How Do Listeners Discriminate Compound Words from Noun Phrases Using Syllable Duration?

Tamami KATAYAMA
Prefectural University of Hiroshima

Abstract

The purpose of this study was to determine the degree to which native English speakers and Japanese English learners employ the cues of syllable duration to discriminate compound words from nominal phrases. Two compound words (greenhouse and goldfish) were selected as the target words, and the first syllables were lengthened by 20 ms, 40 ms, 60 ms, 80 ms and 100 ms (2 words x 6 steps). Twenty-two native English speakers (ES) and 24 native Japanese speakers (JS) were asked to decide whether or not 12 kinds of sound stimuli constituted a compound word. The results showed that the difference between the groups in their responses to greenhouse and goldfish was significant and that responses to the stimuli as compound words also differed depending on the duration of the first syllables. Although neither groups tended to show an innate or intuitive sense of the absolute value of syllable duration that would enable them to discriminate compound words from nominal phrases, they appeared to take into account the ratio of the duration between the first syllable and the second syllable. The native Japanese speakers appeared to be more sensitive to the given acoustic variables.

1. Background

One of the main questions at issue in studies on second language (L2) acquisition is whether speech perception is an acquired linguistic ability or a basic property of the auditory system. While some studies have set out to prove that L2 learners have an innate ability to develop perception in their target language (Flege & Schmidt, 1995), other studies have shown that the learner’s perception can be attributed to contextually determined acoustic variations (Sheldon & Strange, 1982). Yet, so far, studies have mainly focused on the development of segmental rather than suprasegmental features despite the important role that prosody is known to play in speech communication.

1.1 A linguistic ability vs. a basic property of the auditory system

The question of whether speech perception in a speaker’s second language is a linguistic ability (and is therefore capable of being trained and developed) or whether it is simply a basic property of the auditory system (and therefore a feature of anyone’s ability to hear) remains a controversial one. Rochet (1995) conducted an imitation task and a perceptual task using the three high vowels /i/, /y/, and /u/ in
three language groups (French, Canadian English, and Brazilian Portuguese). The results of both tasks tended to be parallel, and thus Rochet concluded that the location of the perceptual boundary of vowels stemmed from linguistic experience rather than from some basic property of the auditory system. Flege, Bohn, and Jang (1997) provided further evidence of a link between an L2 learner’s perception of English vowels and the learner’s L2 experience in general. They reported that adult L2 learners were able to perceive certain L2 vowels more accurately as they gained experience of their L2 language but that the L1 background did play a role in the accurate perception of L2 vowels. They also argued that the variance of performance in English vowels may be attributable to the cross-language differences in the perceived relationship between vowels in English and those closest to them in the L1. According to Leather and James (1991), learners adjust their perception of the target language by subsequently creating their own prototypes. Thus, they do not generally need training for perception of sound patterns of the target language in order to be able to perceive them.

On the other hand, some studies have suggested that L2 speech perception is a basic auditory aptitude and that L2 experience does not affect a learner’s perception of the target phonemes. Obler (1982) gave bilinguals of English and Hebrew a dichotic test in which different words were presented to either ear. Bilingual speakers with native competence in both languages listened to words with common phonemes across the two languages spoken by monolinguals of either language (e.g., English bite and Hebrew /bayt/) and dictated them. While the perception of the bilingual participants appeared to employ a unitary system, they did not turn off or ignore the language not in use. In addition, Sheldon and Strange (1982) investigated Japanese speakers’ perception of /r/ and /l/ and examined their location within a word. The results showed that perceptual difficulties were marked from most to least difficult in the following order: /r/-/l/ contrast in prevocalic stop + liquid clusters, in medial position, in word-initial position, and word-final position. According to Sheldon and Strange (1982), spectral and temporal features vary considerably with phonological context, which in turn affects listeners’ perception. Second language learners show different perceptual performance depending on whether acoustic-phonetic factors are perceptually salient.

We have set out to review two opposing standpoints with regard to L2 speech perception, for it is still a matter of dispute whether this faculty develops with a learner’s experience of an L2 or whether it is an inbuilt and automatic faculty of the auditory system itself. Although researchers of second language acquisition have provided some evidence of the development of speech perception for phonemes by L2 learners, studies on suprasegmentals have been limited due to their complexity despite their significant role in speech. Consequently, our study has focused on how L2 learners of English perceive suprasegmental features and whether their perception differs in any way from that of native English speakers. In particular, syllable duration was examined in this study.

