【論文】

Perception of Vowels Following Obstruents by Native English Speakers and Native Japanese Speakers

Tamami KATAYAMA

要旨（Abstract）

This study was conducted to examine whether the vowel duration in different consonantal contexts affect the category perception of a syllable by native Japanese speakers with varying levels of proficiency in English. Voiceless and voiced fricative (/su/ and /zu/) and stop consonants (/ku/ and /gu/) followed by a vowel were used as sound stimuli, with the duration of the vowel was respectively apprehended at 50% and 100 % by the native Japanese speakers in Katayama's (2014) study. Native Japanese speakers at differing levels of English proficiency and native English speakers transcribed them. Their transcriptions were categorized using the perceptual assimilation model (PAM; Best, McRoberts, & Goodell, 2001). The results revealed that most English speakers assimilated the sound stimuli into a single unit, while the native Japanese speakers tended to classify them as two distinct categories. The native Japanese-speaking participants categorized syllables according to the vowel duration and were also influenced by the type of consonant the vowel followed.

キーワード（Keywords）：speech perception, syllable, the PAM model, fricatives, stops

1. Introduction

When we acquire our first language, we receive a great amount of acoustic information and then attune our auditory perception to recognize phonemes used in the ambient language. It has been reported that there is no absolute acoustic value to identify particular phonetic segmentals in every speech context (Jusczyk, 1997). The same acoustic clue for consonants is perceived differently according to the adjacent vowels. Despite this fact, listeners have the ability to discriminate the sounds and interpret them as phonemes. "This capacity of the psychoacoustic system to perceive sounds discontinuously and in the form of discrete units is known as *categorical perception*" (Boysson-Bardies, 1999, p.19) and is considered to be part of the biological abilities of humans.

Research on perception of a second language has shown that our native language strongly affects the perception of non-native speech. One of the main perception models in L2 perception is the perceptual assimilation model (PAM) proposed by Best (1995). According to this model, the perception of L2 phonemes depends on how phonetically similar they are to phonemes of the listeners' first language (L1). An model is the speech learning model proposed by Flege (1995), which claims that L2 speech perception develops as the listener gains experience in the L2 even after a learner passes puberty. However, previous studies have mainly focused on phoneme levels and research on how L2 learners perceive a syllable is required.

1.1 Best's PAM

Best, McRoberts, and Goodell (2001) mooted the PAM to offer a systematic explication of variations in the perception of non-native speech. According to the PAM, two non-native phones may be separately assimilated as two native phones because of their respective phonetic similarities due to a process termed *Two Category assimilation* (TC). Two non-native phones may also assimilate into a single native phoneme since both could equally well or poorly fit a given category. This process is labeled *Single Category assimilation* (SC). An instance in which both non-native phones could assimilate into a single native phoneme, but one may fit better than the other is called a *Category Goodness difference* (CG). One of two non-native phones may be classified and the other many not; this eventuality is designated as an *Uncategorized Categorized* pair (UC). When both non-native phones are uncategorized as speech, the occurrence is termed *Uncategorized speech segments* (UU). Finally, *Non Assimilable* (NA) nonspeech sounds indicate instances in which both non-native phones may not be perceived as native phonemes.

Best et al. (2001) predicted the discrimination level using the categories above as TC>CG>SC. They tested the perception of native English speakers using voiceless versus voiced lateral fricatives (/ɬ/-/ɮ/), voiceless aspirated versus ejective velar stops (/k^h/-/k'/), and plosive versus implosive voiced bilabial stops (/b/- /ɓ/) in Zulu. The results supported the PAM prediction: the listeners assimilated the lateral fricatives as a TC contrast and the velar stops as a CG difference within a single native phoneme. More than two-thirds of the participants assimilated the bilabial stops into an SC of the English /b/. The listeners identified the presence or absence of phonological contrasts; they also perceived the phonologically irrelevant phonetic and nonlinguistic information in detail. Thus, they could distinguish three types of speech information: phonological, phonetic, and nonlinguistic. Best et al. claimed that the PAM generated systematic comparisons of various types of non-native contrasts within the broader context of the phonological system. In so doing, it considered the phonological distinctions between them, and phonetic variations within them, and the native equivalence classes.

