【論文】

Effects of L1 Phonotactic Constraints and Orthography on L2 Word Segmentation

Tamami KATAYAMA

要旨（Abstract）

The aim of this study was to determine the effects of L1 phonotactic constraints and orthography on L2 word segmentation. Native Japanese speakers with a high level of English proficiency (JH), those with a low level of English proficiency (JL) and native English speakers (ES) performed a word spotting task and detected CVN (e.g., *pen*) targets and CVNC (e.g., *pend*) targets in bisyllabic non-words either in a context with clear boundaries (e.g., *pinkfem*) or in a context with ambiguous boundaries (e.g., *pinklem*). The results revealed that the ES and the Japanese speakers responded to CVNC target words faster and more accurately than to CVN targets both in Experiment 1 and Experiment 2. They syllabified the target words to identify them while being affected by the Possible Word Constraints (PWC). In addition, ES did not respond to non-words with /ŋ/ to detect CVN words. The differences of the nasal contexts affected the responses by ES and JH.

キーワード（Keywords）：phonotactics, syllable, second language, speech processing

1. Introduction

One of the issues in the study of speech processing is how listeners separate one word from another to recognize words in the stream of speech. It has been reported that listeners process speech on the basis of the minimum unit of the language to which they are exposed (Cutler & Butterfield, 1992). These word segmentation strategies have been reported to be language-specific (Cooper, Cutler, & Wales, 2002) and to include a stress-timed strategy (Cutler & Carter, 1987), a syllable-timed strategy (Cutler, Mehler, Norris & Segui, 1986), and a mora-timed strategy (McQueen, Otake, & Cutler, 2001; Otake, Hatano, Cutler, &Yoneyama, 1996). Thus, another issue is how knowledge of L1 and L2 is involved in L2 speech perception (Sebastian-Galles & Diaz, 2012). Investigating the relationship between first language and second language speech perception will shed light on the mechanism of development of second language acquisition (Sebastian-Galles & Diaz, 2012).

It has been reported that phonotactic constraints vary depending on the language (Ernestus, Kouwenhoven, & Mulken, 2017; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Parlato, Frota, Hirose & Peperkamp, 2011), which can lead to perceptual repair of illegal phonotactics. For instance, Spanish listeners repair L2 English [#sC] and perceive it by inserting a vowel before the illicit consonant cluster and recognize *sports* as [ɛspɔrts]. Carlson (2018) tested learnability of [#sC] in late L2 English learners and the weakening of perceptual illusion for early bilinguals. He conducted an AX discrimination task in two groups of Spanish-

English bilinguals and Spanish monolinguals to examine whether they could detect acoustic details. The two Spanish-English groups included a group of early bilinguals in an English-dominant environment and a group of late bilinguals in a Spanish-dominant environment. The results showed a negative correlation between English proficiency and strength of the perceptual illusion. Early bilinguals performed the best and late bilinguals outperformed the monolinguals. Learning an L2 can weaken L1 perceptual illusion even for late adult learners. Carlson argued that learning an L2 means not only having an additional linguistic system but also reorganizing both L1 and L2.

Since revealing a unit of speech is one of the issues in psycholinguistics, Mehler, Dommergues, Frauenfelder, and Segui (1981) examined whether native French speakers segment speech by phoneme or by syllable. They conducted a monitoring task using five pairs of bisyllabic French nouns (e.g., *pa-lace* and *pal-mier*) that were composed of either the first two or three phonemes of the word (e.g., *pa* and *pal*). The participants were asked to respond as fast as possible when they detected either the CV (consonant and vowel) targets or the CVC (consonant, vowel, and consonant) targets in experimental sequences that consisted of the target and bisyllabic filler words. When the participants responded to a CVC word in a bisyllabic word with a CVC target, their response was faster than when the same word did not correspond to the first syllable of the targets. The results of this study supports the hypothesis of syllabic segmentation; that is, a syllable is a unit of speech processing.

In addition to the unit of speech in a language, suprasegmental factors such as stress affect speech processing. Cutler and Norris (1988) examined a word segmentation strategy by native English speakers and found that stress facilitates their word segmentation (Cutler, 1986; Norris, McQueen, & Cutler, 1995). Mattys and Samuel (2000) further investigated the effect of the position of stress using initial-stress words and non-initial-stress words and reported that non-initial-stress words required additional processing time because early information has to be retained until the late information is incorporated for the decoding. They also reported that the listeners distinguished primary stressed syllables and secondary stressed syllables and activated them to different degrees. On the other hand, White, Mattys, and Wiget (2012) reported that listeners are more sensitive to durational cues than stress-based categorization. They created a sequence of *sasasa* syllables on the basis of the recordings of English sentences and Spanish sentences spoken by native speakers. Each /s/ and each /a/ vowel had the same duration as the intervals between boundaries of the corresponding consonant and vowel of the original recordings. The participants were informed that they would listen to unknown languages and were asked to identify which language they heard. As a result, they distinguished the given syllable using temporal cues alone rather than categories of rhythm class.

Phonotactics also affects speech processing. McQueen (1988) examined the effects of phonotactics on speech segmentation by Dutch speakers, who use stress in their speech. In his experiments, the participants failed more frequently to detect words that were misaligned with syllable boundaries cued by phonotactic constraints than words that were aligned with such boundaries. McQueen suggested that phonotactic legality is taken into consideration and helps listeners to segment words. He also stated that phonotactic constraints are likely to be one of the sources of information, such as silence and metrical cues, when listeners segment words. Norris, McQueen, Cutler, and Butterfield (1997) proposed a Possible Word Constraint (PWC) that regulates locations of word boundaries and helps listeners segment words indirectly. In this regulation, listeners disfavour word boundaries at points which would leave impossible residue words. Weber and Cutler (2006) reported that German speakers of English exploited English phonotactics to spot English words. In English, words that begin with /s/ are legal, while this is not the case in German. On the other hand, words may begin with /ʃ/ in German, though this is not the case in English. Thus, /s/-initial and /ʃ/-initial clusters are aligned with clear boundaries in German and in English. Their results showed that German listeners exploited a boundary constraint given only to English, but a phonotactic boundary given only to German was not an advantage for English listeners to spot words. From the results showing that the participants in their study were also sensitive to the L1 boundary cue, they argued that L2 listeners were able to make use of L2 phonotactic constraints for their listening but were not able to suppress L1 lexical activation at the same time.

