10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Perception of English Intonation by Japanese Learners of English

Tomoko Hori[1], *Michiko Toyama*[2], *Mari Akatsuka*[3]

[1]Juntendo University, Japan
[2]Bunkyo University, Japan
[3] Nagoya University of Foreign Studies, Japan

t-hori@juntendo.ac.jp, toyama3@shonan.bunkyo.ac.jp, akatsuka@nufs.ac.jp

Abstract

This study investigated the extent to which Japanese learners of English (JLE) correctly perceived English intonation contours (falling, rising, falling-rising), and whether training could improve their performance. The influence of an acoustic cue (duration) on learners' perception errors (confusions), especially for rising contour, was also explored. In total, 245 JLE completed a forced-choice intonation identification task. To examine the influence of the durational cue, some stimuli were manipulated by shortening the duration between the onset and the starting point of pitch rise in a stressed syllable. Analysis of the performance of JLE in terms of correct identification rate and perception errors revealed that falling tones were easiest to perceive correctly, followed by rising and then falling-rising tones. These differences across the three contours remained unchanged after training. While confusions were also observed between rising and falling-rising contours, correct identification of rising contour increased significantly for manipulated rising stimuli—that is, JLE exhibited sensitivity to duration in distinguishing rising from falling-rising contours, indicating that durational cues may cause perceptual confusion of rising and falling-rising contours.

Index Terms: L2 perception, English intonation, Japanese learners of English, durational cue

1. Introduction

A wealth of previous research has shown the effect of native language on perception of nonnative speech at the segmental level. Perception of nonnative speech contrasts is thought to be constrained by both the phonological and phonetic properties of the native language [1]; for instance, Japanese native speakers' difficulty in perceiving English /l/-/r/ [2], [3] is often explained by phonological differences between Japanese and English. In contrast, phonetic properties such as long and short durational contrasts in consonants of Japanese geminates (e.g., "kitte" vs "kite") are known to cause difficulties for English native speakers [4].

According to theories of cross-language perception, perceived similarities between nonnative and native languages predict difficulties in perceiving nonnative speech segmentals [5], [6]. According to the Speech Learning Model [5], the greater the acoustic dissimilarity between an L2 and the closest L1 sounds, the easier it becomes to discern the L2 sounds.

Unlike segmentals, the relationship between perception of suprasegmentals and L1 is not clearly understood [1]. Research findings on suprasegmental features have been mixed. Some areas of perception exhibit similarity while others show cross-linguistic differences. In a study of how L1 affects non-native perception of Mandarin lexical tones by Hong Kong Cantonese, Japanese, and Canadian English listeners [1], Cantonese listeners were expected to perform as accurately as native speakers because of their L1 experience of using lexical tone. Their difficulty in perceiving tonal contrast was thought to be caused by L1 influence. Some studies of L2 intonation perception have also reported L1 influence on learners' knowledge of intonation patterns [7], [8].

To investigate perception of intonation contours, Grabe, Rosner, García-Albea and Zhou [9] tested L1 effects among English, Spanish, and Chinese listeners using 11 different intonation contours of one phrase. The results confirmed that listeners divided the stimuli into two groups (falling and rising pitch), but found significant cross-linguistic differences related only to falling pitch-stimuli.

Akatsuka, Hori, and Toyama [10] investigated the effects of explicit teaching on intonation perception of non-native language (English) with 20 Japanese listeners. In a four-choice intonation-identification task (falling, rising, falling-rising, don't know), performance (correct response score for all three patterns together) improved significantly after training, confirming the effects of explicit instruction. Identification of three intonation patterns differed significantly; falling tone was most often perceived correctly, followed by rising and then falling-rising tones.

Japanese is a pitch accent language—that is, an accent is realized by a sharp decline from a high to a low tone around the accented mora [11]. Pitch accents are a lexical property of a given word [12], and intonation is determined in Japanese by the pitch contour of the last mora of an intonation phrase [13]. As English is neither a pitch accent language nor a tone language [1], pitch accents highlight words or syllables [12], and intonation is determined by the pitch movement, starting from the accent. Despite these phonological and phonetic differences, the two languages share some similarities; for instance, both employ a rising tone to ask a question. Pitch direction (low to high) is also the same, although there are differences in how pitch rises [14]. As Japanese does not typically use a falling-rising tone [15], perceiving this pitch contour may prove difficult. In addition, as Roach [16] points out, falling-rising can be difficult to identify when tones fall over not only nucleus but also tails (following syllables). As for falling and rising tones, it remains unclear why JLE find it more difficult to perceive a rising tone, and more research is needed to address this issue.

