



The Kinesthetic Effect on EFL Learners' Intonation

N. Yamane¹, B. Teaman², A. Fujimori³, I. Wilson⁴, and N. Yoshimura⁵

¹ Hiroshima University, Japan, ² Osaka Jogakuin University, Japan,

^{3,5} University of Shizuoka, Japan, ⁴ University of Aizu, Japan,

¹yamanen@hiroshima-u.ac.jp, ²teaman@wilmina.ac.jp,

³fujimori.atsushi@shizuoka.ac.jp, ⁴wilson@u-aizu.ac.jp, ⁵yoshimun@u-shizuoka-ken.ac.jp

Abstract

This study investigates the effects of three instruction modalities on English intonation learning. (i) watching haptic video, (ii) doing a punching gesture, and (iii) viewing intonation contours. The intonation of their oral reading was compared in a pretest/posttest design. The results suggest that all instructions were effective in moving towards the canonical pitch shape. However, pitch range expansion was significant only for males.

Keywords: Intonation, EFL, prosody, kinesthetics, pronunciation.

1. Background

Intonation plays a vital role in conveying important information in both monologues and dialogues, but methods of teaching and learning intonation are far from being established in EFL classrooms. Research has not provided an extensive list of what to teach in pronunciation, but we could find key features of what students need to learn [1]. For adult L2 learners, mastery of appropriate intonation requires the integration of the L2 linguistic knowledge and pronunciation skills, since intonation is interrelated with phonology, syntax and semantics in a non-random way. Thus, wrong intonation brings misunderstandings or misinterpretations, which can lead to issues of 'comprehensibility'; listeners find an utterance is difficult to understand [2].

It has been noticed that flat intonation is a tendency/typical characteristic of Japanese EFL learners' spoken English, but its relation to perception and comprehension has not been fully understood. Previous research on prosodic focus [3, 4], however, found that Japanese EFL learners' ability to perceive and comprehend exceeded that of production; and showed high scores on a perception test (i.e., in listening, they were assigned to identify which word was prominent) and comprehension test (i.e., in silent reading, they were assigned to identify which word be prominent semantically), but production test (i.e., they were assigned to read the text orally) was marked

low; their vocalization of focus words were not prominent enough to stand out in sentences. The results suggest that during L2 acquisition there may be a stage when learners fail to map their comprehension of focus words onto phonetic implementation; we speculate that learners at this stage have the concept of focus in discourse, but its conceptual understanding is not delivered in a proper way.

In EFL settings like in Japan, where the necessity of using English is extremely limited, students need appropriate pedagogical support and experiential learning in the classroom. In fact, various types of visual and auditory instructions with technological aids have been found helpful for EFL learners to improve their English intonation [5~8]. However, few studies have dealt with use of 'kinesthetic' instruction - involving a specific physical movement - in pronunciation training [9~11]. Furthermore, multimodal instructions are intuitively attractive, while little research has compared different types of instructions [12].

This study examines the effects of three instruction modalities including kinesthetic instructions on L2 English intonation patterns. Our research question was: Which type of instruction is the most effective for Japanese EFL learners to improve their English intonation contours? To answer this question, we gave them different instructions: (i) watching kinesthetic gesture, (ii) doing kinesthetic and (iii) viewing intonation. The intonation in their oral reading was compared on pretest/posttest design.

2. Procedure

2.1. Participants

Three English classes for first year students (G[r]oup1, G2, G3) at a Japanese university were used for this study, and 86 (average TOEIC level 380-420) signed the consent form. Twenty six students' data were excluded from the research due to not participating in either the pretest and posttest. Therefore, the total number of students analyzed was

60. Group 1 (n=17, 12 males and 5 females) was assigned to watch a kinesthetic video. Group 2 (n=27, 11 males and 16 females) was assigned to do kinesthetic practice while watching a kinesthetic video. Group 3 (n=16, 12 males and 4 females) was assigned to view computer images of intonation contours. Seven native speakers of North American English (n=7, 4 males, 3 females) recorded their oral reading as a control group.

