Perception of Foreign Accentedness in L2 Prosody and Segments: L1 Japanese Speakers Learning L2 French

Kamiyama, Takeki

Laboratoire de Phonétique et Phonologie (UMR 7018) CNRS / Sorbonne Nouvelle, Paris, France
takelkk@phil.c.u-tokyo.ac.jp

Abstract

In order to examine the production of Japanese-speaking L2 learners of French, a series of perception tests were conducted with 17 native speakers of French (from mainland France). The subjects listened to French short phrases 1) synthesized with Mbrola using (European standard and Canadian) French and Japanese segments, combined with duration and F0 found in the recording of phrases read by Japanese learners and French native speakers, 2) read by Japanese learners and French native speakers, and then re-synthesized with manipulation of local duration and F0. The results indicate that duration and F0 play an important role in the perception of foreign accent.

1. Introduction

The role played by the suprasegmental features in the intelligibility and the naturalness of L2 learners' production has long been underestimated in the practice of language teaching. Recently, however, we have come to attach more importance to it (Dalton and Seidelhofer [11]) and have seen works that demonstrate the importance of suprasegmental features in L2 learners' speech (Suzuki [2] for F0, duration and intensity, Tajima et al. [3] for duration). This paper deals with the production of Japanese-speaking L2 learners of French.

Japanese is usually classified as a pitch-accent language. The words, except for some particles, are either accented or unaccented at lexical level. If the word is accented, the lexical accent is realized as a fall in F0 anchored to a specified mora in the word (Kubozono [4], among others). There is generally an F0 rise between the first two morae in the word. Intonation does not change much the underlying F0 pattern of the word: the accented words preserve the fall in F0, and there is usually no continuation rise (which is dominant in French; realized but less extensively in English).

Japanese is also classified as a mora-timed language from the rhythmic point of view. Although the acoustic reality of the isochrony of morae in its strict sense is denied (Beckman [5]), there are claims that some compensation mechanism is at work at the mora level (Kurematsu [6], Sagisaka [7]). At least, native speakers tend to segment, when they are asked to, at mora level, and it is the number of morae that count in poetry.

The target language French is considered as an intonation language, and there is no pitch pattern associated to words at lexical level. When embedded in a sentence, a word may be found with a rising, neutral or falling F0 contour, depending on its position in the sentence (the first word tends to be rising, the penultimate word rising, and the last one falling) and on its relation with the next word (the more independent, the higher the rising is). There is a rise at the end of each group, a major rise at a major node, and a minor one at a minor node (Vaissière [8], for example). This feature gives the language an acoustic impression of rising intonation, which contrasts with Japanese. The rising tendency in French and the falling one in Japanese are also observed in infants’ productions (two-syllable utterances of 18-month-old children: Hallé et al. [9]).

As far as rhythm is concerned, French falls into the “syllable-timing” category. Here again, the acoustic reality of syllable isochronicity is not found (Wenk and Woiland [10]), and final lengthening disrupts syllable isochronicity. Syllable is the basic unit in poetry and it gives at least an isochronous acoustic impression. It is syllable (and not mora) that corresponds to a strong cognitive reality in French.

There exists an important difference between the two languages in the duration domain, namely, the degree of sentence-final lengthening. Final lengthening is a widely observed phenomenon. In French the last syllable in rhythmic groups (or phrases) is lengthened, and in general, a (major or minor) continuation rise is superimposed on it. A longer duration, and an F0 peak (minor phrase) or an F0 rise (major rise) lead to the perception of an “accent de groupe” (Delattre [11]). In Japanese, a slight phrase-final lengthening is observed, but there is no sentence-final lengthening (Takeda et al. [12]). The difference in the importance of final lengthening thus observed in adults' speech in the two languages is also present in infants' speech, just as the case of F0 (disyllabic utterances of 18-month-old French and Japanese: Hallé et al. [9]).

Therefore, when considering L1 speakers of Japanese learning L2 French, it can be expected that the learners do not fully reproduce 1) the rising tendency in pitch at the end of internal phrases, more precisely, continuation rises, and 2) phrase- and sentence-final lengthening. These characteristics are indeed observed in some learners’ short read utterances (Kamiyama [13]).

The present study deals with the French native speakers’ perception of French produced by Japanese learners. In order to approach the issue, we conducted a series of perception tests using stimuli synthesized 1) with the diphone synthesizer Mbrola [14], modifying the quality of the segments (into European Standard French, Canadian French and Japanese phonemes) while keeping the prosody intact, and 2) with Praat [15] (PSOLA algorithm) for F0 and duration manipulations of original sentences, while keeping the segments intact.

2. Methodology

2.1. Stimuli

8 short phrases in French (Table 1) were read by 11 L1 Japanese speakers learning L2 French and by 4 native speakers of French. The learners were second-year university
students who had learned French for one academic year at a university in Japan. For each test 5 recordings were selected from those learners with non-native-like prosody, as well as the corresponding 5 phrases read by a native speaker.

