【論文】

Effects of Closure Duration of Obstruents and Intensity on Identification of Syllables by L2 Speakers of English

Katayama, Tamami

要旨（Abstract）

The purpose of this study was to examine the extent to which closure duration of obstruents (CDO) followed by a vowel (C_1) and intensity of words affect the perception of a vowel in legal ($V_1C_1UC_2V_2$) and illegal ($V_1C_1C_2V_2$) phonotactic contexts by native Japanese speakers. Four sets of non-words that include legal and illegal phonotactic contexts (e.g., /ebzo/ and /ebuzo/) were manipulated with respect to the closure duration of C_1 in four steps (25%, 50%, 75% and 100%) and intensity of the whole word in four steps (+3 dB and +6 dB, -3 dB and -6 dB), and a total of 200 sound stimuli were thus created for a forced-choice task. Twenty native Japanese speakers identified whether a given sound stimulus had a vowel or not. The results showed that the Japanese participants tended not to perceive a vowel when CDO was completely deleted and when intensity was low, and the effects on perception by the native Japanese speakers were even greater when the two factors were combined.

キーワード（Keywords）：intensity, syllable, closure duration of obstruents, vowel perception, second language acquisition

1. Introduction

Linguists who have examined the phonological aspects of speech production have agreed on the importance of the mora as a unit of Japanese language (Kubozono, 2002). Warner and Arai (2001) offered several hypotheses to account for timing of the rhythmic beat by morae (rather than by stress); one of them is that morae are utilized to normalize word duration. Port, Dalby, and O'Dell (1987) defined timing by mora not as an overall isochronic tendency but as affording predictability of word duration deduced from the number of morae in each particular word. They found a strong positive correlation between the number of morae in a word and the word's total duration, and they concluded that speakers keep word duration constant over a given number of morae despite the inherent differences of segmental duration. Port et al. reported that speakers adjusted both the initial stop in the word and the following vowel in order to keep word duration constant, and they therefore considered the mora as an abstract isochronous unit of timing in Japanese that captures many of the most salient features of timing in this language. Although evidence of the mora hypothesis has been reported with respect to production, whether the minimum unit of speech perception by native Japanese speakers is in fact the mora has been disputed (Cutler & Otake, 1994; Katayama, 2015; Otake, Hatano, Cutler, & Mehler, 1993). The purpose of this study is to investigate the minimum unit of speech perception by native Japanese speakers and how listeners recognize syllables with legal and illegal structures.

Samuel (2020) attempted to demonstrate that linguistic units of language should be abandoned to identify perceptual units. He examined the effects of adaptations of /b/ and /d/ by using words with released final stop consonants (i.e., /b/ or /d/) and words with unreleased final stop consonants. In the first condition of an adaptation

task, 60 native speakers of American English listened to 25 words with final-/b/ and 25 final-/d/ words (e.g., *club* and *cord*) recorded by a native speaker of British English. In the second condition, the original words used in the condition were cut just before any release. They tested whether the original sound stimuli led the listeners to shift their cross-position adaptation on an eight /ba/-/da/ continuum. They also tested whether the listeners would show the same shift as that in the first condition when given sound stimuli without any acoustic cues of release. The results showed that there was a significant effect of release of final stops on adaptation of the initial-position stop consonants, though this effect was not significant when the adaptors did not have released stops in the final positions. Samuel argued that the results did not support the linguistic theory that adaptation is driven by an abstract phoneme that is invariant regardless of its position in a word and that the pre-lexical unit of speech perception is the allophone. Although Samuel's study showed that the participants were unable to identify phonemes that were not in the identical positions of words, this does not mean that the unit of speech processing is the allophone.