2.2. Selection of Oral Reading Script

The material of the oral reading test was adapted from a Japanese folktale called *Momotaro* “Peach Boy”. An English version [13] exists, but in this study we removed all occurrences of ‘peach’, and replaced it by ‘melon’ for phonetic reasons.

Sugito [13] used the English version of the *Momotaro* story for a narrative production experiment to two groups of native speakers of Japanese and native speakers of English, to see how focus words were phonetically realized. Her primary finding is that English speakers raised pitch at the head of the noun phrase (e.g., ‘She found a big PEACH’), while Japanese speakers raised pitch at the adjective of the noun phrase (i.e., She found a BIG peach). She commented that Japanese pitch raising on the adjective may have resulted from a transfer from Japanese phonology; In a sequence of an adjective and a noun, the adjective (i.e., OOKINA momo [BIG peach]) has higher pitch in Japanese. Although both groups recognize the noun phrase as the focus in the sentence, higher F0 appeared on different words.

However, the head noun ‘peach’ starts with an aspirated voiceless stop consonant [p], which native speakers of English regularly aspirate at word-initial stressed position. The aspiration comes with a glottal pulse, which potentially raises the F0 on the following vowel. Thus, it is not certain what part of the pitch is due to prosodic considerations and which is a phonetic artifact. Furthermore, because Japanese tend to aspirate much less, it is difficult to tease out the relevant facts. For this reason we replaced ‘peach’ with ‘melon’ yielding the text:

Once upon a time, there lived an old man and an old woman. The old man went to the mountain to gather twigs, and the old woman went to a stream to do the washing. When the old woman was washing clothes, she saw a big melon floating towards her on the water. She picked up the melon and came home with it. The old man and woman cut open the melon and found a boy inside. They named him *Melontaro*.

It was typed in one paragraph, printed out in black ink on A4 paper, and we had students read it for the pretest and posttest.

2.3. Materials and Instructions

An experiment was performed using a pretest/posttest design. Students belonged to either G1 (watching kinesthetic video), G2 (doing air-boxing), or G3 (viewing intonation on Praat). The students were asked to read aloud the story of *Melontaro* before and after training. The three classes were assigned to go through a series of three different training sessions (see sec. 3 for the details).

Thus we created four kinds of material with each group accessing a slightly different set (Figure 1).

	G1 watching gesture	G2 Air- boxing	G3 Viewing intonation
Script (black & white)	✓	✓	✓
Audio (wav)	✓		✓
Video (mp4)	✓	✓	
Annotated Intonation on Praat (ppt)			✓

Figure 1: Summary of treatments for the 3 groups

For all classes, a raw script of *Melontaro* was distributed for pretest (recognition test and oral reading) and posttest (oral reading).

A audio (.wav) file of an oral reading was played through a classroom audio speaker to G1 and G3. It was read by a native speaker of North American English who had recorded it with a condenser microphone in a sound-attenuated room.

A video showing gestures was presented on a projector screen for G1 and G2. A same speaker in the audio acted in the video, so that all students in different groups hear the same voice. We adopted an air boxing gesture called “Fight Club” [9, 10, 12], in order to show the prominence of those focus words (see Figure 2).

<i>big</i>		<i>mel</i>	<i>on</i>
peak	retraction & preparation	peak	retraction

Figure 2: Video demo of air-boxing with *Melontaro*

The air boxing gesture is ideal for this task because each punch can coincide with a specific stressed syllable. In Acton’s system there is another haptic instruction called “Touchinami”, a blend of the English word “touch” and the Japanese word “tsunami”, where a speaker uses a right hand to draw intonation shapes in the air. It has several movement

patterns used to express different intonation contours. Although this approach could potentially contribute more to the performance of intonation, it was not used at this time because the researchers did not have the materials prepared for this activity. In order to identify the kinesthetic effect on vocalization of focus and non-focus words, it was sufficient to use different gestures for focus and non-focus words. The air-boxing consists of clear stages of (i) preparation stage, (ii) peak (i.e., maximum extension of an arm) and (iii) retraction stage [14], so that the fists should be able to hit the peak vowel of the stressed syllable of focus words. Keeping this in mind, we created a 45 second video clip [15], where a native speaker of English (i.e., the second author) punches the air at every focus word in all sentences, while reading a narrative of Melontaro.

