Effects of Pitch Cues on the Identification of Vowel Length in L2 Japanese

Izumi Takiguchi

Graduate School of Foreign Studies, Sophia University, Japan/ JSPS Research Fellow
i_takiguchi[AT]sophia.ac.jp

Abstract

This study explored the effects of pitch cues on the identification of the word-final Japanese vowel length, which is primarily cued by vowel duration. Native speakers of English (NE), Chinese (NC) and Japanese (NJ) participated in the experiment. Learners, who do not use duration distinctively in their L1, utilize duration as a cue for the contrast and they can approximate boundary location to NJ's. In addition, pitch cues did not affect NE's perception but it did affect NC's identification. These results suggest that the role of cues in learners' L1 relates to the use of cues in their L2.

Index Terms: L2 speech perception, Japanese, vowel length

1. Introduction

What differences are there between the perceptions of native speakers (NS) and non-native speakers (NNS) when perceiving the contrasts? They possibly use different cues and/or give different weight to cues in categorizing sounds ([1] and [2]). Munro and Bohn [3] point out that second language (L2) learners tend to focus on the irrelevant cues and they state that the perceptual development can be understood as native-like use of perceptual cues in the L2. In line with these discussions, Guion and Pederson [4] claim that learners can discriminate non-native contrasts if they adopt the appropriate weighting of acoustic cues for the target language (TL).

One question here is whether L2 learners are able to utilize TL cues that do not exist in their L1. According to the feature (prominence) hypothesis ([5] and [6]), L2 features/cues which are not distinctive/prominent in learners' L1 are expected to be difficult to employ and attend to; therefore, L2 learners fail to use those cues appropriately in the production and perception of the L2. The desensitization hypothesis by Bohn [7], on the other hand, proposes that L2 listeners use durational cues if they are desensitized to a distinctive spectral difference of L2 vowels which also exhibit a durational difference. Both hypotheses suggest that L2 cues are accessible to all L2 learners to some extent but it is unclear whether the utilization of cues in learners' L2 may or may not relate to the role of the cues in their L1.

This study investigated the effects of pitch cues on NS' and NNS' perception of the word-final vowel length in Japanese, which is primarily cued by duration ([8]). Specifically, it explored the effect of pitch cues on the identification of vowel quantity by native speakers of English (NE) and Chinese (NC). Native speakers of Japanese (NJ) also participated in the experiment as a control group; therefore, the comparison of the data will reveal the differences and similarities in the utilization of durational and pitch cues between NS and NNS. Furthermore, English and Chinese are different in the roles of pitch and duration in each language. The current study explored how this difference functions in L2 perception.

The present study shows the following points. First, listeners were more likely to identify a target vowel as a long vowel when the vowel was longer in duration, and the boundary values exhibited by learners were similar to that of NJ's in general. Second, effects of pitch cues on the perception of vowel length varied with learners' L1, NE's perception was not affected by pitch cues but NC's perception was influenced. These results suggest that NNS who do not use duration distinctively in their L1 can utilize it as a cue for length in the L2, it is possible for them to approximate boundary location to NS's and the role of cues in NNS's L1 relates to the use of cues in their L2 perception.

2. Backgrounds

2.1. Linguistic background

The roles of duration and pitch vary in Japanese, English and Chinese. Japanese exhibits length contrasts and a lexical pitch accent, that is, both pitch and durational cues are phonemic in Japanese. Chinese is in common with Japanese and distinct from English in that it uses pitch distinctively. English and Chinese are different from Japanese in that both do not have length contrasts, that is, duration is not phonemic.

Although duration is not phonemic in English and Chinese, both use duration for the realization or perception of some linguistic aspects. In English duration is a secondary cue and acoustic correlate between tense and lax vowels and it is rather prominent ([6]). In addition, duration accompanies stress, and vowel duration differs depending on whether the following consonant is voiced or voiceless. In Chinese, duration accompanies tone, but it does not seem to be crucial ([9]). The durational difference is especially significant between full and neutral tones and can be considered parallel to English lexical stress ([10]).

