



## The Development of Visual CALL Materials for Learning L2 English Prosody

<sup>1</sup>Atsushi Fujimori, <sup>2</sup>Noriko Yoshimura, <sup>3</sup>Noriko Yamane

<sup>1</sup>Shizuoka University (Japan), <sup>2</sup>University of Shizuoka (Japan), <sup>3</sup>University of British Columbia (Canada)

### Abstract

*This study aims at developing CALL materials to facilitate the acquisition of prosody by EFL learners, particularly examining whether the recent technology of speech visualization can be a medium for teaching prosody. Forty Japanese EFL learners at CEFR level A2 participated in our experiment. One group was given oral instructions explicitly telling where to put the highest pitch in a question-answer sequence. The other group was shown Praat images of model speech where the highest pitch was clearly visible, while listening to the audio stimuli. Before and after the three series of 10-minute instruction sessions, they were asked to read several question-and-answer dialogs aloud in pairs. Their utterances were recorded in Audacity and their pitch was measured in Praat. The results of the pretest indicate that prosodic prominence was placed inappropriately in sentence-initial position and the pitch range of the entire utterance is much smaller compared to native speakers of English. At the posttest, the students' performance greatly improved in producing prosodic focus marking and in pitch range. These results suggest visual instructions are as effective as oral instructions in learning L2 prosody. Such visual instructions can be used on ubiquitous devices to facilitate L2 learners' self-access learning.*

### 1. Introduction

Prosodic properties of speech, including pitch, intensity and duration, help listeners identify semantically and pragmatically salient elements of an utterance such as question and focus (Healey, 2003) [1]. However, L2 learners have difficulty in producing prosodic focus marking due to crosslinguistic variations in implementing information focus (Gut & Pillai, 2014) [2]. The recent development of computer technology such as speech visualization can help L2 learners properly produce not only segmental features (Wilson, 2005) but also prosodic features (Hardison, 2004; Levis & Pickering, 2004), as a means of giving them feedback of their own speech [3][4][5]. In the present study, we examine whether such technology can also be a medium of teaching prosody, focusing on Japanese-speaking learners of English.

### 2. Background

#### 2.1. Focus marking and prosody

Information focus is a non-presupposed part, and semantically the most salient constituent, of a sentence, observed as an answer to the *wh* constituent in a *wh*-question (Zubizarreta, 1998) [6]. How focus is marked differs between English and Japanese. English focus is encoded phonologically and the placement of prominence is flexible and context-dependent, as in (1) and (2).

- (1) a. What did John eat?  
b. He ate the [cake]<sub>F</sub>.
- (2) a. Who ate the cake?  
b. [John]<sub>F</sub> ate it.

The sentence structures of (1b) and (2b) are same, but the highest pitch is placed on the post-verbal noun in (1b), and the pre-verbal noun in (2b). In either case, focus is marked with high prominence. On the other hand, Japanese focus can be encoded morphologically with a case marker *ga* (Kuno, 1973; Heycock, 2008), as in (3b) [7][8].

- (3) a. (Kyodai-no naka de) dare-ga dokushin desu ka  
'(Among your brothers), who is single?'  
b. [Taroo]<sub>F</sub>-GA dokushin desu.  
'Taro is single.'



The focus can also be marked prosodically. In Japanese, the highest pitch tends to be placed on the sentence-initial word by default and the pitch goes down towards the end of a sentence (Pierrehumbert & Beckman, 1988) [9]. In the case of (3b), this “down-stepping” (H\*L) contour overlaps with the sentence-initial word in focus.

## 2.2. L2 studies

We first take a brief look at the results of a study on the acquisition of prosodic focus marking in L2 English. Nava (2008) investigates ten L1 Spanish-L2 English participants' oral production in a question and answer experiment [10]. In Spanish, focus appears with prosodic prominence at the right edge of the sentence, as in (4b).

- (4) a. ¿De qué te ríes?  
At what you laugh-PRS-PROG  
'What are you laughing at?'
- b. ¡Un pingüino está [bailando]!  
A penguin be-PRS-3SG dance-PROG  
'A penguin is dancing.'