For G3 (viewing intonation on Praat), the material was created in Praat, with the use of the audio (.wav) file and its annotated textgrid (Figure 3).

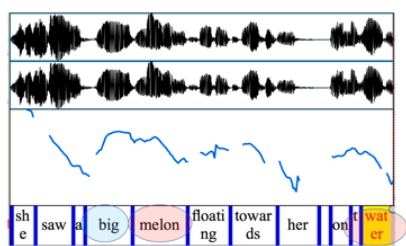


Figure 3: Praat image with annotation

The temporal frequencies of sound waves and pitch contours were made visible. In addition, the signals were segmented into words at the interval tier on the Praat images, and the intervals between boundaries were labeled with words. After this annotation, focus words were highlighted. Then each sentence (or a clause) was zoomed in, and screenshot of the whole frame was taken. This procedure was done for the whole wav file, and each screenshot was pasted onto slides of a Powerpoint (.ppt) file.

3. Experimental Design

On Week 1, a recognition test was conducted before a production test, in order to get students to familiarize themselves with the story of Melontaro before their oral recording. Students were given the script on a sheet of paper (without any marks), and were asked to underline words which they thought would be emphasized in oral reading. Once it was done, all the papers were returned to the instructor. Each student was asked to sit in front of a desk where a microphone (Blue Yeti USB microphone) stood, and orally read the story of Melontaro. Their production was saved in .wav format.

On Weeks 2-4, all the groups were given the same 10 minute explanation, followed by different oral practice.

The training sessions consisted of 3 weeks. In order to get students engaged in practice for a short period, the story was divided into 3 sections in advance, and only two sentences (in each section) were presented to them on each week (see sec.3).

In each sentence, students were given an explanation about the following points:

- (i) pauses and phrasing, focus words, and its relation to the loudness, duration and pitch
- (ii) recurrent mistakes Japanese students make, such as for “picked” they use *[-kd], or *[-kudo] instead of the desired [-kt])
- (iii) linking, reduction, and deletion

This was done to raise their phonological awareness, and to get them ready to read it orally.

Then, each group was given a different instruction by the same instructor. G1 and G2 were asked to just watch the gesture and facial expressions given in the video once, and discussed when they think the speaker punches, and how his facial expression changes and when, and were notified that punches occur at every focus word. Then G1 students orally shadowed 3 times, while G2 students mirrored the speaker’s gesture and utterance simultaneously 3 times. G3 was asked to view pitch contours shown in Praat first, and were notified that pitch rises and falls at every focus word. Then G3 students watched it and simultaneously pronounced it for 3 times.

On week 5, a production test was conducted and participants’ oral reading was recorded in the same way we did in the pretest.

4. Analysis

The target sentence for analysis was “... she saw a big melon”, where Sugito pointed out that there is a difference in the location of pitch raising between English native speakers and Japanese speakers.

4.1. Recognition Test

The result of recognition test showed that 100% of native speakers marked “melon” as a focus word of the sentence, but only about 24% of Japanese did so. The answer ranked top (43%) among Japanese was “(a) big melon” that was stretched over the whole noun phrase. Clearly native speakers recognize only the head noun as the focus word, as the Nuclear Stress Rule suggests [16], where nuclear stress should fall on last accented syllables in the phrase. But it seems not many of Japanese students recognize this rule. This fact is not incompatible with Sugito’s intuition that Japanese speakers phonetically emphasize the whole noun phrase while native English speakers

focus on the head noun. It can be expected that L1 prosodic pattern transfers to L2 English production.