Since a unit of timing called mora is used in Japanese language (Kubozono, 2001; Port, Dalby, & O'Dell, 1987), it was reported that native Japanese speakers segment words on the basis of mora (Culter & Otake, 1994). Otake, Hatano, Cutler, and Mehler (1993) tested a mora-based segmentation strategy for native Japanese speakers. A mora is formed by a vowel itself and a consonant followed by a vowel. The Japanese mora is often idetnfied with 'haku (拍)', an isochronal unit, more precisely the interval between the releases of the onset C, which is not available in non-Japanese utterances. N is also counted as morae, although it does not function in the word initial position. They asked the Japanese participants to detect CV (e.g., *ta*) targets and CVC (e.g., *tan*) targets in CVCVCV words (e.g., *tanishi*) and CVNCV (consonant-vowel-nasal-consonant-vowel sequence) words (e.g., *tanshi*). They found that the miss rate of CVN targets with CVCVCV words was extremely high (64.3%) compared to the other three conditions (less than 8%), and in CVNCV words, the response times (RTs) to CVN targets were longer than those to CV targets. The high miss rate is explained not by the syllable hypothesis but by the mora hypothesis because CVN targets and CVCVCV words do not match at the mora level. Otake et al. argued that Japanese listeners did not decompose words into syllables but segmented the spoken words by morae. According to the researchers, the reason why the Japanese listeners took longer in responding to CVN targets in CVNCV words was due to a complex target: two-mora words. In addition, the response pattern to CV targets was identical in CVCVCV and CVNCV words, being inconsistent with the prediction of syllable-based segmentation but consistent with the mora-based segmentation hypothesis because the initial mora is CV in both CVCVCV and CVNCV words. They also conducted the same experiments using native English speakers and native French speakers and found, as expected, that the patterns of performance by native and nonnative listeners of Japanese differed. That is, the French listeners used syllabic segmentation and the native English speakers used neither morae nor syllables to detect the targets, confirming the theory that speech segmentation is language-specific. However, Otake et al. presented the target words using Roman letters instead of Japanese letters. Japanese hiragana and katakana consist of a vowel, one or more consonants followed by a vowel and exceptional consonants such as /N/. The results showing that the participants missed CVN targets in CVCVCV words may be because Japanese letters are one of the components of their mental lexicon to access Japanese words. If they had been presented English words to detect CV targets and CVC targets, the results might have differed.

Although Cutler and Otake (1994) reported that native Japanese speakers segmented speech based on morae, their methodology raised the question stated above. Katayama (2022) examined whether L1 phonotactic constraints affect L2 auditory word segmentation. L2 speakers with different levels of English proficiency and native English speakers were instructed to spot CV (a consonant followed by a vowel; e.g., *pea*) target words with a legal syllable structure in Japanese and CVC (a sequence of a consonant, a vowel and a consonant; e.g., *peak*) target words with illegal phonotactics that were embedded in a context with clear syllable boundaries (e.g., *peaksom*) and in a context with ambiguous syllable boundaries (e.g., *peaklef*). The results showed that both L1 and L2 speakers responded faster and more accurately to CVC targets than to CV targets and in the context with ambiguous syllable boundaries than in the context with clear syllable boundaries. When L1 speakers of English detect CV target words (e.g., *pea*) in a context with clear word boundaries (e.g., *peaksom*), an illegal syllable will be left as the second syllable (e.g., *ksom*). This is thought to have caused delayed response time and more errors. It is noteworthy that L2 speakers whose L1 inventory does not possess either [*klef*] or [*ksom*] phonotactics reacted in the same way as the L1 speakers. Katayama's study used materials in L2 and showed that phonotactic constraints did not interfere with L2 processing by native Japanese speakers. Although her study provided evidence that L2 speakers used the segmentation strategy that was used by L1 speakers, the methodology used in her study also left a question about whether L2 speakers used the given spellings rather than their mental lexicon since both spelling and pronunciation of the target words were given. Although it has been reported that segmentation strategy is language-specific, results might differ depending on materials and methods used in different tasks. To confirm this hypothesis, further empirical studies are needed.

2. Research questions

In order to verify the claim in the report by Otake, Hatano, Cutler and Mehler (1993) that word segmentation strategies are language-specific and that native Japanese speakers segment L2 words on the basis of morae, the present study was carried out to determine whether native Japanese speakers segment English words using a mora-based segmentation strategy and whether spelling affects their segmentation. The following research questions were raised: Do native Japanese English learners detect English words with CVN syllable structures faster than those with CVNC structures? Is there a difference among native English speakers and L2 speakers of different levels in English proficiency in identifying target words (i.e., CVN vs. CVNC) when they are given spellings and when they are not giving spellings?

The method used in the present study was based on the method used by Katayama (2022) in a word-spotting task for bisyllabic non-words with different syllable boundary contexts (i.e., clear boundary vs. ambiguous boundary). Although Katayama used CV targets and CVC targets, CVN targets were used in the present study as in the study by Otake et al. (1993), but English words were used instead of Japanese words. In two experiments in this study, a nasal consonant /N/ was employed because /N/ is considered to be an independent mora in Japanese, while /n/ itself is not an independent syllable in English. Target words of CVN (e.g., *pen*) and CVNC (e.g., *pend*) were embedded in non-words with clear syllable boundaries (e.g., *pendklus*) and with ambiguous syllable boundaries (e.g., *pendret*) with stress on the first syllables. If Japanese speakers employ a mora-based segmentation strategy, it is predicted that they will segment CVN words faster than CVNC words because CVN is a legal syllable structure consisting of two morae, while the phonotactics of CVNC are illegal since Japanese does not take a consonant as a coda. On the other hand, if the native English speakers do not syllabify the items, there will be no difference in identifying CVN targets and CVNC targets regardless of the position of syllable boundaries. If they syllabify non-words, they will take longer in responding to CVN targets (e.g., *pen*) in the contexts with clear boundaries (e.g., *pendklus*) than in the contexts with ambiguous boundaries (e.g., *pendret*) because an extraneous phoneme will be produced (e.g., /d/). In the contexts with ambiguous syllable boundaries, there will be no difference between CVN words and CVNC words since both residues can form legal syllables in English (e.g., *ret* and *dret* in *pendret*). Furthermore, if Japanese speakers use the L2 mental lexicon that also includes spelling, there will be no difference in responses when they are given spelling and when they are not given spelling. If the spelling itself is a variable affecting their L2 word segmentation, the results may differ between the contexts with and without spelling.

3. Experiments

Since Otake et al. (1993) used Japanese words such as *tanishi* and *tanshi* as target words, it is assumed that the Japanese participants accessed their L1 mental lexicon and searched for the target words using Roman letters that correspond to Japanese words in their mind. The use of English words in the present study eliminated variables of their L1 mental lexicon and made it possible to examine whether native Japanese speakers detect words on the basis of syllables. In Experiment 1, visual stimuli were presented as in the study by Otake et al. and the results obtained when the target words were L1 words were compared with the results when they were L2 words to reveal whether the Japanese participants use morae in segmenting L2 words as well. In Experiment 2, the results obtained when visual stimuli were given were compared with the results obtained when they were not given to determine whether the spelling affects L2 word segmentation by native Japanese speakers. To confirm whether they use syllables instead of the linear order of phonemes, ambiguous syllable boundary and clear boundary contexts were given. If they rely on the linear order presentation of phonemes, difference caused by the contexts would not be observed.