1.2 Perception of L2 sounds

Iverson et al. (2003) have reported that a language-specific network structure fostered by early language experience affects adults' perception of speech sounds. They tested the hypothesis that early language experience influences perception of non-native speech at a low level and that realization of L2 phonemes is impeded by these changes. They conducted experiments on perception of English /r/ and /l/ by three language groups of Japanese, German, and English native speakers. Eighteen /ra/ and /la/ stimuli were manipulated to vary in the frequencies of the second and third formants during consonant closure, and the participants took part in three tasks: the first task was to rate whether the stimulus was a good exemplar of that category using a scale from 1 to 7, the second task was to rate the acoustic similarity of stimulus pairs on a scale from 1 to 7, and the last task was to discriminate one stimulus from another. The results showed that American listeners and German listeners depended on F3 the most to distinguish /r/ and /l/, while the Japanese group was more sensitive to F2 than F3. The Japanese listeners counted F2 into their formation of /l/ representation, which interfered with their identification of /r/ and /l/. The native Japanese speakers

relied not on a critical acoustic cue but on an irrelevant cue for English /r/ and /l/ categorization. Thus, Iverson et al. claimed that listeners modify L2 speech at an early phonetic level for processing, which makes it difficult for adult learners to acquire the L2 phoneme.

1.3 Unit of timing of Japanese and L2 perception

Timing in Japanese is constrained by an abstract temporally defined mora, not by a tendency to regularize syllable durations and word durations depending on the number of morae in a word for Japanese speakers (Port, Dalby & O'Dell, 1987). According to Ueyama (1996), Japanese speakers adjust the length of words depending on the number of syllables in a word. A mora mainly consists of a vowel or a consonant followed by a vowel, and the duration of a vowel distinguishes the meaning of words in Japanese such as *kite* (CVCV meaning "please come") with a short vowel, *kiite* (CV meaning "please listen") with a long vowel, and *kitte* (CVQCV meaning "stamp") with a glottal stop. Fujimoto and Maekawa (2014) used a corpus of oral Japanese production and compared vowel durations before and after glottal stops. They reported that the duration of /a/ and /e/ followed by glottal stops is shorter than that followed by single consonants and that the duration of /o/ after glottal stops is significantly longer than that followed by non-glottal stops. English differs from Japanese with respect to the use of syllable duration. Stressed syllables are pronounced longer in English, but vowel duration does not affect lexical meaning.

Katayama (2014) examined the effects of consonants on perception of the following vowels, using stimuli of CV items, /ku/ and /su/ pronounced by a native English speaker. A total of 18 sound stimuli were employed, consisting of 9 steps of sounds for which durations were manipulated (2 kinds of stimuli x 9 steps): 160 ms, 140 ms, 120 ms, 100 ms, 80 ms, 60 ms, 40 ms, 20 ms, and 0 ms. Then 24 native Japanese speakers (JS) and 22 native American English speakers (ES) judged whether it has a vowel (e.g., /ku/) or not given a sound stimulus as well as the phonetic symbols on the screen of a computer. As a result, for /k/ and /ku/, the 50% boundaries of vowel duration were 53.3 ms for ES and 107.5 ms for JS. For /s/ and /su/, the boundaries of the two groups showed similar values: 63.3 ms for ES and 66.7 ms for JS. The results indicated that perception of the vowel differed depending on the consonants the vowel follows.

When listeners encounter a phoneme that is not stored in their first language inventory, they are likely to depend on acoustic information to recognize the new sound since their category perception for the phoneme is not stabilized. These results invite the question: did Japanese speakers categorize syllables according to their vowel duration or did they depend on other phonetic information?

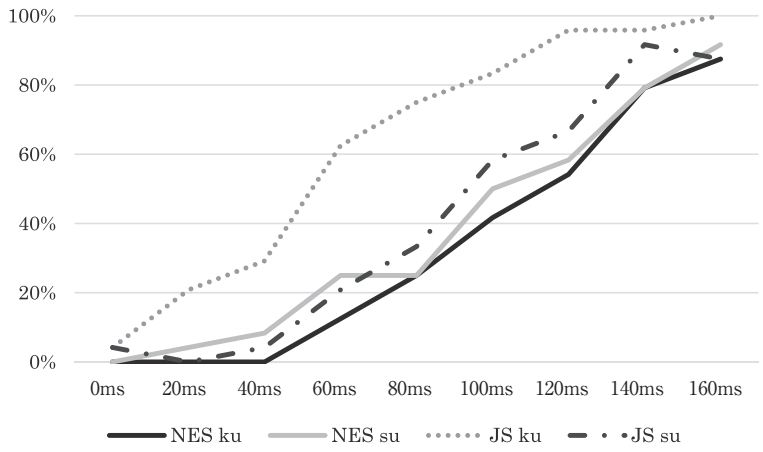

Figure 1. Mean percentages of responses /ku/ and /su/ for the consonant and /u/ vowel continua (Katayama, 2014).