Intensity adjustment: Since Mbrola does not take intensity into consideration, we adjusted the synthesized phrases to the intensity curve of the original recording.

Test 1: The duration of each phoneme was measured, and fundamental frequency (F0) was detected at 10 ms time steps. We then used Mbrola to create 6 types of stimuli combining two factors: 1) segments in French (fr), French Canadian (ca), Japanese (jp); 2) prosody (phoneme duration and F0) of Japanese learners’ productions (Pros JP) and that of French native speakers’ (Pros FR). (cf. Table 2)

Speech rate and pause: Learners tended to pronounce at a slower pace, and with frequent pauses. In order to test if these two factors alone influence the listeners’ judgement, we also selected one phrase and created a set of stimuli by 1) slowing down the whole phrase, 2) inserting a pause, 3) doing 1) + 2).

Speech rate and pause: Learners tended to pronounce at a slower pace, and with frequent pauses. In order to test if these two factors alone influence the listeners’ judgement, we also selected one phrase and created a set of stimuli by 1) slowing down the whole phrase, 2) inserting a pause, 3) doing 1) + 2).

Table 1: Phrases used in the tests.

|-----------------------|---------------------|----------------------|-----------------------------|------------------------|-------------------|-----------------------------|-------------------------|-------------------------------|

Table 2: 6 Combinations of segments and prosody for the stimuli used in Test 1.

<table>
<thead>
<tr>
<th>Segments fr</th>
<th>Segments ca</th>
<th>Segments jp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosody FR</td>
<td>FRfr FRca FRjp</td>
<td></td>
</tr>
</tbody>
</table>

Test 2: Prat was used to manipulate final lengthening and continuation rise of 5 phrases: they were 1) added to the utterances of the learners that lacked them, 2) removed from those of French native speakers.

Table 3: Manipulation of duration and F0 (Test 2)

<table>
<thead>
<tr>
<th>F0 manipulation</th>
<th>duration no duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>both changed</td>
<td>F0 changed</td>
</tr>
<tr>
<td>DUR changed</td>
<td>no change</td>
</tr>
</tbody>
</table>

2.2. Subjects

17 native speakers of French (from mainland France) participated in the tests.

2.3. Experimental protocol

The subjects were asked to answer the following question: “Est-ce qu’il/elle a bien dit ?” (Did he/she say well?). They were provided with a 7-degree scale, ranging from 1 (pas du tout – not at all) to 7 (très bien – very well), and were asked to click on the area on a computer screen corresponding to their answer.

Each test (Tests 1 and 2) was preceded by a training session, and was divided into two sub-parts with a break between them. In order to control the side effect due to the presentation order, the two sets of tests as well as the stimuli in each test were presented in a different order to half of the subjects. Also, each stimulus type was presented twice to check the consistency of judgment.

3. Results

3.1. Test 1 (Mbrola)

3.1.1. Intra-subject response consistency

Every subject shows a high correlation for the two responses to the 60 stimuli (Spearman’s rank order correlation: \( p = 0.61 \)) (median of all subjects), \( p < 0.05 \). Given that the listeners are consistent in their responses, we adopt the mean of the two scores.

3.1.2. Presentation order

Order of the two tests (Test 1 and Test 2): A high correlation (Spearman’s \( p = 0.93 \), \( p < 0.05 \)) is found between the mean scores of the two groups of subjects. The difference in mean score (for all the stimuli) of the two groups is NOT significant (Student’s one sample t test: \( t_{15} = 0.57 \), \( p > 0.05 \)).

Order of stimuli in each test: A high correlation (\( p = 0.68 \), \( p < 0.05 \)) is observed between the two groups. Also, the difference in mean score (of all the stimuli) of the groups is NOT significant (\( t_{15} = 0.23 \), \( p < 0.05 \)).

3.1.3. Factor PROSODY

The mean score for Pros FR (prosody of the French speakers) is 4.3, and that for Pros JP (prosody of the learners) is 3.2 (Figure 1). The difference is statistically significant (Student’s paired t test: \( t_{9} = 9.25 \), \( p < 0.05 \)). If we look at each listener, 11 out of 17 show a significant difference.

When we consider only the stimuli with Segments jp (Seg jp), the difference in mean score (Pros FR: 3.5, Pros JP: 2.7) is also significant (\( t_{11} = 2.94 \), \( p < 0.05 \)). Figure 2), and 11 out of 17 subjects show a significant difference.

Figure 1 (left): Mean scores for Pros FR and Pros JP (30 stimuli each, including all types of segment databases, namely, fr, ca, jp). The large * indicates that the difference between the two conditions is significant. The small * indicates that the score is significantly deviated from 4 (neither good nor bad). The convention also holds for the following figures. Figure 2 (right): Mean scores for Seg jp with Pros FR and Pros JP (10 stimuli each).

