Learning effect of prosodic social affects for Japanese learners of French language

Takaaki Shochi [1], Gwenaëlle Gagnié [1], Albert Rilliard [2], Donna Erickson [3], & Véronique Aubergé [4]

[1] Kumamoto University, Japan
[2] LIMSI-CNRS, Orsay, France
[3] Showa Music University, Kawasaki City, Japan
[4] GIPSA Lab, Grenoble, France

shochi38@gmail.com, gwen@kumamoto-u.ac.jp, albert.rilliard@limsi.fr, ericksondonna2000@gmail.com, veronique.auberge@gipsalab.inpg.fr

Abstract

This paper investigates the differences in the perception of six culturally encoded French social affects for Japanese and native listeners. Half of the Japanese listeners have followed six months of training about both prosodic and facial realization of French social affects. Audio-visual stimuli were presented to listeners, who guess speaker's intended attitude and rate the intensity of the expressiveness. Results showed that the trained Japanese listeners recognized better than the untrained ones; however, culturally specific attitudes (i.e. suspicious irony and obviousness) were confused by Japanese listeners (including trained listeners). Facial information cues seem to be more salient than audio ones.

Index Terms: prosodic social affects, language learning, French.

1. Introduction

Delattre [1, 2] early on advocated the importance of prosody in the expressivity of spoken language and the implication of such variations in intonation for foreign language teaching. Since then, other authors (e.g. [3]) have developed a repertory of expressive prosodic patterns, for the purpose of language teaching. As prosodic variations are linked to language, they are culturally encoded and are not universally recognized [4] – differing from basic emotional expressions [5, 6, 7]. Cross-cultural studies of expressive intonation are therefore important in order to gather precise information about its perception [8].

Audiovisual parameters have an important impact on the perception of prosody [9]. Rilliard et al. [10] show for both Japanese and French social affects, the contribution of audio and visual information. Perception results showed the importance of the speaker's strategy in both modalities for the perception of each attitude. These works insist also on the importance of multimodality during interaction: some expressions are mainly recognized through audio cues or by visual cues, whereas for some others, the synergy between modalities is important.

The current paper investigated the perception of six French attitudes by Japanese listeners with different levels of French language proficiency in order to compare their performances with native listeners' perceptual behavior. The first part of the paper describes the corpus and explains the experimental setting, then the results obtained from Japanese listeners are compared with those of French native listeners in order to identify the correlation between language skill and perception of social affects.

2. Experimental setup

In previous work [10], the acoustic as well as facial characteristics of French audio visual prosody for 6 attitudes were described. Using these objective characteristics, a native French teacher has trained 32 Japanese learners of first year French at Kumamoto University for six months. This training was done during the last 10 minutes of conversation class with 3 different tasks:

1. Explanation of acoustic and facial characteristics for six French attitudes;
2. Perceptual discrimination task (using a different speaker to prevent listeners becoming accustomed to the speaker of our experimental corpus);
3. And production of these six French attitudes in both audio and visual modalities.

This training was conducted systematically for each attitude.

2.1. Corpus

Following the work done by [11] on French prosodic attitudes, based on [12], 6 attitudinal expressions were selected for recording a French audio-visual corpus: *declaration* (DC), *simple question* (IN), *obviousness* (OB), *surprise exclamation* (SU), *doubt-incredulity* (DO), *suspicious irony* (SC). All attitudinal expressions were acted by a native French male speaker for these 6 attitudes: the speaker was instructed to produce each of these sentences in order to express one attitude, as an answer to a statement produced by a partner. He has already been trained to produce these attitudes in a preceding session, and has to behave as naturalistically as possible, without any constraints on his expressive strategy.

This corpus was recorded in a soundproof room at LIMSI. The speaker was standing in front of a video camera, with an omnidirectional AKG C414B microphone placed 40 cm to his mouth. Recordings were digitalized at 44.1 kHz, 16bits. A digital DV camera (Canon XM1 3CCD) recorded the speaker's performances. Video clips were encoded using a cinepack codec with a 784 x 576 pixels resolution, using AVI video file formats. The corpus is based on three sentences respectively of 4, 5 and 7-syllable length. The meaning of sentences did not refers to or forbid the expression one of the 6 attitudes. After the recording and the validation of this corpus [10], the 5-syllable length sentence "*Nicolas revenai*t." [nikola ʁəvnɛ] ("Nicolas was coming back") was selected for the following perception test.