2. Research Purpose

The purpose of this research was to investigate how native Japanese speakers and native English speakers categorize voiced and voiceless velar stops and alveolar fricatives followed by different vowel durations using the PAM model (Best et al., 2001) and whether there is difference of perception of syllables develops between native Japanese speakers with high L2 proficiency and those with low L2 proficiency. The following research question was raised: How do L2 learners with different levels of English proficiency and native English speakers identify voiced and voiceless velar stops and alveolar fricatives followed by different vowel durations?

3. Methods

3.1 Materials

The present study employed the stimuli of voiceless fricative and stop consonants (/su/, and /ku/), which were manipulated in Katayama's study (2014) by varying the vowel duration as well as the corresponding voiced consonants (/zu/ and /gu/) to examine the effects of voicing. The boundaries of the perception of native Japanese speakers vis-à-vis vowels following the consonants /s/ and /k/ were computed at 80 ms and 120 ms, respectively. In Katayama's study, native Japanese speakers could identify a vowel both for /su/ and /ku/ when the vowel duration was 160 ms. Hence, the present study used eight stimuli: the vowel durations for the /su/ and /zu/ stimuli were 80 ms and 160 ms, respectively (su80, su160, zu80, and zu160) and the vowel durations for /ku/ and /gu/ were 120 ms and 160 ms, respectively (ku120, ku160, gu120, gu160). The stimuli were recorded by a female native English speaker from the east coast of the United States. They were subsequently entered into a computer at a sample rate of 48 kHz and were manipulated using speech analyzer software named *Praat*.

Table 1 presents the acoustic features of the eight stimuli and formant frequencies at 50% point of the vowel. The /u/ vowels with high frequency of F2 were employed so that native English speakers would consider them as non-native speech. Since only vowel durations were manipulated, the differences between the original and the manipulated sounds concerned full-syllable durations.

Table 1

Acoustic Values of the Eight Stimuli

	ku120	ku160	gu120	gu160	su80	su160	zu80	zu160
whole syllable duration (ms)	200	240	232	272	244	321	226	306
frequency peak (Hz)	287	287	248	248	243	243	259	258
F1 at 50% point (Hz)	405	369	378	371	432	375	411	372
F2 at 50% point (Hz)	1760	1580	1587	1864	2215	2068	2162	1969
F3 at 50% point (Hz)	2320	2425	2268	2416	2549	2450	2529	2339
VOT(ms)	75	75	-89	-89	-	-	-	-
duration of frication (ms)	-	-	-	-	164	164	145	145

3.2 Participants

Twenty native Japanese speakers with a low level of English proficiency (JL), including graduates and undergraduates of a college in Japan, took part in the study. The mean age of the participants was 19.7 years and their mean score for TOEIC was 359.7. Eighteen of them had no living experience abroad although two of them had visited Canada for two weeks and one month, respectively. Eighteen native Japanese speakers with a high level of English proficiency (JH) who were mainly English instructors at colleges also participated in the study. Their mean age was 38.3 years, mean duration of living experience in English speaking countries was 3.6 years, mean score for TOEIC was 931.8 and mean score for TOEFL iBT was 108. Eighteen native English speakers (ES) who were graduates and undergraduates of the University of Edinburgh also took part in the study as the baseline for comparison with the Japanese groups. Although the sound stimuli were native sounds for ES, the manipulated stimuli were treated as new sounds for them. Fifteen of them were from the U.K. and two were from the U.S.A., and their mean age was 22.6 years.

3.3 Procedure

Each of the participants conducted a dictation task in a quiet room. They were asked to dictate what they heard on a sheet of paper after listening to the eight sound stimuli through a computer. The participants were allowed to use any language to describe the sounds.

