INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Effects of L1 Phonotactic Constraints on L2 Word Segmentation Strategies

Tamami Katayama

Prefectural University of Hiroshima, Japan
katayamat@pu-hiroshima.ac.jp

Abstract

In the present study, it was examined whether phonotactic constraints of the first language affect speech processing by Japanese learners of English and whether L2 proficiency influences it. Seventeen native English speakers (ES), 18 Japanese speakers with high proficiency of English (JH), and 20 Japanese speakers with relatively low English proficiency (JL) took part in a monitoring task. Two types of target words (CVC/CV, e.g., *team*/*tea*) were embedded in bisyllabic non-words (e.g., *teamfesh*) and given to the participants with other non-words in the lists. The three groups were instructed to respond as soon as they spot targets, and response times and error rates were analyzed. The results showed that all of the groups segmented the CVC target words significantly faster and more accurately than the CV targets. L1 phonontactic constraints did not hinder L2 speech processing, and a word segmentation strategy was not language-specific in the case of Japanese English learners.

Index Terms: word segmentation, phonotactics, L2 speech perception

1. Introduction

Segmenting words in the stream of speech is a vital skill of communication. One of the issues in the study of speech processing is how we segment words and access the mental lexicon. Segmentation strategies for words depend on listeners' exposure to a particular language [1]. These language-specific strategies include suprasegmental information and a segmental inventory [2]. For example, native English speakers detect words on the basis of stress, native French speakers depend on syllables, and native Japanese speakers employ mora for word segmentation [3, 4]. Thus, the unit of speech processing is language-specific. However, when we try to master a second language, we need to detect words in speech with unfamiliar sound systems. Perception of L2 phonemes develops as the L2 proficiency improves [5, 6, 7, 8], but previous studies have mainly focused on segmental level, although a unit of speech is the syllable level. Thus, it is not still clear whether perception of L2 syllable structure develops and how it affects L2 communication.

1.1. Word segmentation

Segmentation strategies in continuous speech perception might well be language-specific [9]. [10] examined whether native English speakers and native French speakers perceive the phoneme or the syllable itself directly to segment words. Seven pairs of unambiguous English content words (nouns and verbs) of similar frequencies sharing the same initial three phonemes (a consonant-vowel-consonant (CVC) sequence) were used in that study. Each pair had one member with a syllable boundary after the initial CVC. In the other members, the third phoneme was ambisyllabic: that is, the second consonant of the initial CVC could be considered to belong to both syllables (e.g., balcony/balance). Native French speakers appear to syllabify regardless of whether they are listening to familiar French words which are easy to syllabify, or unfamiliar English words, which are hard to syllabify [9]. However, English-speaking subjects showed no sign of the syllabification effect found with French subjects. It was thus concluded that the syllable's function in speech segmentation differs depending on the speaker's language [9].

In a study using a model of speech segmentation in a stress language, it was found that the occurrence of a strong syllable triggers segmentation of the speech signal, while the occurrence of a weak syllable does not [11]. In that study, participants were instructed to detect real words in nonsense strings. For example, *mint* embedded either in *mintayve* or *mintesh* was presented to the listeners. Because the second syllable (*tayve*) is strong in *mintayve*, the string will be segmented, and lexical access will start at *tayve*. When *mint* belongs partly to both accompanying syllables, this inappropriate intersyllabic segmentation interferes with the detection of the word *mint.* The results of that study showed that the detection of a word is delayed when the word consists of two strong syllables but is not delayed when it consists of a strong syllable followed by a weak syllable. The results suggested that the strong syllable triggered a segmentation effect. Thus, the detection of the word was delayed in intersyllabic segmentation because the listeners needed to collect speech information across a segmentation point. The results suggest that a syllable forms a unit of speech processing. In another study [13], errors of spontaneous misperception were used in order to test the rhythmic segmentation hypothesis [11, 12], and it was found that strong syllables tend to be the initial syllables of lexical words, while weak syllables are likely not to be word-initial syllables. It was concluded from the results of that study that listeners use rhythmic segmentation when perception is difficult.

