Effects of dynamic F0 on the perception of Japanese vowel length by native listeners of Korean and Japanese.

Category: phonetics/phonology, psycholinguistics

ABSTRACT: This study explores effects of dynamic F0 on the perception of Japanese vowel length by native listeners of Korean (NK) as well as Japanese (NJ). We conducted perceptual experiments using speech and nonspeech stimuli. The results show that NJ and NK use dynamic F0 differently for the identification of vowel length and that their use of the cue also varies depending on where a target vowel is in a word (word-initial or word-final). Dynamic F0 generally affects NK's judgments for vowel length in speech and for duration of nonspeech buzz stimuli similarly. In contrast, effects of dynamic F0 emerge differently in NJ's responses to speech and nonspeech stimuli, which indicates language-specific use of dynamic F0 for the perception of vowel length by NJ.

Japanese has a vowel length contrast, which is primarily cued by duration of the target vowel (Fujisaki & Sugito, 1977). Recent studies point out that NJ are more likely to judge a target vowel with dynamic (falling/rising) F0 as a (phonologically) long vowel compared to that with level F0 (Kinoshita et al., 2002; Lehnert-LeHouillier, 2010) and that there is an interaction between effects of dynamic F0 and the position of the target vowel in a word (Takiguchi et al., 2010). Learners of Japanese whose native language is Chinese but not English are also affected by dynamic F0 in the same manner as shown by NJ (Takiguchi, 2010). Takiguchi (2010) states that Chinese listeners are sensitive to dynamic F0 because F0 is distinctive for perception of lexical tone in their L1, Chinese.

We examine whether NK are able to use this dynamic F0 as a cue. Since F0 works as one of the cues for a three-way laryngeal distinction (Kim et al., 2002) even though it is not distinctive, NK are expected to show the dynamic F0 effects. In order to clarify this, we conducted two perceptual experiments, one for the word-initial and the other for the word-final position. In each experiment, 11 NJ and 11 NK identified length of the target vowel as "short" or "long". Stimuli were created from a token /mamama/ uttered by a female speaker of Japanese. By manipulating the target vowel duration and F0 pattern of the original token, 33 stimuli (11 durations × 3 F0 patterns) and 55 stimuli (11 durations × 5 F0 patterns) were prepared for word-initial and word-final condition, respectively (see Table 1 below). The stimuli were presented to the participants 10 times in the word-initial condition and 8 times in the word-final condition.

Table 1: Manipulation of the stimuli

Word-initial (/m**a**mama/)		Word-final (/mamam**a**/)	
Vowel duration	F0 pattern	Vowel duration	F0 pattern
88 ms to 188 ms (in 10-ms steps)	Rising: 150 Hz to 200 Hz Level: 200 Hz Falling: 267 Hz to 200 Hz	122 ms to 272 ms (in 15-ms steps)	Rising: 200 Hz to 267 Hz Later-rising*: 200 Hz to 267 Hz Level: 200 Hz Falling: 200 Hz to 150 Hz Later-Falling*: 200 Hz to 150 Hz

** Later-rising (-falling) and rising (falling) series differ in that F0 starts to move at the midpoint in the former and at the onset in the latter of the third vowel, respectively.*

Figure 1 & 2 and Figure 3 & 4 show the results of the word-initial condition and the word-final condition for NJ and NK, respectively. A series of logistic regression analyses revealed that in each word position, an interaction between listeners' L1 and F0 pattern was significant [*Wald* $\chi^2 = 81.188$, $df = 2$, $p < 0.001$ for word-initial condition; *Wald* $\chi^2 = 54.569$, $df = 4$, $p < 0.001$ for word-initial condition], suggesting that F0 affects NJ and NK differently. In word-initial condition, rising F0 elicited less "long" responses than level or falling F0 from NJ, while NK gave less "long" responses to level F0 than the others (throughout this manuscript we omit the results of multiple comparisons using log odds ratios due to the space limitation). In word-final condition, vowels with level F0 were more likely to be judged as "short" than those with dynamic F0 by NJ, replicating the results in the previous studies. NK's results were in common with NJ's in that vowels with level F0 had a tendency to be judged as "short," but the differences between F0 patterns were greater and boundary location (50%

