INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Mental Representation of Japanese Mora; focusing on its intrinsic duration

Kosuke Sugai

Kindai University, Japan

ksugai@kindai.ac.jp

Abstract

Japanese is one of the typical languages in which vowel quantity plays a key role. In Japanese, a phonological structure called "mora" is a fundamental rhythmic unit, and theoretically, each mora is supposed to have a similar duration (isochronicity). The rhythm of a native language has great importance on spoken language processing, including second language speaking; therefore, in order to get a clear picture of bottom-up speech processing, it is crucial to discern how morae are mentally represented. Various studies have been conducted to understand the nature of speech processing as a cognitive construct; however, most of this research was conducted with the target stimuli embedded in words or carrier sentences to clarify on specifically the relative duration of morae. In this study, two reaction-time experiments were conducted to investigate whether morae are mentally represented and how long the duration is. The isolated vowels /i/, /e/, /a/, /o/, /u/, and syllable /tan/ were chosen as target stimuli, and the first morae were digitally manipulated into 15 durations with 20 ms variations in length, from 150 ms to 330 ms. The results revealed the existence of a durational threshold between one and two morae, ranging around 250 ms.

Index Terms: mora, perception, duration, Japanese

1. Introduction

Many languages have phonemic distinction between long and short vowels, such as Czech, German, Finnish, Japanese, Thai, and many others [1]. In some of these quantity languages, the primary cue is vowel duration [2, p. 294]. For example, in Japanese, the phonemic distinction is realized mainly by vowel duration, such as /'toki/ (a bird called *nipponia nippon*) and /'to:ki/ (pottery). Other phonetic features also contribute to the distinction in many languages [3–8]. In languages such as Thai, it is reported that vowel quality largely affects distinction of the contrast in addition to vowel duration [8], and pitch contour plays some role in some languages [4, 8]. Among those languages, Japanese is one of the typical languages, in which the duration of the vowel is the dominant factor in perceiving phonological contrast [1].

In Japanese, this phonemic distinction is a key feature of the language, and this basic smallest structure is called "mora." Mora is a rhythmic unit of Japanese, and Japanese is categorized as mora-timed language [9–13]. Furthermore, mora is classified into two types. The first type is called normal morae, which have three types of structures: simple vowel, consonant and vowel, and, semi-vowel and vowel. The second type, called special morae, comprises long vowel, chocked sound, and syllabic nasal, and syllables including these special morae or diphthongs are counted as two morae. Extensive research has been conducted to establish a clear picture of the nature of Japanese mora from its acoustical and perceptual aspect. Before experimental research was introduced, Block [10] observed that every mora has almost the same duration in Japanese. Trubetskoy [14] stated, from a phonological viewpoint, that Japanese language rhythm is based not on syllables but on morae; in other words, Japanese morae have isochronicity. Japanese researchers also supported the role of morae, stating that speakers of Japanese can analyze their language not smaller than a mora [11–13]. However, as McCawley [15] argued, mora can only be defined phonologically, such as a long syllable consisting of two morae and a short syllable consisting of one mora; therefore, in order to closely reveal the nature of the mora, a variety of experimental studies have approached this issue. The chief concern of this present research is the durational feature of mora, and previous research related to this specific feature is introduced below.

Several acoustical analyses revealed the relative duration of mora [16–18]. Han [16] conducted an acoustical research asking participants to pronounce target words embedded in carrier sentences, and reported that adding one mora increases the target words' duration by 120 ms. She also reported the number effect of morae on word duration, and the duration of mora in her study ranged from 104.2 ms to 143.5 ms. Campbell and Sagisaka [19] demonstrated the compensation effect between consonant and vowel within each mora; that is, a longer vowel follows a shorter consonant, and a shorter vowel follows a longer consonant. However, no definite duration of morae has been reported. Furthermore, Port et al. [17] claimed that adjacent morae compensate the duration with each other to keep their duration in multiple morae level. These acoustic studies basically addressed mora duration in relation to their phonetic environment, and did not show the intrinsic duration of mora.