Phonotactic constraints can be one of the sources of information, such as silence and metrical cues, when listeners segment words [14]. In that study, Dutch speakers failed more frequently to detect words that were misaligned with syllable boundaries cued by phonotactic constraints than words that were aligned with such boundaries. The results of that study suggested that phonotactic legality is taken into consideration and helps listeners to segment words. In another study [15], perception of the segmentation of spoken Japanese words by native and nonnative listeners was examined. The results showed that the response patterns to CV targets were identical in CVCVCV and CVNCV (a consonant-vowel-nasal-consonant-vowel sequence) words, being inconsistent with the prediction of syllable-based segmentation but being consistent with the mora-based segmentation. This is because the initial

http://dx.doi.org/10.21437/Interspeech.2016-182

mora is CV in both CVCVCV and CVNCV words. The mora hypothesis can explain why the response pattern to CVN targets differs across CVNCV words and CVCVCV words. The Japanese listeners responded to CVN targets in CVNCV, but the response time was comparatively long. According to the authors, this is due to a complex target: two-mora words. However, CVN targets in CVCVCV words received a mixed response because the target and stimulus do not match at the mora level. The authors concluded that Japanese listeners decompose words not by syllables but by mora. Thus, mora-based segmentation is a language-specific effect, as are French listeners' syllabic segmentation and English listeners' stress-based segmentation. In addition, it was reported that Japanese listeners map their moraic pattern of speech processing onto whatever foreign languages they are learning [16], and the authors of that report suggested that their findings about language-specific processing have critical implications for clarifying the processes of second language acquisition.

Consequently, it was examined in another study how French-English bilinguals segment language strings in both English and French [17]. Once the participants had been judged as having native-speaker competence in both languages by native speakers of French and by native speakers of English, they were asked about their language preference in order to divide them into English-dominant bilinguals and French-dominant bilinguals. The results showed that English-dominant bilinguals produced exactly the same pattern as that produced by monolingual English speakers not only when listening to English but also when listening to French. On the other hand, French-dominant participants produced results exactly the same as those of French speakers when they listened to French and exactly the same as those of English monolingual speakers when they listened to English. Based on the results, it was proposed that syllabic segmentation is a restricted procedure and can be "switched off" when it is inefficient, whereas stress-based segmentation (i.e., English in this case) is an unrestricted (general) procedure and is generally available to all speakers. Thus, while French-dominant bilinguals employed stress-based segmentation when they were presented with input in English, the reverse process, according to the authors [17], is not possible, since English does not encourage the development of syllabic segmentation. The authors explained that English-dominant bilinguals operate only those processing procedures for segmentation that are generally available. According to the authors, only those speakers who received the right input at the right time are able to develop the restricted processing procedure. Consequently, even the most skilled bilingual speakers will, according to this view, be restricted to the one segmentation procedure. The authors therefore concluded that rhythmically based segmentation procedures are mutually exclusive as well as language-specific and that they are restricted in their availability.

1.2. Previous studies

It was examined in a previous study whether phonotactic constraints of the first language affect speech processing by Japanese learners of English and whether L2 proficiency influences it [18]. Native English speakers (ES) and L2 speakers with a high level of language proficiency (JH) and those with a low level (JL) took part in a monitoring task. They were given two kinds of sound stimuli as target syllables (i.e., CV and CVC) and were asked to detect them in lists of words that have stress on the first or second syllable (e.g. *bi* and *bis* in *biscuit* and *beside*). The results showed that both stress and phonotactics facilitated segmentation strategies by the three groups. The Japanese groups did not rely on either phonotactics or mora to segment the target syllables. They rather used stress to detect the target syllables in the English words, which is a different segmentation strategy from that of their L1. It was shown in another study that native Japanese speakers segmented speech based on mora but that their methodology employed Japanese words as material [16]. Syllables without meanings were used in that study and it was shown that phonotactic constraints did not interfere with L2 processing by native Japanese speakers. Although that study provided evidence that L2 speakers used the segmentation strategy that was used by native speakers of the target language, the methodology had some limitations. Since the target syllables were recorded separately, the target syllables that were embedded in either the stressed syllables or the unstressed syllables in the target words were not acoustically identical. Thus, in the present study, identical sound stimuli of the target words were used instead of syllable stimuli recorded separately and it was investigated whether native Japanese speakers use mora for the word segmentation strategy.