2.2. Previous studies

Toda [11] analyzed boundary values of word-final vowels using stimuli by manipulating the duration of the first and the second vowel (V1 and V2, respectively) of disyllabic real words. The stimuli were presented to NE in ascending or descending order and listeners indicated when they perceived length change. According to Toda, beginners showed greater values than NJ and advanced learners but the values of advanced learners were similar to those of NJ. She says that NNS can approximate their perception function to NS.

Uchida [12] investigated NJ's and NC's threshold values with stimuli by editing the duration of V1 and V2. The procedure and task was the same as in [11]. He found that beginners had greater values than NJ's when stimuli were presented in ascending order but smaller values when presented in descending order. Based on the findings he states that beginners can perceive length correctly only when the vowel is very short or very long. Advanced learners, on the other hand, showed greater values in descending than in ascending order, which was opposite shown by NJ and beginners. He proposes that advanced learners identify length by their internal standards and do not have a boundary value. He concludes that NC's problem is not to set the boundary but the categorical perception itself.

Kurihara [13] also explored NC's boundary location using stimuli which V2 duration was edited from short to long. She reported that NC's boundary values were smaller than NJ's, which were similar among the learners at three levels. Besides, she showed that NC's judgment was less categorical and stable than NJ's in general, but advanced learners were better in the two aspects than beginners were. From these findings, she claims that L2 learners can establish a boundary and perceive length categorically as learning progresses.

On the one hand, the study by [11] reports that NNS's boundary value can approximate to NS's norm as the learning proceeds; on the other hand, it is said that the boundary value does not change over time ([13]) or NC may not own a boundary ([12]). Therefore, it is still unknown whether NNS have a boundary and how far they can approximate their perception to NS's. Furthermore, none of the above studies has examined the effects of pitch cues on the boundary location for the identification. Therefore, the primary purpose of the current study is to explore whether the boundary values differ depending on the accent patterns.

3. Method

3.1. Participants

NJ, NE, and NC participated in this experiment. No participants reported any hearing problems. There were nine participants in the NJ group, 1 male and 8 females, who speak Tokyo dialect of Japanese. The mean age of the NJ group was 37.0 (19-61). There were 13 participants in the NE group, 8 males and 5 females. The mean age of the NE group was 25.4 (19-41). They have received instruction in Japanese for 35 months on average (7-95 months) and have lived in Japan for 34 months on average (3 weeks-128 months). Finally, the NC group consisted of 12 participants, 1 male and 11 females. The mean age of the NC group was 27.9 (22-41). They have received instruction in Japanese for 55 months on average (12-127 months) and have lived in Japan for 32 months on average (11-141 months). Participants were paid for the participation.

3.2. Sound stimuli

All stimuli were created from a token of a nonsense word /mamama/ with HHH accent pattern produced by a male Japanese native speaker in his twenties, who is a researcher in the field of phonetics and phonology. Table 1 shows the segment duration and F0 of the token.

Table 1. *Duration and F0 of the original token.*

segments	m	a	m	a	m	a
duration (ms)	35.1	100.4	63.6	124.8	76.5	153.1
F0 (Hz)	156.1		156.5		155.2	

He uttered /mamama/ with HHH three times at his normal speaking rate in isolation. They were recorded using a linear PCM recorder (SONY PCM-D1) at a 44.1 kHz sampling rate and 16-bit quantization. A token with the most similar F0 values on the three syllables was selected as the original token for further manipulation.

A total of 44 stimuli (11 durations × 4 pitch patterns) were created by manipulating the final vowel duration and the pitch contour of the original token with Praat ([14]). First, duration of the final vowel was edited by deleting or copying a pitch period around the center of the final vowel at zero-crossings and it ranged from 133 ms to 333 ms in 20 ms steps (11 durations). After manipulating the vowel duration, the pitch contour of each token was edited (HHH→HLL, LHL, LHHL, LHH) by reference to the F0 data collected from the production of the same speaker who uttered the original token (underlined portions ranged from 133 ms to 333 ms). The four patterns are possible in Tokyo dialect of Japanese except a contour tone (falling) was realized on a vowel even when its duration was 133 ms (a falling contour tone can only occur on a long vowel).

