Pitch and duration as auditory cues to identify Japanese long vowels for Japanese learners *

☆Chung Ting Justine Hui (Sophia Univ.), Takayuki Arai (Sophia Univ.)

1 Introduction

Japanese distinguishes between long and short vowels, where duration is an obvious auditory cue to distinguish between the two. Using a spontaneous speech corpus, Arai [1] reported that the ratio between average durations of long and short vowels is much smaller than previously reported ratio obtained from read speech. Minagawa [2] investigated the differences in durations of Japanese long/short vowels in different pitch-accent, syllable positions and speaking rates and found faster speaking rate has a greater effect on long vowel duration, and observed phrase-final vowel lengthening. This is supported by Hirata's observation [3] on the absolute duration differences being greater in normal and fast speed than in slow speaking rate. These studies suggest that there is a great variation on the absolute length of long vowels in Japanese depending on the speaking environment, and duration does not seem to be a reliable cue to distinguish between long and short vowels in spontaneous speech. In addition to duration, previous studies have concluded that pitch contour is used by native speakers to identify vowel quantities e.g. [4, 5]. By using nonsense words with Japanese pitch patterns and a duration continuum, Kinoshita [4] established that native speakers use duration as their primary cue to differentiate between vowel quantities, and pitch as their secondary cue when the duration becomes ambiguous.

On the other hand, Japanese learners have trouble differentiating between the two [6]. Takiguchi then explored how Japanese learners compare with native Japanese speakers in terms of their categorical perception boundaries in separating a long vowel from a short vowel [5]. She found that while the distinction is particularly hard for learners with native languages that do not have pitch differences, the perceptual boundary for advanced learners and learners with a tonal native language such as Mandarin approaches that of a native Japanese speaker.

1.1 Aim and Motivation

As shown by the various studies where duration varies depending on speaking environment, we may need to focus on the role of pitch contours to distinguish vowel length in real words, especially for Japanese learners. While previous investigations have focused on using nonsense words to examine learners' categorical boundaries e.g. [5], it feels counter-intuitive in the current study to use meaningless words to investigate pitch and duration cues to distinguish between the vowel quantities. As the words do not hold any meaning, the participants may only respond to the duration cues when asked whether or not a vowel is short or long, neglecting the effect from the pitch cue. The current study explores how duration and pitch are being utilised in a semantic context for learners of Japanese. As reported by Takiguchi, where native language and length of study affected their durational boundaries [5], We will also investigate the effect of native language and length of study of the participants, as well as the length of time the participant has lived in Japan and their Japanese level according to the Japanese Language Proficiency Test (JLPT).

2 Methodology

The minimal pair "obaasan" (grandmother) and "obasan" (aunt) were chosen as stimuli for this study as they differ in the length of /a/ and both words are expected to be taught at a similarly early stage for Japanese learners. Recordings were made by a native speaker from the Tokyo region speaking standard Japanese in a sound treated room with a sound recorder (Marantz PMD 671). The following sentences were recorded: 'asokoni irunoha obasan/obaasan desu" (It is [my] grandmother/aunt over there.) The minimal pair was put into a carrier sentence to avoid effects of speech rate, allowing listeners to gauge the relative duration for long and short vowels. The experiment has been approved by the ethics committee at Sophia University.

---

*日本語学習者の長母音を識別するピッチと長さの聴覚的手がかり

We would like to thank Keiichi Tajima and Masako Fujimoto for their guidance during our pilot study.

2.1 Stimuli Manipulation

The words "obasan/obaasan" within the sentence pair were manipulated using praat [7] where the duration, pitch or both were modified. Referring to Figure 1, we have the original obasan/obaasan at the two diagonal ends of the grid. The "a" in the bracket indicates that the stimulus was manipulated from the "obasan(aunt)" original sound file, and the "g" from the "obaasan(grandmother)" original file.