The aim of this study was to determine the effects of phonotactic constraints of the first language on second language acquisition. The following research question is raised: How do L2 speakers of different levels in English proficiency identify target English words with a legal syllable structure (CV) and an illegal syllable alignment (CVC)? It was reported that native Japanese speakers used mora to segment words in speech [15], and CVC is an illegal segmental sequence in Japanese except for the case in which the coda is nasal. Therefore, I predicted that the Japanese groups would segment CV syllables faster than CVC syllables. Thus, the following hypothesis was raised: Since most of the Japanese moras consist of CV units, Japanese learners of English will recognize CV words faster and more accurately than CVC words. Based on this hypothesis, native English speakers and L2 speakers with a high level of language proficiency and those with a low level took part in a monitoring task.

2. Experiment

In this study, a monitoring task was performed in a group of native English speakers and two groups of native Japanese speakers with different levels of English proficiency.

2.1. Materials

Ten target bisyllabic non-words with two types of target words (CVC/CV) were created (see Table 1). The target words in the non-words have the possibility of being monosyllabic words that consist of either a CV or CVC syllable structure. Thus, stimuli under the following two conditions were created.

1) The target words have stressed target syllables whose structure is the same as that of the first syllables: a target non-word is teamfesh and its target word is team (CVC).
2) The target words have stressed target syllables whose structure is partly the same as that of the first syllables: a target non-word is teamfesh and its target word is tea (CV).

In addition to the target non-words, 220 distractors that also include monosyllabic words were created. All stimuli were recorded into a computer with sampling at 48,000 Hz using a high quality microphone in a recording studio. The target non-words and the distracters were embedded in a carry sentence, "please say...," and spoken by a female speaker of American English. Then the target non-words and the distractors were taken out of the recordings using *praat*. Both CV and CVC target words (e.g., *bay* and *base*) were created from the original target non-words (e.g., *basetove*) in order to match acoustic factors.

Table 1. *Target syllables and target words for each version*

List no.	non-words with target word	target word (ver.1)	target word (ver.2)
1	basetove	bay	base
2	peaksom	peak	pea
3	cuteklef	cue	cute
4	laceklus	lace	lay
5	tightroz	tie	tight
6	dukeramp	duke	due
7	looptrep	loo	loop
8	teamfesh	team	tea
9	needfem	knee	need
10	dimenen	dime	dye

Two versions of a monitoring task were created with the aid of E-prime software, in which different types of segmental sequences (i.e., CV and CVC) were presented as target syllables. For example, when a target non-word was *basetove*, *bay* was presented in version I and *base* was presented in version II. The proportions of different types of target words and (i.e., CV and CVC) were counterbalanced in each version (see Table 1). Additionally, the instructions for each version were in English for the English speakers and in Japanese for the Japanese speakers. Each version had 20 lists including 10 positive lists with the target words and 10 negative lists without them, and each list contained eight to twelve non-words. In each positive list, one target word appeared and the rest of the spaces were occupied by distractors, and negative lists consisted of distractors.

A target word was presented auditorily and visually followed by a blank screen for 750 ms. A written target word (e.g., *bay*) was set up to appear in the center of the screen on the computer for 750 ms while the target sound was being provided. Then sound files of the non-words in the list were programmed to run while the mark "+" appeared in the center of the screen. In the list, each word was presented followed by 1000 ms blank. If a response was made while the words in a particular list were being presented, this list was programmed to jump to the next list. If there was no response by the end of the list, the screen informed the participant of the end of list and instructed the participant to press the button to continue.

