Jimcts

Suprasegmental Effects on Word Segmentation by Native and Non-native English Speakers

KATAYAMA Tamami

abstract

The present research has examined how native English speakers and L2 speakers of different levels of language experience make use of the cues of syllable duration and pause to segment words. Selected groups of native English speakers, Japanese speakers with extensive experience of English, and Japanese speakers whose experience of English was limited took part in the monitoring task. The results suggest that the native English speakers use syllable duration to segment words, while the groups of Japanese speakers tend to segment words by marking the pauses. The results also indicate that the Japanese speakers with more extensive experience of English segment words more accurately and more quickly than the Japanese speakers whose experience of English is less, or limited. We can therefore say that L2 experience affects the speech processing of the target language, but that word segmentation strategy appears to be language specific.

1 Introduction

The work of psycholinguistics concerns itself, among other issues, with how a listener recognizes and accesses an individual entry in the lexicon, and it considers whether the access code is composed of phonetic segments, of syllables, or of spectral templates. Psycholinguists have learned that the physical acoustic signal and the sounds conveyed by the signal have a complex relationship. This complexity stems from an aspect of "the lack of identity or isomorphism between the acoustic and phonetic levels of language" (Harley, 2001: 220). This is known as the segmentation problem: when sounds run together, they cannot be separated easily.

Turk and Shattuck-Hufnagel (2000) have reported that, in English speech, duration is a prominent phonetic cue, and that the duration patterns change depending on boundary locations. The Japanese language, however, uses a phonological unit called a mora, and every mora takes the same length of time (Ladefoged, 2001); consequently, differences in the length of words affect the meaning of those words. It is therefore possible, if not likely, that different habits of marking duration in the learner's native language will affect speech processing in a second language. In order to examine the differences in speech processing, it therefore becomes important to compare how native English speakers and non-native speakers make use of cues to produce and perceive duration in speech. Although a number of reports suggest that the segmentation strategies are language-specific, it is not yet clear how native Japanese speakers segment English speech; what is clear, however, is that listeners with different L1 backgrounds adopt different segmentation strategies.

2 Background

The segmentation strategies employed to recognize words depend on the listener's exposure to a particular language (Cooper, Cutler, & Wales, 2002). These language-specific strategies include suprasegmental information and a segmental inventory (Dupoux, Pallier, Sebastian, & Mehler, 1997); Cutler and Norris (1988) took strong syllables as the basis on which to test their model of segmentation. In the model of speech segmentation in a stress-timed language, the occurrence of a strong syllable triggers segmentation of the speech signal, while that of a weak syllable does not. Cutler and Butterfield (1992) reported that listeners pay a special attention to rhythmic segmentation when perception is difficult. On the other hand, the pattern of the response by Japanese listeners can be described in terms of mora-based segmentation (Otake, Hatano, Cutler, and Mehler, 1993). Otake et al. found that Japanese listeners do not decompose morae but segment the spoken words by means of mora. Cutler and Otake (1994) go on to argue that Japanese listeners inevitably apply their native moraic pattern of speech processing to their processing of foreign languages, and they believe that the finding of language-specific processing has critical implications for uncovering the process of second language acquisition.

One of the major differences between spoken English and spoken Japanese is the use that is made of syllable duration. Native English speakers, for instance, adjust the duration of stressed syllables and inter-stress intervals to mark what is important in the message, while Port (1981) argues that many abstract timing rules govern speech production in English. His study revealed the tendency for the duration of a stressed syllable to shorten as more syllables were added to the word, and that the effect is greater with words of one and two syllables than with words of two and three syllables. Japanese speakers, on the other hand, adjust the length of the words depending on the number of morae in a word and use pitch instead of duration to produce a word accent. In Japanese, as Ueyama (1996) argues, duration and word accent are independent properties. Japanese speakers are therefore predicted to make less use than native English speakers do of duration in identifying English word boundaries.

It is reported that in Japanese, a pause affects the listener's speech comprehension. Kohno (2001) took an example of Japanese speech and an example of English speech, and inserted pauses between every word, phrase, clause and sentence, to examine how the frequency of pause influenced the ability of native Japanese speakers to comprehend the message. He reported that in both languages, Japanese listeners found it easiest to comprehend the content when the pauses were placed between every phrase rather than between every clause or sentence. They found it difficult, however, to understand the speech when pauses were inserted between every word. Kohno suggests that pause plays a role as an acoustic signal that is used to indicate semantics units. By contrast, Henderson (1980) reported that in English pause is not an acoustic cue for marking word boundaries.