2.2. Listeners

3 groups of listeners, selected according to their French language level, took the experiment. The first group consisted of 32 French native listeners (FR) (mean age = 32, 15 females, 17 males); the second was composed of 31 Japanese, French language learners in Japan, who were trained to produce and perceive audio-visual French attitudes during six months (JP1) (mean age = 19 years, 26 females, 5 males); the third group was composed of 50 Japanese listeners who had never heard French (JP0) (mean age = 18 years, 34 females, 16 males). The total number of subjects for this experiment was 113. French subjects took the experiment in a quiet room, individually using a computer and headset. All Japanese subjects took the experiment in a computer room with individual headsets for each subject. The task was described and after explanations and questions given in listener's mother tongue, they took the experiment. A few listeners did not complete the experiment (they are not included in the above description). No listener reported any perception trouble.

2.3. Paradigm

Stimuli were presented to the subjects in three experimental conditions, according to the modality of presentation: audio-only (A), video-only (V), audio-video (AV). The first two conditions are audio-only and video-only, and listeners perceive audio-video stimuli in the third one. In order to balance the effect of the presentation order, half the subjects began with the video-only stimuli, the other half with the audio-only ones. Each sound could be listened to only once (in the A condition), as were each video played only once in the V and AV conditions.

For each stimulus, they had to select the attitude they perceived in the stimulus as well as its intensity on a scale ranging from "hardly perceptible" to "very marked" (encoded on a 1-100 scale, with the 0 score for the 5 not selected attitudes). Subjects had to fill the questionnaire without any time constraint. Each subject took the three experimental conditions (corresponding to the three modalities of presentation) during the same sitting.

3. Results

3.1. Effect of experimental factors on perception

Results given by all participants (including the results given by French native listeners [10]) were analyzed by an analysis of variance (ANOVA) in order to investigate the effect of the different factors on the perceptual behavior of listeners. There were two between-subject fixed factors: French language skill (FrLv, 3 levels: native listeners, Japanese learners of French and Japanese with no skill in French) and the order of presentation of modalities during the test (Grp - 2 levels). Two within-subject fixed factors were used: the presented attitudes (Att - 6 levels) and the modality of stimuli (Mod - 3 levels).
According to figure 1, a significant effect of language skill on the perception of social affects was observed. The interaction between language skill and modality was not significant ($F(4,214) = 1.19$, $p>.05$). This may indicate an evolution of listener's perception of French social affects depending on their language competence, whatever the modality of presentation.

The interaction between language level and attitudes was significant ($F(10,535) = 6.89$, $p<.01$). A post-hoc test (simple main effect on the different level of factor FrLv) is significant for Obviousness ($F(2,642) = 26.34$, $p<.01$) and Suspicious irony ($F(2,642) = 27.24$, $p<.01$). As shown in figure 1, the recognition rate of Japanese subjects (both JP0 and JP1) are obviously lower than French native listeners' score. These two attitudes seem be more difficult to recognize for Japanese listeners.

Figure 1 shows the 3 groups' recognition rate for each attitude in the three different modalities. French listeners show generally higher recognition score than Japanese listeners. Within the Japanese listeners, JP1 recognition scores are higher than those of naïve listeners. This increase of identification rate acknowledges the benefit of the training on affective prosody. In addition, the perception in the audio-visual modality shows the best score for almost all attitudes in all three groups. However, some different combinations of both audio and visual information for the perception of each attitude was observed. For instance, two attitudes (doubt and surprise) seem primarily influenced by visual cues, whereas interrogation was deeply influenced by audio information. Moreover, a mutual contribution of audio and visual information was also found for the perception of suspicious irony.

As a second step, separate ANOVA were run for each groups of Japanese listeners in order to investigate how non-native listeners perceive French attitudes in various modalities. There was one between subject fixed factor: the order of presentation of modalities during the test (Grp - 2 levels), and two fixed within-subject factors were used: the presented attitudes (Att - 6 levels) and the modality of stimuli (Mod - 3 levels). The results are shown in Table 1 (c, d).