1.2 Studies of word recognition by means of syllable duration

One of the major differences between spoken English and spoken Japanese is the use that is made of syllable duration. Duration plays an important role in determining prominence and word boundary in
speech processing in English. Turk and Shattuck-Hufnagel (2000) examined the nature of the word boundary-related duration information that the speaker provides, focusing on duration patterns at the word level. In polysyllabic shortening, a stressed stem syllable will be shortened as more syllables in the same word are added to its right. For example, the syllable stick was longest when produced as a monosyllabic word, was shorter in the 2-syllable word sticky, and shorter still in the 3-syllable word stickiness. Turk and Shattuck-Hufnagel reported that speakers controlled the relative duration of a polysyllabic word depending on the boundary location. Unlike Japanese, English has a durational structure that appears to be determined by the number of syllables rather than by the number of phonemes.

What we do know is that English speakers adjust syllable duration according to the number of syllables within a word (Port, 1981) and that this in turn facilitates speech processing of the listeners. Native English speakers can distinguish monosyllabic words and polysyllabic words at the offset of the first syllables (Davis, Marslen-Wilson, and Gaskell, 2002). Davis et al. conducted a gating task in native English speakers and presented fragments of test sentences. Their participants were asked to identify all of the words in each fragment. The results showed that they used acoustic differences to discriminate a monosyllabic word (e.g., cap) from the first syllable of a polysyllabic word (e.g., captain). When it was a one-syllable word, the syllable duration was significantly longer than that of the same syllable when it was part of a longer word. Davis et al. suggested that detailed acoustic information plays a critical role in lexical access.

Katayama (2012) examined how speakers of English and Japanese use syllable duration and stress to identify compound words and nominal phrases. Twenty native English speakers and 20 native Japanese speakers were given compound words (e.g., blackboard) for which the first syllables were lengthened by 100 ms, and they were asked to judge whether they were compound words or nominal phrases. The results suggested that syllable duration affects the ability of both native English speakers and Japanese speakers to discriminate compound words from nominal phrases in English. Katayama (2013) subsequently examined whether the perception of syllable duration is language-specific and conducted a discrimination task in groups of native English speakers and native Japanese speakers. A compound word (redwood) was used in the experiment, and the duration of the first syllable (red) was lengthened by 20 ms, 40 ms, 60 ms, 80 ms, or 100 ms to create six steps of sound stimuli. Then two of the sound files were provided in sequence and the participants were asked to judge whether the two sounds were the same or not. The results showed that the two groups tended to perceive syllable duration differently and that the perception of syllable duration was affected by the participants’ first language. Thus, Katayama assumed that the speech perception of syllable duration can be regarded as language-specific and a linguistic ability rather than a basic auditory facility or property. Katayama’s study (2013), however, left the possibility that the participants, especially non-native English speakers, used only acoustic information rather than their linguistic knowledge.

Thus, in this study, an identification task was carried out to examine the following questions: When identifying compound words, do native English speakers and Japanese speakers have an intuitive sense
of the threshold value of the syllable duration of every word they hear? Or do they estimate the relative value of syllable duration on the basis of the duration of each segment of a word? The threshold value indicates the point when syllable duration has satisfied the function of signifying a word boundary. The present study was therefore designed to investigate whether native English speakers and native Japanese speakers recognize intuitively a threshold value of syllable duration that signifies a word boundary.

2. Purpose of the study

The purpose of this study was to determine whether native English speakers and L2 learners of English are equally able to possess a threshold value for syllable duration that in the listener’s mind naturally constitutes a word boundary. This assumption raises the following research question: Do native English speakers and L2 learners of English intuitively recognize syllable duration as entailing a threshold value that affects their word recognition?

3. Methodology

3.1 Participants

Twenty-two native English speakers (ES) who lived in the Boston area and 24 native Japanese speakers (JS) who were undergraduate students in the Life Science Department participated in the experiment. ES included adults over 27 years old (M = 52 years old) who had not learned Japanese. The JS group had never lived in English-speaking countries and their English proficiency ranged from primary to intermediate; eight had the pre-level 2 of the English proficiency test (Eiken), while four had level 3. Six of them had a TOEIC score, the mean score being 411.5 (SD = 65.6). Although the rest of the students had never taken an official English test, they took required English classes at college (i.e., reading classes and speaking classes). ES received five dollars and JS were given a gift voucher for books to the value of 500 yen for their participation. None of the participants in the two groups reported any hearing impairment.

