Problems of macro-prosodic control by Japanese EFL learners when reading an English passage

Toshiko Isei-Jaakkola¹, Keiko Ochi², Keikichi Hirose³

¹Department of Education, Chubu Gakuin University, Japan, ²Kyoto University, Japan, ³Center for Robust Intelligence and Social Technology, NII, Japan
¹toshikoij@gmail.com, ²ochi.rh.02011@gmail.com, ³hirose@gavo.t.u-tokyo.ac.jp

Abstract

Our aim in this study is to determine how well correlated macro-prosodic features such as pitch and intensity movements are for Japanese learners of English as a foreign language (JEFL) when reading a passage, in comparison with English speakers (EL1). For this purpose, we recorded four EL1s and five JEFLs and analyzed a total of 37 story utterances after discarding the speakers and utterances with hesitation, general instability, etc. Pitch and intensity movements were determined using only the voiced parts of each utterance. The results showed that JEFLs’ macro-prosodic pitch and intensity controls were not statistically correlated. JEFLs’ registers were much narrower than those of EL1s, and F0 and intensity resets occurred much more frequently among JEFLs than EL1s. JEFLs’ prosodic boundaries were set at the word level and thus their prosodic units were much shorter than those of EL1s.

Keywords: Macro-prosodic control (pitch, intensity), an English passage reading, American EL1, Japanese EFL learners.

1. Introduction

1.1. Background

English and Japanese are different in many aspects: phonetically, phonologically (syllable structure), morphologically, and syntactically (word order). When spoken, their rhythm and timing are also very different, though Japanese and English both have accentuation. Japanese is a pitch-accented language with mora-timed rhythm, while English is a language with stress-timed rhythm. “Stress accent” often used for English may be determined by subjective auditory impressions, but the physical facts do not always agree with such a classification. In general, “stress accent” can be realized by both pitch and intensity. There are numerous comparative studies between English and Japanese of language learning and acquisition related to phonetic education. Often only pitch (F0) and duration are used in these studies when analyzing accentuation, even in English. Also, the correlation between pitch and intensity has not been sufficiently explained.

1.2. Aims

Although this is a pilot study, our aim is to determine how well correlated macro-prosodic features such as pitch and intensity movements are for Japanese learners of English as a foreign language (JEFLs) compared to English speakers (EL1s) when reading a passage. This analysis may clarify aspects of the difficulty JEFLs face in macro-prosodic realization of reading English and suggest how to overcome such difficulties.

2. Experiments

2.1. Experimental procedures

The text material used was an English fable entitled “The North Wind and the Sun.” (There are several versions of this story; we use the IPA version [1].) The JEFL subjects were five male university students (19–22 years old), majoring in English. All had studied English for four to eight months in the United States or Australia. There was no significant difference in their English proficiency, such as TOEIC test scores (500–600). There were four EL1 subjects: two male university students and two university instructors (20–45 years old), included together based in the judgment that there was no difference between the students and the instructors.

All subjects’ mother tongues were standard variants of their respective languages. We selected only males because we could not collect sufficient female subjects for both languages. The recording was conducted in the recording studio or in a quiet room of the first author’s university. All subjects read aloud the text materials repeatedly more than five times. A headset microphone was used to record the subjects’ voices. The distance (approx. 10 cm) between the participants’ mouths and the microphone was fixed to ensure that the intensity would remain constant during the recording (sampling rate: 48 kHz, 10.21437/ISAPh.2021-7
The subjects were given the material before the experiment and asked to read the story fluently. If they committed any error while reading, they were requested to start over in order to minimize the influence of speech style, speech rate, and general instability. Utterances with hesitation and speakers whose speech rates were too different from the others within the same language were discarded, leaving a data set that included 25 voice recordings by JEFLs and 12 by EL1s. We thus analyzed a total of 37 story utterances.