3.1. Experiment 1

3.1.1. Methodology

Materials

Ten pairs of monosyllabic words sharing the same initial three phonemes were selected with the first three phonemes (CVN) making up the first member of the pair and the initial four phonemes (CVNC) forming the second member of the pair: *pen-pend*, *pan-pant*, *ten-tent*, *pin-pink*, *ton-tong*, *den-dent*, *Dan-dank*, *tan-tank*, *ban-bank*, and *Ben-bend*. The first members of the pair (CVN) are legal syllable structures in Japanese, while the second members are not. Non-words with clear syllable boundaries and those with ambiguous syllable boundaries were prepared by adding monosyllables to the target pairs (e.g., *pendklas* and *pendret*) (see Table 1). The structures of the added syllables were different between non-words with clear syllable boundaries and those with ambiguous syllable boundaries. In clear boundary contexts, added syllables consisted of CVC, CCVC, and CVCC (e.g., *fem*, *klus*, *monf*), while in ambiguous boundary contexts, CVC (e.g., *ret*) syllables were added in order to create ambisyllabicity. In addition to the target words, 400 distractors were created by embedding monosyllabic words in non-words. The syllable boundaries were clear in half of these and ambiguous in the rest. These non-words were embedded in carrier sentences "Please say ___" and recorded by a female native English speaker from Canada at a normal rate with Pro Tools 12, so that stress should fall on the same position and intonation should be consistent. The speaker put a short pause between sentences and a long pause every 4 to 6 sentences to prevent the speaker from making her own rhythm when reading. All of the non-words were taken out from the sentences using *Praat*, a speech analyzer. The researcher checked whether the target words were pronounced as intended, especially whether the consonants following the nasals (i.e., **C** in CVN**C**) were articulated. As shown in Table 1, the ratio of mean durations between the first syllables and the second syllables of CVN words with clear syllable boundaries was 1: 2.3, while that for CVN words with ambiguous syllable boundaries was 1: 2.1. The ratio for CVNC words with clear syllable boundaries was 1: 1.6, while that for CVNC words with ambiguous syllable boundaries was 1: 1.4. Although there was a slight difference in the mean durations between the ambiguous contexts and the clear contexts, the ratios between them were not noticeably different.

Table 1

Phonetic Features of Legal (CVN) and illegal (CVNC) target syllables in Japanese embedded in non-words with clear word boundaries and those with ambiguous word boundaries

boundary context	target non-words	Whole duration (ms)	vowel quality in the final syllable F1 (Hz)	F2 (Hz)	F3 (Hz)	length of CVN	length of second syllable	length of CVNC	length of second syllable
clear	pendklus	1022	555	1279	2520	314	708	378	644
	pantmonf	900	714	1127	2736	222	678	350	550
	dentklep	755	649	1441	2476	273	482	325	430
	tentgok	887	724	1222	2429	206	681	263	624
	bendpref	791	639	1591	2111	185	607	287	505
	pinkfem	792	601	1627	2318	357	547	390	514
	tongplem	904	585	1543	2480	303	488	336	455
	dankminf	952	436	1989	2561	285	667	383	569
	tanktok	856	668	1169	2388	269	587	320	536
	bankfem	855	634	1605	2326	232	623	284	571
ambiguous	pendret	625	619	1673	2489	214	411	244	381
	pantrep	644	637	1567	2361	216	428	295	349
	dentrim	528	445	1584	2123	170	358	253	275
	tentrot	670	710	1287	2381	172	498	268	402
	bendrim	646	482	1704	2344	173	307	230	250
	pinklem	480	539	1332	2450	350	516	382	484
	tongret	866	595	1684	2385	212	434	253	393
	danklef	672	641	1523	2408	198	474	262	410
	tanklep	734	661	1522	2449	241	493	284	450
	bankrosh	740	684	1208	2352	195	545	265	475

Two versions of a monitoring task were created with the aid of *E-Prime* software in two languages, in which instructions were in English for the English speakers and in Japanese for the Japanese speakers. Each version had 40 experimental sequences including 20 positive sequences that had one target and 20 negative sequences that did not have a target. Each experimental sequence consisted of 8 to 12 items. The 20 target words containing both members of a pair (i.e., CVN and CVNC) were each placed as the first to tenth items in respective positive experimental sequences and the rest of the spaces were occupied by fillers. For example, when a target word was *pinkfem*, *pin* was presented in Version I and *pink* was presented in Version II. The participants were divided into two groups: one group took part in Version I and the other group took part in Version II. Thus, the same participants did not listen to the same target word twice, and eight of the sixteen participants in each group responded to each target word. The proportions of the different types of word boundaries (i.e., clear vs. ambiguous) and target words (i.e., CVN and CVNC) were counterbalanced in each sequence. The 20 negative sequences were composed entirely of fillers.

A Toshiba computer, CF-sz6, was used for the experiments together with a multifunctional response and stimulus device, Chronos. A target syllable was set up to be presented auditorily and visually followed by a blank screen for 750 ms, and an experimental sequence was programmed to run while the mark "+" appeared in the center of the screen. If a response was made while the words in a particular list were being presented, this list was programmed to jump to the next experimental sequence. If there was no response by the end of the sequence, the screen informed the participant to press the button to continue.

Participants

Sixteen native English speakers (ES), sixteen native Japanese speakers with a high level of proficiency in English (JH) and sixteen native Japanese speakers with a relatively low level of English proficiency (JL) took part in the experiment. ES consisted of mainly English instructors in Japan, of whom eleven were from the U.S.A., 2 from Canada, 2 from U.K., and 1 from Australia. Their mean age was 44 years and the mean duration of living experience in Japan was 18.3 years. JH were mainly English instructors at colleges and their mean age was 42 years. Their mean duration of living experience in English speaking countries was 5.1 years and their mean score for TOEIC was 927. JL included undergraduate students from Kumamoto University whose mean age was 19 years, and their mean score for TOEIC was 533.8. None of them had no living experience abroad, although two of them had visited Canada for two weeks and one month, respectively.) A book voucher was given to each participant to compensate for their time. All of the participants reported no hearing impairment.

Procedure

The experiment was conducted individually in a quiet room. First, the participants filled in a questionnaire about their age, experience living in English speaking countries (or Japan for native English speakers), and their English proficiency (Japanese proficiency for native English speakers), and they signed a consent form confirming their right to discontinue participation and maintenance of their privacy. Then each participant was instructed to sit in front of the computer, put on the headphones and follow the instructions presented on the screen. They were asked to respond as quickly as possible by pressing button "1" on Chronos when they heard the target syllable in a word. The response time from a target word was recorded. Each word in the experimental sequence was followed by a 1000-ms interval, and the next words were presented automatically. After two experimental sequences for practice, the participants completed 40 real experimental sequences with a voluntary interval after 20 sequences if the participant wanted it.

3.1.2. Results

The response time (RT) was defined as the duration from the end point of a CVN/CVNC target to the time point when a participant made a response. Then, mean RTs were determined and the number of missed responses was counted for each subject (F1) and each item (F2). Table 2 shows the mean RTs and miss rates in Experiment 1. Figure 1 presents the mean RTs by the three groups and Figure 2 shows the miss rates.

Table 2
Mean response time (ms) and miss rate of responses as a function of the syllable boundary for CVN targets and CVNC targets

	target	CVN	CVNC
ES	clear	627.9 (30.0%)	453.9 (13.8%)
	ambiguous	534.1 (23.8%)	453.5 (0%)
JH	clear	596.3 (13.8%)	401.1 (8.8%)
	ambiguous	549.4 (7.5%)	435.3 (3.8%)
JL	clear	467.8 (12.5%)	368.3 (3.8%)
	ambiguous	468.8 (10.0%)	391.4 (5.0%)

Note. ES (n = 16): native English speakers; JH (n = 16): native Japanese speakers with a high level of English proficiency; JL (n = 16): native Japanese speakers with a low level of English proficiency.

