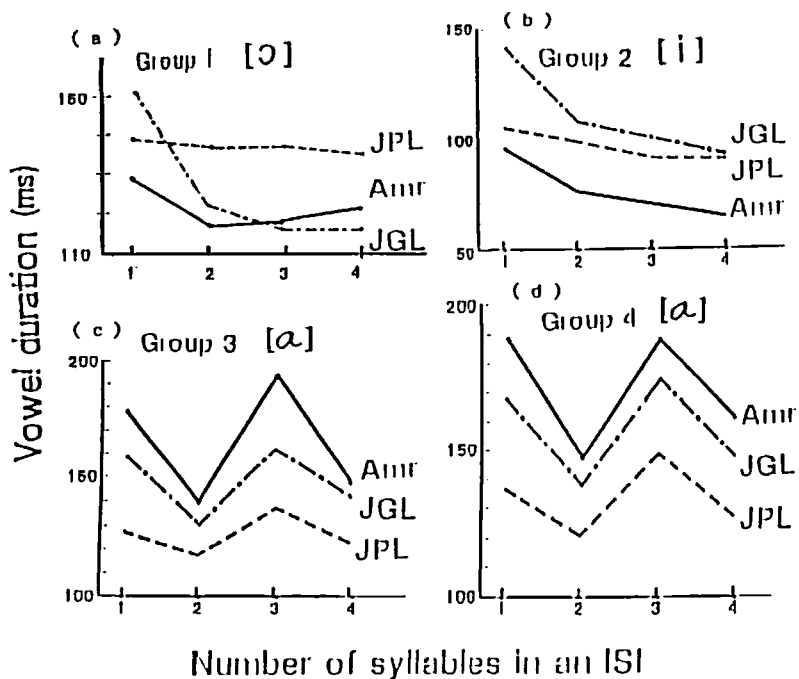


Figure 2. Durations of the target vowel.

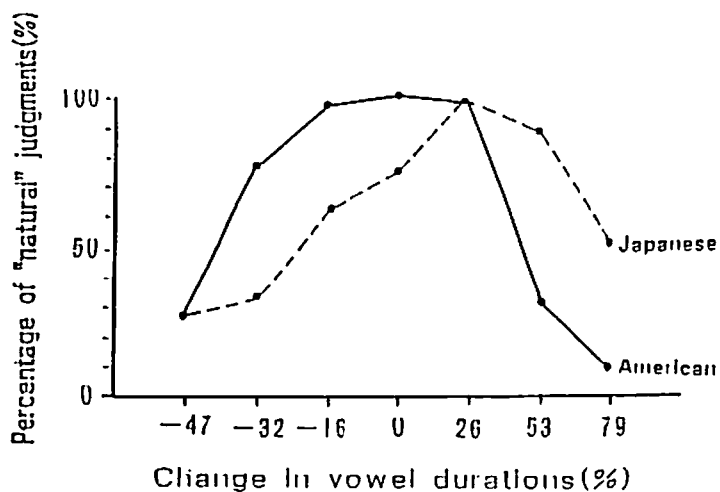


Figure 3. Percentage of "natural" judgments for the sentence group "Pete plays the piano".

Table 1. Estimates for a 50% threshold of "natural" judgments. Percentage change in vowel durations.

Subj.	Sentence	1	2	3	4	5	6	7	Average
		(Dog)	(Doggy)	(Dog)	(Doggy)	(Pete)	(Peter)	(Peterson)	
American	lengthening	35	43	25	39	53	42	39	39
	shortening	-34	-40	-39	-48	-41	-55	-118	-54
	range	69	82	64	87	94	97	157	93
Japanese	lengthening	24	61	30	53	123	96	75	66
	shortening	-82	-51	-51	-46	-23	-49	-64	-52
	range	106	114	81	99	146	145	139	119

(%)