3.2 Materials

Two compound words (greenhouse and goldfish) were selected as the target words for the following reasons. First, the Japanese speakers were more likely to know the meanings of these compound words (In fact, a questionnaire to the Japanese participants revealed that 21 of the 24 subjects knew the meanings of both target words.). Second, the boundaries between the first syllables and the second syllables were clear when their formants were examined using a speech analyzer, Praat, and thus it was technically easy to manipulate the sound. A male native English speaker from Canada read the words aloud in carry sentences to ensure that the primary stress was placed on their first syllables. The words were embedded in the carrier sentences “It’s not in the garden. It’s in the greenhouse.” and “It’s not the tie with the puppies. It’s the tie with the goldfish.” The recordings were put into a computer
at a sampling rate of 48 kHz and the target words were spliced from the sentences using Praat. The first syllables were then lengthened by 20 ms, 40 ms, 60 ms, 80 ms and 100 ms (In Katayama’s earlier study (2012), the stimuli were lengthened by 100 ms.) (see Figures 1 and 2). In total, 12 kinds of stimuli (2 words x 6 steps) were used for the experiment. I devised a forced choice task for both English and Japanese speakers with the aid of E-prime software designed for psychological experiments. The instructions were written in English for the English participants and in Japanese for the Japanese participants. The Japanese instructions were translations of the English instructions, and the content was identical.

Figure 1. Spectrogram and sound wave of “goldfish”

Figure 2. Spectrogram and sound wave of “goldfish” for which the first duration was lengthened by 100 ms
3.3 Procedure

The participants were instructed to sit in front of a computer in a quiet room and asked to put on headphones and follow the instructions on the screen. They were presented with 12 kinds of stimuli as well as two pictures of the target words. For example, on hearing one of the stimuli while the mark “+” appeared in the center of the screen, the participants were simultaneously given pictures of “a goldfish” and “a gold fish” (see Figure 3) and were asked to press a button on the response box to select the picture that indicated their choice. After responding to the first stimulus, the next words were then presented automatically.

![Figure 3. The computer image for selection](image)

4. Results

Table 1 presents the number of responses identified as compound words. A three-way ANOVA was performed both between and within the groups. There was a significant difference in the grand mean of responses between the two groups \((F(1, 44) = 9.1, p < .01)\). That is, the ES were more likely to consider the stimuli as compound words \((M = 0.87, SD = 0.50)\) than were the JS \((M = 0.65, SD = 0.48)\). The main effects of the stimuli (i.e., whether either greenhouse or goldfish) \((F(1, 44) = 17.3, p < .001)\) and duration \((F(5, 220) = 7.1, p < .001)\) also differed. In effect, the difference between the groups in their responses to “greenhouse” and “goldfish” was significant. In addition, responses to the stimuli as compound words also differed, depending on the duration of the first syllables. There were no significant interactions between the three factors.
Table 1. *Response as compound word for each step of syllable duration*

<table>
<thead>
<tr>
<th>group</th>
<th>stimuli</th>
<th>+0 ms</th>
<th>+20 ms</th>
<th>+40 ms</th>
<th>+60 ms</th>
<th>+80 ms</th>
<th>+100 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>greenhouse</td>
<td>20 (91%)</td>
<td>22 (100%)</td>
<td>22 (100%)</td>
<td>18 (82%)</td>
<td>18 (82%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>goldfish</td>
<td>20 (91%)</td>
<td>20 (91%)</td>
<td>18 (82%)</td>
<td>17 (77%)</td>
<td>17 (77%)</td>
<td>14 (64%)</td>
</tr>
<tr>
<td>JS</td>
<td>greenhouse</td>
<td>19 (79%)</td>
<td>20 (83%)</td>
<td>18 (75%)</td>
<td>19 (79%)</td>
<td>17 (71%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>goldfish</td>
<td>17 (71%)</td>
<td>16 (67%)</td>
<td>13 (54%)</td>
<td>14 (58%)</td>
<td>10 (42%)</td>
<td>7 (29%)</td>
</tr>
</tbody>
</table>

N of ES = 22, N of JS = 24

As further analyses, a two-way ANOVA was conducted for both groups. ES showed a significant main effect of duration ($F(5, 105) = 5.1, p < .001$), but other significant effects or interaction were not found. On the other hand, the results for JS had main effects of both the stimuli ($F(1, 23) = 17.4, p < .001$) and duration ($F(5, 115) = 3.2, p = .01$) and an interaction was not seen as was the case for the ES. There results indicate that JS were affected by the kind of stimuli (i.e., whether it is *greenhouse* or *goldfish*) but that ES were not. The variance of syllable duration affected the perception by both ES and JS.

