Effects of a Dynamic F0 on the Perceived Vowel Duration in Japanese

Izumi Takiguchi [1], *Hajime Takeyasu*[1, 2], *Mikio Giriko*[1]

[1] Graduate School of Humanities, Kobe University, Japan
[2] Kansai University of Nursing and Health Sciences, Japan
rikusatoizunatsu AT live.jp, thonsei AT yahoo.co.jp, tellallslant AT gmail.com

Abstract

This study examined the effect of a dynamic and a level F0 on the perceived duration in Tokyo Japanese. It was found that a rising F0 as well as a falling F0 affect the perceived vowel duration. Specifically, a falling F0 increases the perceived vowel duration regardless of its position in a word compared to level F0; on the other hand, a rising F0 shows the same effect only on the word final vowel and it decreases the perceived vowel duration in the word initial syllable.

1. Introduction

It has been pointed out that duration is the primary cue for Japanese vowel length contrast. However, previous studies found that various factors, duration of the preceding vowel ([1] and [2]), the speaking rate [3], the presence or absence of carrier phrases [3] and/or F0 contour of the target vowel ([4] and [5]), also affect its identification. The present study aims to investigate the interaction between duration and F0 in identifying Japanese vowel length.

Kinoshita, Behne, and Arai [4] examined the role of vowel duration and decrease in local fundamental frequency (F0) for Japanese listeners' identification of Japanese vowel quantity. The results show that the steeper the slope of F0 contour descends, the more a vowel is identified as having a long vowel quantity. They conclude that vowel duration serves as a primary perceptual cue for identification of vowel quantity in Japanese but Japanese listeners utilize the F0 information when the duration is ambiguous as its cue.

Lehnert-LeHouillier [5] investigated the effect of a dynamic F0 on the perception of vowel duration cross-linguistically. Native speakers of four languages, Thai, Japanese, German and Spanish, which differ in the role of duration and pitch in its phonology, were asked to judge vowel duration of monosyllabic CV non-sense words with a level F0 on the one hand and with a falling F0 on the other hand using an AXB identification task. The results show that a dynamic F0 (i.e. a falling F0) leads Japanese listeners but not the other language listeners to an increase in perceived duration of a vowel. Based on the finding, she [5] states that neither the phonemic use of duration (i.e. vowel length contrast) nor that of F0 (i.e. tone or pitch accent) in listeners' native language correlates with the effect of a dynamic F0 on the perception of vowel duration and concludes the effect of F0 is not a universal but a language specific phenomenon in the speech perception.

These two previous studies ([4] and [5]) show that a vowel with a falling F0 is likely to be identified as a long vowel at least by Japanese listeners. Lehnert-LeHouillier [5] claims that Japanese listeners were biased toward a long vowel when judging stimuli with a falling F0 due to Japanese phonology, in which a falling F0 occurs on a long vowel but not on a short vowel. She suggests that a dynamic F0 triggers an increase in perceived duration only in languages associating such a dynamic F0 with longer vowel duration.

However, the claim made in [5] is based on the experiment using only monosyllable words; therefore, it is unknown whether the same result can be obtained with multi-syllable words. For instance, van Dommelen [6] showed that monosyllable and disyllable words (in isolation) receive different influence from F0 on the perceived duration. Kinoshita et al. [4] carried out the experiment using disyllable words but they considered only the word-final vowel. Thus, it seems to be premature to conclude that a falling F0 lengthens the perceived vowel duration in Japanese under any conditions. The first aim of the present study is to examine whether F0 affects perceived duration of the vowel in the word-initial syllable as well as in the word-final syllable using multi-syllable words as stimuli.

In addition, both Kinoshita et al. [4] and Lehnert-LeHouillier [5] examined only a falling F0 as a dynamic F0, thus the same effect might not be observed for a rising F0, which is also considered to be a dynamic F0. Based on the hypothesis in [5], it is predicted that a rising F0 leads to a longer perceived duration in Japanese as well because a rising F0 can also occur only on a long vowel word-initially, but neither examined the effect of a rising F0. The second purpose of this study is to explore the effect of a dynamic F0 on the perceived duration of a vowel duration using words with three patterns of F0, falling, rising and level.

2. Experiments

We conducted the experiment using three syllable words as stimuli, which allow us to explore the interaction between the effect of F0 and the word position. Manipulating duration and F0 of the first and third vowels, we examined the effect of F0 on perceived vowel duration. In addition to the falling and level F0, this study included words with a rising F0 in the stimuli.