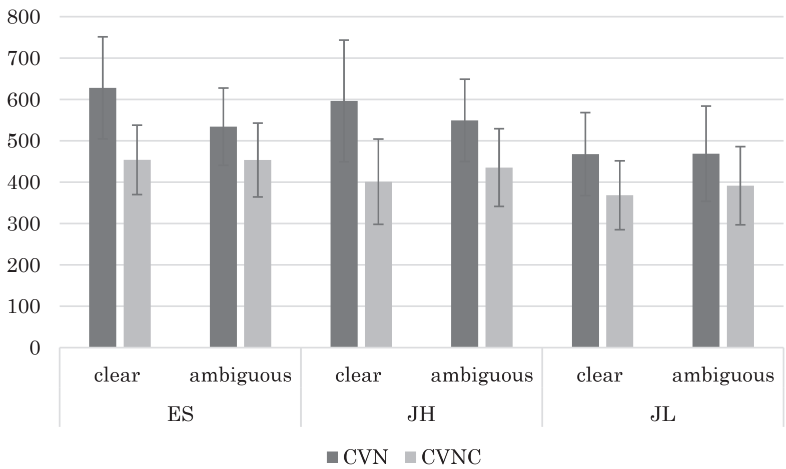

Figure 1. The mean response time (ms) for the three groups as a function of the syllable boundary (clear, e.g., *pendklus*, versus ambiguous, e.g., *pendret*) for CVN targets (e.g., *pen*) and CVNC targets (e.g., *pend*) targets in Experiment 1.

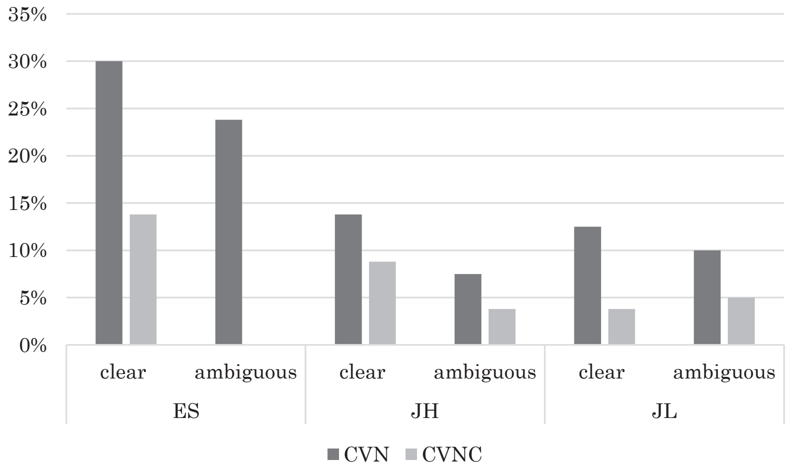

Figure 2. The miss rate of responses for the three groups as a function of the syllable boundary (clear, e.g., *pendklus*, versus ambiguous, e.g., *pendret*) for CVN targets (e.g., *pen*) and CVNC targets (e.g., *pend*) in Experiment 1.

Since none of the participants in the ES group responded to *tongret* with *ton*, responses to *ton* and the corresponding word, *tong*, were deleted from the analysis of RT for ES. The results of three-way ANOVA for RT showed main effects of the target word ($F1(1,182) = 17.2$; $p < 0.001$; $\eta^2 = 0.09$. $F2(1,111) = 34.3$; $p < 0.001$; $\eta^2 = 0.2$.) and the group ($F1(2,182) = 3.6$; $p < 0.05$; $\eta^2 = 0.04$. $F2(2,111) = 15.6$; $p < 0.001$; $\eta^2 = 0.2$.), and there was no other main effect or interaction. All of the groups responded significantly faster to CVNC targets than to CVN targets in both contexts of syllable boundaries. ES took the longest time to detect CVN targets in non-words with clear syllable boundaries. In miss rate analysis, a significant main effect of the target word was found in both analyses by subject and by item ($F1(1,45) = 19.0$; $p < 0.001$; $\eta^2 = 0.3$. $F2(1,111) = 11.4$; $p < 0.01$; $\eta^2 = 0.1$.). There was an effect of the group only in analysis by item ($F2(2,111) = 3.6$; $p < 0.05$; $\eta^2 = 0.06$.), and there was an effect of the syllable boundary ($F1(1,45) = 10.4$; $p < 0.01$; $\eta^2 = 0.2$.) only in analysis by subject. A significant interaction between the target word and the group ($F1(2,45) = 4.1$; $p < 0.01$; $\eta^2 = 0.2$.) was found only in analysis by subject. All of the groups missed the CVN targets more frequently than the CVNC targets, and ES missed the CVN targets both with clear syllable boundaries (30%) and with ambiguous syllable boundaries (23.8%). According to the guidelines provided by Mizumoto and Takeuchi (2008), the effect size is considered small when the η^2 value of a two-way ANOVA and three-way ANOVA is greater than 0.01, medium when it is greater than 0.06, and large when it is greater 0.14. Judging from the guidelines, the effect sizes of the analyses range from small to large. This may be due to the number of cases being different between the analyses by item and those by subject. In addition, ES and JH were affected by the position of the syllable boundaries since they took longer and missed more frequently when the target words did not match the corresponding words at the syllable level.

As subsequent analyses, two-way ANOVA was conducted on RTs and missed responses by subject and by item for each group. The results for RT showed a main effect of the target word for all of the groups (ES:

$F1(1,15) = 21.3$; $p < 0.001$; $\eta^2 = 0.6$. $F2(1,36) = 6.0$; $p < 0.05$; $\eta^2 = 0.14$. JH: $F1(1,15) = 73.1$; $p < 0.001$; $\eta^2 = 0.8$. $F2(1,36) = 27.7$; $p < 0.001$; $\eta^2 = 0.4$. JL: $F1(1,15) = 26.8$; $p < 0.001$; $\eta^2 = 0.6$. $F2(1,36) = 12.1$; $p < 0.01$; $\eta^2 = 0.25$.). Only in ES was a main effect of the word boundary found in analysis by subject ($F1(1,15) = 6.7$; $p < 0.05$; $\eta^2 = 0.3$.). For the miss rate, the main effect of the target words was significant in ES ($F1(1,15) = 18.5$; $p < 0.01$; $\eta^2 = 0.6$. $F2(1,36) = 6.2$; $p < 0.05$; $\eta^2 = 0.15$.) in analyses both by subject and by item and in JL only in analysis by item ($F2(1,36) = 4.5$; $p < 0.05$; $\eta^2 = 0.1$.). The effect sizes range from medium to large. ES missed the CVN targets more frequently than did the other groups. None of the eight participants in the ES group responded to ton in *tongret*, five participants missed *dan* in *dankminf* and *tan* in *tanktok*, four participants missed *pin* in *pinkfem* and *pan* in *pantmonf*, and two participants missed *ban* in *bankfem*. Some participants informally reported after the experiment that they did not respond to CVN targets followed by velar stop consonants (i.e., /k/ and /ɡ/) since the nasal is realized as a different phoneme /ŋ/ from /n/ that is pronounced at the end of the priming words. The codas of these words were affected by the following consonants, and ES tended to perceive the acoustic difference and did not respond to the targets on purpose. In the contexts with ambiguous syllable boundaries, *dan* in *danklef* was missed by five of the eight participants in the ES group, but *ban* in *bankrosh* was responded to by all of the participants, indicating that velar stops did not always affect their perception of the nasals. Five out of the eight participants in the JH group missed *dan* in *dankminf*, four participants missed *pend* in *pendklus*, and three participants missed *ton* in *tongret*, but they missed zero or only one target in other words. For *ton* in *tongret*, since all of the participants in the ES group and three participants in the JL group missed it, the difference in vowel quality may have affected detection of the word by JL. JL responded faster than the other groups did and the difference between detection of CVN targets with clear boundary contexts and CVN targets with ambiguous boundary contexts was less than that in the other groups. Unlike ES and JH, JL did not take into account the position of word boundaries since they responded to the CVN targets in the context with clear word boundaries, in which the target words and the corresponding words in non-words did not match at the syllable level. JL tended to depend on acoustic information and they tried to match the presented phonemes of target words and those embedded in non-words without considering syllable structures. The reason why they responded to CVNC targets faster and more accurately is thought to be that the CVNC targets were acoustically closer to the corresponding words in non-words than were the CVN targets with respect to their duration and vowel quality.