2.2. Participants

Seventeen native English speakers (ES), 18 Japanese speakers with high proficiency of English (JH), and 20 Japanese speakers with relatively low English proficiency (JL) took part in the experiment. ES consisted of graduates and undergraduates of University of Edinburgh. Fifteen of them were from U.K. and the rest were from the U.S.A., and their mean age was 22.6 years. JH mainly consisted of English instructors at colleges and their mean age was 38.3 years. Their mean duration of living experience in English speaking countries was 3.6 years. Their mean score for TOEIC was 931.8 and that for TOEFL ibt was 108. JL were graduates and undergraduates from Prefectural University of Hiroshima. Their mean age was 19.7 years. Eighteen of them had no living experience abroad (Two of them had visited Canada for two weeks and one month, respectively.), and their mean score for TOEIC was 359.7. A book voucher was given to each participant. All of the participants reported no hearing impairment.

2.3. Procedure

The participants were instructed to sit in front of the computer in a quiet room and put on the headphones. Instructions were presented on the screen, and they were asked to respond as quickly as possible by pressing button "1" on the response box when they heard the target words. The response time from the onset of each target word was recorded.

3. Results

Response times of the three groups were calculated by subtracting duration of each target word from the recorded response time (see 2.3.). Figures 1 and 2 show the mean response times and percentage of errors in the monitoring task, respectively. Two-way repeated measures ANOVA was performed for both the response times (RTs) and error rates separately across subjects (*F*1) and across items (*F*2).

In the RT three was a significant main effect of phonotactis ($F1\ (1, 52) = 76.2$, $p < .001$; $F2\ (1, 27) = 20.2$, $p < .001$). The interaction between groups and phonotactics was significant in the subjects analysis ($F1\ (2, 52) = 3.5$, $p < .05$) but not in the items analysis ($F2\ (2, 27) = .7$, $p > .1$). There was no effect of group ($F1\ (1, 52) = .22$, $p > .1$; $F2\ (2, 27) = .38$, $p > .01$). As further analyses, t-tests were conducted between RTs of CV and CVC for the three groups. All the groups showed significant differences both in the subjects analysis and in the items analysis (ES: $t1\ (16) = 6.2$, $p < .001$; $t2\ (9) = 2.9$, $p < .02$, JH: $t1\ (17) = 3.9$, $p = .001$; $t2\ (9) = 2.2$, $p = .05$, JL: $t1\ (19) = 4.8$, $p < .001$; $t2\ (9) = 3.0$, $p < .05$).

In the error rates both the subjects analysis and items analysis showed a significant main effect ($F1\ (1, 52) = 18.2$, $p < .001$; $F2\ (1, 27) = 18.4$, $p < .001$). The interaction between groups and phonotactics was significant in the subjects analysis ($F1\ (2, 52) = 3.5$, $p < .05$) but not in the items analysis ($F2\ (2, 27) = 3.1$, $p = .06$). There was no effect of group ($F1\ (1, 52) = 2.5$, $p = .1$; $F2\ (2, 27) = 2.4$, $p > .1$). As further analyses, t-tests were conducted between RTs of CV targets and CVC targets for the three groups. The results of ES and JL showed significant differences both in the subjects analysis and in the items analysis (ES: $t1\ (16) = 3.4$, $p = .004$; $t2\ (9) = 3.4$, $p = .008$, JL: $t1\ (19) = 2.5$, $p < .05$; $t2\ (9) = 3.0$, $p < .05$), but there was no significant differences in the results of JH: $t1\ (17) = 1.0$, $p > .1$; $t2\ (9) = 1.0$, $p > .1$,).