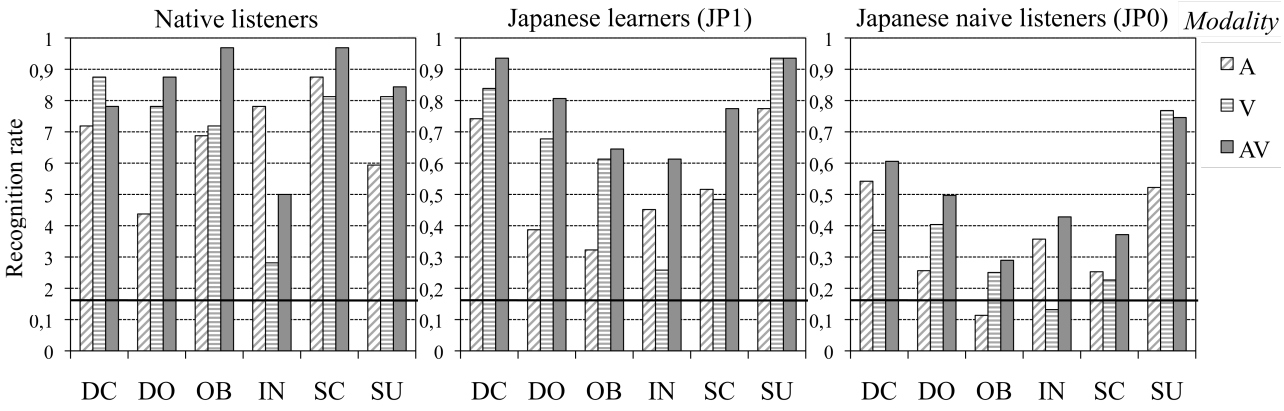

Figure 1: *Recognition rate for each attitude, in each modality: FR (left), JP1 (center) and JP0 (right). Results are presented for declaration (DC), doubt (DO), obviousness (OB), interrogation (IN), suspicious irony (SC) and surprise (SU).*

Table 1. *Results of the ANOVA: table (a) presents the ANOVA on the 3 groups of listeners, while tables (b, c, d) present results for one group (respectively natives, JP0 and JP1). Only significant factors at the 5% level are reported.*

a. 3 groups of listeners

Factors	d.f.	F	p
FrLv	2	21.4	0.000
Mod	2	30.9	0.000
Att	5	28.6	0.000
FrLv:Att	10	6.9	0.000
Mod:Att	10	8.6	0.000

	b. Native			c. JP0		d. JP1	
Factors	d.f.	F	p	F	p	F	p
Mod	2	7.8	0.001	11.1	0.000	12.8	0.000
Att	5	8.0	0.000	21.3	0.000	17.6	0.000
Mod:Att	10	6.0	0.000	4.0	0.000	2. 1	0.028

A significant effect of modality as well as of the presented attitude for both identification and intensity of French social affects was observed. In addition, a significant interaction between the modality and the presented attitude was observed for each group of subjects. This results shows that each group of listeners retrieve and interpret different information from each modality. For example, native listeners use AV cues in synergy for obviousness, while Japanese use mainly visual cues; and Japanese learners perceive declaration from visual cues also, while naïve Japanese don't.

We also found different perceptual behavior among the three groups for each attitude. French listeners perceive obviousness thanks to the contribution of both audio and visual information, but JP1 listeners seem to rely mainly on visual information for this attitude. JP0 listeners perceived this attitude in the same way as JP1, but their recognition rate was significantly lower than JP1.

3.2. Categories of French social affects

A hierarchical cluster analysis was run on the dispersion matrix obtained from each group of listeners. It measures the perceptive distances between each stimulus and allows the identification of wider perceptual categories amongst the six French social affects for each groups of listeners and each modality. Results are presented in Figure 2.