Syllable structures vary according to the language and they affect a listener's perception of a syllable. Dupoux, Kakehi, Hirose, Pallier & Mehler (1999) reported that native Japanese speakers perceived illusory vowels when they heard an illegal syllable structure. They examined the effect of phonotactics on perception of vowels inside consonant clusters in vowel-consonant-consonant-vowel (VCCV) stimuli. Ten Japanese and 10 French native speakers listened to five different stimuli that were created from the original sound file of /ebuzo/ by splicing out the vowel duration in steps. The participants judged whether a vowel [u] was present or not in the stimuli. The results showed that the French speakers were able to judge the presence of the vowel in the /ebuzo/ stimuli and the absence in the /ebzo/ stimuli, whereas the Japanese speakers predominantly recognized vowels in all levels of stimuli. In a subsequent experiment, they investigated whether coarticulation information affected the results by using two additional kinds of stimuli: one was a naturally produced /ebzo/ and the other was one with a different vowel /ebizo/. In their subsequent experiment, 16 triplets were employed for an ABX discrimination task, which conformed to the model $V_1C_1C_2V_2$ - $V_1C_1UC_2V_2$ - $V_1C_1UUC_2V_2$ (e.g., /ebzo/-/ebuzo/-/ebuuzo/). Ten Japanese and 10 French people living in Paris took part in the task with two contrasts: an epenthesis contrast (/ebzo/-/ebuzo/) and a vowel length contrast (/ebuzo/-/ebuuzo/). The results showed that the Japanese group made more errors in identifying the former, whereas the French group had more difficulty with the latter, supporting Dupoux et al.'s claim that phonotactics of the listeners' first language affects speech perception and that proficiency of the foreign language does not influence the effect of epenthesis, a claim based on results showing that the Japanese group with high proficiency of English or French showed the same pattern as that in the low-proficiency group. Dupoux et al. concluded that L2 sounds are not only assimilated into the inventories of listeners' first language but are also distorted to conform to their L1 phonotactic regulations.

Dupoux, Pallier, Kakehi, and Mehler (2001) examined the role of phonotactics and lexical knowledge in prelexical processing. Japanese speakers perceive an illusory vowel when given illegal consonant clusters in their language. For example, they hear a /u/ vowel between /b/ and /z/. Dupoux et al. tested the hypothesis that this vowel epenthesis arises from "top-down" lexical knowledge. The stimuli used in their study contained two sets depending on the vowel that produces a Japanese word: in the u-Set, the vowel that yielded a word was /u/ (e.g., *sokdo→sokudo*), and in the non-u-Set, the vowel was /a/, /e/, /i/ or /o/ (e.g., *mikdo→mikado*). They performed a

transcription task in which 14 native Japanese speakers were asked to transcribe stimuli into the Roman alphabet and a lexical decision task in which the participants were asked to decide whether the given stimuli were Japanese words or not. The results showed that Japanese listeners heard the vowel /u/ between the consonants in both sets (i.e., *mikdo→mikudo*), and that the lexical knowledge of the listeners has no effect on pre-lexical processing. Thus, they concluded that it is not the lexical context but the phonological context that causes the illusory perception of /u/ and that it is likely that phonotactic constraints are modulated prior to lexical access.

In order to investigate how and when L1 phonotactics affects speech perception, Dehaene-Lambertz, Dupoux and Gout (2000) performed a mismatch detection task for native Japanese speakers and native French speakers in which four similar precursor stimuli were presented followed by a fifth stimulus that was similar to or different from the previous stimuli. The stimuli consisted of minimal pairs of nonwords and the only difference between them was whether they had the vowel /u/ between two consonants (e.g., *igumo vs. igmo*). They tested six sets of stimuli (i.e, *igmo, igna, ikino, ikma, okna*, and *ogma*) under three conditions: $V_1C_1C_2V_2$ (*igmo*), $V_1C_1UC_2V_2$ (*igumo*), and $V_1C_1IC_2V_2$ (*igimo*). The results showed that the Japanese speakers were not able to discriminate items such as *igmo* and items such as *igumo*. Dehaene-Lambertz et al. (2000) argued that the native language strongly affected the performance and that L1 phonotactics modified speech perception. Although the error rate of the Japanese group was high compared to that of the French group, the Japanese speakers were still able to distinguish a real vowel from an illusory vowel, suggesting that "this residual capacity is slower and that it relies on a different network than the one used with native contrasts" (p. 644). If their results are applied to the concept of category perception suggested by studies on language acquisition (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), the units of speech processing may be larger units that contain several phonemes.