3 Research questions and hypotheses

The purpose of this study is to investigate the acquisition of L2 suprasegmentals. In particular, this research examines how native English speakers and L2 learners of English of different levels in language experience segment words using the cue of syllable duration and pauses. In order to classify the participants' English levels, the study first established their living experiences in English-speaking countries[1]. This research raises the following research questions:

1) How do syllable duration and length of pause affect word segmentation?
2) Does L2 experience of syllable duration and pause affect perception of word segmentation?

The experiment tests the following hypotheses:

H1: Native English speakers use syllable duration to segment words, while Japanese learners of English segment words through their management of pause.
H2: Japanese speakers with extensive experience of English segment words as native English speakers do, whereas Japanese speakers with limited experience of English do not.

►1 We explain how we classified learners with "extensive experience of English" and learners with "limited experience of English" in section 4.1, *Participants*.

4 Methodology

4.1 Participants

Seventeen native English speakers (NES), seventeen Japanese speakers with extensive experience of English (JEE), and eighteen Japanese speakers with limited experience of English (JLE) took part in the experiment. After we had discarded data that frequently omitted the values, twelve participants were selected from each group. The group of NES consisted of eight British English speakers and four American English speakers. Their ages ranged from eighteen to thirty-seven. The groups of Japanese speakers were classified as having extensive experience or as having limited experience of English according to the length of time that they had lived in English-speaking countries. The former consisted of native Japanese speakers who had lived in English-speaking countries for more than one year and the latter consisted of native Japanese speakers who had spent less than five months in a country where English was the native language. None of the participants in the three groups reported any hearing impairment. Table 1 shows the attributes of the Japanese participants.

4.2 Materials

Twenty compound words (e.g. "blackboard") and twenty corresponding

■ Table 1　*Attributes of the Japanese Participants*

Group	JEE	JLE
Age	18- over 38	18-32
Living duration	3years 7months	7weeks
Position	postgraduate students, etc.	students in summer English course, etc.
English use	very frequently/frequently	no professional use

Note. JEE refers to the Japanese speakers with extensive experience of English, while JLE signifies the Japanese with limited experience of English. Living duration indicates the mean duration of the participants' domicile in English-speaking countries.

two-separate words (phrases)(e.g. "black board") were selected as targets. The phrases include adjectives and nouns. Twenty foils were created, one for each target compound word or phrase. The foils have the same phoneme as the target at the onset of the second syllables or words, such as "seatbelt" and "tight belt", or "blackboard" and "black board." Table 2 lists the target compound words and phrases. In addition to the target words and foils, 180 fillers were selected for the compound words and phrases. The onsets of the second syllables or words in fillers are different from those in the targets and foils. All the adjectives and nouns in the fillers are different. Thus, 656 words were selected for this study, consisting

■ Table 2　*Target Compound Words and Phrases*

	Target Compound Words	Target Phrases	Foil Compound Words	Foil Phrases
1	blackboard	black board	seatbelt	tight belt
2	greenhouse	green house	cowhide	small hide
3	redwood	red wood	downwind	mean wind
4	goldfish	gold fish	pitchfork	Dutch fork
5	flatcar	flat car	bedclothes	baby clothes
6	wildcat	wild cat	suitcase	fancy case
7	whitecap	white cap	cupcake	large cake
8	safeguard	safe guard	headgear	third gear
9	blacktop	black top	turntable	brown table
10	backfield	back filed	bullfrog	blue frog
11	darkroom	dark room	crossroad	wide road
12	backlog	back log	dateline	straight line
13	dumbwaiter	dumb waiter	password	nice word
14	redcoat	red coat	popcorn	ripe corn
15	hardcover	hard cover	racecourse	police course
16	blackmail	black mail	milkman	wise man
17	bluegrass	blue grass	ballgame	night game
18	hardship	hard ship	snowshoe	comfy shoes
19	yellowhammer	yellow hammer	falsehood	purple hood
20	oldschool	old school	moonstone	rough stone