Native listeners. At the higher level of clustering, native subjects perceptually grouped attitudes in three generic groups: one composed of "dubitative" expressions (IN, SU, DO), one with "assertions" (DC, OB) and one with suspicious irony (SC). This broad grouping is observed on audiovisual stimuli and on audio only stimuli. The main differences between modalities concern (1) obviousness, regrouped with declaration with audio cues, but with suspicious irony on the basis of video cues, and (2) surprise, that is mainly perceived from video cues

Japanese learners (JP1). A similar analysis made on perception results obtained by JP1 subjects show the same opposition found for French between dubitative and assertive sentences but with only the declarative expression forming the assertive group: obviousness is systematically grouped with suspicious irony, whatever the modality. Another difference between Japanese learners and native listeners concerns the expression of surprise, which is discriminated by Japanese in any modality, even from audio cues only.

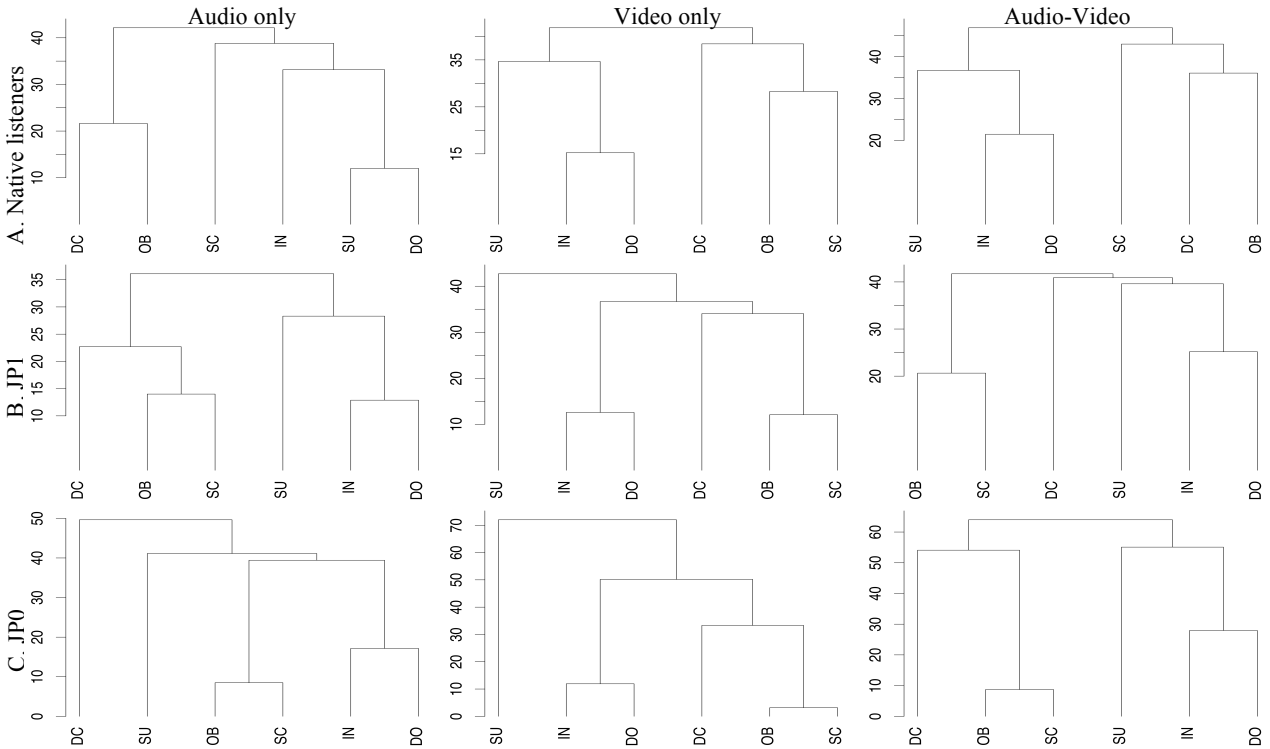

Figure 2. *Hierarchical clustering of the French attitudes, obtained from the results of French (A), JP1 (B) and JP0 (C) listeners, for the audio (left), video (center) and audio-video (right) modalities. The scale indicates the perceptive distance between each attitude. Results are presented for declaration (DC), doubt (DO), obviousness (OB), interrogation (IN), suspicious irony (SC) and surprise (SU).*

Naïve Japanese listeners (JP0). The clustering analysis performed on the results of JP0 listeners show similar results to those of JP1 listeners: they also oppose dubitative to assertive expressions, but for the audiovideo modality only. The confusion between obviousness and suspicious irony is also more important for JP0 subjects, and suggests that these French social affects performances may be difficult to be distinguished for Japanese listeners. The expression of surprise is the better-recognized expression, here also whatever the modality.