2.1. Methods

2.1.1. Subjects

10 adult native speakers of Tokyo Japanese with normal hearing participated in this experiment. 1 out of 10 participants took only experiment (a) and the others took both experiment (a) and (b). Each participant was paid for participating.

2.1.2. Stimuli

All stimuli were created from a token of a nonsense word /mamama/ with an unaccented pitch pattern (LHH), which was uttered by a female native speaker of Japanese. The speaker uttered the word 5 times at her natural speaking rate without any carrier phrases and one token was selected. Table 1 shows

Table 1. *Segment durations of the original token and stimuli in the experiments.*

		Original token	Experiment (a)	Experiment (b)
C1	m	23	23	23
V1	a	88	88-188 (10-ms steps)	88
C2	m	57	57	57
V2	a	121	121	121
C3	m	66	66	66
V3	a	167	167	122-272 (15-ms steps)

the segment duration of the token.

The experiment consisted of two blocks: (a) in which the duration and F0 of the first syllable vowel was manipulated orthogonally, and (b) in which the duration and F0 of the third syllable vowel was manipulated orthogonally. For each block, there were three F0 patterns: falling, rising, and level.

For the experiment (a), duration of the first vowel ranged from 88 ms to 188 ms in 10-ms steps, and for the experiment (b), it varied from 122 ms to 272 ms in 15-ms steps. The duration of the vowels was edited by copying or deleting a pitch period around the center of the vowels at zero-crossings.

In the experiment (a), F0 in the falling-series was 300 Hz at the beginning of the first vowel and descended linearly to 200 Hz toward the end of the vowel; F0 in the rising-series was 150 Hz at the beginning and ascended linearly to 200 Hz toward the end of the vowel; and F0 in the level-series was held constant at 200 Hz. F0 in the second and the third syllables remained 200 Hz. In the experiment (b), F0 in the falling-series began to descend linearly at the onset of the third vowel to 150 Hz toward its end; F0 in the rising-series started to ascend linearly at the onset of the vowel to 300 Hz toward its end; F0 in the level-series was held constant at 200 Hz. F0 in the first and the second syllables was kept to be 200 Hz. In addition, there were two more F0 patterns in the experiment (b): the later-falling- and later-rising-series. These two series had the same F0 loci as the falling- and rising-series, respectively except that the F0 change occurred at the midpoint, not at the onset of the third vowel. These two series were included to ensure that listeners would perceive the pitch movement (i.e. falling or rising) within a target vowel because some participants in a pilot study heard such movement between the second and the third syllable rather than within the target vowel, especially when the duration of the target vowel got longer. This kind of phenomenon was not observed in the initial position (the experiment (a)). Consequently, the later-falling- or later-rising-series were not included in this position. The F0 resynthesis was made using Praat [7]. Figure 1 shows schematic representation of the F0 manipulations.

There were a total of 33 stimuli (3 F0 patterns × 11 durations) in the experiment (a) and a total of 55 stimuli (5 F0 patterns × 11 durations) in the experiment (b).

2.1.3. Procedure

The participants were tested individually in a quiet room. In the experiment (a), each stimulus was presented 10 times in random order and listeners identified the first vowel as a long or a short vowel while in the experiment (b), each stimulus was presented 8 times in random order and listeners judged the vowel length of the third vowel. The ISIs were set at 4s for each experiment. Participants had 12 trials and 10 trials prior to the experiment (a) and (b), respectively as a practice phase in which two endpoint stimuli were presented from all the series. The experiment (b) was carried out more than one month later than the experiment (a).

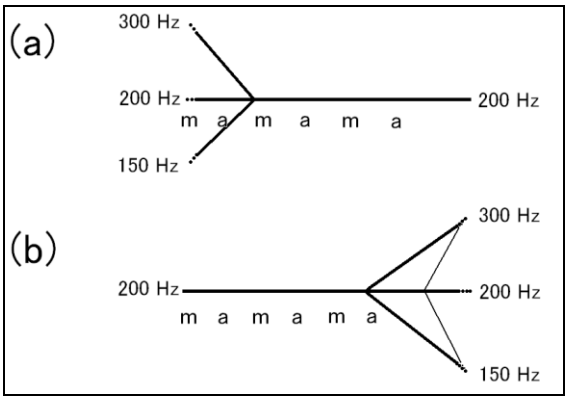

Figure 1: *Schematic representation of the F0 manipulation.*

2.2. Results

Probit analyses were performed on each listeners' identification functions to estimate the location of the boundary between the two response categories (long vs. short). Two separate ANOVAs were carried out on the individual boundary values (see Appendix), one for the experiment (a) and the other for the experiment (b).