Because of the belief in the difficulty to define mora from an acoustical perspective, researchers have tried to examine its nature as a cognitive construct [19–27]. Even though, not few studies have directly tried to clarify the duration of mora. Fujisaki and Sugito [21] conducted a research using target non-words without carrier sentences. Moreover, the duration between vowels were manipulated gradually: (1) from /oi/ to /o:i/, (2) from /ise/ to /isse/, (3) from /ita/ to /itta/, and (4) from /ama/ to /amma/. Furthermore, they analyzed the responses whether the participants judged them as two morae or three morae, and their result showed each boundary: (1) 156 ms, (2) 166 ms, (3) 164 ms, and (4) 152 ms. Minagawa-Kawai, Mori, and Sato [25] used a stimulus /mama/, changing the duration of vowel /a/. Their result showed that the durational boundary of one or two phoneme was 198 ms (SD = 6.9, n = 8); However, the final vowels of words presented without a phonological environment had 1.5 to 2.0 times longer duration than words pronounced in sentence. However, their data should not be acknowledged as a universal duration of mora [25, 28–29]. Kinoshita, Behne, and Arai [24] focused on the

http://dx.doi.org/10.21437/Interspeech.2017-1720

duration of five vowels in Japanese and used /z_za/ as a carrier word. Their analysis was based on the duration of vowels and range of F0 declinations, and they reported a 50 % judgment threshold between short and long vowels (table 1).

Table 1: *Long-short contrast of vowel [24]*

vowel	short	long	duration (ms)
i	i	i:	140 - 170
e	e	e:	171 - 186
a	a	a:	158 - 177
o	o	o:	149 - 177
u	u	u:	137 - 151

These previous studies exhibit a sort of varying durational threshold between short and long phonemes, caused by the difference in experimental paradigms and contexts. In order to get a clearer picture of the perceptual duration of mora, it is necessary to determine the intrinsic duration of mora, that is, the duration without affected by surrounding phonological environment. For the purpose of providing this basic data of mora duration, this study tried to exclude any effect of the phonological environment as much as possible by using stimuli without any carrier sentence and also even any adjacent syllable with the possibility of affecting the relative duration of sound. Along this line, the following research questions were raised: (1) does vowel duration alone contribute to the distinction of long/short vowel?; in other words, is mora mentally represented or not?; and (2) if there is a clear mental boundary between one and two morae, where lies the durational threshold between them.

2. Experiment 1

2.1. Participants

Nineteen native speakers of Japanese (4 female and 15 male) participated in this perception experiment. Their ages ranged from 18 to 21, and most of them were speakers of the Kinki dialect of Japanese. Although some of the participants were from other parts of Japan, their responses did not show any inter-dialect difference; therefore, all the participants were treated as members of a single group. In addition, no participant reported any hearing impairment.

2.2. Materials

The focus of this study is to clarify the perceptual threshold between one and two morae without any phonological environment, and the typical Japanese five vowels (/i/, /e/, /a/, /o/, /u/), which can independently constitute a mora, were chosen as stimuli. Those vowels were pronounced by a male, 39 years old, native speaker of Kinki-Japanese, and were digitally recorded with an OLYMPUS PCI recorder LS-11 through a SONY ECM360 condenser microphone.

The durations of each five vowel were manipulated into 15 variations from 150 ms to 430 ms, with approximately 20 ms step, and a total of 75 stimuli were prepared. Praat (5.3.39) was used for this sound manipulation.

2.3. Procedure

In order to investigate durational boundary between one mora and two morae, participants were asked to perform a self-paced binary forced-choice identification task. The psychology experiment software SuperLab mounted on a laptop computer (Apple, MacBook Pro, Retina, 13-inch, Late 2012) and the response pad RB-830 were used for stimuli presentation and reaction data collection. In each trial, two words that differ only in vowel durations, such as “ア” (/a/) and “アー” (/a:/) were visually presented on the monitor screen, with 100 ms tone for attention at the beginning. After a 600 ms silence, the target sound was presented through monitor headphones (Audio-technica, ATH-SX1a). Participants judged which word on the screen represents the sound and pressed the left or right button to corresponding with the position of the word presented on the screen as fast as possible. A 500 ms interval was inserted after each response from subjects. Each of 75 stimuli was presented 10 times, and a total of 750 stimuli were randomly presented. In order to counterbalance the effect of their handedness, among the 10 times of each target sound presentation, five of them were with the correct mora word on the left side, and the other five with the correct mora on the right side of the screen. The total duration of the experiment exceeded 10 minutes; therefore, the experiment was divided into three sessions to maintain participants' attention.