3.4 Analysis

Following Best et al. (2001), English and Japanese spellings and descriptions were categorized in terms of whether pairs of stimuli (e.g., ku120 and ku160) were the same or different, including detailed descriptions of the stimuli. If a pair was identically spelled, the assimilation pattern was categorized as SC. If a contrast was noted in a pair by the use of a common Japanese or English letter for one phonetic aspect, but the addition of a mark to denote the other phonetic feature (e.g., Kh), and/or acoustic differences were reported between the contrasts, then the assimilation pattern was considered a CG difference. The acoustic description suggested the participant's perceived goodness-of-fit to a phoneme in L1 or L2. If the contrasts were spelled with different letters in Japanese or English or one member of the pair was noted in English spelling and the other in Japanese, the assimilation pattern was identified as TC. If one or both stimuli were described to represent a space between two or more phonemes (e.g., between 'k' and 'ku'), then the assimilation patterns were respectively categorized as UC (uncategorized-categorized) or UU (uncategorized-uncategorized).

4. Results

The spellings and descriptions noted in the writing of 18 participants were categorized using the method posited by Best et al. (2001) and were found to include SC, TC, or CG. Table 2 exhibits the transcription categories according to the three classifications.

Table 2

Frequency for the PAM Model's Categories of Transcriptions by Three Groups

	SC			TC			CG		
	ES	JH	JL	ES	JH	JL	ES	JH	JL
su 80 - 160	6	2	3	9	12	12	3	4	3
zu 80 -160	7	4	5	8	14	10	3	0	3
ku 120 - 160	16	9	13	2	8	4	0	1	1
gu 120- 160	9	7	6	7	10	9	2	1	3

Figures 2-4 indicate the frequencies of each category marked by ES, JH, JL, respectively. Overall, the native Japanese groups classified the contrasts more frequently as TC than ES, especially in the case of fricative stimuli. Transcriptions of the /zu/ contrasts were categorized as TC by 77% of JH and 66% of JL, while 66% of both JL and JH categorized the /su/ contrasts as TC. Thus, the Japanese groups tended to distinguish the contrasts as two distinct phonemes. Conversely, ES were more likely to identify them as SC. The /ku/ contrasts were grouped as SC by 89% of ES, 72% of JL, and 50% of JH participants. ES considered the contrasts of the voiceless velar stop as a single category, and JH participants also

demonstrated this trend. ES categorized the /gu/ contrast as SC more frequently than as TC, and the Japanese groups evidenced the opposite trend. A few CG categories were observed by each group, signifying some attention to phonetic details.

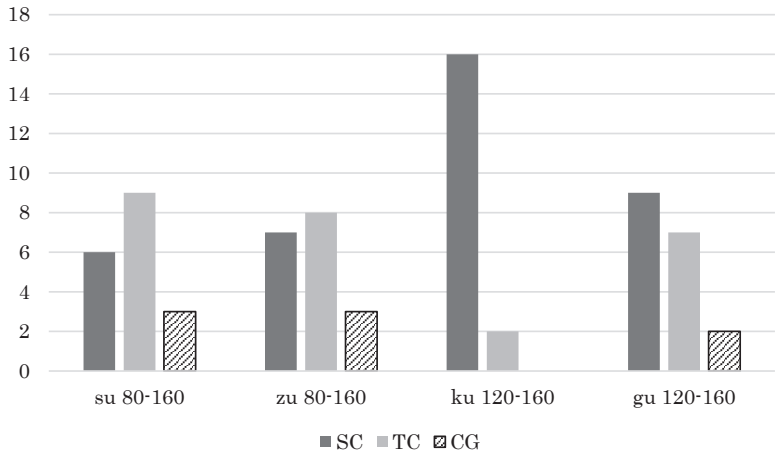

Figure 2. Categories of transcriptions by native English speakers. Given stimuli with different vowel durations (i.e., vowels with durations of 80 ms and 160 ms for /su/ and /zu/ and vowels with durations of 120 ms and 160 ms for /ku/ and /gu/), transcriptions were categorized into *Single Category assimilation* (SC), *Two Category assimilation* (TC), and *Category Goodness difference* (CG).