3.1.4. Factor SEGMENTS

The mean scores for Seg fr, ca and jp are 4.5, 3.6, and 3.1 respectively. A global effect of Segments on the score is found (ANOVA: \( F_{2, 57} = 15.84 \), \( p < 0.05 \)), the difference between each one of the categories (ca, fr, jp) contributing the
global effect (Figure 3). Out of the 17 listeners, 14 show a global effect of Segments. A posteriori comparison between two categories (Student’s paired t test) reveals an individual difference of the subjects in the hierarchy of the three conditions. For example, about half the subjects judged Segments fr significantly better than the other two (Table 4).

Table 4: Hierarchy of ranking for fr, ca and jp. “>>” means that the difference is significant.

<table>
<thead>
<tr>
<th></th>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>fr &gt;&gt; ca, jp</td>
<td>8 subjects</td>
</tr>
<tr>
<td>fr &gt;&gt; ca &gt;&gt; jp</td>
<td>5 subjects</td>
</tr>
<tr>
<td>fr, ca &gt;&gt; jp</td>
<td>2 subjects</td>
</tr>
<tr>
<td>others</td>
<td>2 subjects</td>
</tr>
</tbody>
</table>

Figure 3 (left): Mean scores for Segments ca, fr and jp (20 stimuli each, including Pros FR and JP).
Figure 4 (right): Mean scores for 1) Pros FR + Seg jp and 2) Pros JP + Seg fr (10 stimuli each).

3.1.5. Prosody and Segments

Let us compare now the two factors. What interests us here the most is the comparison between the following two combinations: 1) Pros FR + Seg jp and 2) Pros JP + Seg fr (Figure 4). The mean scores are 3.5 and 3.7 respectively. The difference is not significant (t = 0.83, p > 0.05). Out of the 17 listeners, 9 judged “Pros JP + Seg fr” to be better (the difference being significant for 3 of them), and the others judged the other combination to be better (none of them shows a significant difference).

Also, the results of ANOVA (on the mean) with these two factors indicate that both of them (Prosody: F(1, 36) = 42.93, p < 0.05; Segments: F(1, 36) = 59.67, p < 0.05) as well as the interaction between the two (F(1, 36) = 4.91, p < 0.05) are significant. (Figure 5) As for inter-subject difference, factor Prosody is significant for 16 listeners, Segments for 15, and the interaction for 8 subjects.

3.1.6. Intensity adjustment

- Pros FR: The mean scores with and without adjustment are 4.2 and 4.4 respectively. The global difference is significant (t(14) = 2.31, p < 0.05. Figure 6). However, only 2 subjects show a significant difference.
- Pros JP: The mean scores with and without adjustment are 3.0 and 3.4 respectively. The global difference is again significant (t(14) = 3.52, p < 0.05. Figure 7). 9 out of 17 listeners show a significant difference.

3.2. Test 2 (duration and F0 manipulation)

3.2.1. Intra-subject response consistency

The two responses to the 44 stimuli show a high correlation (Spearman’s r = 0.88 (median of all the subjects), p < 0.05). The subjects can be considered consistent in their responses.

3.2.2. Presentation order

Order of the tests (Tests 1 and 2): A high correlation is obtained between the two groups (r = 0.97, p < 0.05). Also, the difference in score is NOT significant (t(15) = 2.01, p > 0.05).

Order of the stimuli: Again, a high correlation is obtained (r = 0.96, p < 0.05) between the two conditions, and the difference in score is NOT significant (t(15) = 0.16, p > 0.05).

3.2.3. Speech rate and pause

Neither the manipulation of speech rate (ANOVA: F(1, 64) = 0, p > 0.05) nor the insertion of pause (F(1, 64) = 0.12, p > 0.05) influenced the scores significantly.

3.2.4. Duration manipulation

The mean scores for French native speakers’ productions with and without final shortening are 5.3 and 5.8 respectively. The difference is not significant (t(16) = 1.42, p > 0.05. Figure 8). As for Japanese learners, the mean scores are 2.8 and 3.0, with and without final lengthening respectively. The difference is not significant (t(18) = 0.52, p > 0.05).

3.2.5. F0 manipulation

The mean scores for French native speakers’ productions with and without F0 manipulation are 5.4 and 5.7 respectively. The difference is not significant (t(16) = 1.06, p > 0.05. Figure 10). As for the learners’, they are 2.9 and 2.8, with and without manipulation. The difference is not significant (t(18) = 0.19, p > 0.05).
native-like prosody could improve significantly the naturalness of utterances with non-native-like segments (3.2.6). Further experiments will have to be conducted using longer phrases and sentences with more careful manipulations, in order to examine the possibility of improvement by adding final lengthening and continuation rise to learners' productions.

Acknowledgments

I would like to express my gratitude to Jacqueline Vaissière for her guidance, Jean-Yves Dommergues for his advice on the experimental design and statistical analyses, as well as to Cécile Fougeron, Bernard Gautheron, Chakir Zeroual, other laboratory members, and all those who spent their time participating in the recording and the perception tests.

5. References