10.21437/SpeechProsody.2020-171

Since Akatsuka et al.'s study [10] had only 20 participants, and they did not analyze perception errors, the present study investigated the perception of JLE with a large number of participants, focusing on their perception errors. Experiment 1 expanded on Akatsuka et al.'s study [10] to investigate intonation perception and errors. Based on those results, Experiment 2 explored the effect of the phonetic environment, with particular reference to durational cues.

2. Experiment 1

2.1. Method

For the purposes of this study, the data of 94 participants were added to that of the 20 participants in Akatsuka et al. [10]; all 114 participants were Japanese first-year college students majoring in English with Common European Framework of Reference for Languages (CEFR) B1 level. Participants listened to the same set of 27 stimuli prepared for Akatsuka et al. [10], which included 18 (monosyllabic, bisyllabic, trisyllabic) one-word sentences (six falling, six rising, six falling-rising tones) and nine sentences of two or more words (five falling tones, four rising) before and after training. The training sessions (200 minutes in total) took place over five weeks. All stimuli were recorded by an American native speaker. After a familiarization session in which stress and pitch movement were explained to participants, they completed a four-choice identification test (falling, rising, falling-rising, don't know). Target stimuli were presented on a sheet of paper, with nuclear stress words underlined to direct listeners' attention to pitch contours. Each participant clicked computer-based buttons on a Google form to indicate their responses while listening to the stimuli through headphones.

2.2. Results and discussion

2.2.1. Intonation identification rate

Performance on the identification test was assessed in terms of correct identification rate expressed as a percentage. Paired sample t-testing to compare identification rates found a significant difference between pre-test (M = 68.1, SD = 1.1) and post-test (M = 76.4, SD = 1.8) conditions; *t* (113) = 6.75, *p* < 0.01. These results indicate that perception of English intonation patterns as a whole improved after training. However, the degree of improvement for each pattern was found to be different. Significant improvements between pre-test and post-test were confirmed for falling (*p* < 0.01) and falling-rising (*p* < 0.01) tones, but not for rising tones (*p* = 0.37). The results show that it is not easy to improve perception of rising tone.

Two separate one-way ANOVAs with repeated measures were performed to evaluate the effect of intonation pattern on pre- and post-test identification rates, revealing a significant effect for the pre-test (*F* (2, 112) = 287.07, *p* < .001). Bonferroni post-hoc testing revealed that the identification rate for falling tones (M = 89.06, SD = 1.19) was significantly higher than for rising (M = 68.68, SD = 1.72) and falling-rising tones (M = 28.94, SD = 2.33, *p* < .001). For post-test, intonation pattern was also found to have a significant effect (*F* (2, 112) = 116.87, *p* < .001). Post-hoc testing revealed that the identification rate for falling tones (M = 90.05, SD = 8.67) was significantly higher than for rising (M = 70.61, SD = 22.43) and falling-rising tones (M = 55.85, SD = 29.57, *p* < .001). These results align with Akatsuka et al. [10], indicating differences in ease of perceiving the three intonation contours. It is unsurprising that the lowest correct identification rate related to falling-rising tones, as these are not common in Japanese. However, the results for rising tones warrant further investigation, as these were less easily identified despite their frequent occurrence in Japanese, and the practice sessions failed to improve participants' performance (Fig. 1).

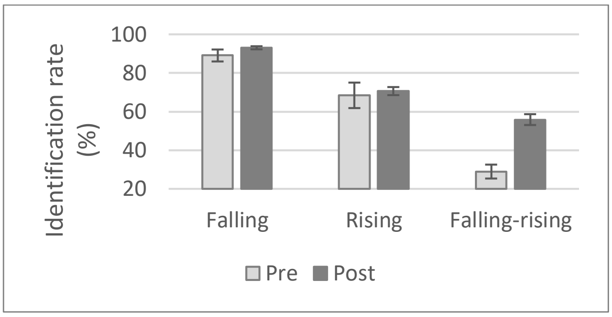

Figure 1: *Identification rates for intonation patterns in pre- and post-tests.*

2.2.2. Intonation pattern confusions

Participants' errors in identifying intonation patterns were also used to investigate the pattern and frequency of confusions. Table 1 shows responses for each target intonation pattern, indicating that confusions related mainly to falling-rising and rising stimuli. In the pre-test, falling-rising tones were correctly identified in only 28.99% of cases while incorrect identification accounted for more than 70% of responses. In particular, there were notable confusions with rising tones (57.69%). In relation to rising stimuli, 68.42% of responses were correct, but 28.51% of responses misidentified rising as falling-rising. The post-test revealed a similar pattern of confusion among falling-rising and rising stimuli.