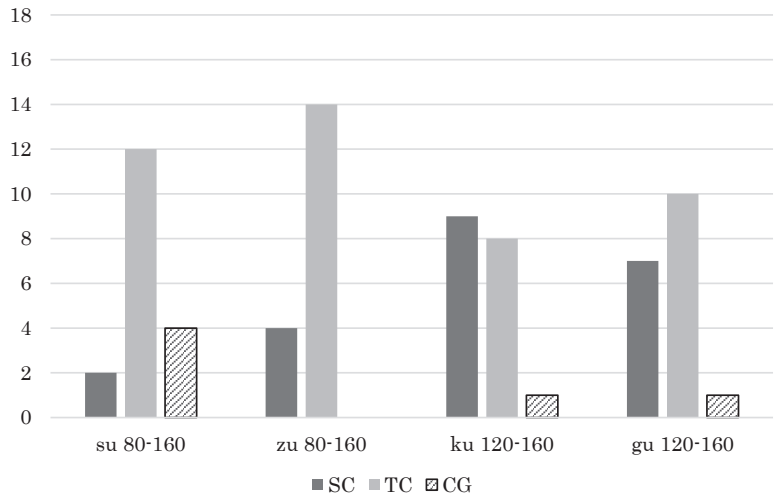

Figure 3. Categories of transcriptions by native Japanese speakers with a high level of English proficiency.

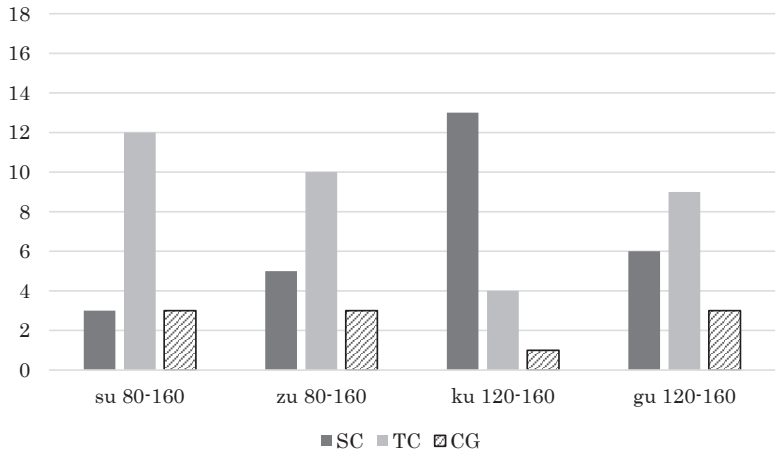

Figure 4. Categories of transcriptions by native Japanese speakers with a low level of English proficiency.

Table 3

Frequency for Each Category of Transcriptions by Three Groups

group	vowel of stimuli	C	CV [+tense]	CV [-tense]	others	short vowel	long vowel	glottal stops
ES	shorter	13	46	8	5	0	0	0
	longer	0	51	19	2	0	0	0
JL	shorter	20	4	0	0	36	5	7
	longer	16	6	3	0	14	32	1
JH	shorter	25	25	6	3	11	2	0
	longer	6	25	25	3	7	6	0

Note. The categories denoting consonants (C), consonant and vowel (CV) combinations including and excluding tense, (i.e., CV[+tense] and CV[-tense]), as well as other classifications were transcribed using the English alphabets, while groupings of short vowels, long vowels, and geminates were transcribed using Japanese *kana*. The category labeled *glottal stops* indicated the transcription of a short vowel followed by a grottal stop consonant.

Table 3 shows the frequency for each category of transcriptions by the three groups. Since most of the ES transcribed the stimuli as CV, they are assumed to have perceived a vowel in general regardless of the vowel duration. While JL tended to categorize the stimuli into a phoneme with a short vowel (e.g.,ぐ) and that with a long vowel (e.g.,ぐう) depending on the vowel duration, JH overlapped categories of C, CV, and CVV. Although JH were still affected by the vowel duration, they tended to transcribe the stimuli using the English alphabet and 35% of them did not perceive a vowel when the vowel duration was short. A notable difference in perception of a vowel between voided consonants and voiceless consonants was

observed. Although some variabilities can be seen, most of the answers given by ES are a combination of a consonant and a vowel. ES tended to identify a vowel regardless of the duration and type of the consonants which the vowel follows, while both JL and JH paid more attention to the vowel duration than ES did.

JL were likely to discriminate the stimuli with different vowel durations by using *kana* letters with long vowels or short vowels followed by glottal stops (e.g., くっ). Most of the JL wrote down "q" after listening to both /ku/ sounds (120 ms vowel and 160 ms vowel). On the other hand, JH showed less variability in their transcriptions than did JL, but JH still tended to be affected by the vowel duration and categorized the stimuli depending on the vowel duration. Most of the JH did not perceive the vowel in the voiced fricative consonant /zu/ for which the vowel duration was 80 ms, while they perceived the 160 ms vowel in /zu/ as a long vowel (see Table 2). JH identified both of the /gu/ stimuli (120 ms vowel and 160 ms vowel) as /gu/, and five participants in this group recognized /gu/ with a 160 ms vowel as "*goo*." Although JH showed a variety of perceptions for both of the /ku/ stimuli with a 120ms vowel and a 160 ms vowel, most of the transcriptions included a vowel following a consonant written in the English alphabet. No transcription of a geminate consonant was observed in JH.

