Stimulus-related effects on the perception of syllables in second-language speech

田嶋, 圭一 / TAJIMA, Keiichi

(出版者 / Publisher)
法政大学文学部
(雑誌名 / Journal or Publication Title)
Bulletin of Faculty of Letters, Hosei University / 法政大学文学部紀要
(巻 / Volume)
49
(開始ページ / Start Page)
139
(終了ページ / End Page)
158
(発行年 / Year)
2004-03-02
(URL)
https://doi.org/10.15002/00002919

Stimulus-related effects on the perception of syllables in second-language speech

Keiichi TAJIMA*

1 Introduction

Syllables exist in most, if not all, languages as a basic unit of prosodic and rhythmic structure. For native English speakers, the number of syllables a given word contains is often intuitively clear (e.g., Liberman et al., 1974). This is so even though it has been extremely difficult to agree on a phonetic definition of syllables (Ohala, 1990).

While syllables are found in all languages, at the same time, its structure shows considerable language-specific variation. Because of this, native-speaker intuitions about the number of syllables in a word are not likely to be shared by non-native speakers. In fact, it is known that perception and production of syllables in a second language (L2) with relatively complex syllable structure, such as English, is difficult for native speakers of a language with relatively simple syllable structure, such as Japanese (Tarone, 1980; Tajima et al., 2003; Otake et al., 1996; Dupoux et al., 1999).

Recently, a series of studies have been carried out to examine the ability of native Japanese listeners to perceive syllables in spoken English words by using a syllable-counting task in which listeners identify the number of syllables in spoken English words (Erickson et al., 1999; Tajima and Erickson, 2001; Tajima et al., 2002). The

* Data collection for this study was conducted while the author was a researcher at ATR Human Information Science Laboratories. The author wishes to thank Reiko Akahane-Yamada and other collaborators at ATR. This research was conducted as part of "Research on Human Communication" with funding from the Telecommunications Advancement Organization of Japan.

primary concern of these studies has been to evaluate the extent to which adult listeners' perception of L2 syllables can be modified through perceptual identification training, which has been shown to be effective in the training of L2 phonetic contrasts (Lively et al., 1994) as well as lexical tone contrasts (Wang et al., 1999). Results have shown that native Japanese listeners significant improve in their ability to count syllables in spoken English words as a result of training.

The purpose of the present study is to closely examine how Japanese listeners' perception of English syllables is affected by characteristics of the speech stimuli such as the length of the stimuli (e.g., whether the stimulus consists of a single syllable or a series of syllables) and the complexity of the syllables (e.g., whether the stimulus contains relatively simple CV-type syllables or contains syllables with complex consonant clusters). Another objective of the present study is to reveal the kinds of errors listeners tend to make. That is, do listeners tend to over-estimate or under-estimate the number of English syllables? As a related question, do Japanese listeners show evidence for a tendency to count moras instead of syllables in English words? Focus is placed on finding out which stimulus-related factors present the greatest problems for Japanese listeners' performance prior to any extensive training on English syllables[1]. Possible answers to these questions may provide important clues not just for second-language training, but also for understanding mechanisms underlying perception of speech prosody.

As will be discussed below in more details, the results revealed that some stimuli were indeed more difficult for Japanese listeners than others. In general, Japanese listeners were less able to count syllables in longer words (those containing more syllables) than shorter ones, with the exception of monosyllabic words, which showed a large variation in accuracy depending on the syllable structure of the word. Longer polysyllabic words (those containing more than three syllables), however, seemed to show an opposite trend from shorter words, in the sense that performance was better for

---

(1) See Tajima and Erickson (2001) and Tajima et al. (2002) for data on the effect of identification training on native Japanese listeners' performance.

items with many consonants than those with fewer consonants. Performance of native Japanese listeners is compared with a control group of native English listeners. As is discussed below, some native English listeners turned out to have difficulties performing the syllable-counting task.

2 Methods

Stimulus materials consisted of 116 English words and 32 nonwords. The 116 English words varied in syllable count from one to six syllables[2], and are listed in the appendix. The number of consonants that each word contained, as it would appear in a typical phonemic transcription, varied across words. As for the 32 nonwords, 16 of them were monosyllabic and were of the form /C_1eC_2/ where /e/ was the vowel in English words such as "head" and "bet", and C_1 and C_2 each varied from zero to three consonants in a factorial manner (C_1 = {none, /s-/, /sp-/, /spl-/}, C_2 = {none, /-p/, /-ps/, /-mps/})[3]. The other 16 nonwords were disyllabic and were of the form /C_1edeC_2/ in which stress was placed on the second syllable[4]. These materials were each spoken by two male and two female native speakers of American or Canadian English (between 31 and 46 years of age). The talkers did not consider themselves fluent in a language other than English. Recordings were made in am anechoic chamber at ATR Laboratories, Kyoto, Japan, and were digitized at 22.05-Hz sampling frequency and 16-bit resolution. Individual words were amplitude-normalized so that the peak amplitude was constant across all items.