Figure 4 shows a graph of the percentages of responses to the stimuli as compound words by the two groups. The curve of responses to goldfish by JS is different from the other curves. When the stimuli were lengthened by 80 ms and 100 ms, Japanese speakers were more likely to identify them as nominal phrases. Other curves are above 60% even when they were lengthened by 100 ms. The results show that the native English speakers did not recognize any identifiable boundary when discriminating between a compound word and two separate words. The Japanese speakers, on the other hand, were more likely to identify the stimuli as two separate words the longer the syllable lasted.

One-way ANOVA tests were performed on responses for each step of the stimuli by JS and ES (i.e., the responses for *greenhouse* made by ES and JS, and those for *goldfish* made by the two groups). There were significant differences among the four kinds of stimuli for all of the steps except the original stimuli that were not manipulated (+20 ms: $F(3, 88) = 3.9, p < 0.05$; +40 ms: $F(3, 88) = 5.3, p < 0.01$; +60 ms: $F(3, 88) = 4.2, p < 0.01$; +80 ms: $F(3, 88) = 4.4, p < 0.01$; +100 ms: $F(3, 88) = 5.8, p < 0.01$). Tukey’s post-hoc tests revealed that the responses for goldfish by JS were significantly different from those for greenhouse by ES for the steps of stimuli lengthened by 20 ms ($p < 0.01$), 40 ms ($p < 0.01$), and 60 ms ($p < 0.01$). They were also different from the results of the others in the step of stimuli lengthened by 80 ms and the responses for greenhouse by JS and ES in the step which has 100 ms manipulation. In short, the responses of goldfish by JS were significantly different from the others in all steps of the stimuli.
5. Discussion and Conclusion

In this study, it was investigated whether native English speakers and L2 learners of English have a threshold value of syllable duration in their mind that will mark or signify a word boundary. The purpose of the investigation raises the following research questions: Do native English speakers and L2 learners of English possess a threshold value (a clear boundary) of syllable duration that affects their word recognition? The answer to this question is “Neither group appeared to show a boundary of syllable duration that would enable them to discriminate compound words from nominal phrases.”

The results for the native English group and the Japanese group showed a different tendency than the results of the study by Katayama (2013). When the two groups were given two sound files in which the difference was only syllable duration of the first syllables, they were able to discriminate one sound from the other. The larger the difference between them was, the more they tended to consider them as different. Katayama’s study (2013) showed a significant difference in the perception of syllable duration between native English speakers and native Japanese speakers, but this study did not show a clear difference depending on the group. This may be because this experiment used pictures and examined lexical knowledge. Although both groups perceived the difference of duration, this does not mean that they recognized words. The study by Davis et al. (2002) suggested that native English speakers can identify whether a word is monosyllabic or polysyllabic by detecting the first syllable duration, but this result needs the condition of word stress falling on the first syllable. Since two kinds of words for which positions of stress were different (one having stress on the first syllable and the other having stress on the second syllable) were used in this study, the native English speakers might not have been able to identify words by using only syllable duration.
Table 2. Ratio of syllable duration between “gold” and “fish”

<table>
<thead>
<tr>
<th>duration</th>
<th>gold</th>
<th>fish</th>
<th>total</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>+0 ms</td>
<td>193</td>
<td>490</td>
<td>683</td>
<td>0.39</td>
</tr>
<tr>
<td>+20 ms</td>
<td>213</td>
<td>490</td>
<td>703</td>
<td>0.43</td>
</tr>
<tr>
<td>+40 ms</td>
<td>233</td>
<td>490</td>
<td>723</td>
<td>0.48</td>
</tr>
<tr>
<td>+60 ms</td>
<td>253</td>
<td>490</td>
<td>743</td>
<td>0.52</td>
</tr>
<tr>
<td>+80 ms</td>
<td>273</td>
<td>490</td>
<td>763</td>
<td>0.56</td>
</tr>
<tr>
<td>+100 ms</td>
<td>293</td>
<td>490</td>
<td>783</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 3. Ratio of syllable duration between “green” and “house”