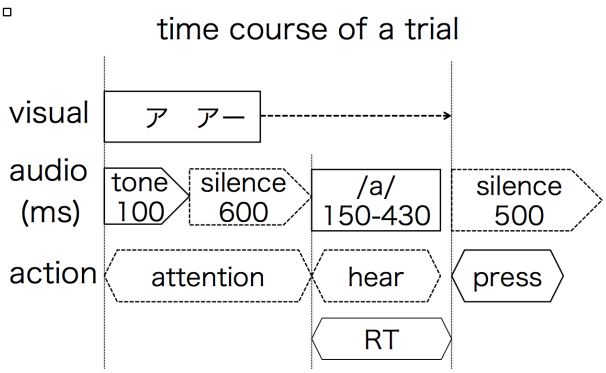

Figure 1: *Schematic illustration of a trial.*

2.4. Result

All the analysis was performed by R.3.2.2 (R Core Team, 2016) and the lme4 package (Bates, Maechler, Bolker, & Walker, 2016). As for one-or two mora judgment, the responses were analyzed by generalized liner mixed model (GLMM) with binomial distribution. The model included vowel types and vowel durations as fixed variables and the participants as a random factor. The result showed that only vowel duration was the significant factor ($z = 7.79$, $p < .001$). Figure 2 shows a typical category perception curve with a threshold between 250 ms and 270 ms. More than 90 percent participants judged 150, 170, and 210 ms vowels as a single mora word. Furthermore, along with vowel duration, the rate greatly shifted to two morae toward 310 ms. Subsequently, more than 90 percent of the participants judged the vowels as two morae from 330 ms.

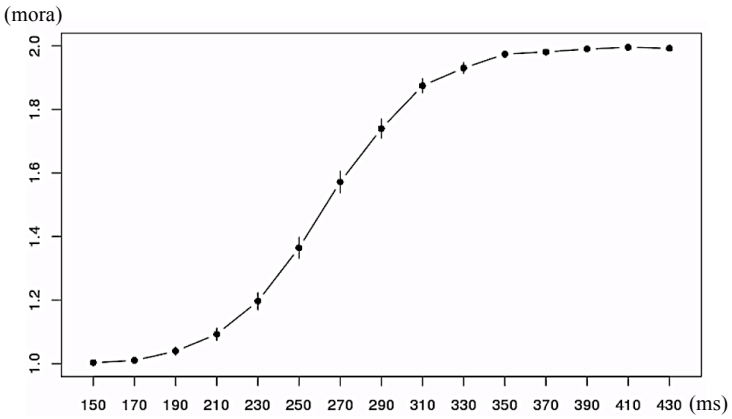

Figure 2: *Judgment data for the five vowels. Horizontal axis shows duration of the stimuli, and vertical axis shows the averaged judgment (one or two) of all subjects.*

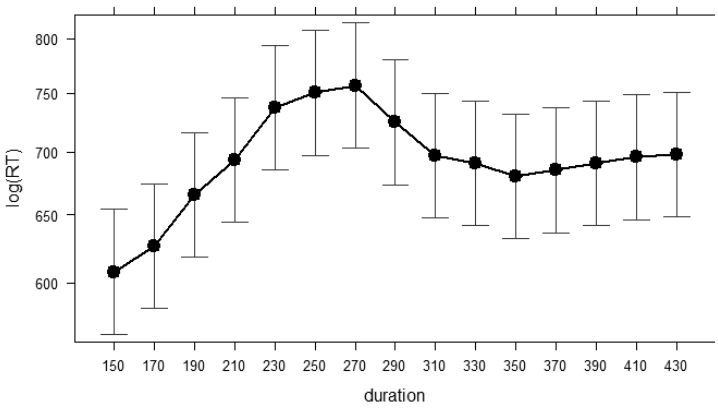

Figure 3: *Averaged reaction time for five vowels. Horizontal axis shows duration of the stimuli, and vertical axis shows averaged reaction time of all subjects.*

RT data were log-transformed and analyzed by linear mixed modeling (LMM) with participants as a random factor and vowel types and vowel durations as fixed variables. The result revealed that only vowel duration was the significant factor ($z = 16.10$, $p < .001$). Along with the stimulus duration, the participants' RTs were delayed sharply from 150 ms to 230 ms. The data clearly reflects participants' decreasing confidence to judge the sound as one-mora. On the other hand, there was a sharp drop in the RT means from 270 ms, thereby indicating that the participants' judgment of the words as two morae become automatic as the word lengthened. The RTs for the stimuli that were longer than 310 ms do not seem to be associated with the stimuli duration. This implies that the sound might be too long be perceived as two morae and their judgments are not as automatized as those of 150 ms and 170 ms.

