Perception of the Kelantan Malay Word-Initial Singleton/Geminate Contrast by Vietnamese Speakers

Mohd Hilmi Hamzah[1], *Kimiko Tsukada*[2,3], *John Hajek*[3], *Đích Mục Đào*[4]

[1]Universiti Utara Malaysia, Malaysia, [2]Macquarie University, Australia, [3]The University of Melbourne, Australia, [4]University of Social Sciences and Humanities, Vietnam National University - Ho Chi Minh City, Vietnam

hilmihamzah@uum.edu.my, kimiko.tsukada@gmail.com, j.hajek@unimelb.edu.au, dich.daovns@gmail.com

Abstract

This study investigates the perception of the Kelantan Malay (KM) word-initial consonant length contrast among Vietnamese speakers who are naïve to such a contrast. One Vietnamese group and a native KM control group participated in the AXB discrimination experiment. As expected, KM listeners outperformed Vietnamese listeners in discriminating KM consonant length, with the overall discrimination accuracy being 88% and 53% for the KM and Vietnamese groups, respectively. While there was a clear between-group difference in discrimination accuracy, there was also greater variability among the KM group, suggesting genuine difficulty in successfully identifying word-initial geminates, particularly those beginning with voiceless stops.

Index Terms: speech perception, consonant length contrast, singleton, geminate, Kelantan Malay, Vietnamese

1. Introduction

In KM, length (i.e., short vs long) is lexically contrastive across all consonants, as empirically established in [1-4]. For example, /kabo/ 'beetle' contrasts word-initially with /kkabo/ 'blurry'. It has been reported that closure duration is the most robust acoustic correlate of the singleton/geminate contrast in KM, with non-durational parameters such as root mean square (RMS) amplitude and fundamental frequency (F0) playing important secondary roles in enhancing the word-initial length contrast in this language. Consonant length in KM is claimed to be cross-linguistically rare and more marked as it only occurs in word-initial position [e.g., 5]. It is well known that the word-initial length contrast is more perceptually indiscernible, particularly for utterance-initial geminates involving voiceless stops, such as /kk/ in /kkabo/. In this particular utterance position, there is ostensibly insufficient acoustic information available for listeners to discern the length contrast [e.g., 6].

Unlike KM, consonant length is not contrastive in Vietnamese [e.g., 7-9]. Vietnamese speakers, however, are familiar with contrastive duration of vowels that is employed to distinguish some tense-lax vowels in the Vietnamese language [10]. It is thus of theoretical interest to examine whether Vietnamese listeners, whose L1 uses duration differently, might adopt varied perceptual strategies when faced with tasks involving consonant length contrasts in a non-native language, such as KM in this study. This is evident in a related study on cross-language perception of phonological length contrasts by [11] who compared German and Italian sounds. It was found that German speakers, who have experience with vowel length contrasts in their L1, were less accurate in distinguishing consonant length in Italian. Conversely, Italian speakers were as proficient as German speakers in discerning vowel length in German.

With regard to Vietnamese speakers, it has been reported in [12] that it was problematic for Vietnamese listeners to perceive word-medial consonant length in Japanese. This was supported by [7] and [13] who found that Vietnamese speakers had a tendency to misidentify singletons and geminates in Japanese. As for KM speakers, who are familiar with consonant length in their L1, it was reported in [14] that they are able to perceive Japanese consonant length at a high level of discrimination accuracy. Similarly, speakers from other languages with L1 contrastive length, such as Korean and Mongolian speakers [15], are also able to perceive Japanese consonant length above chance level. However, those who lack experience in consonant length are less accurate in their perception of Japanese consonant length, such as American English [16] and Mandarin [15] speakers.

In this study, the perception of KM consonant length by native and non-native speakers, i.e., Vietnamese speakers, was compared to examine the extent to which word-initial consonant gemination in KM is processed accurately by speakers of other languages who are unfamiliar with consonant length contrasts. We hypothesise that Vietnamese speakers may have difficulties in perceiving the unfamiliar, rare linguistic event of word-initial consonant gemination in KM. Given the complexities of the word-initial length contrast, we also expect potential variation in discrimination accuracies among KM native listeners due to effects of voicing and manner of articulation. Our findings will provide some theoretical insights into this potential cross-linguistic phonetic transfer and add to our current knowledge about the perception of difficult sounds such as word-initial consonant gemination.