4.2. Intonation Pattern

Pitch was measured using Praat. We searched for the stable portion of the vowel formants of each word, and selected the area from 25% to 75% of each of these stable portions of the vowel, and got the pitch average. We set semitones re 1Hz to enable comparison across speakers regardless of individual differences (see [17] for the justification of the use of semitone scale). It turned out only 2 out of 7 native speakers of English show the highest prominence on “melon” and the rest of them show on “big”. In spite of this variation, their intonation unanimously presents as pitch declining from the verb ‘saw’ to the object article ‘a’ and rising from the article ‘a’ to the adjective ‘big’, showing a trough pattern (Figure 2).



Figure 4: A trough pattern of native speakers' samples.

Destressing of the article ‘a’ is expected due to the typical reduction of function words, and the higher prominence of NP than V in VP is also expected from the Complement Law [18], which ensures that in a head and complement pair of words, main stress falls on the complement.

The number of students who satisfied those two conditions – “a” is lower than both sides, and “big” is higher than “saw” – was counted in pretest and posttest, to see whether this prosodic pattern was produced after each instruction group. All the groups showed a significant increase in the number of students who achieved this intonation pattern, as shown in Figure 3.

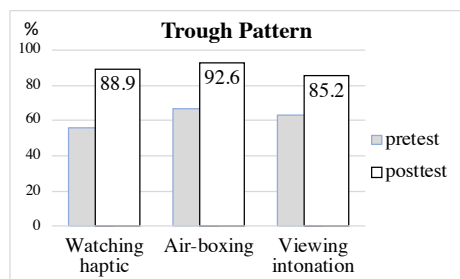


Figure 5: Achievement of the trough pattern.

The results indicate that the number of students who passed this criteria significantly increased in all groups.

4.3. Pitch Range

4.3.1. Across Different Instructions

In order to see whether there is any difference among those groups, pitch range was calculated between the article ‘a’ and the adjective ‘big’. In general, the larger the pitch range is, the more prominent the focus word should sound. Thus the dependent (paired) sample t-tests were conducted for all the groups. Only G1 and G3 showed a significant improvement after the training.

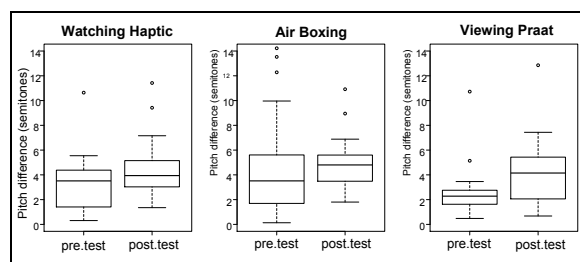
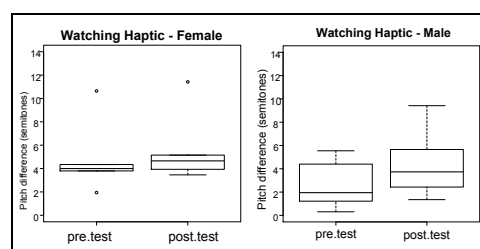


Figure 6: Pitch range in “a big melon” for G1 Watching haptic, G2 Air boxing and G3 Viewing Praat

As for the group of watching haptic, there was a significant difference in the pitch range for pre-test ($M=3.27$, $SD=2.51$) and post-test ($M=4.56$, $SD=2.69$) conditions; $t(16)=-2.07$, $p = 0.03$. As for the group of air-boxing, there was no significant difference in the pitch range for pre-test ($M=6.02$, $SD=4.4$) and post-test ($M=4.73$, $SD=1.9$) conditions; $t(15)=0.94$, $p = 0.18$. As for the group of viewing Praat, there was a significant difference in the pitch range for pre-test ($M=2.55$, $SD=1.89$) and post-test ($M=4.28$, $SD=2.59$) conditions; $t(26)=-5.66$, $p < .0001$. These results suggest that viewing haptic and viewing Praat, but not air-boxing, do have an effect on the expansion of the pitch range.