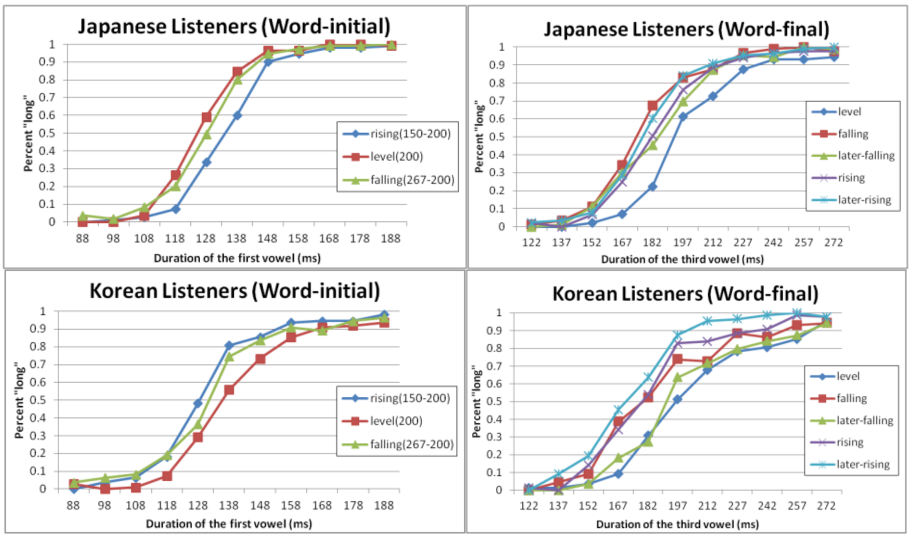

Fig. 1 (upper left panel); Fig. 2 (lower left panel); Fig. 3 (upper right panel); Fig. 4 (lower right panel) crossover point) of later-falling, which is one of the dynamic F0 patterns, was not significantly different from that of level. That NK gave more "long" responses to vowels with falling F0 than those with level F0 is consistent with the results of native listeners of Chinese in Takiguchi (2010).

We further conducted perceptual experiments using AX discrimination tasks with nonspeech stimuli in an attempt to find out what gave rise to this difference between NJ and NK. The participants were the same listeners in the above experiments. Nonspeech materials consisted of a sequence of three nonspeech buzzes, mimicking the duration of the three syllables in the speech stimuli. Listeners were asked to decide whether the first or third buzz was longer in the standard stimulus than in the comparison stimulus for the word-initial and for the word-final condition, respectively. Buzzes in the standard stimulus always had level F0 (held at 200 Hz) and comparison stimulus had three patterns of F0 in the word-initial condition and five patterns of F0 in the word-final condition as in the experiments using speech stimuli. The duration of the target buzz in the standard stimulus was kept to be 140 ms in the word-initial condition and 200 ms in the word-final condition. The duration of the target buzz in the comparison stimulus varied from 80 ms to 200 ms (in 20-ms steps) in the word-initial condition and varied from 140 ms to 260 ms (in 20-ms steps) in the word-final condition.

Figure 5 & 6 and Figure 7 & 8 show the results of the word-initial condition and the word-final condition for NJ and NK, respectively. A series of logistic regression analyses revealed that in each word position, an interaction between listeners' L1 and F0 pattern was significant [*Wald* $\chi^2 = 11.123$, $df = 2$, $p < 0.01$ for word-initial condition; *Wald* $\chi^2 = 28.156$, $df = 4$, $p < 0.001$ for word-initial condition], suggesting that F0 affects NJ and NK differently. In word-initial condition, falling F0 elicited more "long" responses than level or rising F0 from Japanese listeners, which is different from the results of the experiment using speech stimuli. In contrast, NK gave less "long" responses to level F0 than the others, which is parallel to the results in the above experiment. These results indicate that NJ perceive vowel length with a language-specific manner, while NK do not. In word-final condition, vowels with falling or later-falling F0 were more likely to be judged as "short" than those with other F0 patterns in both language groups. Vowels with rising or later-rising F0 tended to elicit more "long" responses than those with level F0 and the difference between them was significant only for NK. NJ again showed a language-specific use of F0: while buzzes with falling and later-falling F0 were *less* likely to be judged as "long" than those with level F0 in the nonspeech experiment, vowels with falling and later-falling F0 were *more* likely to be judged as "long" in the speech experiment. This was also true for NK as well except that boundary location of level and later-falling were not significantly different.