For the F0 data for the manipulation (pitch points and the F0 values), the speaker who uttered the original token was asked to pronounce /mamama/ and /mamama:/ with three (i.e. HLL, LHL and LHH) or four accent patterns (i.e. HLLL, LHLL, LHHL and LHHH) in isolation three times. One token for the four patterns (HLL, LHL, LHH and LHHL) was selected respectively from the three utterances. Three out of four were the tokens having a final short vowel but LHHL, which has a final long vowel, was used because Tokyo dialect of Japanese does not have a short vowel with a contour tone. The F0 data on the pitch points and the F0 values were extracted through line-approximation of the pitch contour of each token using Praat.

Six pitch points were found in total as in Figure 1 by the line-approximation. Pitch point (1) and (3) were at the onset of the first and second vowel, respectively, pitch point (2), (4) and (6) were at the offset of each vowel, respectively. Pitch point (5) was not at the onset but was at one pitch period after the onset of the final vowel. The number of pitch points varied by accent pattern and HLL had three, LHL and LHH had four, and LHHL had five pitch points. Table 2 below demonstrates the F0 values of each pitch point for each accent pattern. For the manipulation, the pitch contour of each token was first stylized for line-approximation, pitch points were added or removed and F0 values were set as shown in Table 2.

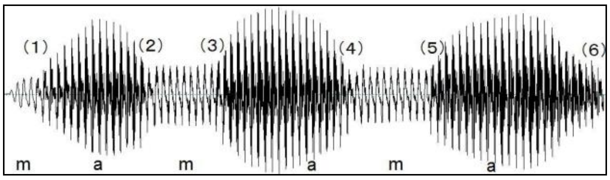

Figure 1: *Pitch points.*

Table 2. *The F0 values at each pitch point (Hz).*

pitch points	(1)	(2)	(3)	(4)	(5)	(6)
HLL	157.4	173.0	n/a	n/a	n/a	96.8
LHL	131.8	124.0	n/a	159.0	n/a	103.5
LHHL	132.0	127.5	152.4	n/a	157.7	99.7
LHH	131.7	129.0	148.4	n/a	n/a	147.0

3.3. Procedure

The participants were tested individually in a quiet room. They identified the final vowel length by clicking the button, "MA" or "MAA" on a computer screen. The responses were automatically recorded. Each stimulus was presented ten times in random order over loudspeakers, and the test consisted of

440 trials (44 stimuli × 10 times) divided into 22 blocks. Participants were allowed to take breaks between blocks. The participants took the practice section, which contained 16 trials (2 endpoints) prior to the test section. (29 minutes + break)

Probit analyses were performed on each listener's identification function to estimate the location of the boundary (50% crossover point). The data on the boundary values were subjected to a two-way repeated-measures ANOVA (between factor: L1, within factor: accent pattern) and if there were any significant differences, Bonferroni correction for multiple comparisons was used as a post hoc test. In order to specify the effects of pitch cues, the pairs in Table 3 were considered. In addition, another set of ANOVAs (between factor: L1) were conducted for each accent pattern to examine in which accent patterns the three groups of listeners had differences if they did.

Table 3. *Pairs for the specification of the pitch effects*.

pitch effects	the pairs in consideration
PITCH	LHL vs. LHH
ACCENT LOCUS	HLL vs. LHL
CONTOUR TONE	LHL vs. LHHL, LHHL vs. LHH
OTHERS	HLL vs. LHHL, HLL vs. LHH

4. Results and discussions

Figure 2, Figure 3 and Figure 4 show percentages of responses as a long vowel (% long) to the stimuli exhibited by NJ, NE and NC, respectively. The leftmost stimulus on X-axis is the shortest sound (133 ms) and the rightmost one is the longest (333 ms). As it is clear from the figures, their % long increased as the vowel duration became longer. This means that listeners in the three groups attend to durational cue to identify length. The figures also show that NJ's and NC's % long varied with accent patterns, but not NE's. This shows that the effects of pitch cues may differ among the participants with different L1 backgrounds.