In Experiment 1, it was focused on how native Japanese speakers and native English speakers detect target words that were presented with pronunciation and spellings in a series of non-words. Still, the possibility remained that the Roman writing system might have affected native Japanese speakers' processing of English words rather in detriment to the their phonological knowledge. In Japan, Roman letters are commonly used to express a Japanese word and a loan word from a foreign language. In the Roman letter system, the English alphabet is used and each mora has corresponding Roman letters. For example, 峰 (*summit*) is written as "*mine*" and ミント(*mint*) is written as "*minto*". Because of the use of Roman letters in Japan, it is possible that native Japanese speakers use the writing system to identify English words. In Experiment 2, in order to control for the effects of Roman letters, the target words were presented only through sounds and without showing letters representing the target words.

3.2. Experiment 2

3.2.1. Methodology

Materials

All of the materials were the same as those in Experiment 1 except that the letters of the target words were presented visually on the screen of the computer along with the sounds.

Participants

The same participants as those in Experiment 1 took part in Experiment 2 with an interval of more than one week to avoid practice effects.

Procedure

The same procedure as that in Experiment 1 was used.

3.2.2. Results

Mean RTs were determined in the same way as in Experiment 1, and the number of missed responses were counted for each subject and each item. Due to a technical error, the responses for *bend* in *bendpref* were deleted from the analyses for RT and miss rate. In addition, because one participant in the ES group missed more than 60% of the responses, his data were deleted from the analysis. None of the participants in the JH group responded to *dankminf* with *dan*, and the data were also deleted from the analysis of RT for this group. Table 3 shows the mean RTs and miss rates in Experiment 2. Figure 3 presents the mean RTs for the three groups and Figure 4 shows the miss rates.

The results of three-way ANOVA for RT showed main effects of the target word in both analyses by subject and by item ($F1(1,182) = 14.4$; $p < 0.001$; $\eta^2 = 0.08$. $F2(1,111) = 15.3$; $p < 0.001$; $\eta^2 = 0.12$.) and the syllable boundary only in analysis by item ($F2(1,111) = 6.4$; $p < 0.05$; $\eta^2 = 0.06$.), and there was no other main effect or interaction. Since η^2 values fall in the range from 0.06 to 0.12, the effect sizes of the analyses were medium. All of the groups responded significantly faster to CVNC targets than to CVN targets and they found it easier to identify the targets in contexts with ambiguous syllable boundaries. In miss rate analyses by subject and by item, there were significant main effects of the target word ($F1(1,44) = 22.4$; $p < 0.001$; $\eta^2 = 0.3$. $F2(1, 109) = 12.6$; $p < 0.01$; $\eta^2 = 0.1$.) and the syllable boundaries ($F1(1,44) = 23.4$; $p < 0.001$; $\eta^2 = 0.4$, $\mathrm{F2}(1,109) = 7.2$; $p < 0.01$; $\eta^2 = 0.06$.). There was a significant interaction between the syllable boundary and the group only in analysis by subject ($F1(2,44) = 6.2$; $p < 0.01$; $\eta^2 = 0.1$.). The effect sizes ranged from medium to large. All of the groups missed the CVN targets more frequently than the CVNC targets and missed more targets in the context with clear syllable boundaries than in the context with ambiguous syllable boundaries. Whether the target words match the corresponding words at the syllable level affected their responses.

Table 3
Mean response time (ms) and miss rate of responses as a function of the syllable boundary for CVN targets and CVNC targets

	target	CVN	CVNC
ES	clear	570.6 (25.0%)	457.4 (13.8%)
	ambiguous	523.1 (20.0%)	404.9 (3.8%)
JH	clear	530.9 (20.0%)	423.7 (15.0%)
	ambiguous	455.9 (10.0%)	380.5 (7.5%)
JL	clear	528.2 (23.8%)	372.4 (5.0%)
	ambiguous	439.2 (13.8%)	340.6 (5.0%)

Note. ES (n = 16): native English speakers; JH (n = 16): native Japanese speakers with a high level of English proficiency; JL (n = 16): native Japanese speakers with a low level of English proficiency.

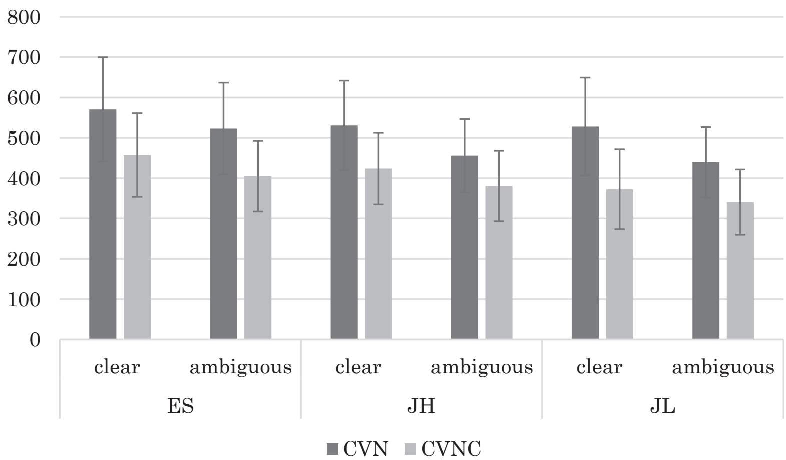

Figure 3. The mean response time (ms) for the three groups as a function of the syllable boundary (clear, e.g., *pendklus*, versus ambiguous, e.g., *pendret*) for CVN targets (e.g., *pen*) and CVNC targets (e.g., *pend*) in Experiment 2: (ES) native English speakers; (JH) native Japanese speakers with a high level of proficiency in English; (JL) native Japanese speakers with a relatively low level of English proficiency.