Participants in the experiment consisted of 23 native Japanese listeners and 18

(2) An attempt was made to avoid English words that vary in syllable count depending on speaking style or dialect, e.g., "fire", "dictionary".

(3) The nonword /pep/ turned out to be a real English word, but it was expected to be unfamiliar to native Japanese listeners.

(4) Some of the nonwords did not obey English phonotactics, e.g., /e/, /spe/. Prior experiments had used nonwords that contained either /ej/ or /e/ in order to test the effect of "long" vs. "short" English vowels on listeners' performance. However, to fully study the effect of word-initial vs. final consonants, the vowel /e/ was used rather than /ej/ because English does not exhibit sequences in which /ej/ is followed by a 3-consonant cluster containing labials such as /mps/.

native English listeners. The Japanese listeners were monolingual college students from Doshisha University, Kyoto, Japan (mean age = 20.6, between 18 and 24 years of age). None of them had stayed in an English-speaking community for more than three months. The English listeners were native speakers of American or Canadian English (mean age = 28.1, between 18 and 39 years of age), who did not consider themselves fluent in other languages. The listeners had lived in Japan between one month and 3.5 years at the time of the experiment.

The experimental task was a syllable-counting task, conducted in a sound-treated booth using a computer program. In the program window, participants saw ten buttons numbered "1" through "10". On each trial, participants listened to a stimulus, counted the number of syllables in it, and responded by clicking the appropriate button. Each stimulus was played back once, but subjects were able to click the "replay" button and listen to the stimulus again, although they were discouraged from doing so frequently. For the Japanese participants, there were two blocks of 148 trials each. In each block, the 116 real words and the 32 nonwords as spoken by one talker was presented once each in a random order. The choice and order of the two talkers was counterbalanced across participants. For the English participants, there were four blocks of 390 trials each. In each block, the 116 words, the 32 nonwords, and an additional set of 300 English words, as spoken by one of the four talkers, were presented in a random order. The 300 additional words were used in a different study and their results will not be reported here. The four talkers were presented in a random order. Prior to the experiment, the following brief description about English syllables was given to the Japanese participants: "There are 'units' in speech. In Japanese, we tend to count the hiragana character as a unit. For instance, the word 'hiragana' has four units, 'hi-ra-ga-na'. In English, however, 'syllables' serve as a unit of rhythm. A syllable generally has one vowel, optionally surrounded by consonants. So 'papa' has two syllables, while 'pin' has one syllable, not two." Several practice trials were given to all participants in order to familiarize them with the task.

3 Results and discussion

In order to analyze listeners' responses as a function of various stimulus-related factors, this section partitions the stimulus set into smaller subsets, each of which typically contains only a few stimuli. Because of the relatively small set size, the discussion of the results will place emphasis on reporting observable tendencies in the data rather than on performing rigorous statistical tests.

3.1 English listeners' performance

Native English listeners were expected to perform almost perfectly in the task. However, the results revealed that a few listeners had difficulties counting syllables in the stimuli, especially in the nonword stimuli. Figure 1(a) plots individual listeners' performance by plotting the identification accuracy for real words on the x-axis and accuracy for the nonwords on the y-axis.

Most data points in Figure 1(a) are clustered in the top right corner of the scatterplot, indicating that most English listeners performed well on both real words as well as nonwords. However, three English listeners scored below 80% on the nonwords. Two of these listeners performed somewhat better on real words (87% and 92%) than on nonwords, while the third listener performed poorly on both words (63%) and nonwords (73%). In addition, a different listener performed well on nonwords (95%) but poorly on real words (71%).

It is not clear why a small subset of native English listeners performed poorly in the task. Biographical data from the listeners did not suggest any peculiar features about these listeners. Perhaps these listeners misinterpreted the task, or perhaps they were less aware of the phonological structure of their native language than other listeners, making it difficult for them to reliably identify syllables in the stimuli. In any case, the data seem to suggest the presence of at least two types of populations, one group of native listeners who can perform the task at close to ceiling level, and another population who have difficulties performing this meta-linguistic task for one reason or another.