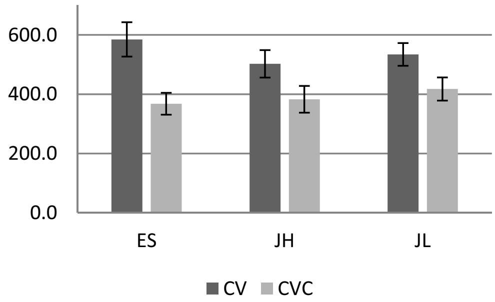

Figure 1: *Mean RTs of the monitoring task*

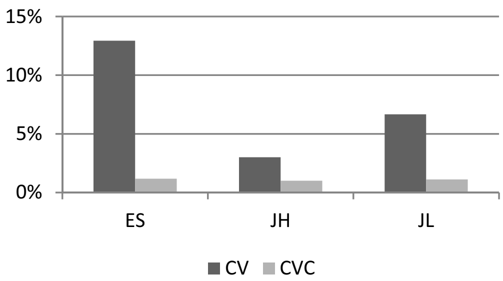

Figure 2: *Percent errors of the monitoring task*

All of the groups segmented the CVC target words significantly faster and more accurately than the CV targets, though the difference in accuracy was not statistically significant in JH. ES took longer to detect and made more errors in detecting the CV target words (mean RT = 584.9, mean error rate = 0.65) than did the Japanese groups (JH: mean RT = 502.5, mean error rate = 0.30; JL: mean RT = 534.1, mean error rate = 0.17). JL showed the same tendency as ES did with respect to high rate of errors in the target word CV. It is notable that JH identified target words accurately regardless of the type of syllable structure.

4. Discussion

The findings of this study do not support the notion that native Japanese speakers segment words on the basis of mora even in a foreign language. The following hypothesis was raised: since most of the Japanese moras consist of CV units, Japanese learners of English will recognize CV words faster and more accurately than CVC words. However, the Japanese groups responded faster and more accurately to the CVC targets than to the CV targets as did the native English speakers. Thus, the hypothesis was rejected.

L2 proficiency did not affect the segmentation strategies of the Japanese groups in this study, but the Japanese groups used a strategy different form that for L1. Both JL and JH used a unit of syllables to detect L2 words. The only difference between JH and JL was the error rates. Contrary to expectations, JL made fewer errors in identifying CVC target words than CV targets, while JH showed high accuracy to segment both types of targets. This means that JH used not only syllable units but also other acoustic information and processed the speech at a phoneme level as well as at a syllable level. As for JL, the possible reason is that they divided the target non-words with two syllables into two units based on the duration. If we consider that JL used a timing unit to divide the words into two parts, it makes sense they responded to CVC targets faster and more accurately than CV targets. Further study is needed to confirm this prediction. Unlike the previous study [15], L2 material was used in this study and the target words were presented visually and auditorily. It was obvious that the Japanese groups heard the targets as English words, and thus these explicit targets might have led the participants to segment words in a way different from that for L1.

The results for ES are consistent with the results of a previous study showing that native English speakers found it harder to detect CV targets with syllable residues, while CVC targets made detection harder with consonant residues [19]. According to that study, detection of a syllable required the entire target to be processed successfully. Thus, the final phoneme affected the response times to detect the words. Given the CVC targets, ES would delete other CVC word candidates with different codas from the targets and would have fewer candidates of the target words than CV target words, which would make it easier for them to access the targets. On the other hand, ES had to wait and find what would come after the CV syllables to process the syllable and identify CV target words. In addition, since the CV target stimuli were technically taken from non-words with CVC words, it does not necessarily match their mental lexicon of the target CV words. These are possible reasons why the number of errors for CV targets was significantly higher than for CVC targets.

In a previous study, syllables without meanings (e.g., *bi* or *bis*) were used instead of words and it was found that native Japanese speakers used stress to spot the target syllables regardless of their level of English proficiency [18]. In the present study, though the factor of stress was not examined, the Japanese groups did not use mora to segment English words. From these results, it is assumed that native Japanese speakers employ stress and syllable to detect L2 syllables and words regardless of their L2 proficiency. Interestingly, JL showed the same tendency with respect to errors they made. In this sense, development of L2 speech segmentation differed from L2 speech perception, which gradually develops as learners gain L2 experience. Since JH showed a high level of accuracy in detecting words regardless of syllable alignment, further research is required to examine how advanced learners of English process L2 speech.