If an L1 transfer effect occurs, it is predicted that the participants would incorrectly put prosodic focus on the final word in the L2 utterance. The results show that Spanish learners of both high and low proficiency preferred placing prosodic prominence sentence-finally in L2 English, as shown in (5b).

- (5) a. Why are you looking out the window?  
b. Madonna just walked [by]<sub>F</sub>! (L1 Spanish-L2 English)  
c. ([Madonna]<sub>F</sub> just walked by! (L1 English)

If such context-dependent prosody is affected by L1 transfer, it is natural to ask whether there is an effective way of instruction for L2 prosody. Hardison (2004) conducted an experiment with a pretest-posttest design to examine the effectiveness of computer-assisted prosody learning. Sixteen English-speaking low-intermediate learners of French participated in 13 practice sessions where they were individually asked to read a set of French sentences aloud at a conversational rate. The students received auditory (hearing their utterance) and visual (seeing their pitch contour on a screen) feedback in real time. The pitch contour of a model's speech was also displayed on the same screen, so that the participants could compare the model pitch contours with their own. The pretest-posttest comparison indicated that their French prosody significantly improved after the training with audio-visual feedback, and their utterances sounded intelligible to native speakers of French.

Given the discrepancies between English and Japanese on focus marking, as in (1-2) vs. (3), the present study investigates:

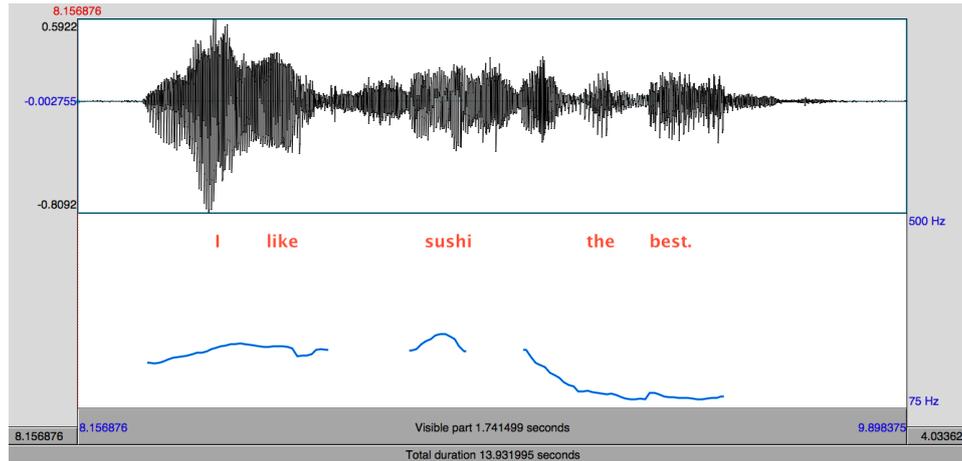
- I. Whether there is an L1 effect on production of prosodic focus marking,
- II. Whether speech visualization is effective in teaching prosodic focus marking
- III. To what extent speech visualization helps improve L2 prosody.

To examine these research questions, we conducted a production task in a pretest-posttest design.

## 3. Methodology

Forty Japanese university students whose major was science participated in this experiment. They had learned English for six years at school before entering university and their English proficiency was at CEFR level A2. They were divided into two groups of twenty participants each. Their average TOEIC scores were 467.0 and 460.5, respectively, and there was no significant difference between the groups ( $t(48)=-.076, p<.785$ ).

The first group was given traditional oral instructions. The instructor confirmed that the participants correctly identified focus in comprehension, namely, which word was an answer to the *wh* element a question-answer sequence. Then the instructor told the participants the rule that the focused word was phonetically prominent in English. The other group was shown Praat images of model speech (e.g. figure 1) after confirming their comprehension of focus. In the images, pitch curves clearly indicated that the focused word was phonetically salient. Both groups participated in three 10-minute sessions where they practiced reading test dialogs (6)-(8) aloud in pairs while listening to recordings of model speech. Both groups were also given a pretest and a posttest. They were asked to play the role of speaker A or B in the dialogs and to read them aloud at a conversational rate. Their utterances were recorded in Audacity and the intrinsic frequency (F0) of each vowel was measured in Praat. There could be more than one highest peak in an utterance if their difference is less than 5 Hz.