Pitch						
	Obasan (aunt)	D1(a)	D2(a)	D3(a)	D4(a)	D5(a)
	P1(a)	P1D1(a)	P1D2(a)	P1D3(a)	P1D4(a)	P1(g)
	P2(a)	P2D1(a)	P2D2(a)	P2D3(a)	P2D4(a)	P2(g)
	P3(a)	P3D1(g)	P3D2(g)	P3D3(g)	P3D4(g)	P3(g)
	P4(a)	P4D1(g)	P4D2(g)	P4D3(g)	P4D4(g)	P4(g)
	D0(g)	D1(g)	D2(g)	D3(g)	D4(g)	Obaasan (grandmother)

Duration

Fig. 1 Chart of stimuli organisation

Horizontally, we have the duration manipulation, where the length of /a/ is increased in 5 steps to become the same duration in the /a/ in "obaasan" and vice versa. The /a/ in "obasan" was measured to be 100 ms and the /a/ in "obaasan" was 250 ms. Therefore each step increased or decreased by 30 ms. Vertically, the pitch is modified. We know that the correct pitch accent for "obasan" is LHHH and LHLLL for "obaasan", which can be observed from the original files, where the pitch contour lies flat between "oba-" and "-san", while in "obaasan", there is a decreasing slope in pitch in "-asan". This is indicated in Figure 2 where the top and the bottom line represents "obasan" and "obaasan" respectively.

Again, five step continual stimuli were created to match the pitch contour of "obasan" to "obaasan" and vice versa. Instead of using Hz as the steps, the unit of cent was used to create the pitch continuum to allow comparison between duration and pitch in a linear way. The schematic diagram in Figure 2 illustrates the pitch contour manipulation between "obasan" and "obaasan". The pitch of the original files were normalised prior to the manipulation to limit the variation in the sound files. To make the sentence sounds more natural, the "desu" was also modified accordingly. The audio files were also normalised in terms of their intensity to avoid effect of loudness on the perception of the words.

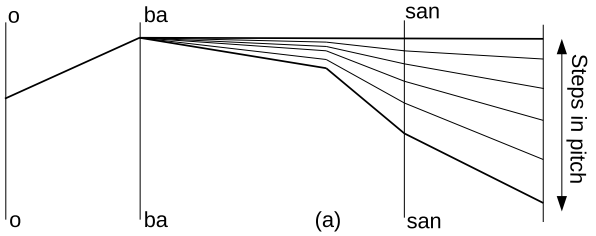

Fig. 2 Schematic representation of the pitch manipulation of stimuli (not to scale)

2.2 Test procedure

The test was carried out in a sound-treated room and in the form of a web survey where participants accessed the survey via their own computer and headphones in a quiet environment. The purpose of utilising a web survey is to capture participants who have not lived in Japan. In the first section, the participants answered questions regarding whether or not they have lived in Japan, how long they have lived in Japan, where in Japan they have lived the longest, how long they have studied Japanese for, their native languages, and their Japanese ability via the Japanese Language Proficiency Test (JLPT).

The participants then listened to the 36 tokens through headphones in random order and answered the following question: "Who is it over there?", by choosing either "Grandmother" or "Aunt". The answer was provided in English to focus on the semantics of the word and avoid effect from the orthographic form in Japanese, where grandmother has an extra "a" and aunt does not. For non English native participants, the meanings of the words were confirmed with them prior to the test. Participants recruited for the web survey listened to each token three times and participants who came to the sound treated room listened to each token six times.

3 Results

Thirty nine Japanese learners participated in the test, where twenty of them completed the web survey, and nineteen were invited to the sound treated room. No significant difference was found between the two groups and therefore the results were combined for analysis. Fifteen native speakers were also recruited to show how native speakers behave with ambiguous stimuli. Figure 3 and Figure 4 shows the average results in terms of the pitch continuum. We can see that while there is a clearer cut boundary when duration is ambiguous (D=2) for native speakers, pitch cues are still being used by non-native

speakers.