Iverson et al. (2003) showed how language experience affects low levels of processing of non-native phonemes by using three language groups: German, Japanese and American English. Eighteen English /ra/ and /la/ stimuli were manipulated with the fundamental frequencies of the second and third formants during the consonant closure. In their first experiment, the participants rated whether the stimulus was a good exemplar of that category using a scale from 1 to 7, in the second experiment, they rated the acoustic similarity of stimulus pairs on a scale from 1 to 7, and in the last experiment, they discriminated one stimulus from another. The results showed that the most reliable acoustic cue to distinguish /r/ and /l/ for the American listeners and the German listeners was F3, while that for the Japanese group was F2. The Japanese listeners relied on F2 to identify /l/, which prevented them from distinguishing /r/ from /l/. What the native Japanese speakers depended on for categorization of /r/ and /l/ is an irrelevant cue for the English speakers to identify /r/ and /l/. Iverson et al. (2003) claimed that a listener's network structure for speech perception is fostered by his/her early language experience and thus it is language-specific and affects his/her speech perception afterwards. Listeners modify L2 speech input at an early phonetic level for processing, which causes difficulty in identification of L2 phonemes by adult learners. According to Iverson et al., early language experience influences perception of non-native speech at a low level, and realization of L2 phonemes is impeded by these changes.

It has also been reported that low-level phonetic detail, as well as native language, affects production and perception of non-native sequences. Wilson, Davidson, and Martin (2014) examined the relationship between perception and production of L2 consonant clusters by manipulating three facets of non-word stimuli (CCaCV):

the presence or absence of pre-voicing (POV), stop burst duration, and burst amplitude. They found that an initial interval of voicing that preceded the onset of the fricative, stop clusters, or high amplitude voicing caused prosthesis. Longer stop bursts with higher amplitudes are likely to be perceived as burst and vocoid sequences, resulting in epenthesis, whereas weak bursts lead to deletion and other modifications of initial consonant clusters since less information is provided about the presence and features of stops.

Katayama (2020) examined whether epenthesis of a vowel by L2 learners is caused by L1 phonotactic constraints or the subsegmental information conveyed by the syllable. She also investigated whether there is a difference between groups of Japanese speakers with different levels of English proficiency in identification of L2 words that possess an illegal syllable structure in Japanese. First, she used stimuli with a $V_1C_1C_2V_2$ structure (/ebzo/, /ebdo/, /ebgu/, /egzo/, /egdo/, /egbu/, /ezgu/, /ezbu/, /ezdo/, /edbu/, /edgu/, /edzo/) and stimuli with a $V_1C_1UC_2V_2$ structure (/ebuzo/, /ebudo/, /ebugu/, /eguzo/, /egudo/, /egubu/, /ezugu/, /ezubu/, /ezudo/, /edubu/, /edugu/, /eduzo/), and 18 native English speakers (ES), 18 Japanese speakers with high proficiency of English (JH) and 18 Japanese speakers with relatively low English proficiency (JL) identified whether the stimuli included a vowel in the middle (e.g., /egzo/ or /eguzo/). The results showed significant differences among the three groups in identification of the stimuli. The accuracy level of ES was almost 100%, that of JH was the second highest and that of JL was the lowest. Notably, the Japanese speakers were not able to identify whether the /u/ vowel exists or not in a certain illegal phonotactic context. Epenthetic vowels listeners perceive are different depending on their L1 (Dupoux et al., 1999), but the perception of L2 consonant clusters is likely to develop as that of phonemes does. In her subsequent experiment, she examined the effects of pre-obstruent voicing (POV) and amplitude on perception of a vowel both in legal ($V_1C_1UC_2V_2$) and illegal ($V_1C_1C_2V_2$) phonotactic contexts by native Japanese speakers with different levels of English proficiency since Wilson et al. (2014) reported that amplitude and POV are likely to affect the perception of the vowel in a syllable. Four sets of non-words were selected from the stimuli used in the first experiment for which native Japanese speakers showed a low level of accuracy in identifying the vowel (/ebzo/, /edzo/, /egubu/ and /ezgu/). As for the stimuli manipulated with POV, a portion of silence between the first vowel and the following consonant was taken out by steps (25%, 50%, 75% and 100%). To examine the effect of ambiguity, the original files were manipulated by increasing the amplitude of the sound stimuli by 2 steps (i.e., +3 dB and +6 dB) and by decreasing the amplitude by 2 steps (i.e., -3 dB and -6 dB). Sixteen native Japanese speakers with a high level of proficiency in English (JH), 16 native Japanese speakers with a relatively low level of English proficiency (JL) and 16 native English speakers (ES) identified 72 sound stimuli. The results showed that both Japanese groups were not likely to perceive a vowel when POV was deleted, while manipulation of POV or amplitude did not significantly affect perception by ES. The perception of an L2 syllable by native Japanese speakers is likely to be affected by the phonetic information it conveys rather than the phonotactic constraints in their first language, and the acoustic information was not a critical cue for the native English speakers to identify the syllables. POV was the most influential factor for the native Japanese speakers to identify the vowel in Katayama's study (2020), and it was shown that L2 perception develops at the syllable level as the level of L2 proficiency increases. The question remains, however, as to how the combination of amplitude and POV affects L2 perception by native Japanese speakers.