4. Discussion & Conclusions

This paper investigates the perceptual behavior of 32 French native listeners vs. 82 Japanese listeners for six prosodically expressed French social affects. Moreover it also examines the relationship between the language skill and the listeners' perceptual behavior. According to the results, the listeners change their perceptual behavior with the different modalities. The perception for the audio-visual modality shows the best scores for almost all attitudes in all three groups of listeners. As shown in previous work [4, 10], three different combinations of audio and visual information for the perception of each attitude were observed: (1) one perceptual style gives priority to visual cues (ex. doubt and surprise); (2) another perceptual style gives more importance to audio information (ex. interrogation); and (3) finally, the contribution of audio and visual information allows listeners to decode the attitude (ex. SC).

An important influence of language skill on the perception of social affects was also identified. This training effect was clearly identified for the perception of attitudes. The effect of this training is particularly important at a global level for the perception of dubitative expressions: they are perceived by Japanese learners as different plausible strategies for similar overall kinds of expressions, but this is not seen with naïve Japanese listeners. At a more detailed level, the perceptual behavior for two culturally specific attitudes (obviousness and suspicious irony) show that Japanese listeners rely on similar cues and interpretations, whatever their French language level and despite an increasing recognition performance: they mostly rely on visual cues to recognize these social affects and always show confusion between these two expressions, even if French listeners regroup obviousness mainly with declaration and not suspicious irony – except for video only cues. Moreover, French listeners discriminate obviousness very well with audio-visual information, while Japanese subjects hardly discriminate these two attitudes.

5. Acknowledgements

We are deeply grateful to students of French department in Kumamoto University for their participation to our perception test. We also thank Dr. Kaoru Sekiyama who allowed us to use all materials in this experiment. This work was supported by the Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007-2010):19520371 to the second author, as well in part by SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.

6. References

[1] Delattre, P. "Les dix intonations de base du français". The French Review, 40(1):1-14, 1966.

[2] Delattre, P. "La nuance de sens par l'intonation". The French Review, 41(3):326-339, 1967.

[3] Martins-Baltar, M. "De l'énoncé à l'énonciation: une approche des fonctions intonatives". Paris: Didier, 1977.

[4] Shochi, T., Rilliard, A., Aubergé, V. and Erickson, D. "Intercultural Perception of English, French and Japanese Social Affective Prosody". In The role of prosody in Affective Speech, ed. S. Hancil, pp.31-59, Linguistic Insights 97, Peter Lang AG, Bern, 2009.

[5] Zinck, A. and Newen A. "Classifying emotion: a developmental account". Synthese, 161:1–25, 2008.

[6] Scherer, K. R. and Wallbott, H. G. Evidence for universality and cultural variation of differential emotion response patterning. Journal of Personality and Social Psychology, 66(2):310–328, 1994.

[7] Scherer, K. R. and Brosch, T. Culture-specific appraisal biases contribute to emotion dispositions. European Journal of Personality, 23:265-288, 2009.

[8] Pavlenko, A. "Emotions and multilingualism". Cambridge (U.K.): Cambridge University Press, 2005.

[9] Swerts, M. and Krahmer, E. "Audiovisual prosody and feeling of knowing. Journal of Memory and Language, 53(1):81–94, 2005.

[10] Rilliard,A., Shochi, T., Martin, J.C., Erickson, D. and Aubergé, V. Multimodal Indices To Japanese And French Prosodically Expressed Social Affects. Language and Speech 52(2&3):223-243, 2009.

[11] Morlec, Y., Bailly, G. and Aubergé, V. Generating prosodic attitudes in French: Data, model and evaluation. Speech Communication, 33(4):357–371, 2001.

[12] Fónagy, I., Bérard, E. and Fónagy, J. Clichés mélodiques. Folia Linguistica, 17:153-185, 1984.