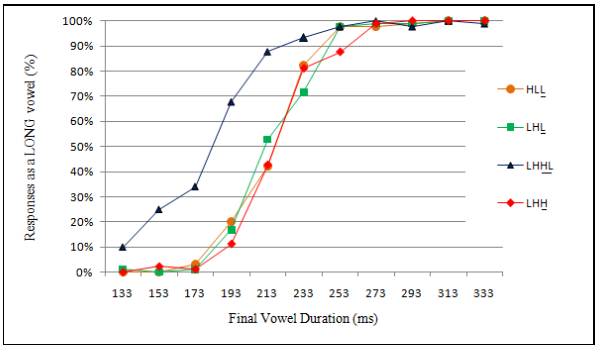

Figure 2: *Percentage of 'long' responses by NJ.*

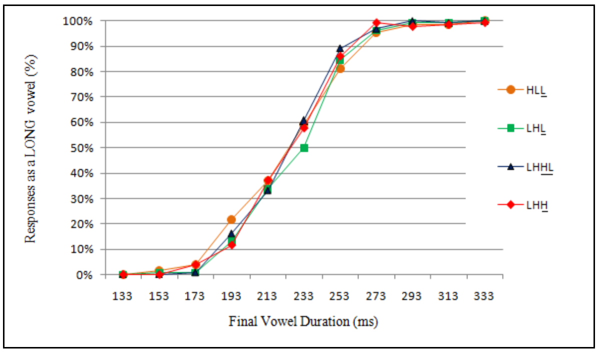

Figure 3: *Percentage of 'long' responses by NE.*

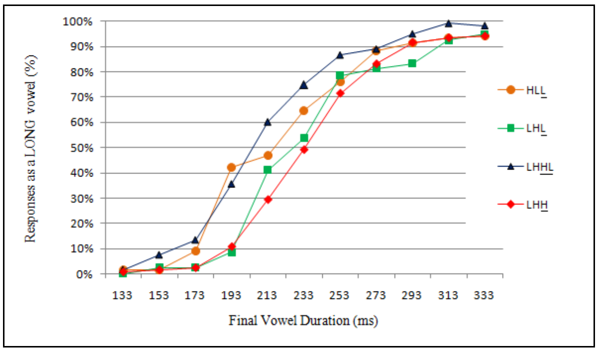

Figure 4: *Percentage of 'long' responses by NC.*

Table 4 demonstrates the mean boundary values for NJ, NE and NC estimated by Probit analyses. The results of an ANOVA showed that there were significant main effects of L1 [$F(2, 31) = 3.51$, $p<0.05$] and accent pattern [$F(3, 93) = 16.06$, $p<0.001$]. In addition, a significant interaction between L1 and accent pattern was found [$F(6, 93) = 4.35$, $p<0.001$]; therefore, ANOVAs (within factor: accent pattern) were carried out further for NJ, NE and NC separately.

Table 4. *Mean boundary values of NJ, NE and NC estimated by Probit analyses (ms).*

		HLL	LHL	LHHL	LHH	Ave.
NJ		214.4	215.3	178.3	217.8	206.5
	SD	8.6	10.7	25.4	13.7	14.6
NE		223.5	227.5	223.7	225.7	225.1
	SD	14.7	16.5	15.7	16.7	15.9
NC		221.8	236.2	210.6	238.0	226.7
	SD	37.8	36.5	23.6	28.9	31.7

The result of an ANOVA for NJ's mean boundary values showed a significant main effect of accent pattern [$F(3, 24) = 22.51$, $p<0.001$], and a post hoc test revealed that the boundary value of LHHL (178.3 ms) was significantly smaller than the other three, HLL (214.4 ms), LHL (215.3 ms) and LHH (217.8 ms) (all $p<0.001$). This result is consistent with the previous findings ([15], [16] and [17]) in that a falling tone increases NJ's perceived vowel duration. It seems clear that NJ's perception of vowel length is affected by a contour tone when the duration is ambiguous for length identification as stated in [15] and [17], and NJ are more likely to judge a vowel with a contour tone as a long vowel than one with a level tone.