**Figure 1** Praat image of model speech (token 2)

- (6) A: Shall we go fishing tomorrow?  
 B: That's a good idea. What time shall we meet?  
 A: Let's meet at [six]<sub>F</sub>. [Token 1]
- (7) A: Do you like Japanese food?  
 B: Yes, I like it very much.  
 A: What kind do you like?  
 B: Oh, I like [sushi]<sub>F</sub> the best. It's popular in Australia these days. [Token 2]
- (8) A: What are you doing?  
 B: I'm making a [windmill]<sub>F</sub> which really works.  
 A: I think windmills are quite useful for our future. [Token 3]

## 4. Results

Table 1 shows the number of utterances where the highest pitch was properly placed on the focused word. The overall correct production rate of the oral instruction (OI) group was 60% at the pretest and it improved to 70% at the posttest. The visual instruction (VI) group performed poorly (20%) at the pretest but their correct production rate almost reached 70% at the posttest.

	OI group		VI group	
	Pretest	Posttest	Pretest	Posttest
Token 1	8	7	1	8
Token 2	8	9	5	9
Token 3	2	5	0	3
Total	18 (60%)	21 (70%)	6 (20%)	20 (67%)

**Table 1** Correct production rates

Nearly 50% of the OI group incorrectly placed the prosodic prominence on the sentence-initial word at the pretest while the incorrect production rate reduced to approximately 20% at the posttest, as seen in table 2. In the VI group, the incorrect production rate was lowered from over 70% to less than 30%.

	OI group		VI group	
	Pretest	Posttest	Pretest	Posttest
Token 1	7	3	10	2
Token 2	5	2	6	2
Token 3	2	0	6	4
Total	14 (47%)	5 (17%)	22 (73%)	8 (27%)

**Table 2** Production rates of sentence-initial prominence



Table 1 shows that some of the participants performed well even in the pretest. Take token 2 for instance. Most of the participants in both groups correctly produced the prosodic prominence. Yet we can observe improvements in their prosody. In the posttest, the focused word was much higher in pitch than that in the pretest. As shown in table 3, the median of the focus F0 difference between the two tests was over 20 Hz while that of the subject F0 difference was less than 20 Hz, which leads to the relative lowering of F0 of the sentence-initial word. Moreover, the post-focal words such as *the* and *best* were produced with a lower F0 in the posttest than in the pretest.

	<i>I</i>	<i>like</i>	<i>[sushi]<sub>F</sub></i>		<i>the</i>	<i>best</i>
Group	/aɪ/	/aɪ/	/ʊ/	/i/	/ə/	/ɛ/
OI	16.9	10.3	45.1	22.7	0.5	-8.23
VI	12.2	12.6	24.5	23.6	-8.2	-5.7

**Table 3** Mean F0 differences (Hz) between the tests (token 2)

As a consequence, the pitch range from the focus F0 to the lowest F0 was expanded across the tokens at the posttest in both groups, as in the model speech, as shown in table 4.

	OI group		VI group		Model speech
	Pretest	Posttest	Pretest	Posttest	
Token 1	27.7	34.6	12.7	48.3	59.3
Token 2	40.3	75.7	25.3	80.6	127.9
Token 3	12.0	40.1	23.1	39.4	116.1

**Table 4** Pitch ranges from the focus F0 to the lowest F0 (Hz)

## 5. Discussion

The results of the pretest indicate L1 transfer effects. The Japanese EFL learners did not have the highest F0 on the focused word within the sentence, while they frequently produced the sentence-initial prominence, which is common in L1 Japanese. The results of the posttest suggest that speech visualization is effective in teaching prosody. The learners who received visual instructions improved in producing prosodic focus marking, like those who received oral instructions. The maximum F0 was increased while the minimum F0 was lowered. Therefore, the pitch range was greatly expanded to mirror English prosody. Furthermore, the sentence-initial prominence was observed less often after the instruction session.

A remaining problem is that the learners had difficulty in uttering token 3 with proper prominence in the posttest, as seen in table 1. We assume that the difficulty is related to structural complexities. In tokens 1 and 2, information focus appears in a simplex sentence while token 3 contains a relative clause, as in (9).