Table 1: *Confusion matrices for pre- and post-test responses.*

Pre-test	Response (%)			
Target	Falling	Rising	Falling-rising	Total
Falling	**89.07**	7.74	2.87	99.68
Rising	2.89	**68.42**	28.51	99.82
Falling-rising	13.03	57.69	**28.99**	99.70
Post-test	Response (%)			
Target	Falling	Rising	Falling-rising	Total
Falling	**93.06**	5.42	1.43	99.92
Rising	0.96	**70.61**	28.33	99.91
Falling-rising	8.33	35.23	**55.85**	99.41

Bold signifies the target identification tones.

These findings indicate that confusion of rising and falling-rising leads to difficulties in correctly identifying rising tones. Rising stimuli were misidentified as falling-rising with more than 35% of the responses including "no" (Fig. 2), "Monday," "wonderful," "tomorrow," and "excuse me." Observation and acoustic examination of pitch, intensity, and duration of rising stimuli suggests that one acoustic cue used to perceive rising stimuli as falling-rising is durational balance before and after the point where pitch rise commences. Confusion often seems to occur when the duration before the pitch rise is about the same length or longer than the duration after pitch rise commencement. If that is so, it can be hypothesized that shortening the duration before the pitch rise would increase correct identification of a rising contour. To explore this possibility, Experiment 2 studied the effect of a shortened duration before pitch rise using manipulated stimuli.

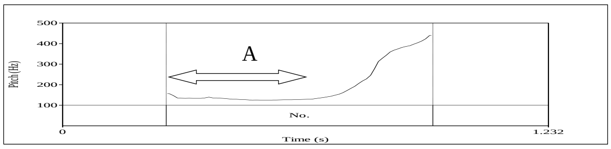

Figure 2: *Pitch of the stimulus "no."*

3. Experiment 2

3.1. Participants

In total, 131 JLE participated in this experiment (31 male, 99 females, 1 N/A). The participants were college students majoring in English with CEFR B1 level, ranging in age from 18 to 21 years (mean = 19 years). Three native speakers of American English (two males, one female) also participated, and their data were used for comparison purposes. None of the participants had hearing problems.

3.2. Materials

In this experiment, only monosyllabic one-word sentences were used as target stimuli to reduce factors that may have influenced perception. We prepared 13 monosyllabic one-word sentences (shown in Table 2) read in three intonation patterns (falling, ring, falling-rising). The words were chosen to ensure their familiarity to participants. Word familiarity in auditory presentation of JLE [17] was 5.91 on average.

Table 2: *Stimuli.*

Target	Down. Noun. Shout. Game. Main. May. Say. Noon. Soup. When. Men. Pet. Now.
Filler	Mickey. Honey. Money. Sorry. Monday. Tuesday. What's your name? Excuse me. Open to page three. How about you? Why don't you come? Is it easy?

Underline signifies the words used for manipulation.

In addition, eight manipulated stimuli with rising contour were added as target stimuli to investigate whether durational change would affect identification rate. Eight stimuli (the underlined words in Table 2) were taken from the target stimuli for manipulation. The duration between onset and starting point of pitch rise (indicated as A in Fig. 2) was shortened by approximately 50% using Praat [18]. Manipulation involved speeding up the designated segment while preserving the original F_0 contour and intensity. The manipulated stimuli were tested on a native English speaker, who found none of them unnatural. Twelve fillers (six bisyllabic one-word sentences and six multi-word) were also prepared. The stimuli read by a native speaker of British English were used for the experiment because of the clarity of the pronunciation. All speech files were normalized to 70 dB using Praat. Participants listened to 56 stimuli in total (16 falling, 27 rising, 13 falling-rising), with a pause of 180 ms between stimuli.