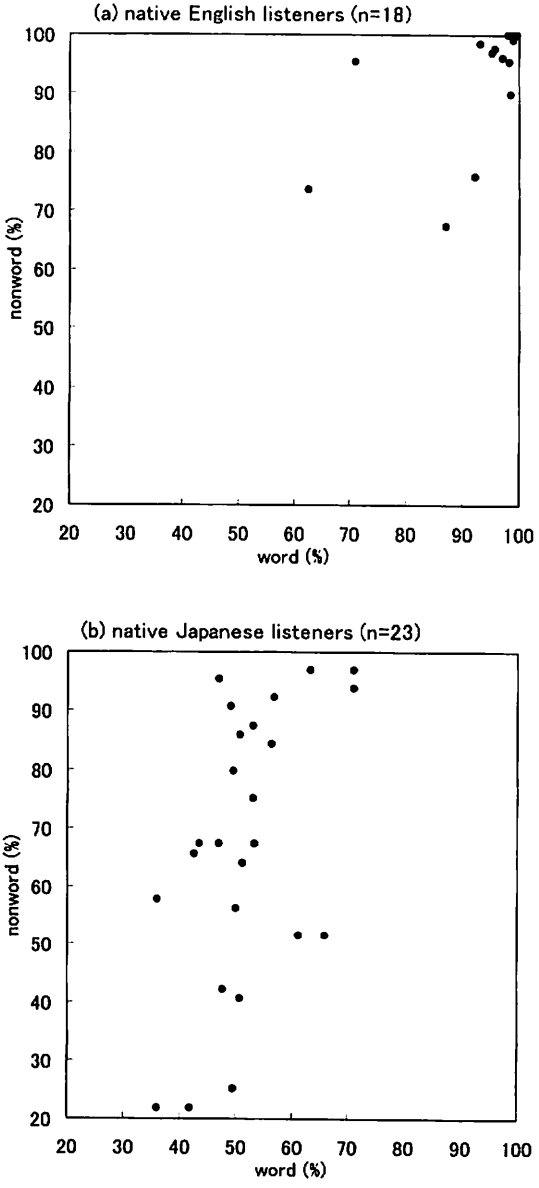


Figure 1: Individual listeners' identification accuracy for words (n = 116) and nonwords (n = 32), for (a) native English listeners (n = 18), and (b) native Japanese listeners (n = 23).

Accordingly, the high-scoring group and the low-scoring group will be analyzed separately.

High-scoring group

The mean identification accuracy of the high-scoring group was 98.0%. The syllable count of over 75% of the items was correctly identified more than 98% of the time. Items that had the lowest accuracy were "helped" (1 syllable, 82%), followed by "pragmatism" (4 syllables, 88%) and "incomprehensible" (6 syllables, 88%).

Low-scoring group

The mean identification accuracy of the low-scoring group was 78.3%. Items whose accuracy was higher than 98% accounted for only 22% of the data, as opposed to over 75% for the high-scoring group. Table 1 shows a confusion matrix of the data, in which the percentages of listeners' correct and incorrect responses are shown for stimuli with various numbers of syllables. The percentages of correct responses are shown in boldface.

Table 1: Confusion matrix of responses by low-scoring English listeners. Each row represents data from stimuli with a particular number of syllables. The correct responses are shown in boldface.

Correct	Listeners' responses (%)							
response	1	2	3	4	5	6	7	8
1	**72.0**	26.6	1.4	0.0	0.0	0.0	0.0	0.0
2	2.5	**89.2**	7.9	0.4	0.0	0.0	0.0	0.0
3	0.3	11.5	**86.2**	2.1	0.0	0.0	0.0	0.0
4	0.3	2.8	27.4	**69.4**	0.0	0.0	0.0	0.0
5	0.0	0.0	12.5	22.9	**64.6**	0.0	0.0	0.0
6	0.0	0.0	6.3	22.9	16.7	**54.2**	0.0	0.0

The confusion matrix in Table 1 shows a general tendency for accuracy to be lower for longer words than for shorter words, except for monosyllabic words whose accuracy was lower than that for disyllabic words. In other words, the more syllables there were

in a word, the harder it was to count them accurately. This may be partly because longer words tend to be produced at a faster speaking rate than shorter words, increasing the chances for segment or syllable reduction processes to take place. The item with the lowest accuracy for this listener group was again "helped" (0%). This was followed by "linked" (1 syllable, 13%) and "glanced" (1 syllable, 25%). Notice that all three low-score items were English words that ended with the past tense morpheme "-ed", suggesting that the low-scoring group listeners treated this inflectional suffix as an extra syllable.

Looking at the error responses in Table 1, for monosyllabic and disyllabic words, listeners showed a tendency to count more syllables than there actually were in the stimuli. By contrast, for 3- to 6-syllable words, listeners tended to count *fewer* syllables than there were in the tokens. In fact, for these long words, the listeners virtually never over-estimated the number of syllables.