5. Discussion

The purpose of this study was to examine whether the vowel duration affected the category perception of a syllable by L2 speakers with varying levels of proficiency in English and by native English speakers. Voice and voiceless velar stops and alveolar fricatives with different vowel durations were used as stimuli on the basis of Katayama (2014). The transcriptions made by the participants were then classified using Best et al.'s (2001) method. The research question was formulated as: How do L2 learners with different levels of English proficiency and ES identify voiced and voiceless velar stops and alveolar fricatives followed by different vowel durations? To answer this research question, native Japanese speakers categorized at two levels of English proficiency and ES transcribed the sound stimuli of a CV syllable structure at discrete vowel durations.

The present study suggested that the listeners categorized syllables depending both on the phoneme and on vowel duration at the syllable level. As predicted, native Japanese speakers most frequently assimilated the stimuli as TC in this study since the vowel duration encompasses distinctive functions in the meaning of a word. However, the assimilation pattern differed according to the consonant that preceded the vowel. In particular, the fricative contrasts were assimilated as TC, while the classification of the stop pairs was split into TC and SC apart from the JL group's transcriptions of voiceless stop contrasts. Like the ES, 72% of JL assimilated them as SC.

In Katayama's study, ES perceived 50% vowel duration at 53.3 ms for /ku/ and 63.3 ms for/su/, while native Japanese speakers recognized the same at 107.5 ms for /ku/ and 66.7 ms for /su/. The results for ES in the present study were congruent with the findings recorded by Katayama (2014) because the value of the boundary for the fricative /su/ approximates the vowel duration for one of the stimuli (i.e., 80 ms). This result accounts for split categories either as SC or TC. However, both JH and JL categorized the /zu/ contrasts as TC more frequently than ES, contradicting the prediction that they would show the same

trends as ES.

The results that 89% of ES and 50 % of JH categorized the /ku/ contrasts as SC can also be explained by the outcome that the boundary of vowel perception by ES was 53.3 ms and was 107.5 ms by the native Japanese speakers. The results obtained for the JH group for the stop contrasts, /ku/ and /gu/, were aligned with the findings of a previous study whose categories were separated into SC and TC since their perception boundary of the vowel approximated the stimuli of shorter vowel durations (i.e., 120 ms). JL participants tended to assimilate the contrast into a single category only for the /ku/ contrast. This result is intriguing because the group with less English proficiency evidenced a trend identical to the ES. The Japanese groups tended to exhibit discrete categorization depending on the consonants that preceded the vowel sound.

Although ES tended to perceive the vowel regardless of its length, Japanese speakers paid more attention to the vowel duration and categorized the sounds depending on the vowel duration. They tended to rely on different acoustic cues. JH did not recognize the vowel in some contexts. Some JL recognized the consonant with a shorter vowel duration as a glottal stop. It is assumed that acoustic cues to perceive the unit of a consonant and vowel might change as their L2 proficiency develops. For responses to the /zu/ stimulus with an 80-ms vowel duration, half of the ES identified it as “zu” and 60% of the JL group recognized it as “ず”, but 66% of the JH did not perceive the vowel. Another possible reason for the responses made by ES to the fricative stimuli were split into SC and TC with those to TC being slightly more frequent is that since ES’s boundaries for /su/ were 63.3 ms (Katayama, 2014), the fricatives with short vowels (i.e., 80 ms) might not have been long enough for them to perceive the vowel. On the other hand, JH categorized the two /zu/ stimuli into /z/ and /zu/ depending on the vowel duration. Since JH are advanced English learners, they are likely to know that /z/ and /zu/ belong to different categories. Figure 5 shows a spectrogram of /zu/ with an 80-ms vowel duration. The acoustic features of /z/ and /zu/ are closer than those of other stimuli. When the vowel duration is short, it is difficult to determine the boundary between the consonant and the vowel. JH might have paid more attention to F1 and F3 since these are common features of /z/ and /u/. On the other hand, ES tended to consider F2 as well since this is a clear difference between the /z/ consonant and the /u/ vowel.