The aim of this study was to determine whether the perceptual unit of native Japanese speakers is identical to its

linguistic unit, mora, by investigating how native Japanese speakers perceive a vowel both in legal ($V_1C_1UC_2V_2$) and illegal ($V_1C_1C_2V_2$) phonotactic contexts when low level phonetic features, including closure duration of obstruents (CDO) and intensity, have been manipulated. Although Katayama (2020) showed that POV is more influential than amplitude in native Japanese speakers' perception of a vowel, closure duration and intensity were used in this study instead of POV and amplitude, respectively. In view of the above rationale for the study, the following research question was raised: Is there a synergistic effect of CDO and intensity of a whole word when native Japanese speakers are given these mixed factors? If a synergistic effect is observed, it is possible that a listener's perception is affected by multiple acoustic factors other than syllable structures. In addition, if the most influential combinations are identified, the priority of acoustic cues the listener relies on will be revealed.

2. Methods

2.1 Materials

$V_1C_1C_2V_2$ and $V_1C_1UC_2V_2$ were used as stimuli in this study. $V_1C_1C_2V_2$ type stimuli included /ebzo/, /ebdo/, /ebgu/, /egzo/, /egdo/, /egbu/, /ezgu/, /ezbu/, /ezdo/, /edbu/, /edgu/, and /edzo/, while $V_1C_1UC_2V_2$ type stimuli had the /u/ vowel inserted between C_1 and C_2 in the same stimuli as those reported above: /ebuzo/, /ebudo/, /ebugu/, /eguzo/, /egudo/, /egubu/, /ezugu/, /ezubu/, /ezudo/, /edubu/, /edugu/, and /eduzo/. A female native English speaker from Canada pronounced the target words embedded in a carrier sentence "Please say _____." and the productions were recorded in a recording studio. The recording was digitized and stored on the computer with 48 kHz of sampling frequency. The words were cut out from sentences and stored in the form of wav files using a free audio editor, Audacity. Then all of the stimuli were manipulated with a silent period between the end point of V_1 and the start point of C_1 by elminating it from the onset of the silence in four steps (25%, 50%, 75% and 100%). In addition, the intensity of the whole word was manipulated by increasing it (+3 dB and +6 dB) and decreasing it (-3 dB and -6 dB) in two steps, respectively. In total, 200 sound stimuli were prepared (type of non-words [4] x phonotactic contexts [2] x CDO [5] x intensity [5] = 200 sound stimuli) for a forced-choice task. Figure 1 shows a spectrogram of the original file of /ebuzo/. Figure 2 shows a spectrogram of /ebuzo/ in which VOT was completely deleted and the arrow in Figure 1 indicates the part in which CDO was manipulated.