The difference in NK's responses to speech and nonspeech stimuli may imply that NK have learned to use F0 cue in perceiving Japanese vowel length as NJ do. More interestingly, this seems to occur only or earlier in the word-final position. It has been reported that it is more difficult for L2 learners to perceive vowel length correctly in the word-final position (Minagawa-Kawai et al., 2002). This difficulty might lead learners to use secondary cue (i.e., F0 in this study), and future research should address whether this interpretation is appropriate for vowel length contrasts and for any other contrasts. It would be interesting to see whether L2 learners learn to use secondary cues more quickly when it is challenging for them to perceive contrasts only with the primary cues.

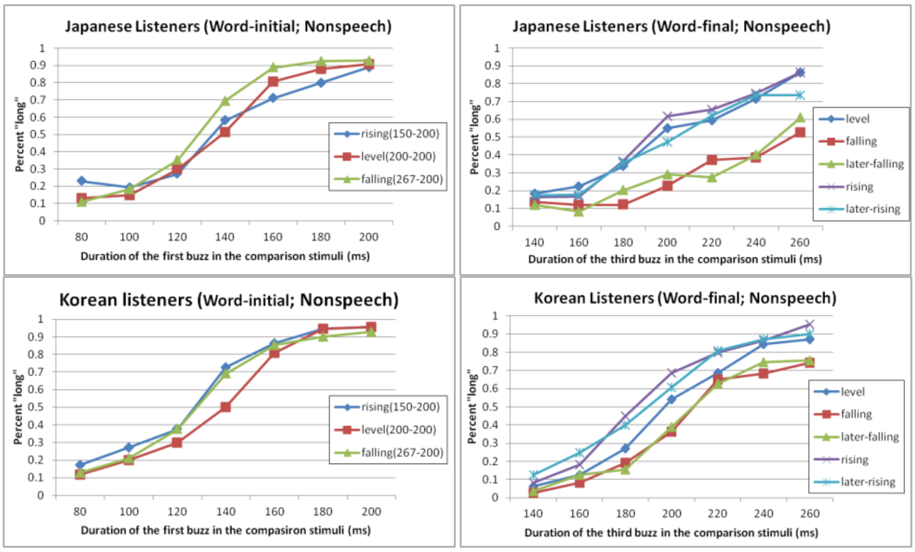

Fig. 5 (upper left panel); Fig. 6 (lower left panel); Fig. 7 (upper right panel); Fig. 8 (lower right panel)

References: Fujisaki, H. and M. Sugito (1977). Onsei no butsuriteki seishitsu. *Iwanami koza nihongo 5 On'in*. S. Ohno and T. Shibata, Iwanami shoten: 63-98. Kim, M.-R., Beddor, P. S., Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30, 77-100. Kinoshita, K., Behne, D. M. Arai, T. (2002). Duration and F0 as perceptual cues to Japanese vowel quantity. *Proc. 7th ICSLP Denver*. Lehnert-LeHouillier, H. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics*, 38, 472-482. Minagawa, K. Y., Maekawa, K., Kiritani, S. (2002). Effects of pitch accent and syllable position in identifying Japanese long and short vowels: Comparison of English and Korean speakers. *J. Phone. Soc. Jap.* 6(2), 88-97. Takiguchi, Izumi (2010). Effects of pitch cues on the identification of vowel length in L2 Japanese. *Proceedings of INTERSPEECH 2010 Satellite workshop on Second Language Studies: Acquisition, Learning, Education and technology*. Waseda University, Japan. September 23rd, 2010. 22-24 Takiguchi, I., Takeyasu, H., Giriko, M. (2010). Effects of a dynamic F0 on the perceived vowel duration in Japanese. *Proc. 5th Speech Prosody Chicago*. 100944.