2. Methods

2.1. Stimuli preparation

2.1.1. Speakers and procedures

The experimental stimuli were obtained from previous production experiments in KM [3, 4] involving sixteen native speakers (8 males, 8 females). Six of them were students from several universities in Melbourne, Australia, and ten were students from Universiti Malaysia Kelantan located in the state of Kelantan, Malaysia. Their age ranged from 20 to 28 (mean age: 22.4). All were born and raised in Kelantan, Malaysia. For the speakers in Melbourne, the experimental materials were

recorded individually on the main campus of the University of Melbourne. As for the speakers in Kelantan, they were recorded individually in a quiet room at Universiti Malaysia Kelantan.

In all sessions, speakers were asked to repeat each token in isolation and in a carrier sentence. The carrier sentence was: /diɔ katɔ (the target word) tigɔ kali/ "he said (the target word) three times", adapted from [17]. All experimental tokens were presented in randomised order using a powerpoint presentation on a computer. The experiment took approximately one and a half hours for each speaker. They were compensated financially for their participation in the experiment.

In the present study, the production data collected in Kelantan, Malaysia were selected as experimental stimuli, which involved six native speakers of KM (3 males, 3 females). Only tokens produced in isolation, i.e., utterance-initial position, were used.

2.1.2. Speech materials

Table 1. *Nineteen pairs of Kelantan Malay words with target sounds underlined and bolded.*

Phoneme pair	Singleton		Geminate	
	Word	Gloss	Word	Gloss
/p/–/pp/	/**p**itu/	door	/**pp**itu/	at the door
	/**p**agi/	morning	/**pp**agi/	early morning
/t/–/tt/	/**t**ido/	sleep	/**tt**ido/	sleep by chance
	/**t**anɔh/	land	/**tt**anɔh/	outside
/k/–/kk/	/**k**iɣi/	left	/**kk**iɣi/	to the left
	/**k**abo/	blurry	/**kk**abo/	beetle
/b/–/bb/	/**b**ini/	wife	/**bb**ini/	married
	/**b**atʃɔ/	read	/**bb**atʃɔ/	is reading
/d/–/dd/	/**d**ike/	song	/**dd**ike/	sing a song
	/**d**apo/	kitchen	/**dd**apo/	at the kitchen
/g/–/gg/	/**g**igi/	teeth	/**gg**igi/	on the teeth
	/**g**adʒi/	salary	/**gg**adʒi/	sawing tool
/m/–/mm/	/**m**isa/	moustache	/**mm**isa/	moustached
	/**m**aɣi/	come	/**mm**aɣi/	cupboard
/n/–/nn/	/**n**ikɔh/	marriage	/**nn**ikɔh/	married
	/**n**anɔh/	pus	/**nn**anɔh/	getting pus
/ŋ/–/ŋŋ/	/**ŋ**aŋɔ/	open the mouth	/**ŋŋ**aŋɔ/	agape
/l/–/ll/	/**l**idɔh/	tongue	/**ll**idɔh/	on the tongue
	/**l**apu/	lights	/**ll**apu/	on the lights

Table 1 shows nineteen KM word pairs used in this study. All tokens were disyllabic words with either C(C)VCV or C(C)VCVC structures. They contained singletons (n = 114) or geminates (n = 114) in word-initial position (underlined and bolded). These phonemes were grouped according to consonant type: voiceless stops (/p/–/pp/, /t/–/tt/, /k/–/kk/); voiced stops (/b/–/bb/, /d/–/dd/, /g/–/gg/); and sonorants consisting of nasals (/m/–/mm/, /n/–/nn/, /ŋ/–/ŋŋ/) and liquids (/l/–/ll/). Each phoneme was followed by two distinct vowels: the high front vowel /i/ and the low central vowel /a/, except /ŋ/–/ŋŋ/ (low central vowel /a/ only).