### 2.2. Analyses

We extracted the fundamental frequency and intensity utilizing Praat [2] as follows. The temporal frame step was 10 ms. The ceiling and floor of the F0 value were automatically set, referring to De Looze et al. [3], for each speaker. The individual differences of the mean and SD of F0 were removed by Z-score normalization within a single-story reading. In order to specify the boundaries between pause and non-pause segments, we carefully confirmed the syntactic and semantic boundaries and phonetic phenomena in the speech.

We first obtained the word boundary using a force alignment using the WebMAUS system [4] for English, and then calculated the root square mean (RMS) value for each segment with a window step of 10 ms and window duration of 20 ms, regarding any segment with RMS below a threshold as a pause segment.

The correlation (CRR) between F0 and intensity was then calculated by dividing the covariance by the root-mean variance using Equation (1), where \(X_i\) and \(Y_i\) are the logarithmic F0 and intensity of each voiced segment, and \(\bar{X}\) and \(\bar{Y}\) are the mean values within one read story.

\[
CRR = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2}(\sum(Y_i - \bar{Y})^2)}
\]

The correlations of the two subject groups (JEFLs and EL1s) were then compared using a t-test.

### 3. Results

#### 3.1. Correlation between pitch and intensity

There was a significant difference between JEFLs and EL1s in terms of the correlation between pitch and intensity \((p < 0.05)\).

#### 3.2. Pitch and intensity movements

Figures 1 and 2 exhibit the relationships between pitch (F0) and intensity in a passage read by an EL1 and a JEFL, respectively, including real time pauses. In Figure 1, the blue line shows pitch, while the same is shown by the red line in Figure 2. The intensity is shown by the yellow line in Figures 1 and 2. The vertical lines show word and pause boundaries in both figures. For Figure 1, an EL1 was selected who was a good model in the EL1 group, and a JEFL with poor pronunciation skills was selected for Figure 2. It is noted that this JEFL had very long pauses, while the EL1 had very short pauses apart from those between sentences. Thus, the pitch and intensity movements are comparable between these two language speaker groups. Figure 3 shows the mean normalized log F0 of each word for all EL1s (blue line) and all JEFLs (red line) of the whole utterance, and Figure 4 the range of normalized log F0 within each word for EL1s (blue line) and JEFLs (red line).

#### 3.2.1. Pitch movements

In terms of pitch movements in a whole passage (see Fig. 3), there was a large register difference between EL1s and JEFLs, such that EL1s had much wider registers than JEFLs. JEFLs’ registers frequently narrowed toward the ending of each word, as the ceiling became lower but the bottom remained unchanged. On the other hand, EL1s’ registers did not change; both the ceiling pitch and bottom pitch became lower.

EL1s made a pitch reset mostly at sentential, clausal, and phrasal boundaries, but JEFLs’ pitch resets were not at all identical to those of EL1s. Individual differences as identified by the standard deviation were larger in EL1s than JEFLs, despite the smaller numbers of EL1 subjects and total utterances.

#### 3.2.2. Intensity movements

In terms of intensity (amplitude) movements (see Figs. 1 and 2), the intensity was often attenuated toward the bottom line at a syllabic or lexical level among JEFLs, whereas among EL1s, intensity did so generally at a lexical level.

#### 3.2.3. Summary

(1) EL1s appeared to reset pitch and intensity at a sentential boundary, but this did not appear to hold for JEFLs.

(2) EL1s raised the pitch and intensity in tandem. Conversely, the JEFLs sometimes raised the intensity but lowered the pitch, or vice versa (see Figures 1 and 2).

(3) The JEFLs regularly controlled both pitch and intensity at an (English) syllabic level, often cutting them apart at pauses and making an F0 reset at the beginning of the following word.
Figure 1: Part of the F0 and intensity movements in NW for an EL1. Blue line = pitch, yellow line = intensity. Vertical lines = word and pause boundaries. P in green = pause.

Figure 2: Part of the F0 and intensity movements in NW for a JEFL. Red line = pitch, yellow line= intensity. Vertical lines = word and pause boundaries. P in green = pause.

Figure 3: Mean normalized log F0 of each word for EL1s (blue line) and JEFLs (red line). The vertical lines show the punctuational boundaries in the text: straight line = periods (sentences), dashed line = commas and semicolon.