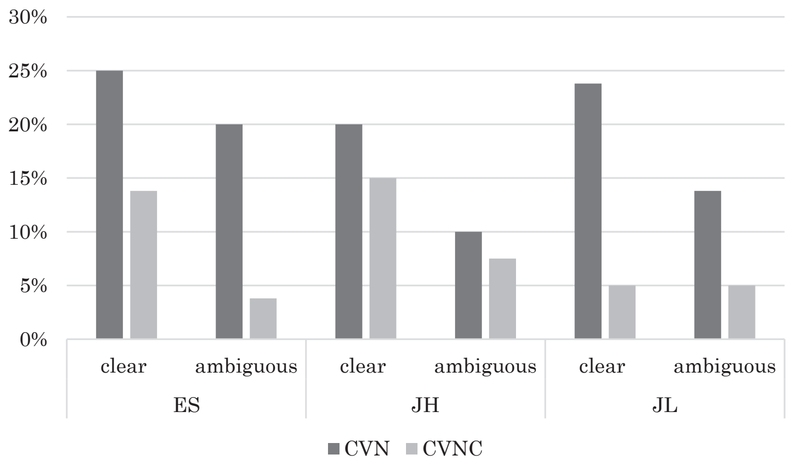

Figure 4. The miss rate of responses by the three groups as a function of the word boundary (clear, e.g., *pendklus*, versus ambiguous, e.g., *pendret*) for CVN targets (e.g., *pen*) and CVNC targets (e.g., *pend*) in Experiment II: (ES) native English speakers; (JH) native Japanese speakers with a high level of proficiency in English; (JL) native Japanese speakers with a relatively low level of English proficiency.

As subsequent analyses, two-way ANOVA was conducted on RTs and missed responses by subject and by item for each group. The results for RT showed a main effect of the target word for all of the groups in analyses by subject (ES: $F1(1,14) = 10.8$; $p < 0.01$; $\eta^2 = 0.4$. JH: $F1(1,15) = 15.9$; $p < 0.01$; $\eta^2 = 0.5$. JL: $F1(1,15) = 39.2$; $p < 0.001$; $\eta^2 = 0.7$.) and for ES and JL in analyses by item (ES: $F2(1,36) = 12.2$; $p < 0.01$; $\eta^2 = 0.3$. JL: $F2(1,36) = 11.7$; $p < 0.01$; $\eta^2 = 0.24$.). A main effect of syllable boundary was found in analysis by subject in JH ($F1(1,15) = 9.7$; $p < 0.01$; $\eta^2 = 0.4$.) and JL ($F1(1,15) = 12.3$; $p < 0.01$; $\eta^2 = 0.5$.). For the miss rate, a main effect of the target word was significant in analyses both by subject and by item for ES ($F1(1,14) = 7.6$; $p < 0.05$; $\eta^2 = 0.4$. $F2(1,35) = 5.4$; $p < 0.05$; $\eta^2 = 0.1$.) and JL ($F1(1,15) = 14.1$; $p < 0.01$; $\eta^2 = 0.5$. $F2(1,36) = 8.0$; $p < 0.01$; $\eta^2 = 0.2$.) and in analysis only by subject for JH ($F1(1,15) = 5.1$; $p < 0.05$; $\eta^2 = 0.3$.). There was a main effect of the syllable boundary for JH in analyses by subject and by item ($F1(1,15) = 25.6$; $p < 0.01$; $\eta^2 = 0.6$. $F2(1,36) = 4.1$; $p = 0.05$; $\eta^2 = 0.1$.) and for ES in analysis only by subject ($F1\ (1,14) = 7.3$; $p < 0.05$; $\eta^2 = 0.3$.). The effect sizes of the analyses for RT were large (range: 0.24 to 0.7) and those for miss rate were from medium to large (range: 0.1 to 0.6). Although all of the groups responded faster to the CVNC targets than the CVN targets, JH were more affected by the syllable boundaries. In addition, JH and ES missed the targets more frequently when responding to the CVN targets in the contexts with clear syllable boundaries than in the rest of the contexts, that is, when the target words did not match at the syllable level.

Overall, ES responded faster and more accurately in Experiment 2 than Experiment 1. They responded to CVNC targets faster in Experiment 2 than in Experiment 1, but their miss rate increased in Experiment 2. While all of the participants missed *ton* in *tongret* in Experiment 1, five of them responded to it in Experiment 2. Both *dan* and *dank* in *dankminf* were missed by five participants, *pink* in *pinklem* was missed by four participants, and *ban* in *bankfem*, *tan* in *tanktok*, *pan* in *pantmonf*, *ten* in *tentdrot*, and *ton* in *tongret* were missed by three participants. The nasal /n/ in the coda position of the target words was affected by the following phoneme: the nasal is represented as /ŋ/ followed by a uvular stop (i.e., /k/ and /ɡ/]), while it is realized as a dental stop nasal [n̪] followed by a dental stop (i.e., /t/ and /d/). The native English speakers tended to perceive the fine acoustic differences and recognized the nasals in the target words as either a different phoneme or its allophone. Moreover, a possible reason why ES responded to more targets in Experiment 2 than in Experiment 1 is that they might have been more lenient in identifying the target words when given only sounds since they were not provided the spelling of words, which would lead them to match their mental lexicon.

JH responded faster in Experiment 2 than in Experiment 1 except when responding to CVNC targets in the context with clear boundaries, but their miss rate increased in Experiment 2. Their miss rate for CVN targets in the syllable context with clear syllable boundaries was 20%, while that for CVNC targets was 15%. Five of the eight participants in the JH group missed *ban* in *bankfem*, four of them failed to respond to *tan* in *tanktok*, *dank* in *dankminf*, and *pend* in *pendklus*, and three did not identify *pan* in *pantmonf*, ton in *tongplem* and *ban* in *bankrosh*. The increase in the miss rate might have been caused by the fact that they found it hard to retain the target words in their mind without being given visual information. In addition, since they were not given spelling to rely on, they were likely to be more perceptive to comprehensive acoustic information including syllable structures.

JL responded faster when the context was ambiguous than when it was clear. Specially, when they responded to the CVN target in target words with clear boundaries, their miss rate almost doubled (23.8%) compared to that in Experiment 1 (12.5%). Seven of the participants in the JL group missed *dan* in *dankminf*, four failed to detect *ton* in *tongplem* and three did not identify *ban* in *bankfem*. Less than two participants missed the rest of the targets. The value of missed responses was close to that of ES and it is assumed that JL paid more attention to phonotactics for segmenting the target words. The results of Experiment 2 did not support the mora-based segmentation hypothesis, according to which native Japanese speakers employ morae to segment even foreign speech. Even less proficient L2 learners took into account syllable structures that do not exist in their first language.

Although the purpose of the present study was to examine whether native Japanese speakers segment CVN words with legal phonotactics in Japanese and CVNC words with illegal syllable structures on the basis of morae, there was the possibility that the consonant following the nasal affected the responses. The nasal consonant /n/ is affected by the following phonological context and is realized into two different phonemes in English, /n/ and /ŋ/, while it is represented as ん/N/ in Japanese. The reason why ES missed more and took longer than the Japanese groups when given CVN targets in the context with clear syllable boundaries might have been a mismatch of visual presentation of the target CVN and the actual realization of [CVŋ]. Further analyses were conducted to examine whether there is a difference when the participants detected a [CVn] target and when they identified a [CVŋ] target.