On average, the VOT durations for voiceless stops were 29 ms and 18 ms for singletons and geminates, respectively. As for voiced stops, the closure durations were 60 ms for singletons and 152 ms for geminates (geminate-to-singleton ratio: 2.51). For sonorants, the closure durations were 58 ms and 158 ms for singletons and geminates, respectively (geminate-to-singleton ratio: 2.69). These durational values are in good agreement with what has been reported in previous research [e.g., 1, 3].

2.2. Participants

Two groups of young adults participated in an AXB discrimination task. The first group consisted of 24 (10 males, 14 females) native speakers of Vietnamese (mean age = 27.1). They were initially divided into two sub-groups, with the first one involving either academics or students at Vietnam National University, Ho Chi Minh City, Vietnam, while the second one involved general workers outside the university setting. Given their similar results (see below), we decided to merge them together. All of them were born and raised in Vietnam and are fluent in Vietnamese. All were naïve to KM.

The second and a control group consisted of 12 (6 males, 6 females) native speakers of KM (mean age = 39.5). They were either academic or non-academic staff members at Universiti Utara Malaysia in Kedah, Malaysia. All KM participants were born and spent the majority of their life in Kelantan, Malaysia. None of them participated in the recording sessions. According to self-report, all had normal hearing at the time of the experimental sessions.

2.3. Procedure

For the Vietnamese participants, they were tested individually in a quiet room at their workplace or on the university campus of Vietnam National University. As for the KM participants, the experiments were conducted in a quiet room on the main campus of Universiti Utara Malaysia in Sintok.

In all sessions, the experiment was self-paced and lasted approximately 15 to 20 minutes. The participants heard the stimuli at a self-selected, comfortable amplitude level over the high-quality speakers on a notebook computer. They completed a two-alternative forced-choice AXB discrimination task, in which they were asked to listen to trials arranged in a triad (A-X-B). The presentation of the stimuli and the collection of perception data were controlled by the PRAAT program [18]. In the AXB task, the first (A) and third (B) tokens always came from different length categories, and the participants had to decide whether the second token (X) belonged to the same category as A (e.g., 'pitu$_2$'-'pitu$_1$'-'ppitu$_3$') or B (e.g., 'kabo$_3$'-'kkabo$_1$'-'kkabo$_2$'; where the subscripts indicate different speakers).

The participants listened to a total of 160 unique trials. The first eight trials were for practice and were not analysed. The three tokens in all trials were spoken by three different speakers. Thus, X was never acoustically identical to either A or B. This was to ensure that the participants focused on relevant phonetic characteristics that grouped two tokens as members of the same length category without being distracted by audible but phonetically irrelevant within-category variation (e.g., in voice quality). This was considered a reasonable measure of participants' perceptual capabilities in real world situations [19]. All possible AB combinations (i.e., AAB, ABB, BAA, and BBA, 38 trials each) were tested.

The participants were given two ('A', 'B') response choices on the computer screen. They were asked to select the option 'A' if they thought that the first two tokens in the AXB sequence were the same and to select the option 'B' if they thought that the last two tokens were the same. No feedback was provided during the experimental sessions. The participants could take a break after every 40 trials if they wished. The participants were required to respond to each trial, and they were told to guess if uncertain. A trial could be replayed as many times as the participants wished in order to reduce their anxiety, but responses could not be changed once given. The interstimulus interval in all trials was 0.5 s.

3. Results

We used R version 4.4.0 for statistical analyses and data visualisation reported below [20]. The packages used include ez [21] and tidyverse [22].