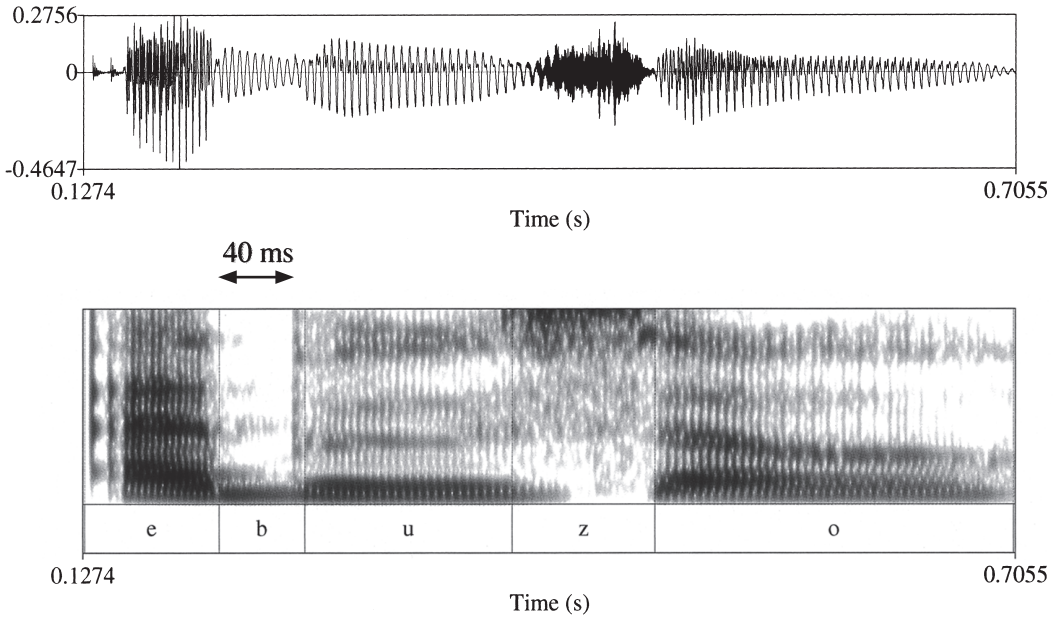

Figure 1. Spectrogram of the original file of /ebuzo/. The arrow indicates the part which was manipulated to create sound files with decreased closure duration of C_1.

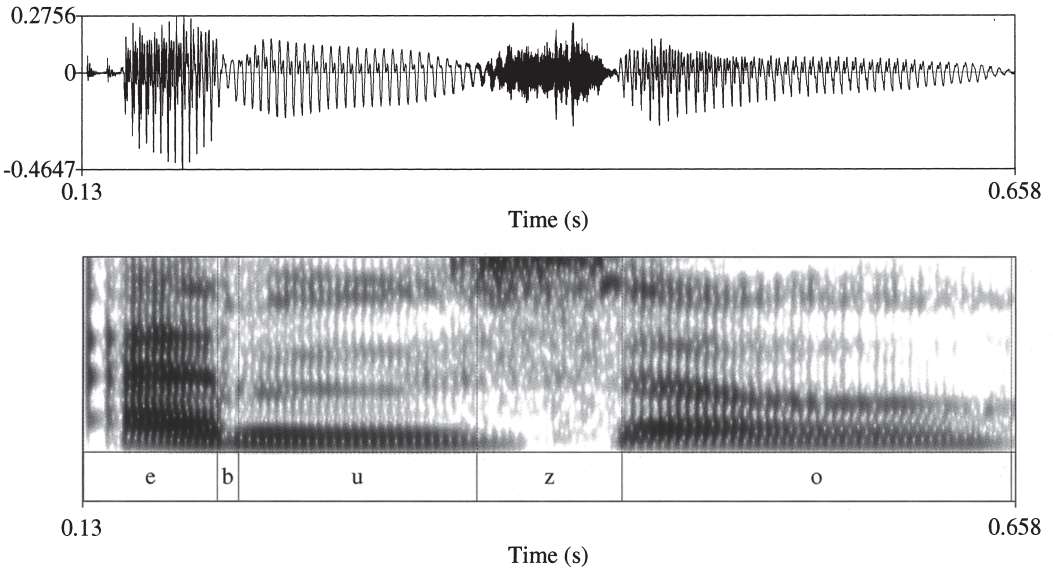

Figure 2. Spectrogram of a file of /ebuzo/ in which closure duration of C_1 was completely deleted.

2. 2 Participants

Twenty native Japanese speakers who were undergraduates of Kumamoto University took part in this study. Their mean age was 19.2 years and their mean score for TOEIC was 572.9. None had never had lived in English-speaking countries although one of them had stayed for two weeks both in Canada and Australia and four of them had visited English speaking countries (Australia, U.S.A. and Singapore) for less than 10 days. None reported hearing impairment.

