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

Takeki Kamiyama

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
takekik@phiz.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 [1]) and have seen some works that demonstrate the relative 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 it 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 jump 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 Wioland [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 in duration between the two languages, namely, sentence-final lengthening. Final lengthening is a widely observed phenomenon. In French the last syllable in rhythmic groups (or phrase) is lengthened and generally a (major or minor) continuation rise is superimposed on it. Lengthened duration and F0 peak (minor phrase) or F0 rise (major rise) leads to the perception of "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 pitch (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.

Un peu de panaché.
Beaucoup de panaché.
L'économie des États-Unis.
L'économie du Canada.
Les mots à éviter.
Une compagnie dénationalisée.
Une photo de chameaux.
Il y a une photo de chameaux.

Table 1: *Phrases used in the tests.*

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)

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

	Segments fr	**Segments ca**	**Segments jp**
Prosody FR	FRfr	FRca	FRjp
Prosody JP	JPfr	JPca	JPjp

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

Test 2: Praat 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.

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).

	duration manipulation	**no duration manipulation**
F0 manipulation	both changed	F0 changed
no F0 manipulation	DUR changed	no change

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

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 tests as well as the stimuli 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-speaker response consistency

The two responses for the 60 stimuli show a high correlation with every subject (Spearman's rank order correlation: $\rho = 0.61$ (median of all subjects), $p < 0.05$). Given that the listeners are consistent in their responses, we take the mean of the two scores.

3.1.2. Presentation order

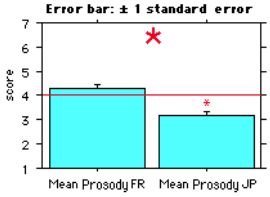
Order of the tests (Test 1 and Test 2): A high correlation (Spearman's $\rho = 0.93$, $p < 0.05$) is found between the mean scores of the two groups of subjects. The difference in mean score (for all the stimuli) of the two groups is NOT significant (Student's one sample t test: $t_{15} = 0.57$, $p >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

Order of stimuli: A high correlation ($\rho = 0.68$, $p < 0.05$) 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 is 3.2 (Figure 1). The difference is statistically significant (Student's paired t test: $t_{29} = 9.25$, $p < 0.05$). If we look at each listener, 16 subjects out of 17 show a significant difference.

When we consider only those stimuli with Segments jp (Seg jp), the difference in mean score (FR 3.5, JP 2.7) is also significant ($t_9 = 2.94$, $p < 0.05$. Figure 2), and 11 out of 17 subjects show a significant difference.



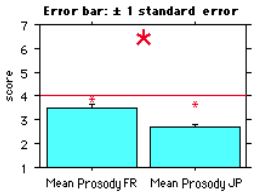


Figure 1 (left): *Mean scores for Pros FR and Pros JP (30 stimuli each, including all types of segment databases, namely, fr, ca, jp). The large * indicates that the difference between the two conditions is significant. The small * indicates that the score is significantly deviated from 4 (neither good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

fr >> ca, jp	8 subjects
fr >> ca >> jp	5 subjects
fr, ca >> jp	2 subjects
others	2 subjects

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

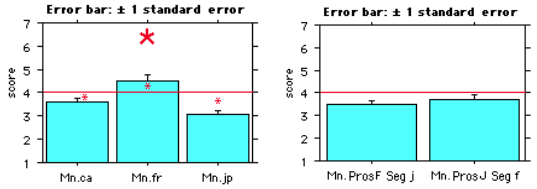


Figure 3 (left): *Mean scores for Segments ca, fr and jp (20 stimuli each, including Pros FR and JP).* Figure 4 (right): *Mean scores for 1) Pros FR+Seg jp and 2) Pros JP+Seg fr (10 stimuli each).*

3.1.5. *Pros and Segments*

Let us compare now the two factors. What interests us here the most is the comparison between the following two conditions: 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 ($t_9 = 0.83$, $p > 0.05$). Out of the 17 listeners, 9 judged Pros JP + Seg fr better (the difference being significant for 3 of them), and the others judged the other combination 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 each listener, factor Prosody is significant for 16 listeners, Segments for 15, and the interaction for 8 subjects.