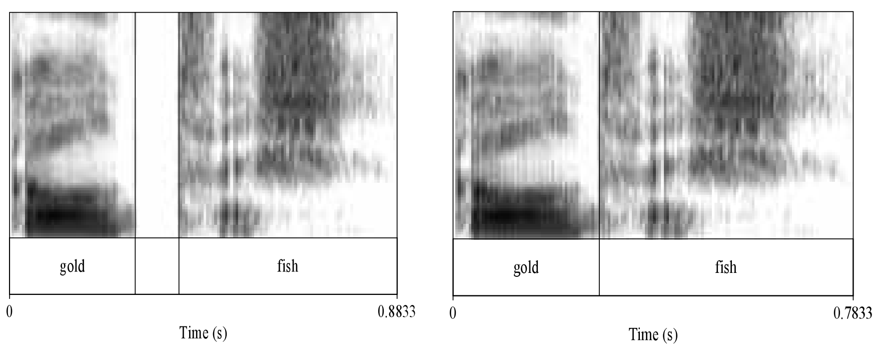
of 220 compound words, 216 adjectives (nine target phrases use the same adjectives), and 220 nouns.

Because this study investigates whether or not listeners are able to make use of the cue of syllable duration to segment words, a carrier sentence was created for each of the compound words and phrases in order to avoid other factors, such as pitch, that listeners would also perceive. The carrier sentences for the paired target compound words and phrases have the same number of words in the sentences. In addition, two sets of sentences were created for each target so that prominence would fall on the first syllables or words. For example, in order to obtain the sounds for "greenhouse" and "green house", two sets of sentences were created: "It's not in the garden. It's in the greenhouse," and "It's not in the red house. It's in the green house." In total, 880 paired carrier sentences were created for 440 words for targets, foils, and fillers.

The recording of all the sentences was carried out twice, on different days, in the recording studio of the Phonetics Laboratory at the University of Edinburgh. A male native English speaker from Canada, judged by a phonetician and psycholinguist to be an English speaker without a regional accent, read the 880 sentences. He was instructed to stop for a break when he saw the symbol ( ) between items: this break was located every six, eight, ten, or twelve sentences so that during his reading he was not able to impose his own rhythm. The recording was digitized and stored on the computer with 4800kHz of sampling frequency.

We used an Adobe Audition version 2.0 to cut out the target compound words and phrases, which we then segmented with the help of a Praat into first syllables/words and second syllables/words at the point of zero crossing. Figure 1 shows an example of segmentation for the compound word "goldfish".

Figure 1. Segmentation for "goldfish" with and without silence. The time unit is a millisecond.



A silence of 100ms was inserted between the first syllables/words and the second syllables/words in all the targets and foils. The insertion of 100ms of silence is obviously perceivable, but it does not render the sound of the words unnatural. In addition to the targets and foils, 100ms of silence was inserted in fifty percent of the fillers of compound words and phrases. Table 3 is a summary of the sound files used in this study.

Table 4 shows the length of the first syllable of the target compound words, their adjectives, and their ratios. Although the *yellow* of *yellowhammer* has two syllables, the length of *yellow* was measured as if it were one syllable only. A related t-test was performed to examine if there were significant differences between the length of the first syllables of the compound words and that of the phrases. The dependent variable is the duration of the first syllables of the target words, and the independent variable is the word type (compound words and phrase). Since these sounds were produced by the same speaker, we were able to use a related measure. There was a significant difference between the two groups [t (19) = -2.95, p < .01]. That is, the duration of the first syllables of the phrases (M = 259.15ms, SE = 18.2) was significantly longer than that of the compound words (M = 218.2ms, SE = 10.2). In addition, we used a speech analyzer to

Table 3 *Sound Stimuli of the Study*

Sound stimuli	Targets	Foils	Fillers
Compound words	20	20	90
Phrases	20	20	90
Compound words with pause	20	20	90
Phrases with pause	20	20	90