4.3.2. Male in G2

In order to see whether there is any difference between genders, the dependent (paired) sample t-tests were conducted for gender-segregated data in each group. The results showed that the male in all groups improved their pitch rise after the training, while female did not.



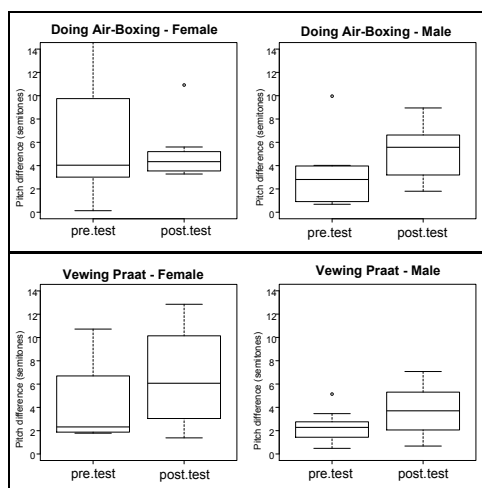


Figure 7: Pitch range expansion for female and male for all groups.

A paired-samples t-test was conducted to compare the pitch range of the sequence “a big melon” in pre-test and post-test conditions. There was *no* significant difference in the pitch range in any of the FEMALE groups: (i) in G1: for pre-test ($M=3.52$, $SD=3.32$) and post-test ($M=5.87$, $SD=3.25$) conditions; $t(3)=-1.46$, $p=0.12$, and (ii) in G2: for pre-test ($M=6.02$, $SD=4.94$) and post-test ($M=4.73$, $SD=1.84$) conditions; $t(15)=0.94$, $p=0.18$, and (iii) in FEMALE: for pre-test ($M=4.38$, $SD=4.31$) and post-test ($M=6.31$, $SD=4.85$) conditions; $t(2)=-1.4$, $p=0.12$.

In contrast, a significant difference was attested in all MALE groups: (ii) in G1: for pre-test ($M=2.63$, $SD=1.83$) and post-test ($M=3.98$, $SD=2.44$) conditions; $t(10)=-2.65$, $p=0.01$, (ii) in G2: for pre-test ($M=3.36$, $SD=2.75$) and post-test ($M=5.66$, $SD=2.11$) conditions; $t(8)=-3.98$, $p=0.01$, and (iii) in G3: for pre-test ($M=2.25$, $SD=1.03$) and post-test ($M=3.9$, $SD=1.90$) conditions; $t(21)=-5.05$, $p<0.0001$.

These results suggest that all treatments contributed to enhancing the pitch range for male. Females had already showed a relatively larger pitch range than males in pretest, and the number of females are smaller than males, which may be the reason that we did not get the significant difference. The physical difference in gender and the gender unbalance in participants were beyond our control.

What was interesting is that it is G2 male group who made a largest pitch expansion from pretest ($M=3.36$ semitones) to posttest ($M=5.66$ semitones). In G2 training sessions, an instructor also observed that once they saw the video, some students started to smile, and then according to the instructions, they kept punching through the end with smiling. Also in the second and the third sessions, as soon as the video was played, some students started punching even before they were told to mirror the action in the video.

It has been noted that pronunciation learning may be more advantageous to females, who tend to activate both hemispheres during language processing (see [19] and references therein for MRI evidence of female bilateral processing). Thus in a classroom with mixed genders it may be a challenge for instructors to ensure the learning effective and pleasant for both populations. A good news is that the kinesthetic task of punching seems to help male students' learning.

5. Conclusion

All of the three different instructions, namely, watching haptic videos, doing haptic/kinesthetic were effective in reading English sentences in appropriate F0 contour.

The expansion of the pitch range was attested in males for all groups. The largest pitch expansion was observed in the group of males who received kinesthetic instructions.