3.3. Follow-On Analyses

The target non-words used in Experiment 1 and Experiment 2 were categorized depending on whether the coda of CVN was a velar nasal /ŋ/ or not as shown in Table 4. Three-way ANOVAs (3 groups x 2 types of nasals x 2 types of syllable boundaries) were conducted across items for RT and missed responses in Experiment 1 and Experiment 2. Analyses across participants were not conducted due to the uneven number of responses per participant. For the analysis of RT in Experiment 1, there were significant main effects of group (F (2,48) = 11.4, p < 0.001; η^2 = 0.3) and nasal (F (1,48) = 9.0, p < 0.01; η^2 = 0.2). The interaction between group and nasal was not significant but peripheral (F (2,48) = 3.2, p < 0.051; η^2= 0.1). Responses to [CVn] targets were significantly faster than to responses to [CVŋ] targets. The Tukey test as a post-hoc test revealed that JL responded significantly faster (452.4 ms) than did JH (578.3 ms) and ES (596.6 ms). Two-way ANOVAs for each group revealed main effects of nasal in ES (F (1,16) = 6.0, p < 0.05; η^2 = 0.3) and JH (F (1,16) = 5.2, p < 0.05; η^2 = 0.2). There were no other main effects and interactions. For the analysis of missed responses in Experiment 1, there were significant main effects of group (F (2,48) = 4.7, p < 0.001; η^2 = 0.2) and nasal (F (1,48) = 10.4, p < 0.01; η^2 = 0.2). The number of missed responses to [CVŋ] targets was significantly larger than the number of missed responses to [CVn]. The number of missed responses to [CVŋ] targets by ES was significantly larger than that in the other groups. Two-way ANOVAs for each group revealed that a main effect of nasal was observed only in ES (F (1,16) = 7.3, p < 0.05; η^2 = 0.3). Only ES missed [CVŋ] words significantly more than [CVn] words.

Three-way ANOVA of missed responses in Experiment 2 showed main effects of nasal (F (1,48) = 15.8, p < 0.001; η^2 = 0.3) and boundary (F (1,48) = 6.9, p < 0.05; η^2 = 0.1). The number of missed responses to [CVŋ] targets was significantly larger than that to [CVn] as in Experiment 1, but the effect of syllable boundary was significant. The participants missed words with an ambiguous syllable boundary less than words with a clear syllable boundary. Two-way ANOVAs for each group revealed a main effect of nasal in ES (F (1,16) = 7.1, p < 0.05; η^2 = 0.3) and JH (F (1,16) = 6.8, p < 0.05; η^2 = 0.3). There were no other main effects and interactions. The number of missed responses to [CVŋ] targets was significantly larger than the number of missed responses to [CVn] in ES and JH. Three-way ANOVA for RT showed no main effects or interactions. Table 5 shows summaries of results of three-way ANOVA and two-way ANOVA in Experiment 1 and Experiment 2.

Table 4

CVn target words and CVŋ target words embedded in non-words with clear word boundaries and those with ambiguous word boundaries

clear		ambigous	
[CVn]	pendklus	[CVn]	pendret
	pantmonf		pantrep
	dentklep		dentrim
	tentgok		tentdrot
	bendpref		bendrim
[CVŋ]	pinkfem	[CVŋ]	pinklem
	tongplem		tongret
	dankminf		danklef
	tanktok		tanklep
	bankfem		bankrosh

Table 5

Main effects of missed responses in Experiment 1 and Experiment 2

	Experiment 1		Experiment 2	
	3-way ANOVA	2-way ANOVA	3-way ANOVA	2-way ANOVA
RT	**group (JL <ES, JH) **nasal	*ES: nasal *JH: nasal	no significance	N/A
miss	**group (ES > JL, JH) **nasal	*ES: nasal	**nasal *boundary	*ES: nasal *JH: nasal

Note. *: $p < 0.05$, **: $p < 0.01$

To sum up, [CVŋ] targets were missed significantly more than [CVn] targets both in Experiment 1 and Experiment 2. In Experiment 1, RTs by ES and JH were significantly longer than the RT by JL, suggesting that ES and JH considered the context of non-words with the target words and accessed their mental lexicon and that JL depended on acoustic information to identify words. That is, ES and JH processed the words at a deeper level than did JL. Regardless of whether they were given the spelling (Experiment 1) or not (Experiment 2), ES did not respond to words with [ŋ] to spot the [CVN] words. The number of missed responses by ES was significantly larger than the numbers of missed responses by JL and JH.

In Experiment 2, in which spellings were not provided, the participants considered the syllable context as well as the nasal context. They did not respond to words with /ŋ/ and words with a clear syllable boundary that leave illegal phonotactics in the second syllable. The nasal type was the most influential factor for ES to spot target words, while JH were affected by phonotactic context as well as nasal context. JL did not consider the differences between those phonemes and they identified both /n/ and /ŋ/ in English as /N/ in Japanese in both Experiment 1 and Experiment 2. It is plausible that recognizing the nasal for JL is an issue of the relationship between spelling and its corresponding L1 phoneme. That is, they lack knowledge about acoustic details that each English word possesses in various phonological contexts and they were affected by the Roman letter “N” that corresponds to the Japanese “ん”.

4. Discussion

The aim of this study was to determine the effects of phonotactic constraints of the first language and orthography on word recognition in a second language. Experiments were carried out to determine whether word segmentation strategies are language-specific and whether native Japanese speakers segment L2 words on the basis of L1 phonotactics when using English words given both pronunciation and spellings. Since word stress was placed on the first syllables of all of the words used in the experiments in this study, only the effect of phonotactic constraint was examined through the monitoring tasks. The following research questions were the focus of the study:

Do native Japanese English learners detect English words with CVN syllable structures faster than those with CVNC structures?

Is there a difference among native English speakers and L2 speakers of different levels in English proficiency in identifying target words (i.e., CVN vs. CVNC) when they are given spellings and when they are not giving spellings?

With respect to the first question, the evidence from this study suggests quite strongly that they do not segment L2 speech on the basis of L1 phonotactics. Regardless of English proficiency, native Japanese English learners responded faster and more accurately to target words with illegal phonotactics in Japanese (CVNC) than to those with a legal syllable structure (CVN). Regarding the second research question, all of the groups syllabified the non-words to detect the target words, and they missed more and took longer to respond when the target words did not match those embedded in non-words at the syllable level both when they were given spellings (visual presentation) and when they were not given spellings (auditory presentation). However, with respect to the miss rate, the presence of Roman letters had a greater effect for JH and JL than for ES. For CVN targets with clear syllable boundaries and when not given the spelling, JH missed 16.2% more responses than when they were given the spelling, while the miss rate for JL was 11.3% higher when they were not given spellings.

In the two experiments in this study, the same target words (e.g., *pin*) were used in the context with clear boundaries (e.g., *pinkfem*) and in the context with ambiguous boundaries (e.g., *pinklem*), and the frequency of the tokens was thus identical in non-word contexts. However, the results of this study showed that the position of the word boundaries in non-words affected response times and miss rate for detecting the target words. The increases of RT and miss rate in detection of the CVN targets in the context with clear boundaries indicate that the listeners considered the structure of the second syllable. If they segment the CVN targets (e.g., *pin*) in the context with clear boundaries (e.g., *pinkfem*), an extraneous phoneme [k] will be produced since the rest of the phonemes are not able to form a legal syllable (i.e., [kfem] is illegal phonotactically as the onset in English). On the other hand, in the context with ambiguous boundaries (e.g., *pinklem*), segmenting the CVN targets did not cause delayed response times and high miss rates as in the context with clear boundaries since the rest of the phonemes can form a legal syllable (e.g., *klem*).