An ANOVA for NE's mean boundary values revealed that there was no main effect of accent pattern [$F(3, 36) = 0.76$, $p=0.52$ *n.s.*] and this means that NE's perception was not affected by pitch cues. This suggests that they are sensitive to duration and identify length based on duration. It is important here that NE attend to the durational cue without being affected by pitch, which is not distinctive, and they are sensitive to duration, which is not distinctive but prominent in their L1.

NC's result of an ANOVA indicated a significant main effect of accent pattern [$F(3, 33) = 4.09$, $p<0.05$]. A post hoc test revealed that the value of LHHL (210.6 ms) was significantly smaller than those of LHL (236.2 ms) and LHH (238.0 ms) (both $p<0.05$). It became clear that pitch, which is a distinctive cue in Chinese, affects NC's perception and a contour tone can increase NC's perceived duration of vowels. Experiments investigating effects of pitch cues on the percentage of correct identification (% correct) of word-final vowel length, which were carried out prior to the present study, furthermore, showed that NC's % correct of LHHL pattern was significantly higher than that of HLLL and LHHH patterns ([18]). It seems possible to understand that the longer perceived duration on LHHL allows NC to be highly accurate in the identification of a long vowel with a contour tone.

The previous study ([16]) suggested that the influence of a dynamic F0, i.e. a contour tone, on the perceived duration appears only in native speakers/listeners of languages which associate a dynamic F0, i.e. a contour tone, with longer vowel duration. Chinese, however, does not seem to have such contributing factors. One possible explanation for the current result is that NC have learned to utilize a contour tone for the identification. Pitch is a distinctive cue for NC so it is not unreasonable to think that it makes them attend to pitch cues in the L2. That might help them learn to use a contour tone for the identification of Japanese vowel length. Further study should be conducted to find out whether the use of pitch cue has been learned or relates to their L1, Chinese.

Finally, another set of ANOVAs for each accent pattern revealed that there was a significant main effect of L1 only on LHHL [$F(2, 31) = 12.2$, $p<0.001$] and a post hoc test indicated that NJ's boundary value (178.3 ms) was significantly smaller than NE's (223.7 ms, $p<0.001$) and NC's (210.6 ms, $p<0.01$). This implies that NNS can establish similar boundary values to NJ's for the length identification at least under very controlled conditions. This study only investigated the boundary value of trisyllabic words in isolation; therefore, it is necessary to carry out experiments using disyllabic words as stimuli and using stimuli embedded in a sentence for better understanding. In addition, although the difference was not significant, NNS's boundary values tend to be somewhat larger than those of NJ's consistently. This indicates that NNS need longer duration to perceive a long vowel than NJ do and NJ will perceive a vowel as a long vowel but NNS will perceive it as a short vowel.

5. Conclusions

% long exhibited by NJ, NE and NC increased as a function of vowel duration. This suggests that listeners in the three groups use durational cue to identify length. The boundary location was similar between NS and NNS except for one for LHHL. This indicates that NNS can approximate their perceptual function to NS' as claimed in [11]. However, the results of the current study come from very limited conditions; therefore, further study is needed to attest NNS's perceptual approximation to NS. In addition, it showed that NNS's boundary values were somewhat larger than NJ's, and this implies that NNS would judge a vowel as a short vowel even when NJ identify it as a long vowel.

The effect of pitch cues varied among the participants with different L1 backgrounds. NJ's and NC's perception of duration was affected by the pitch cues in the same manner and it became clear that a contour tone increases their perceived duration. On the contrary, there was no effect of pitch cues on NE's perception and it demonstrates their sensitivity to duration. It remains unclear why a contour tone has effect on NC's perceived duration; however, there is a possibility that they have learned to use it and the distinctive role of pitch in their L1 helps them attend to the cue in the L2.

6. Acknowledgements

This is a part of my MA thesis and my deepest gratitude goes to Prof. Haruo Kubozono. I would like to thank Mr. Hajime Takeyasu for his technical support on the experiment and the statistic analyses. I am also thankful to all the participants.