This paper showed haptic effect for air-boxing on the F0 pattern only, but its effects on other prosodic factors such as duration, intensity, pause, and voice quality should also be looked at. Also, these different prosodic features may be enhanced by different types of gestures such as iconic gestures of mimicking the aspects of the prosody as in the “touchinami” mentioned above. A survey of students' needs, research of individual differences, and strategies of raising phonological awareness should also be integrated in L2 pronunciation learning program.

6. Acknowledgement

The authors would like to thank the audience at ISAPh and the organizing committee members for their invaluable comments, reviews and assistance.

7. References

- [1] Derwing, T. 2018. Comprehensibility. Ms. University of Alberta.
- [2] Derwing, T. M., Munro, M. J. 2015. *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research* (Vol. 42). John Benjamins Publishing Company.
- [3] Yoshimura, N., Fujimori, A., Shirahata, T. 2015. Focus and prosody in second language acquisition. *Journal of International Relations and Comparative Culture* 13, University of Shizuoka, 21-36.
- [4] Yamane, N., Yoshimura, N. Fujimori, A. 2016. Prosodic transfer from Japanese to English: Pitch in focus marking, *Phonological Studies* 19, Phonological Society of Japan, 97-104.
- [5] Anderson-Hsieh, J. 1992. Using electronic visual feedback to teach suprasegmentals. *System*, 20(1), 51-62.

- [6] Levis, J. and Pickering, L., 2004. Teaching intonation in discourse using speech visualization technology. *System*, 32(4), 505-524.
- [7] Wilson, I. 2008. Using Praat and Moodle for teaching segmental and suprasegmental pronunciation. In 3rd International WorldCALL Conference: *Using Technologies for Language Learning* (WorldCALL 2008).
- [8] Fujimori, A., Yoshimura, N., Yamane, N. 2015. The development of visual CALL materials for learning L2 English prosody. In Conference proceedings. *ICT for Language Learning*. libreriauniversitaria. it Edizioni, 249.
- [9] Acton, W., Baker, A. Ann., Burri, M., Teaman, B. 2013. Preliminaries to haptic-integrated pronunciation instruction. In J. Levis & K. LeVelle (eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Ames, United States: Iowa State University, 234-244.
- [10] Burri, M., Baker, A., Acton, W., 2016. Anchoring academic vocabulary with a "hard hitting" haptic pronunciation teaching technique. In T. Jones (ed.), *Pronunciation in the Classroom: The Overlooked Essential*. Alexandria, United States: TESOL Press, 17-26.
- [11] Baker, A. 2012. Integrating pronunciation into content-based ESL instruction. Paper presented at the 4th annual Pronunciation in Second Language Learning and Teaching Conference, Vancouver, BC.
- [12] Teaman, B., Yamane, N. 2017. Haptic techniques for teaching English prosody. *Proceedings of the 31st General Meeting of the Phonetic Society of Japan*.
- [13] 杉藤美代子. 2003. 声にだして読もう!: 朗読を科学する. 明治書院
- [14] Kita, S., Van Gijn, I., Van der Hulst, H. 1997. Movement phases in signs and co-speech gestures, and their transcription by human coders. In *International Gesture Workshop*. Springer, Berlin, Heidelberg. 23-35.
- [15] Yamane, N. (2018, September 30) *MELONTARO* [Video file]. Retrieved from https://www.youtube.com/watch?v=DVe_GMUJI5Y&feature=youtu.be
- [16] Chomsky, N., Halle, M. 1968. *The Sound Pattern of English: Morris Halle*. Harper & Row.
- [17] Nolan, F. 2003. Intonational equivalence: an experimental evaluation of pitch scales. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* (Vol. 39).
- [18] Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., Donati, C. 2008. Different phrasal prominence realizations in VO and OV languages. *Lingue e linguaggio*, 7(2), 139-168.
- [19] Moyer, A. (2016). The puzzle of gender effects in L2 phonology. *Journal of Second Language Pronunciation*, 2(1), 8-28.