3.3. Procedure

Prior to the experimental task, the participants engaged in a familiarization session, in which they listened to three stimuli (falling, rising, falling-rising contours) read by a native speaker of English. The aim was to enable participants to learn how each stimulus corresponded to each intonation contour. For sentences of two or more words, participants were instructed to pay attention to the last (nuclear stress) word to identify the intonation contour.

The participants then completed a forced-choice (falling, rising, falling-rising) identification task, listening to the stimuli through headphones without looking at the list of stimuli. They were allowed to listen only once before clicking a response button on a Questant web-based questionnaire.

3.4. Results and discussion

3.4.1. Intonation-identification rate

Participants' performance (target stimuli only) was assessed in terms of correct identification rate (i.e., correct responses divided by total responses). Two repeated measures one-way ANOVAs were performed to evaluate the effect of intonation patterns on identification rate, revealing a significant effect of intonation pattern (F (2, 129) = 113.29, $p < .001$). Bonferroni post-hoc testing revealed that the identification rate for falling tones (M = 0.93, SD = 0.17) was significantly higher than for rising (M = 0.73, SD = 0.32) and falling rising tones (M = 0.40, SD = 0.36, $p < .001$). These results align with Experiment 1 and Akatsuka et al. [10], indicating differences in ease of perceiving the three intonation contours.

Errors in identification of intonation patterns were also examined. As shown in Table 3, errors related mainly to falling-rising and rising stimuli. As in Experiment 1, identification of rising tones as falling-rising accounted for more than 20% of errors. However, confusions with falling-rising tone differed in Experiments 1 and 2. While participants often misidentified falling-rising as rising rather than falling in Experiment 1 (Table 1), more confusions between falling-rising and falling were observed in Experiment 2. This result may be related to the difference in the stimuli for the experiments. Experiment 1 used multisyllabic one-word and two or more word sentences, whereas Experiment 2 used only monosyllabic one-word sentences.

Table 3: *Confusion matrices for responses to target stimuli by JLE.*

	Response (%)			
Target	Falling	Rising	Falling-rising	Total
Falling	**93.48**	3.54	2.57	99.58
Rising	9.09	**67.21**	23.26	99.56
Falling-rising	34.86	26.59	**37.91**	99.36

Bold signifies the target identification tones.

One notable finding relates to the performance of the three native speakers of English, whose confusions tended to be very similar to JLE. As shown in Table 4, falling tones attracted the highest number of correct responses, followed by rising and then falling-rising tones. The correct response rate for rising tones was about 67%, which is almost the same as for Japanese participants.

Table 4: *Confusion matrices for responses to target stimuli by native speakers of English.*

	Response (%)			
Target	Falling	Rising	Falling-rising	Total
Falling	**93.94**	6.06	0.00	100.00
Rising	1.59	**66.67**	31.75	100.00
Falling-rising	30.56	11.11	**58.33**	100.00

Bold signifies the target identification tones.

3.4.2. Duration effect

Paired sample t-testing to compare identification rates revealed a significant difference for original (M = 0.64, SD = 0.27) and manipulated stimuli (M = 0.72, SD = 0.23); t (130) = 3.70, $p < 0.01$. The result indicates that a shortened duration before pitch rise enhanced correct perception of rising stimuli. It also implies that Japanese speakers may be sensitive to durational cues in perceiving pitch patterns. One possible reason why a long duration before pitch rise in the stimuli confused the participants is the effect of the mora rhythm in the Japanese language, whose speakers keep each mora about the same duration. When they hear stimulus such as Figure 2, a monosyllabic word (L* H-H%) with the same duration of a low pitch and rising pitch, they may perceive it as two moras (counts) consisting of a low level and a high rising pitch. This may lead to the misperception of the rising stimuli as falling-rising. Another reason is that the participants may hear low level pitch before rise as falling pitch.

4. Conclusion

This study investigated the extent to which JLE correctly identified three English intonation contours and explored the influence of phonetic environment (especially durational cues) on errors in perceiving a rising contour. The results for one-word sentence stimuli indicate that performance differed significantly across the three intonation contours; falling tone was most often perceived correctly, followed by rising and then falling-rising tones. Although both falling and rising tones are common in Japanese it seems that, in English, a rising tone is more difficult to perceive than a falling tone. In fact, this difference may apply not only to JLE but also to speakers of other languages, and this issue should be investigated cross-linguistically with a larger sample.