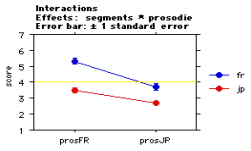


Figure 5: *Interaction between the factors Prosody and Segments (10 stimuli for each condition).*

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.

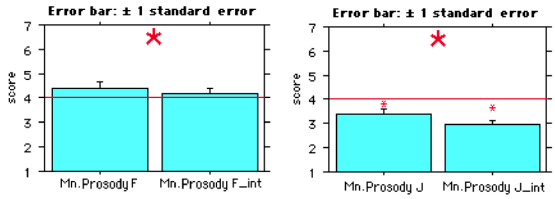


Figure 6 (left): Mean scores for Pros FR with (right) and without (left) intensity adjustment. *Figure 7 (right):* Mean scores for Pros JP with (right) and without (left) intensity adjustment.

3.2. Test 2 (duration and F0 manipulation)

3.2.1. *Intra-speaker response consistency*

The two responses to the 44 stimuli show a high correlation (Spearman's $\rho = 0.88$ (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 ($\rho = 0.97$, $p < 0.05$). 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 ($\rho = 0.96$, $p < 0.05$) 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 ($t_{18} = 0.52$, $p < 0.05$. Figure 9).

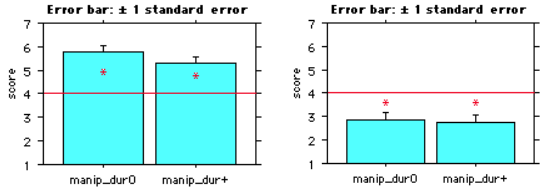


Figure 8 (left): *Mean scores for French speakers' productions with (right) and without (left) duration manipulation (18 stimuli).* Figure 9 (right): *Mean scores for Japanese learners with (right) and without (left) duration manipulation (22 stimuli).*

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 ($t_{18} = 0.19$, $p > 0.05$, Figure 11).

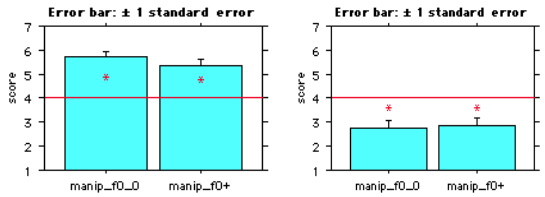


Figure 10 (left): *Mean scores for French speakers' productions with (right) and without (left) F0 manipulation (18 stimuli)*. Figure 11 (right): *Mean scores for Japanese speakers with (right) and without (left) F0 manipulation (22 stimuli).*

3.2.6. Duration + F0 manipulation

The mean score (of the 17 listeners) for French speakers' productions with final-shortening and F0 manipulations (= 5.2) is significantly lower than for those with no manipulation (= 6.1; Student's paired t test: $t_{16} = 6.98$, $p < 0.05$. Figure 12). The learners' productions with manipulation of these two factors were not judged significantly better except for the sentence "Il y a une photo de chameaux." (3.0 and 2.2, with and without manipulation; $t_{16} = 2.47$, $p < 0.05$. Figure 13).

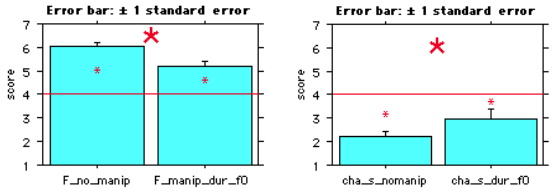


Figure 12 (left): *Mean scores for French speakers' productions with F0 and duration manipulation (right) and no manipulation (left) (9 stimuli).* Figure 13 (right): *Mean scores for Japanese speaker's production of the sentence "Il y a une photo de chameaux." with F0 and duration manipulation (right) and no manipulation (left) (2 stimuli).*