3.1. Overall results

Figure 1 shows the distributions of percentages of correct discrimination by the two groups of participants. The overall mean discrimination accuracy was clearly higher for the native KM control group (88%) than the Vietnamese group (53%). Note that the accuracy score for the Vietnamese group was just above the 50% chance-level. A comparison via the Welch two-sample *t*-test showed that the difference between the KM and Vietnamese groups was significant [$t(15.3) = 34$, $p < .001$]. Nonetheless, as seen in Figure 1, there is greater individual variation in the discrimination accuracies among the KM participants (sd=0.1) as compared to the Vietnamese participants (sd=0.04). The highest discrimination accuracy for the KM group was 97%, while the lowest was 65%, indicating that the discrimination task was somewhat challenging for some native speakers. As for the Vietnamese group, the discrimination accuracy ranged from 47% to 63%, with six of the 24 scoring below the 50% chance-level.

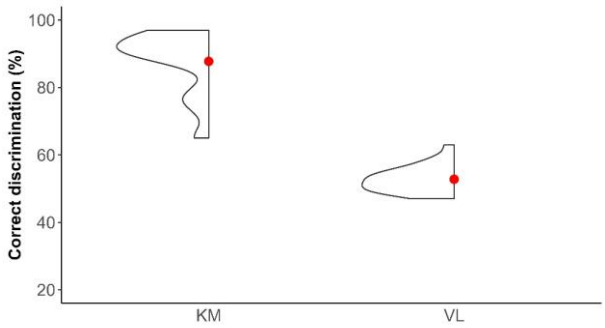

Figure 1: *Accuracy (%) of length discrimination by two groups of participants (KM: Kelantan Malay; VL: Vietnamese). The red circle indicates the mean.*

3.2. Comparison of the length category (Geminate vs Singleton) of the target token (X in AXB)

Figure 2 shows the distributions of percentages of correct discrimination for trials differing in the length category (geminate, singleton) of the target token. The question of interest was if the participants' discrimination accuracy differed between trials in which X in AXB was a geminate and trials in which X in AXB was a singleton. Two-way analysis of variance (ANOVA) results with group (KM, Vietnamese) and length category (geminate, singleton) reached significance only for the main effect of group [$F(1, 68) = 391.7$, $p < .001$, $\eta^2 = .85$]. That is, the KM group was significantly more accurate than the Vietnamese group whether X in the AXB sequence was singleton or geminate. Length category did not yield any significant main effect on discrimination accuracies. Two-way interaction also yielded a non-significance result. As seen in Figure 2, neither group was biased with respect to the length category of the target token, though the Vietnamese group shows a higher discrimination accuracy for geminates (55%) than for singletons (51%). As for the KM group, the discrimination accuracies for geminates and singletons were both similar (88% for both geminates and singletons).

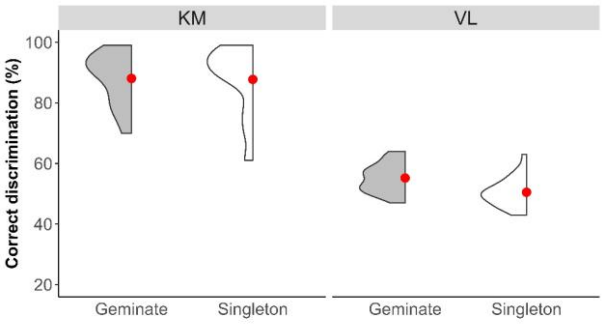

Figure 2: *Accuracy (%) of length discrimination for trials differing in the length category of the target token.*

3.3. Comparison of the consonant type (Voiceless stop vs Voiced stop vs Sonorant) of the target token (X in AXB)

Figure 3 shows the distributions of percentages of correct discrimination for trials differing in consonant type (voiceless stop, voiced stop, sonorant) of the target token. The question of interest was to determine if the participants' discrimination accuracy differed between trials in which X in AXB was a voiceless stop, a voiced stop, or a sonorant. Two-way ANOVA results with group (KM, Vietnamese) and consonant type (voiceless stop, voiced stop, sonorant) reached significance for the main effects of group [$F(1, 354) = 658.3$, $p < .001$, $\eta^2 = .65$] and also consonant type [$F(2, 354) = 3.7$, $p < .05$, $\eta^2 = .02$]. Two-way interaction yielded a non-significance result.