5. Conclusions

In the present study, it was examined whether phonotactic constraints of the first language affect speech processing by Japanese learners of English and whether L2 proficiency influences it. Although previous studies have suggested that a segmentation strategy is language-specific, the results of this study showed that L2 learners segmented words in a different way from their L1 strategy. Thus, L1 phonontactic constraints did not hinder L2 speech processing and a word segmentation strategy was not language-specific in the case of Japanese English learners. Still, L2 proficiency appeared to affect the way of processing because different tendencies were found depending on L2 proficiency. We need to investigate which acoustic cues L2 speakers of low and advanced levels employ to unveil the mechanisms of L2 speech processing.

6. Acknowledgements

This study was supported by Grant-in-Aid for Scientific Research (Grant Number 15K02757).

7. References

[1] N. Cooper, A. Cutler and R. Wales, "Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners," *Language and Speech*, vol. 45, no. 3, pp. 207-228, 2002.

[2] E. Dupoux, C. Pallier, N. Sebastian, and J. Mehler, "A destressing 'deafness' in French?" *Journal of Memory and Language*, vol. 36, pp. 406-421, 1997.

[3] A. Cutler, "Forbear is a homophone: lexical prosody does not constrain lexical access," *Language and Speech,* vol. 29, no. 3, pp. 201-220, 1986.

[4] A. Cutler, J. Mehler, D. Norris, and J. Segui, "The syllable's differing role in the segmentation of French and English," *Journal of Memory and Language,* vol. 25, pp. 385-400, 1986.

[5] B. Rochet, Perception and production of second-language speech sounds by adults-Speech perception and linguistic experience. Baltimore: York Press, 1995.

[6] J. E. Flege, O. Bohn, and S. Jang, "Effects of experience on non-native speakres' production and perception of English vowels," *Journal of Phonetics,* vol. 25, pp. 437-470, 1997.

[7] J. Leather and A. James, "The acquisition of second language speech," *Studies in Second Language Acquisition,* vol. 13, pp. 305-341, 1991.

[8] J E. Flege and A. M. Schmidt, "Native speakers of Spanish show rate-dependent processing of English stop consonants," *Phonetica,* vol. 52, pp. 90-111, 1995.

[9] J. Mehler, J. Y. Dommergues, and U. Faruenfelder, "The syllable's role in speech segmentation," *Journal of Verbal Learning and Verbal Behavior,* vol. 20, pp. 298-305, 1981.

[10] A. Cutler, J. Mehler, D. Norris, and J. Segui, "The syllable's differing role in the segmentation of French and English," *Journal of Memory and Language,* vol. 25, pp. 385-400, 1986.

[11] A. Cutler and D. Norris, "The role of strong syllables in segmentation for lexical access," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 14, no. 1, pp. 113-121, 1988.

[12] A. Cutler and D. M. Carter, "The predominance of strong initial syllables in the English vocabulary," *Computer Speech and Language,* vol. 2*,* pp. 133-142, 1987.

[13] A. Cutler and S. Butterfield, "Rhythmic cues to speech segmentation: Evidence from juncture misperception," *Journal of Memory and Language*, vol. 31, pp. 218-236, 1992.

[14] J. M. McQueen, "Segmentation of continuous speech using phonotactics," *Journal of Memory and Language*, vol. 39, pp. 21-46, 1988.

[15] T. Otake, G. Hatano, A. Cutler, and J. Mehler, "Mora or syllable? Speech segmentation in Japanese," *Journal of Memory and Language,* vol. 32, no. 2, pp. 258-278, 1993.

[16] A. Cutler, A. and T. Otake, "Mora or phoneme? Further evidence for language-specific listening," *Journal of Memory and Language,* vol. 33, pp. 824-844, 1994.

[17] A. Cutler, J. Mehler, D. Norris, and J. Segui, "The monolingual nature of speech segmentation by bilinguals," *Cognitive Psychology,* vol. 24, pp. 381-410, 1992.

[18] T. Katayama, "Effect of Phonotactic Constraints on L2 Speech Processing," *i-Perception*, vol. 6, no. 6, pp. 1-13, 2015. doi: 10.1177/2041669515615714.

[19] R. Kerns, D. Norris, and A. Cutler, "Syllable processing in English," in *ICLSP 2002-7th International Conference on Spoken Language Processing*, Colorado, USA, Proceedings, 2002, pp. 1657-1660