3. Experiment 2

The results of experiment 1 show a clear boundary between one and two morae, regardless of the vowel types. In addition, the 50 % threshold lies between 250 ms and 270 ms. Although this result seems clear, hearing sensitivity differs between individuals, and this may affect the participants' onset and offset vowel perception. In order to minimize the effect of this individual difference, another experiment was conducted using the closed monosyllabic word /tan/ as the stimulus.

3.1. Participant

This experimental study comprised 17 Japanese native speakers (3 female and 14 male), between 19 and 21 years old. Most of them were speakers of the Kinki dialect, whereas a few of them were not. Similar to experiment one, no difference was observed between their responses; therefore, all of them were treated as members of a single group.

3.2. Materials

The closed monosyllabic word /tan/ was chosen as the stimulus. The reason for /t/ to be chosen as an onset is its phonetically clear beginning. The nasal /n/ was chosen because it is the only consonant that allows a closed syllable in Japanese phonology, and it is categorized as a special mora that on its own constitutes an independent mora. This word was pronounced by a 41-year-old male native speaker of Japanese of the Kinki dialect. Similar to experiment one, vowel /a/ was manipulated into 15 durations with 20 ms step to make /ta/ mora from 150 ms to 430 ms.

3.3. Procedure

The experiment was conducted with exactly the same procedure as experiment 1. Each of the 15 stimuli was presented randomly 10 times with handedness counter-balanced. Furthermore, the participants were required to judge each sound as "タン" (/tan/) or "ターン" (/ta:n/) and press the button as quickly as possible.

3.4. Result

The rates of judgment were analyzed as the main data, whereas the reaction times were used as supporting data. As for experiment 2, the same approach as experiment 1 was used for the analysis. However, the only data from tan150 to tan290 were analyzed because almost no one-mora response exists between tan310 and tan430. The result showed that mora duration was significant ($z = 18.28$, $p < .001$).

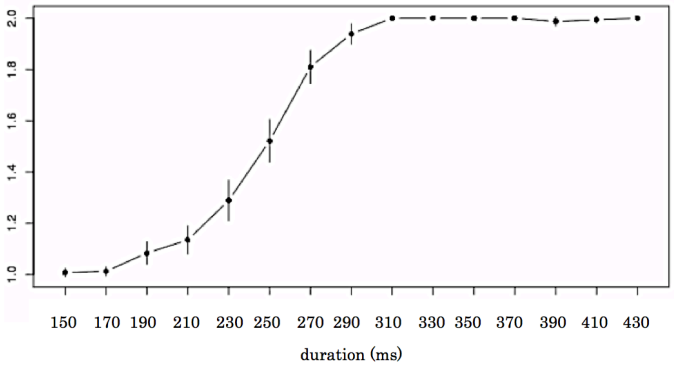

Figure 4: *Judgment data for /ta/ of /tan/. Horizontal axis shows stimulus names, and vertical axis shows averaged judgment (one or two mora) of all subjects.*

Figure 4 summarizes the results of the judgment score. Each number of the stimulus name shows the duration of the first mora /ta/ in each syllable. The line chart in Figure 4

shows almost the same curve as in the result of vowel duration judgment in experiment 1. More than 80% of the responses for tan150 - tan210 are one mora. In addition, the response is sharply shifting to two morae, and over 80% of the judgments were two morae, at tan290. Furthermore, almost no one-mora response exists between tan310 and tan430. Similar to the vowel data in experiment 1, there exists a clear durational boundary between one and two morae around 230 ms and 270 ms.

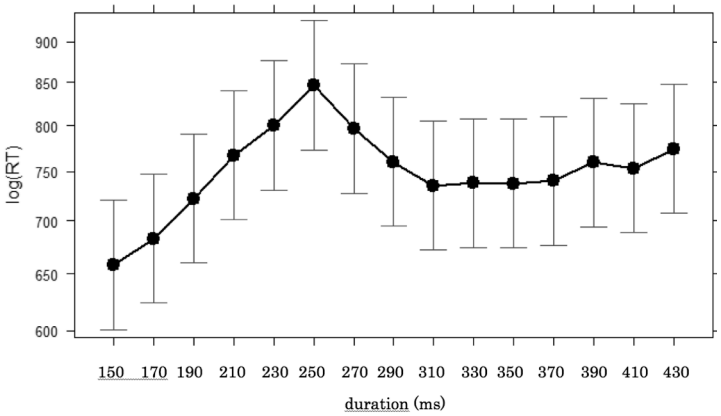

Figure 5: *Reaction time for /tan/. Horizontal axis shows duration of the first mora /ta/, and vertical axis shows averaged reaction time of all subjects.*