2.2.1. Experiment (a)

Figure 2 shows the results of the experiment (a). It clearly demonstrates that F0 affects listeners' responses only when the durational cue is ambiguous. The mean boundary values were 119.2 ms for the falling-series, 133.4 ms for the rising-series, and 129.1 ms for the level-series. An ANOVA revealed that F0 patterns significantly affect the boundary values [$F(2, 18) = 111.8$, $p < 0.001$]. Post hoc analysis with Bonferroni correction for multiple comparisons indicated that the mean boundary value of the rising-series is significantly higher than that of the level-series, which is in turn significantly greater than that of the falling-series. This indicates that a falling F0 elicits more "long" responses and this result is consistent with the previous studies ([4] and [5]). However, a rising F0 causes less "long" responses than a level F0 and this is inconsistent with the prediction from the hypothesis in [5] that a dynamic fundamental frequency (F0) triggers an increase in perceived duration in languages associating such a dynamic F0 with

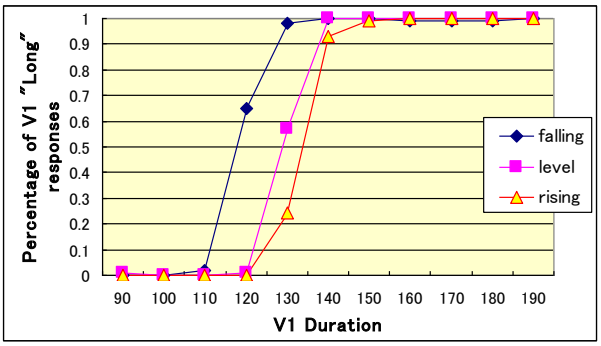

Figure 2: *Mean percentage of "long" responses in the experiment (a).*

longer vowel duration.

2.2.2. Experiment (b)

Figure 3 shows the results of the experiment (b). As in experiment (a), F0 is found to be a secondary cue since it affects listeners' responses only when the durational cue is ambiguous. The mean boundary values were 151.6 ms for the falling-series, 149.4 ms for the later-falling-series, 156.7 ms for the rising-series, 144.4 ms for the later-rising-series, and 195.4 ms for the level-series. An ANOVA revealed that F0 patterns significantly affect the boundary values [$F(4, 32) = 312.6$, $p < 0.001$]. Significant differences were found among all mean boundary values except between those of the falling- and later-falling-series with Bonferroni-corrected multiple comparisons.

The fact that a falling F0 elicits more "long" responses than a level F0 is consistent with the previous studies. More significantly, a rising F0 also increases "long" responses compared to a level F0. These results indicate that a dynamic F0 elicits more "long" responses compared to a level F0 regardless of its direction (falling or rising).

The reason is not clear but there is also a significant difference in those between rising- and later-rising-series, and between (later-)rising- and (later-)falling-series. This needs to be further investigated in future. However, more important finding here is that there is a great difference in mean boundary values between dynamic and level F0 and this result supports the hypothesis in [5].

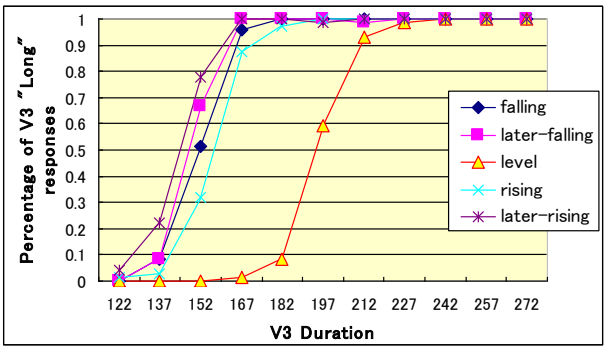

Figure 3: *Mean percentage of "long" responses in the experiment (b)*

3. Discussion

The findings of the current study suggest that a dynamic F0 does not always trigger an increase in perceived vowel duration of the vowel in Japanese. A falling F0 and a rising F0 can have a different effect on the perceived duration depending on its position even though they are both classified as a "dynamic" F0.

The present experiments replicated the findings in the previous studies with respect to the effect of a falling F0. Japanese listeners are more likely to judge the stimuli with a falling F0 as "long" than the stimuli with a level F0 regardless of the word position, word-initial or word-final. This strongly supports the claim in [5] that the effect of F0 is due to Japanese phonology, in which a falling F0 occurs on a long vowel but not on a short vowel and this distributional fact biases Japanese listeners toward "long" responses on the words with a falling F0.