Table 4 *Length of the First Syllables of the Target Words*

	Words	Compound Words	Phrases	Ratio
1	backfield	236	228	1.04
2	backlog	228	239	0.95
3	blackboard	270	315	0.86
4	blackmail	227	249	0.91
5	blacktop	191	242	0.79
6	bluegrass	118	181	0.65
7	darkroom	224	269	0.83
8	dumbwaiter	170	201	0.85
9	flatcar	315	344	0.92
10	goldfish	191	248	0.77
11	greenhouse	170	205	0.83
12	hardcover	250	521	0.48
13	hardship	240	221	1.09
14	oldschool	193	205	0.94
15	redcoat	197	140	1.41
16	redwood	181	210	0.86
17	safeguard	272	334	0.81
18	whitecap	182	224	0.81
19	wildcat	254	283	0.90
20	yellowhammer	255	324	0.79

measure the intensity and the pitch of the first syllables of the target words, after which we conducted a related t-test to examine whether there were significant differences for intensity and pitch between the values of the compound words and the phrases. The results show no significant differences between them either for the intensity [t (19) $<$ 1, p =.51] (compound words: M = 76.59dB, SD = 2.38; phrases: M = 77.79dB, SD = 7.52) or for the pitch [t (19) $<$ 1, p =.54] (compound words; M = 109.73Hz, SD = 7.08; phrases: M = 108.15Hz, SD = 10.27). These results indicate the validity of the sound stimuli.

We used E-prime software to create four versions of a monitoring task for both English and Japanese speakers. Each version has forty lists, including twenty positive lists with target words and twenty negative lists without them. These lists were mixed randomly, but we took care that no more than three positive or three negative items were placed adjacently. As targets, each version of the lists had either a compound word with or without silence, or phrases with or without silence. A Latin square was made to offer equal opportunities of listening to every kind of target. In addition to the target word, the corresponding foil was included in each list. Each list consisted of eight to twelve words, and the rest of the spaces were occupied by fillers. At the same time, each list had the same proportion of the four types of word. The position of the target words varied randomly, but, except for the target words, all the conditions were the same.

The program of the monitoring task, which required visual as well as auditory attention, isolated each word on the screen for 750ms, followed by a blank screen for 750ms. The words in the list were then programmed to run while the mark “+” appeared in the center of the screen. The words selected were the second word/syllable of the target phrases or compound words. For instance, the word “board” was the target word for “blackboard” and “black board”. The words were extracted either from the target compound words/phrases or from the fillers. The reaction of the response was set up to vary depending on whether it was in the middle of the sound lists or not. If the response was made while the words on a particular list were being presented, this list was programmed to jump to the next list. If a participant had made no response by the time the presentation of the list was over, a message on the screen announced that the list had come to an end and instructed the participant to press the button to continue. For the English participants, the instructions were written in English, and for the Japanese participants they were written in Japanese. The Japanese instructions were translations of the English instructions, and the content was exactly the same except that the target words were accompanied with Japanese translation as well.

4.3 Procedure

The participants were instructed to sit in front of a computer in a booth and put on the headphones. The instructions were presented on the screen, and the participants were asked to respond as rapidly as possible whenever they heard the target word by pressing the button "1" on the box. The response time from the onset of the second syllables was recorded. Should the participants fail to hear the target word, they were instructed to press button "5" to continue. The target word and lists of words appeared after the auditory instruction, "Please listen for…" Each word in the list was followed by a 900ms interval of silence, and the next words were presented automatically.

5 Results

We subjected both items and subjects to analysis. For analysis by subject, each participant heard five items for each type of word, and their mean response time was calculated. If they missed any of the target words, the mean was calculated on the basis of those words that they had heard. For analysis by item, three participants heard the same item in each group, and the mean response time for the item was calculated. If all the three participants in a group missed the target word, it was deleted from the analysis: for this reason, two target words, *backlog*and *hardship*, were deleted from the lists. Consequently, the analysis used eighteen target words.

Table 5 shows the accuracy of the response for each group. The Japanese speakers with little experience of English (JLE) missed more items than the other groups. Table 6 and Figure 2 present the mean response time (RT) by subject. For the analysis by subject, we submitted three groups to ANOVA two-way repeated-measures to examine whether there were any major effects or interactions between the type of words and the silences. The dependent variable is the response time, and the independent variables are the word type and the presence or absence of silence. For the analysis by item, we performed a two-way ANOVA test for all the groups. The dependent variable and the independent variables are the same as the analysis by subject.