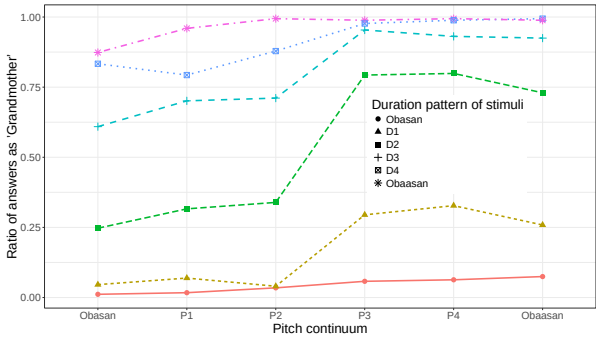

Fig. 3 Averaged results of non-native learners

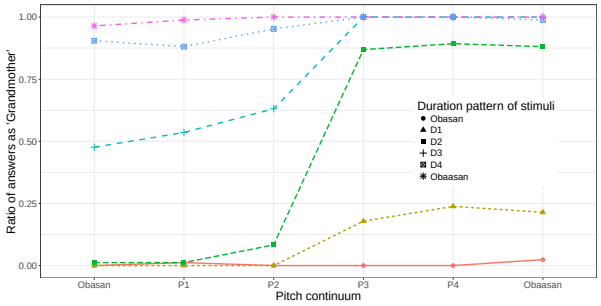

Fig. 4 Averaged results of native speakers

3.1 Effect of length of stay and study

The biggest effect we expected on the perception of long vowels was whether or not the participant has lived in Japan or not, and the length of time they have lived in Japan. Following this, we also expected an effect of the length of study on their interpretation on the words.

The length of stay and study were categorised into four to five groups according to the duration: never lived in Japan (10), 0-1 years (20), 1-2 years (6) and 2-5 years (3) for length of stay, and 0-1 years (18), 1-2 years (7), 2-5 years (16), 5-10 years (6) and 10-20 years (7) for length of study, where the number within the bracket represents the number of participants. We performed a logistic mixed effects analysis using R [8] and *lme4* [9] to examine the effect of length of stay and length of study. Living in Japan is defined to have stayed in Japan for over one month at the time of study. Firstly, as fixed effects, we entered in whether the participant has lived in Japan or not, the pitch level and the duration level into the model. As random effect, we entered in the participants. We then carried out a likelihood ratio test by comparing the model with the factor of interest, that is, whether or not the participant has lived in Japan or not, with a null model without this factor. We found no evidence of significant difference between learners who have lived in Japan and those who have not ($\chi^2(1) = 0.97$, p=0.33).

We then investigated the length of stay by having the factor as one of the fixed effects, and again the participants as the random effect, and found no evidence of any significant effect ($\chi^2(1) = 0.012$, p=0.91). Similarly, for length of study, we found no effect from length of study on the perception of "aunt" and "grandmother" ($\chi^2(1) = 0.05$, p=0.83).

3.2 Regions resided

Due to the different influences of dialects around Japan, we also took into consideration the place the Japanese learners have stayed the longest. The majority of participants have lived in Tokyo region, which includes both Saitama and Kanagawa, the longest (18), followed by Osaka (4) and Kyoto (4), and only one participant for Gumma, Hyogo, and Fukuoka. Applying a logistic mixed model analysis, we found a slight effect of the regions on the results ($\chi^2(5) = 13.5$, p=0.019). Looking into it further by carrying out a Tukey post-hoc test, there is a significant difference between the results from participants who have resided in Kyoto and Fukuoka ($z = -3.08$, p=0.022) and between Fukuoka and Gunma ($z = -2.93$, p=0.03). However, as there was only one participant from Fukuoka, we cannot make any conclusion from the current dataset.

3.3 Native language of participant

Almost half of the participants were native English speakers (19), followed by native speakers of Mandarin (10) and Cantonese (3). The remaining participants were native speakers of Vietnamese (2), Bulgarian (1), Russian (1), German (1) and Azerbaijani-Persian (1). Takiguchi (2011) has reported to find an effect from the native languages of the learners on their use of pitch cues to identify long vowels [5]. However, the current study, using a logistic mixed effect model analysis with the native language as the fixed effect showed that there was no evidence of significant difference between the different native languages ($\chi^2(8) = 4.73$, p=0.79).