When English words were used, the Japanese speakers did not employ a mora-based segmentation strategy as proposed by Otake et al. (1993). Orthographies including *hiragana, katanakana, kanji*, and Roman letters are assumed to be stored in the lemma of a Japanese speaker's mental lexicon, which represents the syntactic properties of words. In the study by Otake et al., Japanese words such as *tanishi* were used as targets and the participants were instructed to spot "tan" presented in Roman letters. Given the Roman letters, the participants had to spell out "たにし" in Japanese to "*tanishi*" in Roman letters in order to spot *tan*. It is considered that failure in detecting the target word is due to the low frequency of appearance of a Japanese word represented in Roman letters. In the present study, since English words were used as targets, the spellings presented as targets might have been consistent with representation in their memory. Therefore, it is arguable that the accuracy rate and RTs of a monitoring task are likely to be determined by how many related items a word has for retrieval and how easy it is to access the word rather than to be determined by mora-based segmentation which is unique to native Japanese speakers. The Japanese groups in this study, as well as the native English speakers, segmented L2 words following PWC in English, although PWC in Japanese is different. Especially, the Japanese group with a low level of English proficiency was more influenced by the PWC with the aid of visual letters. Since they did not have sufficient English phonological knowledge, they were likely to have depended on the spelling to detect the target words while ignoring the word boundaries. However, when these visual cues were not presented, JL paid more attention to the syllable structures of the non-words. From the findings in these two experiments, it is presumed that L2 learners have the potential ability to segment words on the basis of syllables, which can be accessible generally, but further research is required to support this presumption. Although it was reported that native English speakers employ a stress-based segmentation strategy and that native Japanese speakers use a mora-based segmentation strategy, the results of this study revealed that English speakers and Japanese speakers syllabified target words to identify them while being affected by the PWC (Norris, McQueen, Cutler, & Butterfield, 1997), which means that activation of lexical access is reduced when words are misaligned with cued lexical boundaries. Accordingly, word segmentation is facilitated by input consisting of units that are possible words. If words violate the PWC, a path that words lie should be penalized even if the path possesses overlapping word candidates. For instance, detecting *apple* in *fapple* is harder than in *vuffapple* because the residue [f] alone is an impossible word, while *vuff* could be an English word. This results in reduction of the path probabilities.

The results in the present study support the proposal of Mattys and Samuel (2000) that listeners use retroactive processing as well as proactive processing. Phonological knowledge is required to interpret phonetic information given by the target words. When a listener is given a signal produced by a phoneme, what counts is how likely it is for a particular acoustic-phonetic signal to occur rather than how similar the signal and the phoneme are (Norris & McQueen, 2008). Since the prior probabilities of the words will also be altered by the context in which the word appears, one phoneme may be misperceived more often than other phonemes consistent with the claim of Norris and McQueen that lexical decisions are not driven directly by the likelihood of spoken words. As shown by the results of Experiment 1 and Experiment 2, RTs and miss rates varied according to whether the target words were presented visually or not. This difference is plausibly derived from their modulation of the probabilities of the target words being true. That is, it is assumed that JH considered N of the CVN/CVNC targets as [n] in English, while JL considered it as [N] in Japanese Roman letters. Norris et al. claim that phonological knowledge continues to be updated over the listener's lifetime by accumulated experiences with speech sounds, and adults are likely to modulate their perception of phonetic categories. This proposed mechanism can explain the results showing that the Japanese speakers exhibited different trends of word recognition depending on their levels of English proficiency. Since advanced English learners are more likely to have received more input of the target language, the probabilities of activation of the path to English phonemes are also different between native Japanese speakers with a high level of English proficiency and those with a low level. The difference in ways of activating the path is more likely than phonotactic constraints in their first language affect their recognition of English words. Although it was reported that word segmentation strategies are language-specific and determined by a listener's first language, it would be more correct to state that they are dependent on their experiences of the language which enable listeners to calculate the probability of word/phoneme occurrence.

5. Conclusion

When participants were given foreign words with different phonotactics from their L1, they used a word segmentation strategy that also differs from their L1 word segmentation strategy. Regardless of their L2 proficiency, the participants segmented words following the PWC so that the residual syllable would be legal in L2. Although auditory word recognition involves word frequency and acoustic-phonetic signal, a question remains as to whether listeners have the potential capacity to recognize the unit of syllables as a basic property of the auditory system. Further research is required to reveal what is general and what is language-specific with respect to word segmentation.

With respect to limitations of this study, *tong* was categorized as a CVN word to distinguish a CVN word, *ton* in regard to the spelling; however, this caused missed responses by ES and JH as discussed in the previous section. The coda of the target words included two types of phonemes in English, /n/ and /ŋ/, which added another variable for ES and JH, though JL were not affected by this variable. The results indicating that nasal types affected their responses suggests that JH processed the target words using the L2 phoneme /n/, while JL processed the target words using their L1 phoneme inventory since they recognized both /n/ and /ŋ/ as /N/. It can be argued, therefore, that JL were more influenced by the orthography of the Japanese Roman letters. Further study is needed to investigate whether learners process L2 words on the basis of L1 phonology or L2 phonology. If they use L2 phonology, the amount of experience required with the L2 also needs to be investigated.

References

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*(3), 207-228.

Carlson, M. T. (2018). Making room for second language phonotactics: Effects of L2 learning and environment on first language speech perception. *Language and Speech, 61*(4), 598-614. doi: 10.1177/0023830918767208.

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

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

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385-400.

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

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

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptural illusion? *Journal of Experimental Psychology; Human Perception and Performance 25*, 1568-1578.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language 64*, 199-210.

Ernestus, M., Kouwenhoven, H., & Mulken, M. (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics, 62*, 50-64.

Katayama, T. (2022). Effects of L1 phonotactic constraints on word segmentation with clear and ambiguous English syllable boundaries. *Journal of the Phonetic Society of Japan, 26*, 1-9.

Kubozono, H. (2001). *Oninkouzou to akusento* [phonological structure and accent]. Tokyo: Kenkyusha Shuppan.

Mattys, S. L., & Samuel, A. G. (2000). Implications of stress-pattern differences in spoken-word recognition. *Journal of Memory and Language*, *42*, 571-596. doi: 10.1006/jmla.1999.2696.

McQueen, J. M. (1988). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21-46.

McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language, 45*, 103-132.

Mehler, J., Dommergues, J. Y., Faruenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298-305.

Mizumoto, A., & Takeuchi, O. (2008). Kenkyuronbun niokeru kokaryo no hokoku notameni: kihontekigainen to chuiten [Basics and considerations for reporting effect sizes in research papers]. *Studies in English Language Teaching*, 31, 57-66.

Norris, D. & McQueen, J. M. (2008). Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review, 115*(2), 357-395. doi: 10.1037/0033-295X.115.2.357.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*(5), 1209-1228.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*(3), 191-243.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*(2), 258-278.

Otake, T., Hatano, G., Cutler, A., & Yoneyama, K. (1996). Speech segmentation by Japanese listeners. In T. Otake & A. Cutler (Eds.) *Phonological Structure and Language Processing: Cross-Linguistic Studies*, 183-201, Mouton de Gruyter: Berlin.

Port, R., Dalby, P., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America, 81*(5), 1574-1585.

Sebastián-Gallés, N., & Díaz, B. (2012). First and second language speech perception: Graded learning. *Language Learning, 62*(2), 131-147.

Weber, A. & Cutler, A. (2006). First-language phonotactics in second-language listening. *Acoustical Society of America, 119*(1), 597-607. doi: 10.1121/1.2141003.

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language, 66*, 665-679.doi: 10.1016/j.jml.2011.12.010