3.2 Japanese listeners' performance

Native Japanese listeners were expected to have great difficulty identifying syllables in spoken English words. Hence, their syllable-counting accuracy was predicted to be substantially lower than that for native English listeners. The mean identification accuracy of the Japanese listeners as a group was 55.3%, suggesting that they miscounted syllables in roughly one out of two stimulus items. When individual Japanese listeners' performance was plotted in the same manner as Figure 1(a), the results are as shown in Figure 1(b).

Figure 1(b) reveals that the Japanese listeners' identification accuracy showed considerable inter-subject variation, especially for the nonwords. Accuracy for the real words varied between 36% and 71%, while accuracy for nonwords varied from 22% to 97%. The correlation between the two measures was 0.34. According to these results, a safe conclusion to make is that a listener's performance on nonword stimuli is not necessarily a good indicator of his/her performance on real English words. Thus, an experiment based solely on highly controlled but non-existent English words require careful interpretation.

A confusion matrix for the Japanese listeners' responses is shown in Table 2. The confusion matrix shows a tendency for accuracy to be lower for longer words than for shorter words, except for monosyllabic words which had a lower score than disyllabic words. This trend is similar to that in Table 1, but the magnitude of the effect is considerably greater. Looking at the percentage of incorrect responses, the confusion matrix indicates that the Japanese listeners often over-estimated the number of syllables in monosyllabic words. For polysyllabic words, the listeners made both over-estimation and under-estimation errors. To closely examine the results, the data are further split into monosyllabic stimuli, disyllabic stimuli, and longer stimuli in the subsequent sections.

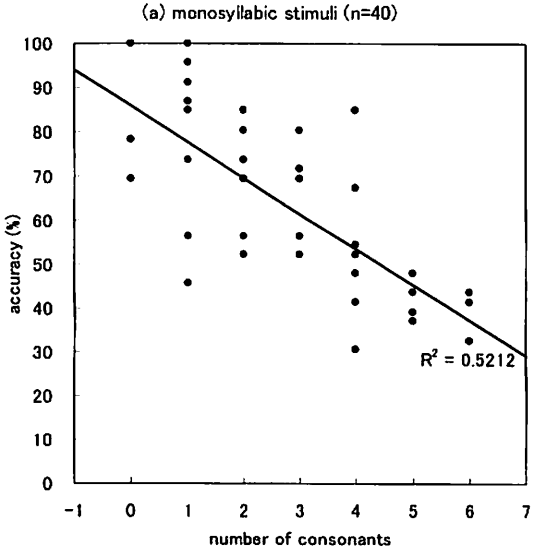
Table 2: Confusion matrix of responses by native Japanese listeners.

Correct	Listeners' responses (%)							
response	1	2	3	4	5	6	7	8
1	**62.8**	30.3	5.2	1.3	0.4	0.0	0.0	0.0
2	8.7	**69.7**	16.2	4.2	1.0	0.2	0.0	0.0
3	1.7	33.4	**49.1**	12.1	3.0	0.6	0.0	0.0
4	0.5	11.2	37.1	**37.8**	10.0	3.1	0.2	0.0
5	0.2	6.2	27.9	27.5	**30.4**	6.2	1.4	0.2
6	0.0	2.9	18.8	27.9	28.6	**16.3**	4.7	0.7

Monosyllabic stimuli

Monosyllabic words varied in complexity from those with no consonants ("owe", "eye", and the nonword /e/) to those with six consonants ("splints", "scripts", and /splemps/). Accuracy was found to vary substantially as a function of syllable complexity. This is illustrated in Figure 2(a), in which the number of consonants in each word is plotted on the x-axis and the mean accuracy for the word on the y-axis. Also shown is the regression line for the data, along with R^2, which indicates the proportion of the variance in accuracy that is accounted for by the number of consonants in the stimulus. The regression analysis indicated that for these monosyllabic words and nonwords, the number of consonants accounted for 52% of the variability in the listeners' syllable-

counting accuracy. Put another way, these results suggest, not surprisingly, that monosyllabic words with more consonants are harder for Japanese listeners than words with fewer consonants. Even though syllable complexity is a good predictor of accuracy for monosyllabic words, it turns out that it plays a smaller role in predicting accuracy for longer words, as will be discussed below. In other words, increased syllable complexity, as measured by the number of consonants in the stimulus, greatly reduces accuracy by Japanese listeners in the case of monosyllabic words, but does not reduce accuracy as strongly for polysyllabic words. This is not a surprising finding. A fixed increase in the number of consonants contributes more to the total number of segments or to the acoustic duration of shorter words than longer words. Furthermore, in general, it is easier to count multiple instances of an unfamiliar object than a single instance, since the fact that there are more than one instance allows the observer to extract common features among the instances that characterize the object. For these reasons, it may be concluded that perceiving syllable structure in monosyllabic words is particularly difficult for Japanese listeners.