We also separated the participants according to their native language being tonal or not, where tonal languages include Mandarin, Cantonese and Vietnamese, and the remaining languages are considered to be non-tonal. Again, no significant difference was observed between the two types of languages ($\chi^2(2) = 0.24$, p=0.89).

3.4 Japanese Language Proficiency Test

The Japanese Language Proficiency Test is offered worldwide and evaluates non-native speaker's ability of the language. Twenty four of the learners have passed the test, with 10 having passed N1 (highest), 5 for N2, 4 for N3, 4 for N4 and 1 for N5. Applying a logistic mixed model analysis on the effect of JLPT level yields no significant difference amongst the different levels ($\chi^2(4) = 4.12$, p=0.39).

4 Discussion

We expected Japanese learners with little experience in Japan to not be able to use the pitch cues. But looking at Figure 3, we can clearly see that once the duration is ambiguous, the answers were motivated by the pitch cues, regardless of the length of stay, length of study, native language of the participant or their Japanese ability. This result contradicts our expectation, potentially suggesting a more innate universal approach for when listeners encounter an ambiguous word despite their experience with the language, they would use the remaining cue to distinguish the words. Investigating on the stimuli where duration is ambiguous however, we are able to separate the learners into three distinct group: where the category boundary resembles native speakers as in Figure 4, where there is a less vigorous slope, and lastly, where the boundary was horizontally flat. We found a significant difference between the native-like group and the group with the flat boundary ($z = -2.51, p = 0.03$). This shows that while some learners are able to use pitch similar to their native counterparts, there were learners who did not realise pitch cues were available. However, we were not able to explain the groupings by any of the information we have regarding the participants. We observed that the participants in the last group seem to have a longer exposure to Japanese, while having very little experience living in Japan. But again, no significant effect was found between the interaction of length of study and length of stay ($\chi^2(1) = 0.05$, p=0.82). Further experiment will be needed to verify this observation.

5 Conclusion

The current study examines how pitch and duration affect Japanese learners to distinguish Japanese long and short vowels. We found that whether or not the learner has lived in Japan, the length of stay in Japan, the length of study, their native language, their Japanese ability measured through the Japanese Language Proficiency Test have no effect on their perception when deciding between the stimuli pair "obasan" and "obaasan". However, we also observed a possible pattern when the duration of the stimulus becomes ambiguous, where some participants knew to use pitch, and others did not. As this observation cannot be explained by the current study, further investigation is needed.

References

[1] T. Arai, N. Warner, and S. Greenberg, "Analysis of spontaneous Japanese in a multi-language telephone-speech corpus," *Acoustical Science and Technology*, vol. 28, no. 1, pp. 46–48, 2007.

[2] Y. Minagawa, T. Kagomiya, and K. Maekawa, "Durational variations of Japanese long / short vowels in different speaking rates : analysis of a spontaneous speech corpus," in *ICPhS*, 2003, pp. 2127–2130.

[3] Y. Hirata, "Effects of speaking rate on the vowel length distinction in Japanese," *Journal of Phonetics*, vol. 32, no. 4, pp. 565–589, 2004.

[4] K. Kinoshita, D. M. Behne, and T. Arai, "Duration and F0 as perceptual cues to Japanese vowel quantity," *Proc. of ICSLP*, pp. 757–760, 2002.

[5] I. Takiguchi, "Perceptual Development on the Identification of Length in L2 Japanese," *ICPhS*, pp. 1950–1953, 2011.

[6] K. Tajima, H. Kato, A. Rothwell, R. Akahane-Yamada, and K. G. Munhall, "Training English listeners to perceive phonemic length contrasts in Japanese," *JASA*, vol. 123, no. 1, pp. 397–413, 2008.

[7] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2016. [Online]. Available: http://www.fon.hum.uva.nl/praat/

[8] R Core Team, "R: A language and environment for statistical computer," 2015. [Online]. Available: https://www.r-project.org/

[9] D. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, "Fitting Linear Mixed-Effects Models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.