- (9) a. Let's meet at [six]<sub>F</sub>.  
 b. I like [sushi]<sub>F</sub> the best.  
 c. I am making a [windmill]<sub>F</sub> [which really works]<sub>RC</sub>.

A pause needs to be inserted immediately before the relative pronoun *which* in (9c). The learners did not learn the fact, nor did they put any pause while uttering the complex sentence. Hence, they could not have an abrupt pitch rise on the focused word, compared with the other tokens. However, a sign of improvement can be seen in the pitch contours of the VI group, as in figure 2. The sentence-initial peak declined in the posttest. The pitch pattern also showed several F0 peaks with an apparent pitch rise on the focused word. We anticipate that the pitch rise on the focused word will be more salient if the learners acquire phrase boundary pauses.

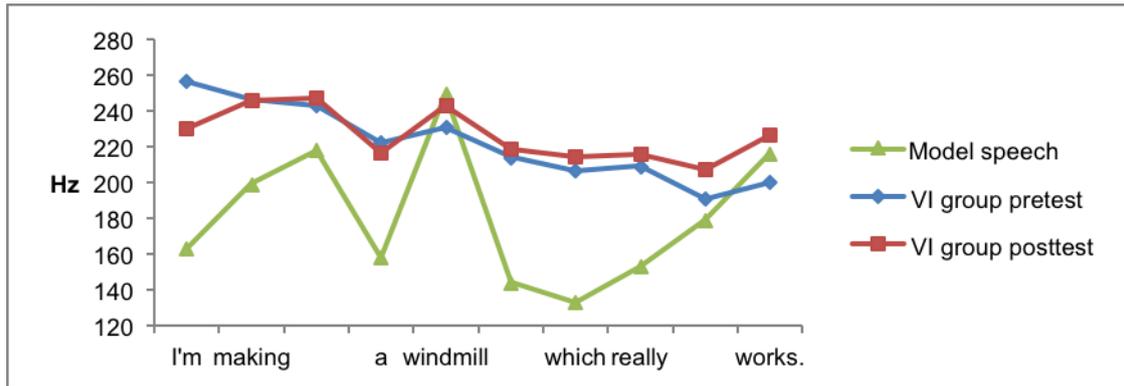


Figure 2 Median pitch contours of VI group (token 3)

## 6. Conclusion

Based on these results, we conclude that speech visualization is an effective way of teaching prosodic focus marking. It should be emphasized that such visual instructions have great potential to be implemented on ubiquitous devices such as laptops and smartphones to facilitate L2 learners' self-directed learning. However, notice that we dealt with a single linguistic phenomenon at the prosody-discourse interface. Given that prosody is also associated with other linguistic components such as syntax, as briefly mentioned in the discussion, we continue to explore effective ways in which computer technology can improve L2 teaching and learning.

## References

- [1] Healey, N. (2003). A CALL system for teaching pronunciation. *JALT Journal* 1: 25-36.
- [2] Gut, U., & Pillai, S. (2014). Prosodic marking of information structure by Malaysian speakers of English. *SSLA* 36: 283-302.
- [3] Wilson, I. (2005). Using Praat and Moodle for teaching segmental and supra segmental pronunciation. *TESOL Quarterly* 3: 33-43.
- [4] Hardison, D. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology* 8: 34-52.
- [5] Levis, J., & L. Pickering. (2004). Teaching intonation in discourse using speech visualization technology. *System* 32: 505-524.
- [6] Zubizarreta, M. L. (1998). *Prosody, Focus, and Word Order*. Cambridge, MA: MIT Press.
- [7] Kuno, S. (1973). *The structure of the Japanese language*. Cambridge, MA: MIT Press.
- [8] Heycock, C. (2008). Japanese -wa, -ga, and information structure. In S. Miyagawa & M. Saito (Eds.), *The Oxford Handbook of Japanese Linguistics*, 54-83. New York: Oxford University Press.
- [9] Pierrehumbert, J., & Beckman M. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.
- [10] Nava, E. (2008). Prosody in L2 acquisition, *Proceedings of the 9<sup>th</sup> Generative Approaches to Second Language Acquisition Conference*, 155-164.