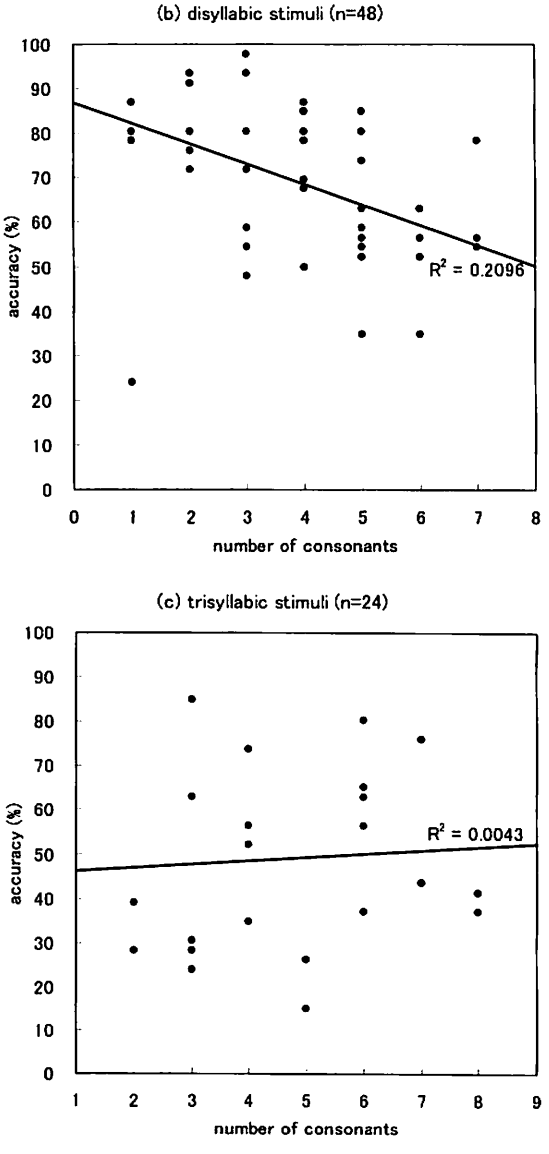


Figure 2: Native Japanese listeners' identification accuracy for individual stimuli plotted as a function of the number of consonants in each item. Data are plotted for (a) monosyllabic, (b) disyllabic, and (c) trisyllabic stimuli. Also plotted on each panel is the linear regression line, along with R^2, which indicates the proportion of variance in accuracy accounted for by the number of consonants.

Several additional comparisons were made on a subset of the monosyllabic stimuli, in order to examine the relative effect of initial vs. final consonants, singleton consonants vs. consonant clusters, and certain inflectional morphemes. Table 3 shows the distribution of listeners' responses for a subset of monosyllabic words grouped according to the syllable structure of the stimuli. Based on the data in Table 3, the following observations can be made.

Table 3: Distribution of native Japanese listeners' responses to a subset of monosyllabic words arranged according to syllable structure. The correct response for all items in this table is "1" (shown in boldface).

Syllable structure	Stimulus	Listeners' responses (%)					
		1	2	3	4	5	6
V	owe eye /e/	**82.6**	17.4	0.0	0.0	0.0	0.0
CV	bee lie boy two /pe/	**89.6**	10.0	0.4	0.0	0.0	0.0
VC	aim ice /ep/	**62.3**	36.2	1.4	0.0	0.0	0.0
CCV	grow spy /spe/	**59.4**	38.4	2.2	0.0	0.0	0.0
CVC	tough these /pep/	**73.9**	25.4	0.7	0.0	0.0	0.0
CVCC	dust belt /peps/	**73.9**	20.3	5.1	0.7	0.0	0.0
CCCVC	scrap /splep/	**46.7**	41.3	10.9	0.0	1.1	0.0
CCVCC	trench clocks /speps/	**51.4**	39.1	7.2	2.2	0.0	0.0
CVCCC	next /pemps/	**76.1**	18.5	5.4	0.0	0.0	0.0
CVCCC	helped linked	**35.9**	62.0	1.1	1.1	0.0	0.0
CCCVCC	strict /spleps/	**43.5**	39.1	12.0	3.3	2.2	0.0
CCVCCC	glanced /spemps/	**40.2**	42.4	13.0	4.3	0.0	0.0
CCCVCCC	splints scripts /splemps/	**39.1**	36.2	13.8	8.0	2.9	0.0

1. There was a general trend for listeners to respond with higher syllable counts as the number of consonants in the stimulus increased. However, it appears that listeners did not respond with the number of moras that they may have heard in the stimuli. For example, CVCC words such as "dust" and "belt" would contain three moras, but Japanese listeners entered "3" only 5% of the time. Instead, the most frequent error was "2" which was observed 20% of the time. This result suggests that Japanese listeners' response patterns is not the result of simple mora-based segmentation, but is instead apparently based

on something between strict mora counting and syllable counting.