2. 2 Procedure

A forced-choice task was created with the aid of E-prim 2.0 in which a sound stimulus was presented while the mark "+" was appearing in the center of the computer screen. A $V_1C_1C_2V_2$ word and a $V_1C_1UC_2V_2$ word (e.g., *egzo* or *eguzo*) then appeared visually on the screen with the instruction to press button "1" for the former and button "2" for the latter. One second after the response, another trial was presented automatically. After a practice session with three trials, 200 trials were programmed to run with a break after every 50 trials. The duration of the break was not controlled and the next trial was programmed to start by pressing any key on the keyboard so that the participants could take as much rest as needed.

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 identify which word they heard and respond by pressing a button on the response device, Chronos.

3. Results

The responses were considered as "correct" when the syllable structure of the stimulus matched the response made by the participant (e.g., an *ebuzo* reponse to /ebuzo/ stimuli was "correct", while the same response to /ebzo/ was "incorrect"). The number of correct responses was counted by item and by manipulated sound stimulus. Table 1 shows the descriptive statistics by item and Table 2 shows the number and the rate of correct responses by type of manipulation.

Table 1

Descriptive statistics of correct responses for each item (N = 25)

non-word	mean	SD	ratio
ebzo	18.8	5.0	0.75
edzo	21.4	3.8	0.86
egbu	20.8	5.3	0.83
ezgu	18.9	6.0	0.75
ebuzo	22.7	3.6	0.91
eduzo	21.8	5.2	0.87
egubu	7.1	6.5	0.28
ezugu	15.4	7.2	0.61

Note. Each item includes 25 types of manipulated files (CDO [5] x intensity [5]).

Table 2

Number of correct responses for the sound stimuli by manipulation type (N = 80)

deleted duration of CDO(%) intensity (dB)		0	-25	-50	-75	-100	average
VCCV	6	58	66	61	68	71	64.8
	3	56	61	65	66	65	62.6
	0	58	57	66	69	67	63.4
	-3	61	60	66	61	71	63.8
	-6	59	65	67	69	66	65.2
	average	58.4	61.8	65	66.6	68	
VCUCV	6	60	55	56	60	52	56.6
	3	58	61	56	56	49	56
	0	58	54	54	51	52	53.8
	-3	49	49	58	47	45	49.6
	-6	54	57	47	52	47	51.4
	average	55.8	55.2	54.2	53.2	49	

Note. Twenty participants listened to four types of non-words for each type of manipulation.

One-way repeated measures ANOVA showed a significant difference among the items (F (7, 133) = 19.5, p < .001). The responses to *egubu* were significantly different from those of the rest (p < .01). The responses to /ezugu/ were also different from those of /ebuzo/ (p < .01) and /eduzo/ (p < .05). The number of correct resonses varied according to the item. It is notable that 72% of the participants did not perceive the vowel of /egubu/ and 40% of them did not identify the vowel of /ezugu/ even though their syllable sturucture is legal. The ratios of accuracy for the two items were lower than those for items with illegal phonotactics. The results of the chi-squared test showed that the participants identified the sounds with illegal phonotactic structures more accurately than the syllables with legal structures (χ^2 (1) = 87.9, p < .001). Despite the fact the participants gave the same correct responses to the original simuli both with legal and illegal syllable structures, manipulation of CDO and intensity affected perception by the native Japanese speakers. In addition, significant effects of CDO were observed both in the responses to VCCV (F (4, 16) = 8.27, p < .001) and in the responses to VCUCV (F (4, 16) = 3.13, p < .05). The longer the duration of deletion of CDO, the less likely it was that the particpants would perceive the vowel. There was a significant main effect of amplitude only in responses to the VCUCV contexts (F (4, 16) = 3.5, p < .05). When the intensity was decreased, the participants tended to identify $V_1C_1UC_2V_2$ words as $V_1C_1C_2V_2$ words.