For the analysis by subject with the group of Japanese speakers with extensive experience of English (JEE), a significant main effect on response time

Table 5 *Accuracy of the Response in the Monitoring Task*

Group	NES	JEE	JLE
Number of missing value	16	15	27
% of correct response	92.6%	93.1%	87.5%

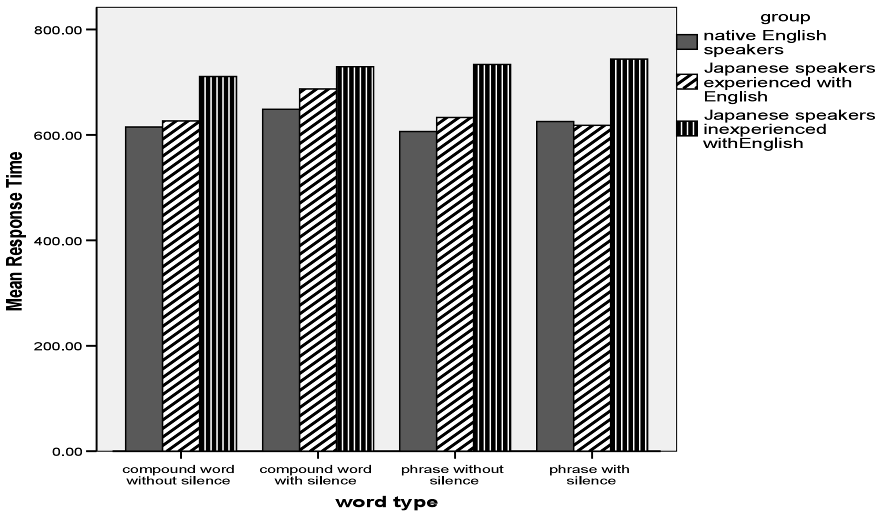
Note. The total number for the analysis is 216 (20 participants x 18 items). NES = the native English speakers. JEE = the Japanese speakers with extensive experience of English. JLE = the Japanese speakers with limited experience of English.

■ Table 6 *Response Time of Monitoring Task by Subject*

Group	Type of Word	*Mean RT*	*SD*	*N*
Native English speakers	Compound word without silence	614.97	101.91	12
	Compound word with silence	648.61	168.68	12
	Phrase without silence	606.39	120.56	12
	Phrase with silence	625.35	109.57	12
Japanese speakers with extensive experience of English	Compound word without silence	626.52	86.45	12
	Compound word with silence	687.15	118.49	12
	Phrase without silence	633.07	104.79	12
	Phrase with silence	618.10	78.20	12
Japanese speakers with limited experience of English	Compound word without silence	710.73	98.02	12
	Compound word with silence	729.29	100.03	12
	Phrase without silence	733.55	121.75	12
	Phrase with silence	743.73	96.15	12

Note. N refers to the number of the participants, RT, the mean response time, and SD, standard deviation.