2. There were surprisingly large percentages of errors on V words and CV words, even though they are structurally simple. It is possible that the incorrect responses was triggered by the diphthongs in these stimuli; that is, the Japanese listeners may have treated each diphthong to consist of multiple syllables.
3. Both CV and VC words have one consonant, but there were more errors for VC words than for CV words, suggesting that Japanese listeners tended to count post-vocalic word-final consonants as an extra syllable rather treating it as belonging to the same syllable.
4. Both CCV and CVC words contain two consonants, but CCV words induced more incorrect responses than did CVC words, possibly suggesting that word-initial consonant clusters present greater difficulty than word-final singleton consonants.
5. CCCVC words, CCVCC words, and CVCCC words all contain four consonants, but there was a trend for accuracy to decline as the number of initial consonants increased. This lends further support for the claim that word-initial consonants exert greater influences on Japanese listeners' responses than do word-final consonants.
6. Accuracy for CVCCC words varied depending on whether the word ended with the past tense morpheme or not, suggesting that Japanese listeners tended to count the past tense morpheme "-ed? as a separate syllable even if it did not phonetically comprise a syllable.

Disyllabic stimuli

Disyllabic words varied in complexity from those containing one consonant, e.g., "any", "allow", to those containing seven consonants, e.g., "snowstorms", "conflicts". The variation in identification accuracy for each word as a function of syllable complexity is illustrated in Figure 2(b). A regression line for the data is also shown. The regression analysis indicated that the number of consonants accounted for 21% of the variance

in identification accuracy, which was much lower than the 52% obtained with monosyllabic words[5]. This suggests that Japanese listeners' identification accuracy is not as strongly affected by syllable complexity for disyllabic words as it is for monosyllabic words.

In Table 4, the distribution of listeners' responses for a small subset of disyllabic words is shown, arranged according to the syllable structure.

Table 4: Distribution of native Japanese listeners' responses to a subset of disyllabic words arranged according to syllable structure. The correct response for all items in this table is "2" (shown in boldface).

Syllable structure	Stimulus	Listeners' responses (%)					
		1	2	3	4	5	6
VCV	ego away allow /ede/	13.6	**81.5**	4.9	0.0	0.0	0.0
CVCV	summer cowboy wider†	6.1	**87.4**	6.1	0.4	0.0	0.0
VCVC	iris assert /edep/	6.5	**85.5**	8.0	0.0	0.0	0.0
VCCV	answer under	13.0	**83.7**	3.3	0.0	0.0	0.0
CCVCV	speedy /spede/	13.0	**56.5**	30.4	0.0	0.0	0.0
CVCVC	begin pudding /pedep/	10.1	**87.7**	2.2	0.0	0.0	0.0
CCVCVC	scallop /spedep/	12.0	**59.8**	23.9	4.3	0.0	0.0
CVCCVC	cocktaiol chances	5.4	**72.8**	17.4	4.3	0.0	0.0

† Also includes: taboo /pede/.

1. Listeners sometimes treated disyllabic stimuli as monosyllabic. This trend was somewhat stronger for VCV stimuli than for other stimuli. Among the VCV stimuli, this trend was especially strong for the word "any", which was correctly responded to only 24% of the time[6]; since this score was much lower than that for other VCV words, it was not included in Table 4. This trend offers one convincing piece of evidence that the Japanese listeners were not

(5) If one of the data points that appears to be an outlier (the data point for the word "any" which contained one consonant and had an accuracy of 24%) was removed from the regression analysis, then the predictor variable accounted for 36% of the variance in accuracy. This percentage was still lower than that obtained for monosyllabic words.

(6) The incorrect responses were all "1" responses.

simply trying to count moras in the stimuli. Since the number of moras in a word is equal or greater, but not smaller, than the number of syllables, Japanese listeners would not have responded with smaller values if they were counting moraic units in the stimuli.