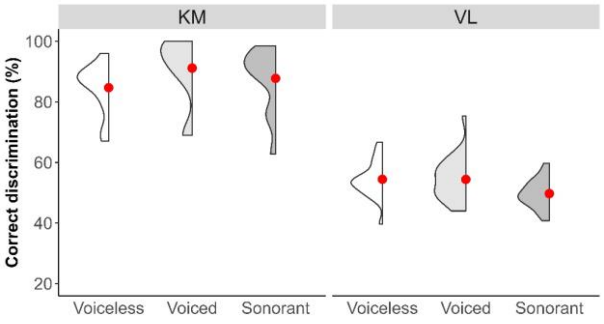

Figure 3: *Accuracy (%) of length discrimination for trials differing in the consonant type of the target token.*

As observed in Figure 3, both groups of participants appeared to be biased with respect to the consonant type of the target token. Note that the order of discrimination accuracy (from the lowest to the highest) was different between the two groups. On one hand, for the KM group, voiceless stops received the lowest discrimination accuracy (85%), while voiced stops were best discriminated (91%). Sonorants were intermediate between the other consonant types (88%). On the other hand, for the Vietnamese group, sonorants were discriminated with the lowest discrimination accuracy (50%) as compared to voiceless stops and voiced stops (both 54%). The comparison of phonemes within these consonant types is described in the following section.

3.4. Comparison of the phonemes of the target token (X in AXB)

Figures 4 and 5 below show the distributions of percentages of correct discrimination for trials differing in the phonemes of the target token for the KM and Vietnamese groups, respectively. The phonemes in these figures are arranged from the lowest (left) to the highest (right) discrimination accuracies. As seen in Figure 4 for the KM group, the voiceless alveolar stop pair (/t/-/tt/) was discriminated with the lowest discrimination accuracy (83%), while the bilabial nasal pair (/m/-/mm/) received the highest discrimination accuracy (93%), followed by the voiced bilabial stop pair (/b/-/bb/; 92%). As for the Vietnamese group (Figure 5), the alveolar nasal pair (/n/-/nn/) was discriminated with the lowest discrimination accuracy (45%), while the voiced bilabial stop pair (/b/-/bb/) received the highest discrimination accuracy (62%), followed by the voiceless bilabial stop pair (/p/-/pp/; 58%).

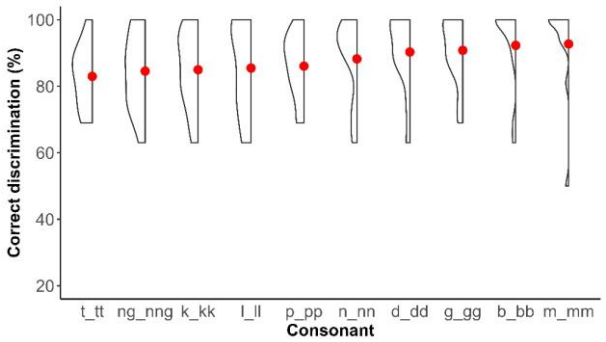

Figure 4: *Accuracy (%) of length discrimination for trials differing in the phoneme type of the target token (KM).*

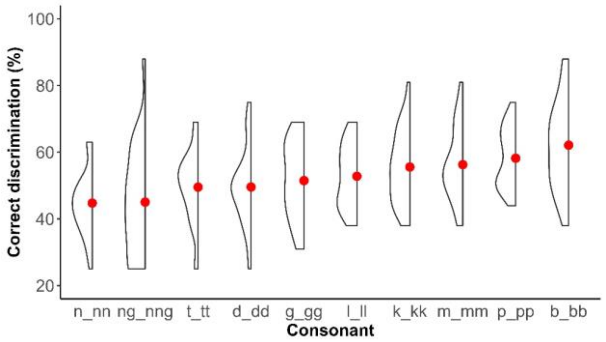

Figure 5: *Accuracy (%) of length discrimination for trials differing in the phoneme type of the target token (Vietnamese).*

4. Discussion

In this study, we examined how Vietnamese speakers may perceive the KM word-initial length contrast that is potentially challenging for non-native speakers, given the cross-linguistically marked status of this particular contrast [e.g., 5]. Consonant length is not contrastive in the Vietnamese language, though duration is employed to distinguish the tense-lax vowel distinction in this language [10]. Thus, we were interested in determining if the Vietnamese speakers, who are naïve to consonant length, are able to perceive the unfamiliar singleton/geminate contrast in word-initial position.