Dupoux, Kakehi, Hirose, Pallier and Mehler (1999) reported that listeners perceive epenthetic vowels in illegal phonotactic contexts in their first language. However, the results of this study showed that JH failed to perceive the vowel even in a legal phonotactic context, CV. Thus, it is likely that the failure in identifying the vowel was caused by acoustic cues they depend on rather than L1 phonotactic constraints.

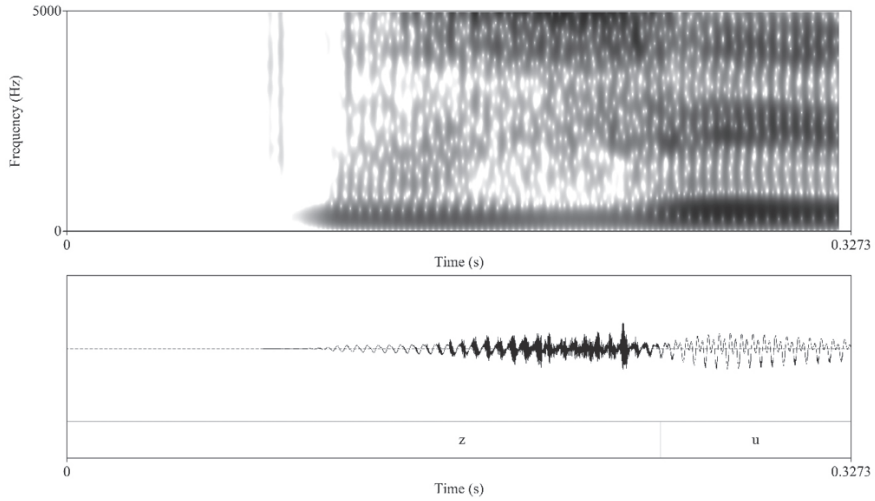

Figure 5. Spectrogram of /zu/ with a vowel duration of 80 ms.

Although the present study is exploratory due to the limited sample data, the results support the results of Fujimoto and Maekawa's study (2014) and suggest that JL recognize sound stimuli with a shorter vowel as a glottal stop in /u/ contexts as well. Fujimoto and Maekawa found that durations of vowels followed by glottal stops are shorter than durations of vowels followed by single consonants when the vowels are /a/ and /e/, but they could not find this trend in other vowel contexts because of the high variance in speech data. Although perception and production of glottal stops may not be necessarily equivalent, the findings in this study could contribute to investigation of how native Japanese speakers produce and perceive speech. The results of this study revealed that native Japanese speakers displayed more variance than did ES in their transcriptions, given the stimulus of a syllable with different vowel durations. English accords categories for phonemes comprising a single consonant (e.g., /k/) and those consisting of consonants and vowel units (e.g., /ku/), but Japanese groups mora that generally encompass consonants followed by different vowel durations (e.g, く: /ku/, くう: /kuu/ and くっ: /kuʔ/) and it is pronounced without a vowel in some contexts. Therefore, the ES might have been more sensitive than the Japanese-speaking groups in identifying a vowel regardless of its duration. Both JH and JL overgeneralized the classification of syllables by dividing them into two categories relying on vowel duration when they were supposed to belong to the same grouping. JL tended to describe the stimuli as Japanese characters or *kana*. Notably, only this group identified speech sounds with short vowel durations as glottal stops, a construct observed in Japanese. The fact that waveforms suddenly disappeared might have caused JL to perceive the sounds as cease of air stream, which resulted in their transcriptions of glottal stops. JL were thus inclined to recognize given speech sounds through their first language. They tended to depend on acoustic cues different from ES to identify a phoneme; this fact could also influence their perception of the second language syllables and words. The reason that JL did not transcribe geminates might derive from their L2

experience since the stimuli were pronounced by a native English speaker. Even though the stimuli were legal in phonotactics, the Japanese speakers tended to perceive subtle acoustic information that can be identified as speech produced by a native English speaker.

6. Conclusion

This study was conducted to examine how native Japanese speakers recognize fricative and stop consonants with different vowel durations and whether speech perception at the level of syllables develops with advancing L2 proficiency. The research question was formulated as: How do L2 learners with different levels of English proficiency and native English speakers identify voice and voiceless velar stops and alveolar fricatives followed by different vowel durations? Most ES participants identified the sound stimuli as SC, while the native Japanese-speaking groups JH and JL were likely to categorize them as TC. The native Japanese speakers were influenced by vowel duration in identifying syllables, regardless of their English proficiency levels. However, a variance was observed in their perceptions of vowels based on the preceding consonant. In sum, although the ability of native Japanese speakers to discriminate between syllables was substantially affected by vowel duration, the length of time required by them to perceive the vowel depended on the type of consonants. The phonetic features of a consonant are thus probably involved in the ability of native Japanese speakers to identify a vowel.