2. Accuracy did not vary much among CVCV, VCVC, and VCCV words, which all had two consonants, but were structurally different. This is in contrast to the difference in accuracy obtained in monosyllabic words between CV and VC stimuli which, like the above stimuli, had the same number of consonants, but were structurally different. This is consistent with the claim that syllable structure plays a smaller role in polysyllabic words than in monosyllabic words.
3. Despite the similarity in accuracy observed among CVCV, VCVC, and VC-CV stimuli, when accuracy was compared between CCVCV and CVCVC stimuli, and between CCVCVC and CVCCVC stimuli, accuracy was lower for the first member of each pair than the second, i.e., lower for stimuli containing word-initial clusters[7]. This suggests that word-initial clusters are consistently problematic for Japanese listeners, whether they appear in monosyllabic or polysyllabic words. Together with the results from CVCV, VCVC, and VCCV stimuli, it appears that stimuli containing word-initial consonant clusters are more problematic for Japanese listeners than words containing word-final consonants or word-medial consonant clusters.

Trisyllabic and longer stimuli

Trisyllabic words varied in complexity from stimuli containing two consonants, i.e., "allergy", "attorney", to those containing eight consonants, i.e., "straightforward", "grandparent". Figure 2(c) plots each word's accuracy as a function of the number of consonants, along with the linear regression function. It is clear from this scatterplot that syllable complexity played virtually no role in predicting listeners' accuracy for

(7) Incidentally, all of the word-initial clusters here were /s/+consonant clusters.

these trisyllabic words. Items that received the highest accuracies were "banana", "Canada", and "employment", while items that received the lowest scores were "strawberry", "honesty", and "ministry", whose syllable counts were more often under-estimated than over-estimated. For trisyllabic and longer words, the syllable structures of the stimuli were too variable across items to allow systematic comparisons like the ones in Tables 1 and 2.

The overall number of 4-syllable, 5-syllable, and 6-syllable stimuli were smaller than the number of shorter words. Nevertheless, syllable complexity, as measured by the number of consonants, varied across stimuli, from three to eight consonants for 4-syllable words, from four to nine consonants for 5-syllable words, and from five to ten consonants for six-syllable words. To analyze the effect of syllable complexity in the face of the small set size, accuracy was not analyzed as a semi-continuous function of number of consonants, but was instead analyzed by dividing each word set into two groups, relatively "little" words and relatively "big" words. For 4-syllable words, "little" words were defined to be those that had five or fewer consonants. Similarly, for 5-syllable words, little words were defined to be those that had six or fewer consonants. Finally, for six-syllable words, little words were those that had seven or fewer consonants. Any other words were treated as "big" words. If syllable complexity exerted an influence on listeners' accuracy even for these polysyllabic words, then listeners' response patterns should vary between these two word types. The distribution of listeners' responses for these word sets is shown in Table 5.

The table shows that for 4-, 5-, and 6-syllable words, accuracy was slightly higher for big words than for little words. That is, words with a greater number of consonants led to a somewhat higher accuracy than words with relatively few consonants. The 4- and 5-syllable words showed about a 6-point difference in accuracy between big and little words, while the 6-syllable words showed only a marginal, 1-point difference. This trend is opposite from the trend observed for monosyllabic and disyllabic words.

It is not clear exactly why a reversal occurs in the effect of syllable complexity on accuracy. Perhaps it is simply an experimental artifact stemming from subjects' tendency to respond with values that are near the middle of the range rather than near the

Table 5: Distribution of native Japanese listeners' responses for stimuli of various numbers of syllables and various relative "size", i.e., "little" words containing relatively few consonants, and "big" words containing many consonants (see text for details).

Stimulus properties			Listeners' responses (%)							
Syllable count	Relative "size"	N	1	2	3	4	5	6	7	8
1	little	18	**75.4**	23.4	1.2	0.0	0.0	0.0	0.0	0.0
1	big	22	**52.6**	36.0	8.5	2.3	0.7	0.0	0.0	0.0
2	little	22	13.2	**78.0**	8.4	0.4	0.0	0.0	0.0	0.0
2	big	26	4.8	**62.7**	22.8	7.4	1.8	0.3	0.0	0.0
3	little	12	2.9	39.3	**50.0**	7.4	0.4	0.0	0.0	0.0
3	big	12	0.5	27.5	**48.2**	16.8	5.6	1.3	0.0	0.0
4	little	8	0.8	14.7	44.0	**34.2**	5.7	0.5	0.0	0.0
4	big	10	0.2	8.5	31.5	**40.7**	13.5	5.2	0.4	0.0
5	little	6	0.4	7.6	30.4	31.2	**27.9**	2.2	0.4	0.0
5	big	6	0.0	4.7	25.4	23.9	**33.0**	10.1	2.5	0.4
6	little	3	0.0	3.6	24.6	26.1	26.8	**15.9**	2.2	0.7
6	big	3	0.0	2.2	13.0	29.7	30.4	**16.7**	7.2	0.7

extreme. Perhaps the reversal took place partly because words may have been produced at a somewhat faster rate than shorter words, which may have made longer words more prone to segment and syllable reduction processes. Perhaps the English lexicon is such that there are systematic differences in syllable structure that typically appear in short vs. long words. Further investigations are necessary to tease apart these possible factors.