The results of this study showed that CDO was more influential than intensity for identifying a vowel in syllables. When CDO was completely deleted and intensity was relatively low (-3 dB), the native Japanese speakers were more likely to identify the stimuli with both phonotactic types of non-words as $V_1C_1C_2V_2$ words rather than the original non-words. In short, the Japanese participants tended not to percieve a vowel when CDO was completely deleted and when intensity was low, and the effects on perception by the native Japanese speakers were even greater when the two factors were combined.

3. Discussion

The participants identified non-words with illegal phonotactic structures more accurately than syllables with legal structures when CDO was deleted. When the intensity was decreased, the participants tended to identify $V_1C_1UC_2V_2$ words as $V_1C_1C_2V_2$ words. Although both CDO and amplitude affected the perception of a vowel in syllables with respect to the native Japanese speakers, CDO was more influential and the combination of CDO and intensity had an even larger effect on their perception. Thus, as an answer to the research question, when CDO was completely deleted and intensity was relatively low (-3 dB), the native Japanese speakers were more likely to identify the stimuli with both phonotactic types of non-words as $V_1C_1C_2V_2$ words rather than $V_1C_1UC_2V_2$ words. When the acoustic infomation is ambiguous, listeners are more likely to depend on the cues they rely on to identify speech sounds.

It has been reported that phonotactic constraints in the first langauge affect a listener's perception of L2 speech, but manipulation of phonetic information influenced the perception of an L2 syllable even though the syllable structure itself was not altered. Perception of a vowel by native Japanese speakers is more likely to be affected by low-level acoustic information than by phonotactic constraints in L1, suggesting that listeners process phonetic information they rely on without modulating it to match their L1 phonotactic constraints. Thus, the results suggested that a perceptual unit for native Japanese speakers was not a mora. It is also intriguing that manipulation of VOT of the obstruents influenced perception of the following instead of the consonants themselves. This indicates that the listeners did not recognize the consonants (e.g., /b/) and the following vowels (/u/) separately and that the perceptual unit was not a phoneme as Samuel (2020) proposed. The fact that the native Japanese speakers recognized non-words with legal syllable structures as those with illegal syllable structures when CDO and intensity were manipulated suggests that listeners do not necessarily process a given phoneme sequentially or linearly but rather recognize speech in potentially larger holistic chunks. Although it has been considered that Japanese is a mora rhythm language, native Japanese speakers are unlikely to perceive auditory input by morae.

The question remains as to whether the mora rhythm is the outcome of education in orthography or a language-specific unit of speech processing that is fostered during first language acquisition. If we presume that the mora rhythm in Japanese is caused by orthography and that the minimum unit of speech perception by native Japanese speakers is the syllable, the introduction of Roman letters at elementary school may, arguably, hinder the learning of a foreign language. As an educational implication of this argument, it would be important to attract a learner's attention to syllable rhythm by exposing them to English rhythm without applying the filter of Roman letters. Finally, further study is needed to reveal the relationship between Japanese orthography and speech perception. Further study is also needed to examine what phonetic cue a listener prioritizes depending on the language, how second language learners develop their perception of L2 speech and what acoustic cues advanced learners use.

References

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

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*(4), 635-647.

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

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes, 16* (5/6), 491-505.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47-B57. Doi: 10.1016/S0010-0277(02)00198-1

Katayama, T. (2015). Effect of phonotactic constraints on L2 speech processing. *i-Perception, 6*(6), 1-13. doi: 10.1177/204166951561571

Katayama, T. (2020). Effects of subsegmental features on vowel perception in L2 phonotactic contexts. *Annual Review of English Language Education in Japan*, *31*, 33-48

Kubozono, H. & Honma, T. (2002). *Onsetsu to mora* [A syllable and a mora]. Tokyo: Kenkyusha Shuppan.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lind-blom, B. (1992). Linguistic experiences alter phonetic perception in infants by 6 months of age. *Science*, *255*, 606-608.

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.

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.

Samuel, G. A. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of memory and language, 111*. https://doi.org/10.1016/j.jml.2019.104070

Warner, N. & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica*, *58*, 1-25.

Wilson, C., Davidson, L., & Martin, S. (2014). Effects of acoustic-phonetic detail on cross-language speech production. *Journal of Memory and Language*, *77*, 1-24.