The results of this study also suggest that the perception of rising stimuli as falling-rising seems to relate to the initial segment before pitch rise. Further studies are needed to verify the present findings, based on variation of the durational environment. For instance, as well as shortening the segment before pitch rise, the effect of lengthening or indeed changing the speed of whole stimuli should be examined. In general, it seems clear that a range of factors influence intonation perception, and further research is needed to assess whether the present results can be generalized beyond the study setting.

5. Acknowledgements

This work was supported by the Joint Research Program of Juntendo University, Faculty of Health and Sports Science.

6. References

[1] C. K. So and C. T. Best, “Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences,” *Language and Speech*, vol. 53, no. 2, pp. 273–293. 2010.

[2] R. A. Yamada, “Age and acquisition of second language speech sounds perception of American English /ɹ/ and /l/ by native speakers of Japanese,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, W. Strange, Ed. Baltimore, MD, USA: York Press, 1995, pp. 305–320.

[3] K. Aoyama and J. E. Flege, “Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers,” *Journal of the Phonetic Society of Japan*, vol.15, no. 3, pp. 5–13, 2011.

[4] D. M. Hardison and M. M. Saigo, “Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy,” *Applied Psycholinguistics*, vol. 31, pp. 81–99, 2010.

[5] J. E. Flege, “Second language speech learning: Theory, findings, and problems,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, W. Strange, Ed. Baltimore, MD, USA: York Press, 1995, pp. 233–277.

[6] C. T. Best and M. D. Tyler. “Nonnative and second-language speech perception: Commonalities and complementarities,” in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, M. J. Munro and O. S. Bohn, Eds. Amsterdam, The Netherlands: John Benjamins, 2007, pp.13–34.

[7] P. Mok, Y. Yin, J. Setter and N. M. Nayan, “Assessing knowledge of English intonation patterns by L2 speakers,” in *Proceedings of Speech Prosody*, Boston, USA, 2016, pp. 543–547.

[8] K. Puga, R. Fuchs, J. Setter and P. Mok, “The perception of English intonation patterns by German L2 speakers of English,” in *INTERSPEECH*, Stockholm, Sweden, Aug. 2017, pp 3241–3245.

[9] E. Grabe, B. Rosner, J. E. García-Albea, and X. Zhou, “Perception of English intonation by English, Spanish and Chinese listeners,”
in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 2003, pp. 763–766.

[10] M. Akatsuka, T. Hori, and M. Toyama, “Nihonjin eigogakushusha ni okeru eigo intonation no hatsuonsido no koka” [Effects of instructing English intonation pronunciation to Japanese learners], *Language Education & Technology*, vol. 55, pp.151–170, 2018.

[11] M. Beckman and J. Pierrehumbert, “Intonational structure in Japanese and English,” in *Phonology Yearbook Ⅲ*, C. Ewen and J. Anderson, Eds. Cambridge, UK: Cambridge University Press, 1986, pp. 255–309.

[12] J. J. Venditti, “The J-ToBI model of Japanese intonation,” in *Prosodic Typology and Transcription: The Phonology of Intonation and Phrasing*, S. Jun, Ed. Oxford, UK: Oxford University Press, 2005, pp.182–200.

[13] T. Makino, *Nihonjin no tameno eigo onseigaku lesson* [English Phonetics Lessons for Japanese Learners]. Tokyo, Japan: Taishukan, 2005.

[14] H. Saito and I. Ueda, “Eigo gakushusha niyoru intonation kaku no gohaichi” [Misplacement of nuclear stress by Japanese learners of English], J*ournal of Phonetic Society of Japan*, vol. 15, no.1, pp. 87–95, 2011.

[15] S. Kori, “Nihongo intonation nitsuiteno ikutsukano tyoushu jikken” [A perceptual study on Japanese intonation], *Studies in Language and Culture*, vol. 43, pp.249–272, 2018.

[16] P. Roach, *English Phonetics and Phonology*. Cambridge, UK: Cambridge University Press, 2000.

[17] H. Yokokawa, Ed., *Nihonjin Eigogakushusha no Eitango Shinmitsudo: Onseihen* [English Word Familiarity for Japanese Learners of English: Auditory Presentation]. Tokyo, Japan: Kurosio, 2009.

[18] P. Boersma, and D. Weenink, "Praat: Doing phonetics by computer" (Version 6.0.41) (2018). [Computer program]. Available: http://www.praat.org/ [Accessed October 27, 2018].