The findings support our earlier hypothesis: the Vietnamese speakers did face considerable difficulties in perceiving the rare word-initial consonant length contrast in KM, irrespective of the length category (singleton or geminate), with all speakers as a group discriminating the trials just above chance level (overall discrimination accuracy=53%). Surprisingly, some native speakers of KM in our study were also found to face difficulties in perceiving the length contrast in their own language, confirming the genuine difficulties of the word-initial length contrast. Our observations on Vietnamese speakers lend some evidence to the prediction by [11] who claimed that cross-language perception of phonological length contrasts would be more successful for speakers who have L1 experience with *consonant* length contrast rather than with *vowel* length contrast. Our data also accord well with the previous findings involving Vietnamese speakers such as [7], [12] and [13]. In this regard, the Vietnamese speakers in our study are comparable to American English [16] and Mandarin [15] speakers who are also less accurate in perceiving Japanese consonant length.

Pertaining to the significant effect of consonant type, the results support our earlier prediction, though different groups faced different challenges in this regard. For the KM group, the discrimination of trials beginning with voiceless stops was relatively poor, which is expected given the absence of acoustic information for listeners to discern the length contrast in this specific utterance context [e.g., 6]. With regard to the Vietnamese group, it is striking to note that trials beginning with voiceless stop pairs (e.g., /p/-/pp/) were relatively better discriminated than those beginning with other consonant types, such as nasals (see Figure 5). It can be speculated that the Vietnamese speakers in our study might have employed a unique perceptual strategy when dealing with a discrimination task involving the word-initial length contrast in KM.

5. Conclusions

The findings of the present study have shown that the speakers of Vietnamese did not match the KM native speakers in discriminating the KM singleton/geminate contrast word-initially, which is clearly due to lack of specific knowledge of the phonetic characteristics of KM singletons and geminates. Both groups were affected by the consonant type factor and discriminated KM consonant length more accurately when a specific consonant type occurred in a target position (e.g., bilabial nasals for KM participants). Our results support earlier research that experience with consonant length contrast in L1 may be helpful in processing word-initial consonant gemination.

In the future, it is obvious that there is a need to examine the perception of KM consonant length among speakers of other languages who use consonant length contrastively, such as Japanese and Italian. Further, given the complexities of word-initial consonant gemination, it would be valuable to include speakers who have experience with this particular linguistic event, such as Tashlhiyt Berber speakers, and examine if there is additional benefit of familiarity with a length contrast that occurs in word-initial position.

6. Acknowledgements

We thank Universiti Utara Malaysia (UUM) and University of Social Sciences and Humanities, Vietnam National University for their generous assistance and for making the data collection possible. We thank reviewers for their time and input.

7. References

[1] Hamzah, M. H., Fletcher, J. and Hajek, J., "Durational correlates of word-initial voiceless geminate stops: The case of Kelantan Malay", Proceedings of the 17th International Congress of Phonetic Sciences: 815-818, 2011.

[2] Hamzah, M. H., Fletcher, J. and Hajek, J., "Word-initial voiceless stop geminates in Kelantan Malay: Acoustic evidence from amplitude/F0 ratios", Proceedings of the 18th International Congress of Phonetic Sciences, 2015.

[3] Hamzah, M. H., Fletcher, J. and Hajek, J., "Closure duration as an acoustic correlate of the word-initial singleton/geminate consonant contrast in Kelantan Malay", Journal of Phonetics, 58: 135-151, 2016.