Figure 4: Range of normalized log F0 within each word for EL1s (blue line) and JEFLs (red line). The vertical lines show the punctuational boundaries in the text: straight line = periods (sentences), and dashed line = commas and semicolon.
3.3. Prosodic boundaries

It was found in EL1s’ English that the punctuational boundaries (period, semicolon and comma), which suggest reading timing guided in a written form in the text, grammatical boundaries (such as sentential, clausal, phrasal, and lexical units), and prosodic units in the utterance focusing on F0 resets were not identical. In Figures 3 and 4, the vertical lines show punctuational boundaries: The straight lines represent periods (= sentential boundaries) and dashed lines a semicolon and commas in the text. Observing Figure 3, we note a very high F0 rise and F0 fall toward the bottom line of the whole F0 of the passage at the sentential boundary; and at the clausal boundary there was similar F0 movement, though smaller than that at the sentential boundary. There were similar pitch resets at clausal boundaries in the phrases or within phrases.

When comparing grammatical boundaries with prosodic boundaries, the following issues arose:

1. The semicolon in “around him; and at last” is realized as either a sentential boundary or a clausal boundary in the story. As there was a very low F0 before the semicolon, we judged there to be a sentential boundary between the preceding sentence and semicolon; (2) “The one” is not a clause but a phrase, as in this complicated sentence structure SV[O]: “They agreed [that the one [who first succeeded in making the traveler take his cloak off] should be considered stronger than the other].” Here there is a clause inside a clause. There was a very high F0 reset like that at a sentential boundary between “one” and “who.” Therefore, we assigned a clausal boundary between “the one” and the following “who”; (3) Before “wrapped” in the participle “wrapped in a warm cloak” there was a small F0 reset, but we judged that it was not sufficient to posit a clausal boundary between the preceding phrase and the participle; and (4) “In making the traveler take his cloak off” is a phrase, but there was an explicit F0 reset in “take his cloak off.” However, we decided it was not large enough to regard it as marking a clausal boundary between “traveler” and “take.”

Therefore, in our judgement, || marks the grammatical and cooccurring prosodic boundaries, and # the other prosodic boundaries in English uttered by EL1s as follows.

The North Wind and the Sun were disputing || which was the stronger, || when a traveler came along # wrapped in a warm cloak. || They agreed that # the one # who first succeeded # in making the traveler # take his cloak off # should be considered stronger than the other. || Then the North Wind blew # as hard as he could, || but the more he blew || the more closely did the traveler fold his cloak around him; || and at last # the North Wind gave up the attempt. || Then the Sun shined out warmly, || and immediately the traveler took off his cloak. || And so # the North Wind was obliged to confess || that the Sun was the stronger of the two.

Thus, there were more prosodic boundaries than grammatical boundaries and punctuation boundaries, i.e., prosodic boundaries > grammatical boundaries > punctuational boundaries in the text.

4. Conclusions

In conclusion, JEFs’ macro-prosodic pitch and intensity controls largely differed from those of EL1s (see Figure 4), and they were not statistically significantly correlated. JEFs’ registers were much narrower than those of EL1s. JEFs’ F0 and intensity resets occurred much more frequently than did ELs’. JEFs’ prosodic boundaries were set at a word level, and thus their prosodic units were much shorter than those of EL1s. We can judge that JEFs’ prosodic boundaries affect their frequent and long pausal manners (cf., for example, [5]). It is suggested that Japanese university students as learners of EFL should be trained to (1) widen their pitch range by raising the F0, particularly that of the primary stress, when the lexical stress of content words serves as a sentential stress for pitch control, (2) consciously move their lower abdominal muscles, as [6], for example, suggested for intensity control, and (3) avoid pauses at word boundaries for continuous expiratory power, resulting in less F0 and intensity resets and no lowering of pitch. For future studies, we will extend the investigation of the present study by adding more EL1s and JEFs (the same number for males and females in the same generation) and possibly by using other English passages.

5. References