<table>
<thead>
<tr>
<th>duration</th>
<th>green</th>
<th>house</th>
<th>total</th>
<th>ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>+0 ms</td>
<td>179</td>
<td>514</td>
<td>693</td>
<td>0.35</td>
</tr>
<tr>
<td>+20 ms</td>
<td>199</td>
<td>514</td>
<td>713</td>
<td>0.39</td>
</tr>
<tr>
<td>+40 ms</td>
<td>219</td>
<td>514</td>
<td>733</td>
<td>0.43</td>
</tr>
<tr>
<td>+60 ms</td>
<td>239</td>
<td>514</td>
<td>753</td>
<td>0.46</td>
</tr>
<tr>
<td>+80 ms</td>
<td>259</td>
<td>514</td>
<td>773</td>
<td>0.50</td>
</tr>
<tr>
<td>+100 ms</td>
<td>278</td>
<td>514</td>
<td>792</td>
<td>0.54</td>
</tr>
</tbody>
</table>

On the other hand, the native Japanese speakers recognized one of the stimuli, goldfish, using syllable duration, which means that they were not likely to be affected by the position of stress. In addition, lengthened syllable duration does not always influence their word recognition, since they showed a significant difference in identification of greenhouse and goldfish. Consequently, the ratios of duration between the first and second syllables were determined (see Tables 2 and 3). The duration of “gold” (193 ms) was longer than that of “green” (178 ms), while that of “fish” (490 ms) was shorter than that of “house” (514 ms). Thus, the ratio between the first syllable and the second syllable is different for “goldfish” and “greenhouse.” For the files that were lengthened by 100 ms, the ratio of gold to fish was 0.6 and that of green to house was 0.54. If JS had considered a syllable as a unit of timing, they might have considered the following syllable as one unit on the basis of the duration of the first syllable. That is, if the duration of the following syllable was shorter than the first one, they could have considered it as a part of the first syllable rather than an independent unit. The closer the following syllable is to the duration of the first syllable, the more they are likely to consider it as a separate unit. This may be influenced by the feature of Japanese language for which the basic timing unit is a mora. Port, Dalby, and O’Dell (1987) stated that “The concept of the mora as an abstract isochronous unit of timing in Japanese captures many of the most salient features of timing in this language.” (Port et al., 1987, p.1584) despite large differences in the inherent duration of the segment type. They defined timing by mora not as an isochronic tendency but as predictability of word duration deduced from the number of morae in each particular word. Because native Japanese speakers recognize a mora on the
basis of timing, they might employ the rhythm of their first language. Although the first syllables were lengthened by 100 ms, the native English speakers were not affected by the added length as much as were the speakers of Japanese.

This result indicates that although the ratio of syllable duration does affect word recognition, native English speakers appear to take account of other information such as stress and recognize words in a comprehensive way. Native Japanese speakers, however, are more likely to be influenced by syllable duration. This might be due to their lack of linguistic knowledge of English as well as sensitivity to syllable duration. For native Japanese speakers, timing may be the basis of speech processing. Further study is needed to determine whether we can adjust our psychological timing when learning a second language.

The results of this study may contribute to the accumulation of evidence about whether L2 speech perception is an acquired linguistic ability or a basic property of the auditory system. This study revealed a significant difference between the ability of native English speakers and that of Japanese speakers to identify compound words using syllable duration. The results also showed that native English speakers appear to give consideration to both stress and syllable duration, whereas native Japanese speakers are likely to focus on the changed acoustic information (in this case, syllable duration). This in turn suggests that speech perception for Japanese speakers is a basic property of the auditory system, but it is highly likely that Japanese speakers process L2 acoustic information by applying their way of L1 speech processing. While previous studies on the development of L2 speech perception have dealt with phonemes and have shown that the perception of phonemes themselves changes, the present study showed that both native English speakers and native Japanese speakers were affected by syllable duration and that Japanese speakers were more sensitive to variation of acoustic information. This indicates that, unlike the perception of phonemes, the perception of syllable duration may not change fundamentally but that L2 learners may learn to integrate various factors at the same time like native English speakers as they develop their proficiency in the target language. Another possibility is that the L1 prosodic feature might govern L2 processing despite the development of L2 proficiency. In any case, we might need an additional speech perception model for suprasegmental features. Further study is needed to determine how advanced learners of English use syllable duration for L2 speech processing.

In conclusion, although neither group tended to recognize a threshold value of syllable duration enabling them to discriminate compound words from nominal phrases, the native English speakers and the native Japanese speakers appeared to take into account stress and the ratio of the duration between the first and second syllables, respectively, for word recognition. The native Japanese speakers were more sensitive to the ratio of syllable duration of the first syllable to the second syllable. This result might stem from the fact Japanese is a mora-timing language and native Japanese speakers use this rhythm for speech processing.

To uncover the system of speech processing in second language acquisition, further study is needed to determine in what way prosodic factors such as sound quantity (syllable length) influence L2
word recognition and how learners with high proficiency of English process L2 speech.

References