RT data were also analyzed with the identical approach with experiment 1. Unlike judgment analysis, all the data from all the durations were used. The results showed that the duration was a significant factor ($t = 6.84$, $p < .001$). Figure 5 shows the reaction time (RT) data for the fifteen variations of the stimulus /tan/. These data include both reaction as one mora and two morae, and should be treated supplementally to support the judgment data. As the duration of the first mora increases to tan250, the RT becomes longer, and from tan250 onwards, their reaction becomes faster. This finding indicates that the participants became increasingly less confident about their judgments of one mora as the mora lengthened to 250 ms. Furthermore, their judgment as two morae became automatic along with the duration until tan310. These RT data clearly support the existence of a durational boundary of mora. That is, a boundary exists between 230 ms and 270 ms, peaking around 250 ms.

4. Discussion

Two research questions were raised at the beginning of this research: (1) does vowel duration alone contribute to the distinction of long/short vowel?; in other words, is mora mentally represented or not?; and (2) if there is a clear mental boundary between one and two morae, where lies the durational threshold between them.

A clear answer to research question (1) is found in the results of both experiments 1 and 2. The judgment data as well as reaction time data show typical features of categorical perception, and it means duration is a key feature for Japanese native speakers to perceive mora numbers.

For research question (2), experiments 1 and 2 show slightly different results. In experiment 1, the threshold between one and two morae lies between 250 ms and 270 ms, but in experiment 2, the threshold lies just around 250 ms. One possible reason for this result is caused by the phonological structure of the stimuli. In experiment 2, the stimulus /tan/ contains the onset consonant /t/, and it lengthens the acceptable duration of one mora slightly longer than a mora with only vowels. Although the reason for this reason is to be analyzed further, it is still reasonable to state that a one and two morae boundary approximately exists between 250 and 270 ms.

5. Conclusions

In order to develop the model of spoken word recognition, it is evidently necessary to discern the perception and recognition of the phonological unit in a larger context, such as sentences. Even so, the result proposed here has the importance of furthering the understanding of bottom up processing of Japanese native speakers.

6. Acknowledgements

This research was partially supported by a Grant-in-Aid for Scientific Research (C) (No. 16K02944) of The Ministry of Education, Science, Sports and Culture, Japan. I am also grateful to Dr. Mitsuhiro Morita (Hiroshima University) for assistance with statistical analysis.

7. Notes

This paper is based on two previous presentations [31, 32].

8. References

[1] I. Maddieson, *Patterns of Sound.* Cambridge: Cambridge University Press, 1984.

[2] O. S., Bohn, "Cross language speech production in adults: First language transfer doesn't tell it all." *Speech perception and linguistic experience: Issues in cross-language research,* edited by in Winnifred Strange, Baltimore: York Press, 1995, pp. 279–304.

[3] A. Abramson, and N. Ren, "Distinctive vowel length: Duration versus spectrum in Thai," *Journal of Phonetics*, vol. 18, pp. 79–92, 1990.

[4] K. Hadding-Koch, and A. Abramson, "Duration versus spectrum in Swedish vowels: Some perceptual experiments," *Studia Linguistics,* vol. XVIII, pp. 94–107, 1964.

[5] I. Lehiste, *Suprasegmentals*. Cambridge, MA: MIT Press, 1970.

[6] S. G. Nooteboom, J. P. L. Brokx, and J. J. Rooij, "Contributions of prosody to speech perception." *Studies in the Perception of Language*, edited by W. J. M. Levelt and G. B. Flores d'Arcais, UK: John Wiley & Sons Ltd, 1978, pp. 75–108.

[7] H. Traunmüller, and D. Krull, "The effect of local speaking rate on the perception of quantity in Estonian," *Phonetica,* vol. 60, pp. 187–207, 2003.

[8] W. A. Van Dommelen, "Does dynamic F0 increase perceived duration? New light on an old issue," *Journal of Phonetics,* vol. 21, pp. 367–386, 1993.

[9] M. E. Beckman, *Stress and Non-Stress Accent.* Dordrecht: Foris, 1986.

[10] B. Bloch, "Studies in Colloquial Japanese IV: Phonemics," *Language,* vol. 26*, no. 1*, pp. 86–125, 1950.

[11] S. Hattori, "Genshi nihongo no akusento [Accent of modern Japnese]," *Kokugo Akusento Ronsho*, pp. 43-65. Hosei University Publishing, 1951.