■ Figure 2. Monitoring task by subject for the three groups

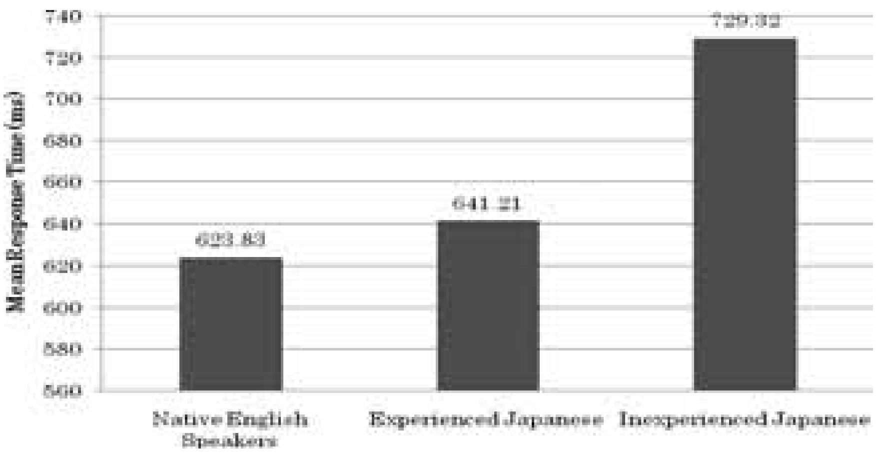


depended on whether the target was a compound word or a phrase [$F_1(1, 11) = 5.84$, $p < .05$]. Silence had no significant effect [$F_2(1, 11) < 1$] nor was there any significant interaction between silence and word type [$F_3(1, 11) = 1.31$, $p = .28$]. The same was true for the other groups (For the native English speakers (NES), the effect of word type: F_1 (1, 11) < 1, the effect of silence: $F_2(1, 11) = 1.06$, $p = .33$, the interaction between word type and silence: $F_3(1, 11) < 1$; the effect of word type for JLE,: $F_1(1, 11) < 1$, the effect of silence: $F_2(1, 11) < 1$, the interaction between word type and silence: $F_3(1, 11) = < 1$). For the analysis by item, there were no significant main effects or interactions between items in any of the groups (For NES, the effect of word type: $F_1(1, 68) < 1$, the effect of silence: $F_2(1, 68) < 1$, the interaction between word type and silence: $F_3(1, 68) < 1$; for JEE, the effect of word type: $F_1(1, 68) < 1$, the effect of silence: $F_2(1, 68) < 1$, the interaction between word type and silence: $F_3(1, 68) = 1.31$, $p = .26$; for JLE, the effect of word type: $F_1(1, 68) < 1$, the effect of silence: $F_2(1, 68) < 1$, the interaction between word type and silence: $F_3(1, 68) < 1$).

We submitted all three groups to a three-way mixed ANOVA test to carry out the analysis between subjects. The dependent variable is the response time and the independent variables are the word type with two levels: compound word and phrase, the silence with two levels, with or without 100 ms of silence, and the group with three levels: NES, JEE, and JLE. We also submitted the three groups to a three-way ANOVA to examine the effects or interactions by item. The dependent variable and the independent variables was the same as for the analysis between subjects.

The results for the analysis by subject show that there was a significant main effect of the group on response time [$F_1(2, 33) = 5.79$, $p < .01$]. A post-hoc Tukey test revealed that on response time JLE differed significantly from both NES ($p < .01$) and JEE ($p < .05$). That is, JLE took significantly longer to respond to the stimuli than either of the other groups (See Figure 3). JEE took longer ($M = 641.21$) to respond than NES ($M = 623.83$), but more quickly than JLE ($M = 729.32$). There was no significant difference in response time between NES and JEE ($p = .87$), nor did the response time signal any other significant main effects or interactions (The effect of word type: $F_2(1, 33) < 1$, the effect of silence: $F_3(1, 33) = 1.82$, $p = .19$, the interaction between word type and group: $F_4(2, 33) = 1.20$, $p = .31$, the interaction between silence and group: $F_5(2, 33) < 1$, the interaction between word type and silence: $F_6(1, 33) = 1.32$, $p = .26$). For the analysis by item, there is also a significant main effect of the group on their response time [$F_1(2, 204) = 12.56$, $p < .01$]. The Tukey test reveals significant differences in response time not only between NES and JLE ($p < .01$), but also

Figure 3. Response time of the monitoring task by the three groups



between JEE and JLE ($p < .01$). That is, NES and JEE responded more quickly than JLE. Although NES responded more quickly ($M = 618.96$) than JEE ($M = 637.99$), the difference between them was not significant ($p = .73$). There were no other significant main effects or interactions on the response time (the effect of word type: $F_2(1, 204) < 1$, the effect of silence: $F_3(1, 204) = 1.25$, $p = .27$, the interaction between word type and group: $F_4(2, 204) < 1$, the interaction between silence and group: $F_5(2, 204) < 1$, the interaction between word type and silence: $F_6(1, 204) < 1$).

In the analysis by subject, the type of word had a significant main effect on the response time in the group of JEE ($p < .05$). The mean response time for compound words was 656.83 ms, while the mean response time for phrases was 625.85 ms. Specifically, this group showed the shortest response time for the phrases with silence ($M = 618.1$ ms), the second shortest response time for the compound words without silence ($M = 626.52$ ms), the third shortest response time for the phrases without silence, and the longest response time for the compound words with silence ($M = 687.15$ ms). When they were compared with the other groups, their mean response time for the compound words with pause was much longer. Since, according to Ueyama (1996), the syllables before pauses are lengthened in Japanese, JEE may consequently have been confused by short syllable duration followed by pauses. Since a lengthened syllable before a pause is actually common in both English and Japanese, native English speakers showed similar confusion.