4. Concluding remarks

The results suggest that prosody plays a highly important role in the evaluation of the naturalness by French listeners, and that a native-like prosody could improve significantly the naturalness of utterances with non-native-like segments (3.1.3). Segments play also an important role in the evaluation of naturalness, but not so much as prosody. It is interesting to note that even native-like segments (Seg ca) could receive a low evaluation if they do not conform to the norm of the listeners' standard (in this case, French listeners from France. cf. 3.1.4). It is difficult, however, to evaluate directly the relative importance of native-like segments and prosody from our data (3.1.5). As for intensity (3.1.6), the intensity curve produced by Japanese learners could be one of the sources of non-native-likeness. However, since the score for Pros FR without intensity adjustment is higher than that with adjustment, we cannot exclude the possibility that the manipulation in itself made the stimuli less natural.

Concerning Test 2, it is suggested that the slow speech rate and the presence of pauses, often observed in learners' utterances, are not in themselves important factors in the naturalness evaluation (3.2.3). Final shortening manipulation combined with continuation rise deletion causes to deteriorate the goodness of native speakers' speech, but on the other hand, final lengthening and creation of continuation rise did not improve significantly the non-native-like prosody of learners' except for one sentence (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 lab members, and all those who spent their time participating in the recording and the perception tests.

5. References

[1] Dalton, C. and B. Seidlhofer, 1994. *Pronunciation*. Oxford: Oxford University Press.

[2] Suzuki, H., 1992. Gengo gijutsu to shite no purosodii (Prosody as a Linguistic Skill). *Gengo* (Language) 21, 38-45.

[3] Tajima, K., R. Port, J. Dalby, 1997. Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 25(1), 1-24.

[4] Kubozono, H., 1993. *The Organization of Japanese Prosody*. Tokyo, Kuroshio shuppan.

[5] Beckman, M., 1982. Segment Duration and the 'Mora' in Japanese. *Phonetica* 39, 113-135.

[6] Kurematsu, A., 1997. Inritsu jouhou no jouhou kagaku eno tekiyou (Applications of prosodic information to information sciences). In *Nihongo onsei 2: akusento, intoneeshon, rizumu to poozu* (Japanese phonetics 2: accent, intonation, rhythm and pause). M. Sugito et al.. Tokyo: Sanseido. 303-318.

[7] Sagisaka, Y., 1999. Nihongo on-in no jikanchou seigyo to chikaku (Duration control and perception of phonemes in Japanese). *Gengo* (Language). 28(9).

[8] Vaissière, J., 2001. Cross-Linguistic Prosodic Transcription: French versus English. In *Problems and Methods of experimental Phonetics ---in honour of the 70th anniversary of Prof. L. V. Bondarko*. N. B. Volskaya, N. D. Svetozarova and P. A. Skrelin.

[9] Hallé, P., B. Boysson-Bardies, M. Vihmanal, 1991. Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech* 34(4), 299-318.

[10] Wenk, B. and F. Wioland, 1982. Is French really syllable-timed? *Journal of Phonetics* 10(2), 193-216.

[11] Delattre, P., 1966. Accent de mot et accent de groupe. In *Studies in French and Comparative Phonetics*. P. Delattre. The Hague: Mouton & co., 69-72.

[12] Takeda, K., Y. Sagisaka, H. Kuwabara, 1989. On sentence-level factors governing segmental duration in Japanese. *Journal of the Acoustical Society of America* 86(6), 2081-2086.

[13] Kamiyama, T., 2002. Les aspects suprasegmentaux des phrases déclaratives courtes en français lues par les apprenants japanophones. Unpublished DEA thesis. DEA de phonétique, Université de la Sorbonne Nouvelle.

[14] Dutoit, T. Mbrola Project, http://tcts.fpms.ac.be/synthesis/mbrola.html.

[15] Boersma, P. and D. Weenink, 1993-2003. Praat, http://www.fon.hum.uva.nl/praat/.