7. References

[1] Cebrian, J., "Experience and the use of non-native duration in L2 vowel categorization", Journal of Phonetics, 34, 372-387, 2006.

[2] Cebrian, J., "The effect of perceptual factors in the acquisition of an L2 vowel contrast", in P. Avery, B. E. Dresher & K. Rice [Eds.], Contrast in phonology: theory, perception, acquisition, Berlin: Mouton de Gruyter, 303-321, 2008.

[3] Munro, M. J. & Bohn, O. S., "The study of second language speech: A brief overview", in O. S. Bohn & M. J. Munro [Eds], Language experience in second language speech learning: In honor of James Emil Flege, Amsterdam: John Benjamins Publishing Company, 3-11, 2007.

[4] Guion, S. G. & Pederson, E., "Investigating the role of attention in phonetic learning", in O. S. Bohn & M. J. Munro [Eds], Language experience in second language speech learning: In honor of James Emil Flege, Amsterdam: John Benjamins Publishing Company, 57-77, 2007.

[5] Flege, J. E., "Second language speech learning: Theory, findings, and problems", in W. Strange [Ed], Speech perception and linguistic experience: issues in cross-language research, Timonium, Maryland: York Press, 233-277, 1995.

[6] McAllister, R., Flege, J. E. & Piske, T., "The influence of L2 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian", Journal of Phonetics, 30, 229-258, 2002.

[7] Bohn, O. S., "Cross-language perception in adults: First language transfer doesn't tell it all", in W. Strange [Ed], Speech perception and linguistic experience: issues in cross-language research, Timonium, Maryland: York Press, 379-410, 1995.

[8] Fujisaki, H. and Sugito, M., "Onsei no butsuriteki seishitsu", On'in (Iwanami Koza Nihongo 5), 63-106, Tokyo: Iwanami, 1977.

[9] Tseng, C. –y., "An acoustic phonetic study on tones in Mandarin Chinese", Taipei: Institute of History & Philology Academia Sinica, 1990, cited in So, C. K., "Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences", Proceedings of the 11th Australian International Conference on Speech Science & Technology, Online: http://www.assta.org/sst/2006/sst2006-109.pdf, accessed on 13 Dec 2009.

[10] Xu, Y. & Wang, M., "Organizing syllables into groups: Evidence from F0 and duration patterns in Mandarin", Journal of Phonetics, 37, 502-520, 2009.

[11] Toda, T., "Second Language Speech Perception & Production: Acquisition of Phonological Contrasts in Japanese", Lanham, Maryland: University Press of America, 2003.

[12] Uchida, T., "Characteristics of auditory cognition of long vowels and double consonants for Chinese students in learning Japanese language", Japanese Journal of Educational Psychology, 41, 414-423, 1993.

[13] Kurihara, M., "Perceptual characteristics of Japanese long vowels by Mandarin Chinese speakers", Journal of linguistic science (Tohoku University) 8, 1-12, 2004. (in Japanese)

[14] Boersma, P. & Weenink, D., "Praat: doing phonetics by computer (Version 5.1.04)", [Computer program]. Online: http://www. praat.org/, accessed on 10 Apr 2009.

[15] Kinoshita, K., Behne, D. M., & Arai, T., "Duration and F0 as perceptual cues to Japanese vowel quantity", Proceedings of the 7th International Conference on Spoken Language Processing, 2002,Online:http://www.splab.ee.sophia.ac.jp/papers/2002/2002_03. pdf, accessed on 7 Jul 2009.

[16] Lehnert-LeHouillier, H., "The influence of dynamic F0 on the perception of vowel duration: Cross-linguistic evidence", Proceedings of the 16th International Congress of Phonetic Sciences, 757-760, 2007.

[17] Takiguchi, I., Takeyasu, H. and Giriko, M., "Effects of a dynamic F0 on the perceived vowel duration in Japanese", Proceedings of the 5th International Conference on Speech Prosody, 2010.

[18] Takiguchi, I., "Second language speech perception: Pitch cues to Japanese vowel length", unpublished MA thesis, Kobe University, 2010.