The response time of JLE was significantly longer than that of NES in both analysis by item and by subject ($p < .01$ for both), and longer than JEE in the analysis by item ($p < .01$) and by subject ($p < .05$). In addition, their percentage accuracy (87.5%) was lower than that of NES (92.6%) and JEE (93.1%): this led to the frequent missing values. Once we have taken account of these facts, we are bound to conclude that this task is too difficult for JLE.

The native English speakers responded more quickly to the items without silence (mean by subject = 610.68 ms, mean by item = 603.76 ms) than to those with silence (mean by subject = 636.98 ms, mean by item = 634.18 ms). This indicates that NES do not appear to rely on pauses to segment the words, and must therefore make use of syllable duration for recognizing word segmentation. Since, as Cutler (1986) reported, duration is related to both stressed and unstressed syllables, NES tended to make use of prosodic rhythms to segment speech. These results support those of previous studies (Mehler, Dommergues, and Frauenfelder, 1981; Cutler, 1986; Cooper, Cutler, & Wales, 2002; Vogel & Raimy, 2002), which argue that lexical stress is unrelated to the lexicon and does not facilitate word segmentation.

The most prominent difference between NES and JEE is the response time for the phrases with and without silence. JEE responded more quickly to the phrases with silence ($M = 618.10$ms by subject, $M = 616.73$ms by item) than those without silence (M = 633.07ms by subject, M = 628.50ms by item). On the other hand, NES responded more quickly to the phrases without silence (M = 606.39ms by subject, M = 597.57 ms) than to the phrases with silence (M = 625.35ms by subject, M = 625.33ms by item). This indicates that, just as Hypothesis One predicted, JEE appeared to segment words by attending to the pause rather than to syllable duration. For the analysis by item, JLE also responded more quickly to the phrases with silence (M = 737.04ms) than to the phrases without silence (M= 740.22ms), though these results are not consistent with the results of the analysis by subject. This may be because JLE frequently failed to press the button to respond to the target words, and consequently, the missing values are a prominent feature of the analysis.

6 Discussion

This study has examined how native English speakers (NES) and L2 speakers with different levels of language experience made use of the cue of syllable duration and pauses to segment compound words and phrases. To accomplish this, the following research questions were raised:

1) How do syllable duration and length of pause affect word segmentation?
2) Does L2 experience of syllable duration and pause affect perception of word segmentation?

Since, in English, syllable duration is related to both stressed and unstressed syllables, and since NES seem to segment words on the basis of stress, they also appeared to use syllable duration for the same purpose. On the other hand, the groups of Japanese speakers used pauses to segment words, and consequently did not appear to segment words in English speech on the basis of stress. These results support Cutler and Otake's study (1994), which claims that word segmentation is language-specific, and that Japanese listeners apply their own moraic pattern of speech processing to their perception of foreign languages.

In answer to research question two, we found that the Japanese speakers with extensive experience of English (JEE) were able to segment English words more accurately and more quickly than the Japanese speakers whose experience of English was limited (JLE). In the monitoring task, missing values were prominent in JLE. This group was unable to recognize the target words to which they were supposed to respond, even though the words appeared with high frequency. In addition, their response time was significantly longer than that of the other groups.

Because they were not able to segment words, they may not have been able to recognize words in the stream of speech. In this sense, the level of L2 experience affected the perception of English words. As for word segmentation, both groups of Japanese speakers appeared to use pause to segment English words, whereas NES tended to use syllable duration. Thus, regardless of experience, non-native speakers of English do not seem to develop the word segmentation strategy that is employed by native speakers of the second language.

Two further hypotheses were therefore examined.

H1: NES use syllable duration to segment words, while Japanese learners of English segment words through their management of pause.

Although the differences were not significant, NES tended to use syllable duration, while Japanese speakers were likely to segment words with a pause. In particular, the duration of the first syllable was longer in some compound words, such as “backfield” and “redcoat”, than in the corresponding phrases. As for *backfield*, the ratio of the duration between the compound words and the phrases was 1.04 (“back” of “backfield” = 236ms, “back” of “back field” = 228ms) and NES appeared to find it difficult to spot the difference. Thus, although they responded more quickly to the phrases than to the compound words, we may count this difference as a random error. Nonetheless, since the first syllable of “redcoat” was much longer than that of “red coat” (“red” of “redcoat” = 197ms, “red” of “redcoat” = 140ms), and as the response time was shorter for the compound word than the phrase, this result is consistent with the main findings: NES make use of syllable duration to segment words.