[4] Hamzah, M. H., Fletcher, J. and Hajek, J., "Non-durational acoustic correlates of word-initial consonant gemination in Kelantan Malay: The potential roles of amplitude and f0", Journal of the International Phonetic Association, 50(1): 23-60, 2020.

[5] Ridouane, R., "Geminates at the junction of phonetics and phonology", in C. Fougeron, B. Kühnert, M. D'Imperio and N. Vallée (Eds), Laboratory Phonology 10, 61-90, De Gruyter Mouton, 2010.

[6] Ridouane, R. and Hallé, P. A., "Word-initial geminates: From production to perception", in H. Kubozono (Ed), The Phonetics and Phonology of Geminate Consonants, 34-65, Oxford University Press, 2017.

[7] Đỗ, H. N., "Issues regarding Vietnamese learners' perception of long vowels and geminate consonants in Japanese (in Japanese)", Journal of Science, Foreign Languages, 31: 31-38, 2015.

[8] Kanamura, K., "Prosodic features in Japanese speech by native Vietnamese speakers (in Japanese)", Studia Linguistica, 12: 73-91, 1999.

[9] Pajak, B. and Levy, R., "The role of abstraction in non-native speech perception", Journal of Phonetics, 46: 147-160, 2014.

[10] Emerich, G. H., "The Vietnamese vowel system", PhD thesis, University of Pennsylvania, US, 2012.

[11] Altmann, H., Berger, I. and Braun, B., "Asymmetries in the perception of non-native consonantal and vocalic length contrasts", Second Language Research, 28: 387-413, 2012.

[12] Sugimoto, T., "Error analysis of Japanese pronunciation by Vietnamese learners III: On the geminate consonant and the syllabic nasal", Departmental Bulletin of College of Humanities, Ibaraki University, 2: 149-164, 2007.

[13] Đỗ, H. N., "The influence of position and accent on the perception of long vowels and geminate consonants: Vietnamese learners majoring in Japanese (in Japanese)", Journal of Science, Foreign Languages, 28: 242-254, 2012.

[14] Hamzah, M. H., Tsukada, K. and Hajek, J., "Perception of the Japanese word-medial singleton/geminate contrast by Kelantan Malay speakers", in R. Skarnitzl and J. Volín (Eds), Proceedings of the 20th International Congress of Phonetic Sciences: 162-166, Guarant International, 2023.

[15] Tsukada, K., Yurong, Kim, J.-Y., Han, J.-I. and Hajek, J., "Cross-linguistic perception of the Japanese singleton/geminate contrast: Korean, Mandarin and Mongolian compared", in H. Hermansky, H. Cernocký, L. Burget, L. Lamel, O. Scharenborg and P. Motlícek (Eds), Proceedings of the 22nd Annual Conference of the International Speech Communication Association (Interspeech): 3910-3914, International Speech Communication Association, 2021.

[16] Tsukada, K., Idemaru, K. and Hajek, J., "The effects of foreign language learning on the perception of Japanese consonant length contrasts", in J. Epps, J. Wolfe, J. Smith and C. Jones (Eds), Proceedings of the 17th Australasian International Conference on Speech Science and Technology: 37-40, Australian Speech Science and Technology Association, 2018.

[17] Abramson, A. S., "The perception of word-initial consonant length: Pattani Malay", Journal of the International Phonetic Association, 16: 8-16, 1986.

[18] Boersma, P. and Weenink, D. Praat: Doing Phonetics by Computer [version 6.4.06], retrieved from http://www.praat.org (Last viewed May 4, 2024).

[19] Strange, W. and Shafer, V. L., "Speech perception in second language learners: The re-education of selective perception", in J. G. Hansen Edwards and M. L. Zampini (Eds), Phonology and Second Language Acquisition, 153-191, John Benjamins, 2008.

[20] R Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, 2019. https://www.Rproject.org/

[21] Lawrence, M. A., "ez: Easy analysis and visualization of factorial experiments", R Package Version 4.4.0., 2016. https://CRAN.Rproject.org/package=ez

[22] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R. et al., "Welcome to the tidyverse", Journal of Open Source Software, 4(43): 1686, 2019.