[12] H. Jōo, “Gendai nihongo no on'in [Phonology of modern Japanese],” *Iwanami Kōza Nihongo,* vol. 5, pp. 107–145. Tokyo: Iwanami Shoten, 1977.
[13] T. Kamei, “‘On’in’ no gainen wa yuuyou nariya [How meaningrul the notion of phonology is.],” *Kokubungakusei,* vol. 15*,* pp. 1–11, 1956.
[14] N. S. Trubetzkoy, *Principles of Phonology* (Translated by C. A. M. Baltaye). Berkeley: University of California Press, 1969.
[15] J. D. McCawley, *The Phonological Component of a Grammar of Japanese*. The Hague: Mouton, 1968.
[16] M. S. Han, “Acoustic manifestations of mora timing in Japanese,” *The Journal of the Acoustical Society of America*, vol. 96, no. 1, pp. 73–82, 1994.
[17] R. F. Port, J. Dalby, and M. O’Dell, “Evidence for mora timing in Japanese,” *The Journal of the Acoustical Society of America*, vol. 81, no. 5, pp. 1574–1585, 1987.
[18] M. Sugito, *Onsei Hakei wa Kataru [Speech waveform talks]*. Osaka, Japan: Izumi Shoin, 2000.
[19] W. N. Campbell and Y. Sagisaka, “Moraic and syllable-level effects on speech timing,” *Acoustical Society of Japan, Trans. Tech. Committee Speech SP90-107*, pp. 35–40.
[20] A. Cutler, and T. Otake, “Mora or phoneme? Further evidence for language-specific listening,” *Journal of Memory and Language*, vol. 844, pp. 824–844, 1994.
[21] H. Fujisaki, and M. Sugito, “Onsei no butsuriteki seishitsu [Physical feature of the speech sound],” *Iwanami Koza Nihongo 5: On’in*, pp. 63–106. Tokyo: Iwanami Shoten, 1977.
[22] Y. Hirata, “Effects of speaking rate on the vowel length distinction in Japanese,” *Journal of Phonetics*, vol. 32, no. 4, pp. 565–589, 2004.
[23] Y. Hirata, and K. Tsukada, “Effects of speaking rate and vowel length on formant frequency displacement in Japanese,” *Phonetica*, vol. 66, no. 3, pp. 129–149, 2009.
[24] K. Kinoshita, D. M. Behne, and T. Arai, “Duration and F0 as perceptual cues to Japanese vowel quantity” In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)*, no. 1, 2002, pp. 757–760.
[25] Y. Minagawa-Kawai, K. Mori, & Y. Sato, “Different brain strategies underlie the categorical perception of foreign and native phonemes,” *Journal of Cognitive Neuroscience*, vol. 17, pp. 1376–1385, 2005.
[26] Y. Nagano-Madsen, “Influence of fundamental frequency pattern on the perception of the vowel mora in Japanese,” in *PHONUM, Report from the Department of Phonetics*.
[27] T. Uchida, “Nihongo tokushuhaku no shinritekina ninchikatei kara toraeta onsetsu to haku: Teijouteki onsei kukan no jizokujikan ni taisuru kategorii teki chikaku [Categorical perception of relatively steady-static speech sound duration in Japanese moraic phonemes],” *Journal of the Phonetic Society of Japan* vol. 2, no. 3, pp. 71–86, 1998.
[28] C. Hoequist Jr, “Syllable duration in stress-, syllable-and mora-timed languages,” *Phonetica*, vol. 40, no. 3, pp. 203–237, 1983.
[29] D. H. Klatt, “Interaction between two factors that influence vowel duration,” *The Journal of the Acoustical Society of America*, vol. 54, no. 4, pp. 1102–1104, 1973.
[30] A. Mizumoto, and O. Takeuchi, “Kenkyuu ronbun ni okeru koukaryou houkoku no tameni: kihon gainen to chuui ten [Basics and considerations for reporting effect sizes in research papers],” *Studies in English Language Teaching*, vol. 31, pp. 57–66, 2008.
[31] K. Sugai, “Nihongo mora no shinriteki jituzai saikou: boin no tandokuteiji niyoru tikaku jikkenn [A re-consideration on the mental representation of Japanese mora],” *Proceedings of the auditory research meeting*, vol. 45, no. 2, pp. 89–91, 2015.
[32] K. Sugai, “Nihongo mora no jizoku jikancho: tan’onsetugo teiji niyoru tikaku jikkenn [Duration of Japanese mora: perceptual experiment with one syllable stimulus],” in *JSLS 2015 Conference Handbook,* pp. 48-51.