There is an asymmetrical effect of a rising F0 on the perceived duration. A rising F0 affects the perceived duration of the third vowel in the same way as a falling F0, that is, it is perceived longer with a rising F0 than that with a level F0, but the first vowel with a rising F0 was the least responded as a long vowel among the three. The result obtained from the exploration on the third vowel supports the hypothesis in [5]. Also, this is consistent with the former findings (e.g., [8]), which pointed out that a dynamic F0 increases a perceived vowel duration compared to a level F0. However, the result regarding the first vowel cannot be interpreted in the same way. The result, rather, indicates that the perceived duration positively correlates with the onset F0, the higher F0 of a vowel is, the longer it is perceived and this fails to support the claim in [5].

A possible explanation is that the functional difference between rising and falling pitch in Japanese causes an unexpected pattern of a rising F0 shown in the first vowel. In the system of Tokyo Japanese pitch accent, rising pitch in word-initial position is not used distinctively; therefore, Japanese listeners were less sensitive to a rising F0 in the experiment (a). On the other hand, rising pitch functions as a marker of an interrogative in word (or utterance)-final position. That is, a rising F0 is not a redundant feature in a word-final position but it is redundant in a word-initial position. A falling F0, unlike a rising F0, can be used distinctively in both word-initial and word-final positions. Thus, according to such a system of Tokyo Japanese, only word-initial rising is redundant and it showed an unexpected pattern in the present experiments. We suggest that "redundancy", as well as co-occurrence of a dynamic F0 and longer vowel duration, is also responsible for the fact that a dynamic F0 lengthens the perceived vowel duration compared to a level F0. Further study should be conducted to confirm this interpretation.

4. Conclusions

This study considered the effect of a dynamic and a level F0 on the perceived vowel duration in Japanese. It was found that a rising F0 as well as a falling F0 affect the perceived vowel duration compared to a level F0. However, it was also found that there is an asymmetry between a falling and a rising F0 in that such F0 effect is not observed when a rising F0 occurs in the vowel in the word-initial syllable. The relationship between F0 and perceived durations may interact with phonology (e.g., the systems of the pitch accent) of listeners' native language.

5. Acknowledgements

We would like to thank the participants in the two experiments.

6. References

[1] Fujisaki, H, Nakamura, K., and Imoto, T., "Auditory perception of duration of speech and non-speech stimuli", in G. Fant and M. A. A. Tatham [Eds], Auditory Analysis and Perception of Speech, 197-219, London: Academic Press, 1975.

[2] Fujisaki, H. and Sugito, M. "Onsei-no butsuriteki seishitsu", in Onin (Iwanami Kouza Nihongo 5), 63-106, Tokyo: Iwanami, 1977.

[3] Hirata, Y & Lambacher, S. G., "Role of word-external contexts in native speakers' identification of vowel length in Japanese", Phonetica, 61, 177-200, 2004.

[4] 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 2 Sep, 2009.

[5] 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.

[6] van Dommelen, W. A., "Does dynamic F_0 increase perceived duration? new light on an old issue", Journal of Phonetics, 21, 367-386, 1993.

[7] Boersma, P and Weenink, D (2009). "Praat: doing phonetics by computer (Version 5.0.46)" [Computer program]. Online: http://www.praat.org/, accessed on 14 Jan, 2009.

[8] Lehiste, I. "Influence of fundamental frequency pattern on the perception of duration". Journal of Phonetics, 4, 113-117, 1976.

Appendix: *Individual boundary values (estimated by probit analyses) in the experiment (a) and (b)*

	Experiment (a): Vowel in the initial syllable			Experiment (b): Vowel in the final syllable				
Listener	falling	level	rising	falling	later-falling	level	rising	later-rising
1	124.4	131.1	131.1	148.3	145.8	194.7	147.8	137.2
2	119.0	128.2	131.6	159.0	160.0	200.6	156.6	147.1
3	119.8	128.2	133.0	145.8	145.0	193.5	157.9	144.6
4	116.3	129.1	135.0	152.3	149.3	197.0	158.4	145.4
5	116.1	129.1	132.0	155.8	151.1	197.3	158.4	151.1
6	122.2	129.1	131.1	151.1	146.6	193.2	161.8	141.6
7	118.2	128.2	134.0	152.9	143.9	193.0	153.9	141.6
8	118.4	131.1	138.1	—	—	—	—	—
9	118.2	126.0	134.0	146.6	153.4	190.2	156.5	151.1
10	119.2	131.1	134.3	152.9	149.3	199.5	159.5	140.3
Mean	119.2	129.1	133.4	151.6	149.4	195.4	156.7	144.4
SD	2.5	1.6	2.2	4.3	5.0	3.4	4.0	4.8