The present study has several limitations. First, since the sample data were limited, inferential statistics were not performed. In order to confirm the results, further study using larger samples is required. Second, since the stimuli used in the study had quite high F2 values, it is reasonable to conclude that JL, JH, ES made unique responses given by the stimuli similar to /u/. Pronunciation of various types of native English speakers would solve this problem in future research. The present study offered additional insights to contribute to the literature on category perception of syllables using the PAM model. Further investigations are required to examine other phonemic contexts pronounced by speakers of other languages as well.

7. References

Best, C. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross Language Research* (pp. 167-200). Timonium, MD: York Press.

Best, C., McRoberts, G.W., & Goodell, E. (2001). Discrimination of non-native consonants varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America,* 109(2), 775-793.

Boysson-Bardies, B. de. (1999). *How language comes to children: From birth to two years*. Cambridge, MA: MIT Press.

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(6), 1568-1578.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross Language Research* (pp. 233-277). Timonium, MD: York Press.

Fujimoto, M. & Maekawa, K. (2014). Effects of sokuon on adjacent vowel duration: An analysis of the corpus of spontaneous Japanese. *Journal of the Phonetic Society of Japan, 18*(2), 10-22.

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

Jusczyk. P.W. (1997). *The Discovery of Spoken Language*. Cambridge: The MIT press.

Katayama, T. (2014). The Effect of stop and fricative consonants on perception of the following vowels: Comparative study of native Japanese speakers and native English speakers. In H. Sato (Ed.), *Proceedings of the 28th General Meeting of the PSJ* (pp. 69-74). Tokyo, Japan: The Phonetic Society of Japan.

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

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.

Appendix

A. *Transcriptions by ES and frequencies*

ES	C		CV		CVV		others	
su 80	s	5	su	10	sue	2	s'ih	1
zu 80	z	5	zu	9	zoo	2	z'ih, z'u'	2
ku 120	k	1	ku(cu)	13	koo(coo)	3	queue	1
gu 120	g	2	gu(ge)	14	goo	1	g'ih	1
su160			su	12	sue(soo)	6		
zu 160			zu	13	zoo	5		
ku 160			ku(cu)	14	koo(coo)	2	queue	2
gu 160			gu	12	goo(gue)	6		

B. *Transcriptions by JL and frequencies*

JL	C		CV		CVV		short vowel		long vowel		glottal stops	
su 80	s	4	su	1			す	11			すっ	2
zu 80	z	6	zu	1			ず(づ)	10			ずぃっ	1
ku 120	q (g)	7	qoo	1			く(きゅ)	4	くぅ(きゅう)	5	きゅっ	1
gu 120	g	3	gu	1			ぐ	11			ぐっ	3
su160	s	1	su	3	Soo (suu)	2	す	4	すう(すぅ)	8		
zu 160	z	4	zu	2			ず	3	ずう(ズー)	9		
ku 160	q	7					く(きゅ)	4	くぅ(きゅう)	7		
gu 160	g	4	gu	1	goo	1	ぐ	3	ぐぅ(ぐう･ぎゅう)	8	ぎゅっ	1

C. *Transcriptions by JH and frequencies*

JH	C		CV		CVV		others		short vowel		long vowel	
su 80	s	5	su (se)	8	sue	1	sh	1	す	3		
zu 80	z	12	zu	2			zh	1	ず	3		
ku 120	k (q)	4	ku (qu)	4	coo (goo, cue)	4	kh	1	kyo (kyu)	3	くう（キュー）	2
gu 120	g	4	gu	11	goo	1			ぐ	2		
su160			su	7	sue(su:, soo)	7	sie	1	す	1	すう	2
zu 160	z	1	zu	7	zoo(zu:)	8	zh	1	ず	1		
ku 160	q	2	ku(cu)	5	koo(coo,qoo cou,cue)	5			く (kyu)	2	くう(クー･キュー･ kyu:)	4
gu 160	g	3	gu	6	goo (gue)	5	gut	1	ぐ(gyo)	3		