4 Conclusion

To summarize, this study investigated native and non-native listeners' perception of syllable structure in English by utilizing a syllable-counting task. Results revealed that not all native English listeners are equally able to perform the task. While most native English listeners performed almost perfectly, a distinct group of listeners had difficulties performing the task. The results also revealed that native Japanese listeners

sometimes under-estimated and sometimes over-estimated the number of syllables in the speech stimuli. As the stimuli increased in syllable complexity, Japanese listeners' identification accuracy generally decreased. This effect was particularly strong for monosyllabic words. Furthermore, Japanese listeners had greater difficulty treating VC syllables as monosyllabic than CV syllables, but even greater difficulty was encountered with stimuli containing word-initial consonant clusters. It appears that an asymmetry exists between word-initial and word-final clusters, the former exerting a greater influence on accuracy than the latter. However, the effect of relatively simple vs. complex syllables (little vs. big words) was reversed for words containing four or more syllables, suggesting that increased syllable complexity is not unconditionally detrimental to accuracy, but could in fact yield better performance depending on the length of the stimulus. Altogether, these results suggest that the processes underlying Japanese listeners' perception of syllables in spoken English are quite complex, and cannot be explained simply as a consequence of mora-based segmentation.

References

Erickson, D., Akahane-Yamada, R., Tajima, K., and Matsumoto, K. F. (1999). Syllable counting and mora units in speech perception. In *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 1479 - 1482.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., and Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 18, 201 - 212.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/ III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*, 96, 2076 - 2087.

Ohala, J. J. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In *Proceedings of the 26th regional meeting of the Chicago Linguistic Society, Volume 2, The Parasession on the Syllable in Phonetics and Phonology.*, pp. 319 - 338, Chicago, IL.

Otake, T., Hatano, G., and Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In Otake, T. and Cutler, A. (eds.), *Phonological Structure and Language Processing: Cross-linguistics Studies*, pp. 187 - 201, Mouton de Gruyter, Berlin.

Tajima, K., Akahane-Yamada, R., and Yamada, T. (2002). Perceptual learning of English syllable rhythm by young and elderly Japanese listeners. In *Proceedings of the 16th general meeting of the Phonetic Society of Japan*, pp. 103 - 108, Tokyo.

Tajima, K., and Erickson, D. (2001). "Syllable structure and the perception of second-language speech." In Spoken Language Working Group (ed.), *Speech and Grammar III*, Kuroshio Publishers, Tokyo, pp. 221 - 240.

Tajima, K., Erickson, D., and Nagao, K. (2003). Production of syllable structure in a second language: Factors affecting vowel epenthesis in Japanese-accented English. In *Speech Posody and Timing: Dynamic Aspects of Speech (Indiana University Working Papers in Linguistics, Volume 4)*, pp. 77 - 92, Bloomington, IN.

Tarone, E. (1980). Some influences on the syllable structure of interlanguage phonology. *International Review of Applied Linguistics in Language Teaching*, 18, 139 - 152.

Wang, Y., Spence, M. V., Jongman, A., and Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649 - 3658.

Appendix

The following is the list of 116 English words used in the present study (see main text for description of the 32 nonwords that were also used in the experiment).

1-syllable words:

owe eye bee lie aim boy two ice tough grow these spy dust belt next helped trench scrap clocks linked strict glanced splints scripts

2-syllable words:

ego away any allow summer cowboy iris answer wider taboo under assert supply begin speedy pudding cocktail scallop accept chances spotlight displayed structure contact blanket scraper playground trademark breakfast sculptures conflicts snowstorms

3-syllable words:

allergy attorney melody banana itemize Canada emotion honesty envelope removal telephone terrific strawberry ministry proposals screwdriver understand countryside employment sentences gymnastics frustration straightforward grandparent

4-syllable words:

radiator alligator delicacy military analysis capacity meditation application professional propaganda development pragmatism subconsciously electronics constitution receptionist demonstrated infrastructure

5-syllable words:

affiliation anniversary economical similarity possibility illumination participating juxtaposition representative extravaganza fundamentalist congratulations

6-syllable words:

autobiography revolutionary peculiarity responsibility indistinguishable incomprehensible