H2: JEE segment words as NES do, whereas JLE do not.

The monitoring task supported Hypothesis Two: JEE showed accuracy and response time more like those of NES than did JLE. The major effect of the group on the response time was significant. A post-hoc test for the analysis by subject and by item revealed that although the difference between NES and JEE was not significant (p = .86 by subject, p = .73 by item), the response time of both NES and JEE were significantly different ($p < .01$ for both) from those of JLE. Thus, although JEE responded more slowly than did NES, they responded significantly faster than JLE.

With respect to the accuracy of the response, NES and JEE were more accurate (92.6% for NES, and 93.1% for JEE) than JLE (87.5%). We may observe in fact that JEE were even more accurate than NES. In summary, JEE were more accurate and quicker to respond than JLE, and were closer to NES. Thus, the results supported Hypothesis Two. Nevertheless, the results of this study suggest that word segmentation strategy is indeed language-specific and will not change even if an individual's accuracy and speed of L2 speech processing continue to develop.

7 Conclusion

This research has attempted to contribute to studies of second language acquisition by investigating whether L2 experience affects word segmentation in the target language. Regardless of their experience of English, both groups of Japanese speakers who participated in the tests tended to use pause in segmenting words. Even so, the Japanese speakers with more extensive experience of English were able to segment English words more accurately and more quickly than those with less or limited experience. One drawback of our study was that, apart from the individual participant's exposure to the target language, we did not take into account such attributes as the age of the Japanese students or their proficiency scores on language tests. More work therefore needs to be done to discover which of a L2 learner's personal attributes most helpfully facilitate the acquisition of L2 speech perception.

References

Cooper, Nicole, Anne Culter, and Roger Wales. (2002). Constraints of lexical stress on lexical access in English: Evidence from non-native listeners. *Language and Speech,* 45*(3)*, 207-228.

Cutler, Anne. (1986). Forbear is a homophone: lexical prosody does not constrain lexical access. *Language and Speech,* 29*(*3*),* 201-220.

Cutler, Anne and Sally Butterfield. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language,* 31, 218-236.

Cutler, Anne and Dennis Norris. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* 14*(*1*),* 113-121.

Cutler, Anne and Takashi Otake. (1994). Mora or Phoneme? Further evidence for language-specific listening. *Journal of Memory and Language,* 33*,* 824-844.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian, and Jacques Mehler. (1997). A distressing "deafness" in French? *Journal of Memory and Language* 36, 406-421.

Harley, Trevor. (2001). Understanding speech. *The Psychology of Language: From data to theory.* 2*nd edition*. (pp. 219-242). Hove: Psychology Press.

Henderson, Alan. (1980). Juncture pause and intonation fall and the perceptual segmentation of speech. In Hans Dechert and Manfred Raupach (Eds.), *Temporal Variables in Speech* (pp. 199-206). The Haugue: Mouton Publishers.

Kohno, Morio. (2001). *Onseigengo-no ninshiki-to seisei-no mekanizumu: Kotoba-no jikanseigyokiko-to sonoyakuwari*［Mechanism of speech recognition and production: linguistic temporal control system and its role］. Tokyo: Kinseido.

Ladefoged, Peter. (2001), *A Course In Phonetics/Fourth Edition*. Boston: Thomas Learning.

Mehler, Jacques, Jean Dommergues, Uli Faruenfelder, and Juan Segui. (1981). The syllable' s role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior,* 20, 298-305.

Otake, Takashi, Giyoo Hatano, Anne Cutler, and Jacques Mehler. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language,* 32*(*2*),* 258-278.

Port, Robert. (1981). Linguistic timing factors in combination. *The Journal of the Acoustical Society of America*, 69(1), 262-274.

Turk, Alice, and Stefanie Shattuck-Hufnagel. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics,* 28, 397-440.

Ueyama, Motoko. (1996). Phrase-final lengthening and stress-timed shortening in the speech of native speakers and Japanese learners of English. In *Proceedings of the Fourth International Conference on Spoken Language Processing.*

Vogel, Irene and Eric Raimy. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language*, 29, 225-250.

（Received: September 24, 2009; Accepted: January 6, 2010）